Article

Defining mitochondrial protein functions through deep multiomic profiling

https://doi.org/10.1038/s41586-022-04765-3

Received: 19 January 2021

Accepted: 14 April 2022

Published online: 25 May 2022

Check for updates

Jarred W. Rensvold[1,2,17], Evgenia Shishkova[3,4,17], Yuriy Sverchkov[5], Ian J. Miller[3,4], Arda Cetinkaya[6], Angela Pyle[7,8], Mateusz Manicki[1,2], Dain R. Brademan[2,4,9], Yasemin Alanay[10,11], Julian Raiman[12], Adam Jochem[2], Paul D. Hutchins[9], Sean R. Peters[9], Vanessa Linke[9], Katherine A. Overmyer[2,3,4], Austin Z. Salome[9], Alexander S. Hebert[3,4], Catherine E. Vincent[4], Nicholas W. Kwiecien[3,4], Matthew J. P. Rush[9], Michael S. Westphall[3,4], Mark Craven[5], Nurten A. Akarsu[6], Robert W. Taylor[7,8,13], Joshua J. Coon[2,3,4,9] ✉ & David J. Pagliarini[1,2,3,14,15,16] ✉

Mitochondria are epicentres of eukaryotic metabolism and bioenergetics. Pioneering efforts in recent decades have established the core protein componentry of these organelles[1] and have linked their dysfunction to more than 150 distinct disorders[2,3]. Still, hundreds of mitochondrial proteins lack clear functions[4], and the underlying genetic basis for approximately 40% of mitochondrial disorders remains unresolved[5]. Here, to establish a more complete functional compendium of human mitochondrial proteins, we profiled more than 200 CRISPR-mediated HAP1 cell knockout lines using mass spectrometry-based multiomics analyses. This effort generated approximately 8.3 million distinct biomolecule measurements, providing a deep survey of the cellular responses to mitochondrial perturbations and laying a foundation for mechanistic investigations into protein function. Guided by these data, we discovered that *PIGY* upstream open reading frame (PYURF) is an *S*-adenosylmethionine-dependent methyltransferase chaperone that supports both complex I assembly and coenzyme Q biosynthesis and is disrupted in a previously unresolved multisystemic mitochondrial disorder. We further linked the putative zinc transporter SLC30A9 to mitochondrial ribosomes and OxPhos integrity and established *RAB5IF* as the second gene harbouring pathogenic variants that cause cerebrofaciothoracic dysplasia. Our data, which can be explored through the interactive online MITOMICS.app resource, suggest biological roles for many other orphan mitochondrial proteins that still lack robust functional characterization and define a rich cell signature of mitochondrial dysfunction that can support the genetic diagnosis of mitochondrial diseases.

Mitochondria are probably remnants of an ancient endosymbiotic event between an alphaproteobacterium and a eukaryotic progenitor[6]. These organelles retain a vestige of their original bacterial genome that encodes just 13 proteins in humans; the remaining proteins that make up mitochondria are encoded by genes that have been transferred or added to the host nucleus across more than 1 billion years of evolution. Hundreds of these proteins remain poorly characterized or entirely uncharacterized. This important knowledge gap has limited basic understanding of mitochondrial function and has hampered efforts to diagnose and treat mitochondrial diseases, for which there are no US Food and Drug Administration-approved drugs[7].

We recently devised an integrative systems biochemistry[8] approach that leverages high-throughput quantitative mass spectrometry (MS)[9–11] to help to assign functions to mitochondrial uncharacterized (*x*) proteins (MXPs) in *Saccharomyces cerevisiae*[12]. Here we advance this methodology and apply it to a set of 203 human HAP1 knockout (KO) cell lines, each with a nucleus-encoded mitochondrial gene disrupted by CRISPR–Cas9 technology. Our targeted genes (1–2 KO lines per target) included 50 encoding MXPs and another 66 encoding sentinel proteins with more established functions, most of which have been directly linked to human disease (Fig. 1a, Extended Data Fig. 1a–c and Supplementary Table 1).

[1]Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO, USA. [2]Morgridge Institute for Research, Madison, WI, USA. [3]National Center for Quantitative Biology of Complex Systems, Madison, WI, USA. [4]Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, WI, USA. [5]Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, Madison, WI, USA. [6]Department of Medical Genetics, Faculty of Medicine, Hacettepe University, Ankara, Turkey. [7]Wellcome Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, UK. [8]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [9]Department of Chemistry, University of Wisconsin–Madison, Madison, WI, USA. [10]Department of Pediatrics, Pediatric Genetics Unit, Faculty of Medicine, Hacettepe University, Ankara, Turkey. [11]Department of Pediatrics, Pediatric Genetics Unit, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey. [12]Department of Clinical Inherited Metabolic Disorders, Birmingham Women's and Children's Hospital NHS Trust, Birmingham, UK. [13]NHS Highly Specialised Services for Rare Mitochondrial Disorders, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [14]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA. [15]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [16]Department of Biochemistry, University of Wisconsin–Madison, Madison, WI, USA. [17]These authors contributed equally: Jarred W. Rensvold, Evgenia Shishkova. ✉e-mail: jcoon@chem.wisc.edu; pagliarini@wustl.edu

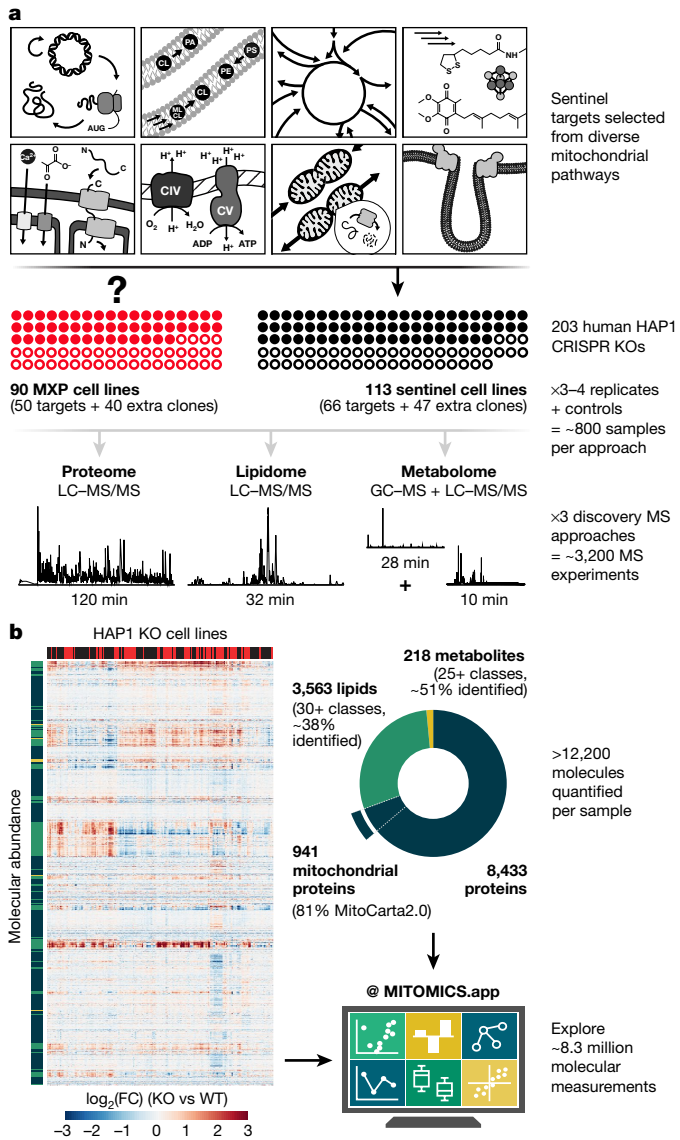


Fig. 1 | MITOMICS experimental design and data resource summary. **a**, Overview of the experimental workflow including KO target selection, analysis strategy and data collection. KO targets were selected to include genes encoding MXPs and sentinel proteins known to be involved in diverse mitochondrial processes. Each cell line was analysed using three distinct MS approaches, including analysis of proteins by LC–MS/MS shotgun proteomics, lipids by LC–MS/MS discovery lipidomics and metabolites by GC–MS and LC–MS/MS untargeted metabolomics. **b**, Hierarchical clustering of biomolecule abundance (proteins, lipids and metabolites) in 203 KO cell lines compared with WT cells (mean, $n = 3–4$) and breakdown of the >12,200 biomolecules quantified in each cell line by class. These data are available on the interactive MITOMICS resource at https://www.mitomics.app.

We monitored the growth rates of each cell line in biological triplicate (Extended Data Fig. 1d, e and Supplementary Table 1) and profiled them in depth using high-resolution and accurate mass discovery MS techniques. Overall, this mitochondrial orphan protein multiomic CRISPR screen (MITOMICS) encompassed approximately 3,200 gas chromatography (GC)– and liquid chromatography (LC)–MS experiments, generating approximately 8.3×10^6 quantitative measurements. Across each cell line, we quantified 8,433 proteins, 3,563 lipids and 218 metabolites (Fig. 1b and Supplementary Tables 2 and 3). These measurements were of high quality as evidenced by their low median relative s.d. of 11.6%, 21.8% and 18.6% for proteins, metabolites and lipids, respectively, and their high dynamic range, with many molecules showing regulation over 3–4 orders of magnitude (Extended Data Fig. 1f, g). Notably, our single-shot LC–MS/MS proteomics methodology incorporating multiple technical advancements[13,14] consistently quantified 5,192 proteins in all 772 experiments so that, in total, only 5.4% of the protein measurements in the final dataset were imputed (Extended Data Fig. 1h). Thirteen cell lines did not pass our stringent proteomics quality control filters and were not included in subsequent analyses here (Supplementary Methods). Using these data, we built an online interactive MITOMICS resource equipped with intuitive analysis tools for exploring mitochondrial protein function. This resource is accessible at https://www.mitomics.app.

Protein-specific molecular phenotypes

Simple molecule-centric analyses across each omic plane of the MITOMICS data can be used to recapitulate known biology and suggest new protein functions. For example, our metabolite measurements showed that disruption of ALDH18A1, a critical enzyme for de novo proline biosynthesis[13], led to the expected proline deficiency (Fig. 2a). Unexpectedly, we observed a comparable reduction in proline after disruption of NADK2, the mitochondrial NAD kinase (Fig. 2b). In agreement with this, NADK2 was linked to this pathway while the current work was under revision[14,15]. Similarly, our lipid data showed the expected alteration of cardiolipin and acylcarnitine levels in cells lacking *TAZ*[16] (Fig. 2c) and *CPT2* (ref. [17]) (Extended Data Fig. 2a), respectively. However, we observed similar changes in these lipids for cells lacking the mitochondrial fusion regulator *MFN2* (ref. [18]) and in specific acylcarnitine levels for cells lacking members of the mitochondrial contact site and cristae organizing system (MICOS)[19] or *PPTC7*, encoding a matrix phosphatase[20] (Fig. 2d and Extended Data Fig. 3a). The connection of MFN2 and MICOS to these lipids suggests that the importance of proper interaction between mitochondrial membranes in the coordination of lipid metabolic processes has been underappreciated. We noted similar results for cysteine, taurine, demethoxy coenzyme Q (DMQ) and DL-4-hydroxyphenyllactic acid (Extended Data Fig. 2b–e).

Our proteomics analyses also identify new biology. A prominent example of this is the analysis of SLC30A9, a putative zinc transporter that has only recently been associated with mitochondria[21,22]. In our study, disruption of *SLC30A9* resulted in substantial loss of mitochondrial ribosome (mitoribosome) and OxPhos proteins (Fig. 2e). Recent cryogenic electron microscopy studies have shown that mitoribosomes have unusual zinc-binding motifs proposed to stabilize the structures and quaternary interactions of subunit proteins[23]. Consistent with this, the abundance of all six mitochondrial DNA (mtDNA)-encoded OxPhos subunits detected by our MS method was strongly decreased in the SLC30A9^{KO} line, which was comparable to the mitoribosome sentinel MRPS22^{KO} line (Fig. 2f and Extended Data Fig. 3a). We further validated this finding with immunoblots of the mtDNA-encoded OxPhos subunit MT-CO2 in cells lacking *SLC30A9*, *MRPS22* or *MTRES1*—which encodes a mitochondrial RNA-binding protein involved in mtDNA expression—and in cells we engineered to harbour an *SLC30A9* mutation identified in a patient[24] (Extended Data Fig. 3b, c). Also supporting this observation, we found that the highest-ranking genes co-essential with *SLC30A9* in the DepMap project[25] were strongly enriched for those encoding mitoribosome subunits (Extended Data Fig. 3d, e). Furthermore, our previous study[26] and ref. [27] both identified SLC30A9 as a binding partner for mitoribosome-related proteins, including MTRES1 (Extended Data Fig. 3f). These data nominate SLC30A9 as a zinc-related transporter with a crucial function in core mitochondrial processes. On the basis of similar profiles generated with the tools available in our MITOMICS resource, we propose new functions for mitochondrial proteins in diverse areas, including redox biology, MICOS integrity, protease function, mtDNA regulation and glycogen metabolism (Extended Data Fig. 4a–e).

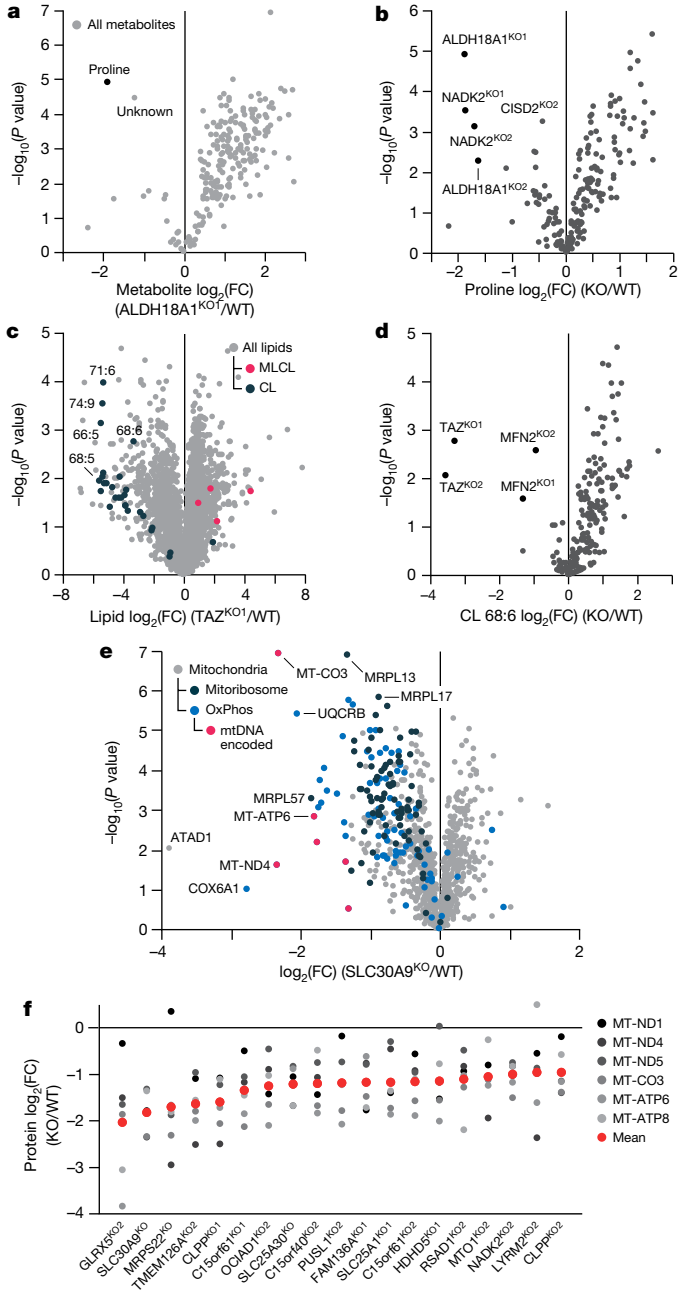


Fig. 2 | Molecule-centric analyses suggest new mitochondrial protein functions. a, Relative metabolite abundance in ALDH18A1^{KO1} cells compared with WT cells versus statistical significance. FC, fold change. **b**, Relative abundance of proline versus statistical significance across all KO lines. **c**, Relative lipid abundance in TAZ^{KO1} cells compared with WT cells versus statistical significance with all cardiolipin (CL) and monolysocardiolipin (MLCL) species highlighted. **d**, Relative abundance of cardiolipin 68:6 versus statistical significance across all KO lines. **e**, Relative protein abundance in SLC30A9^{KO} cells compared with WT cells versus statistical significance with mitoribosome, OxPhos and mtDNA-encoded proteins highlighted. **f**, Relative abundance of all six mtDNA-encoded proteins detected in our analyses across KO lines with ≥3 proteins having $P < 0.05$ (mean rank ordered). Data reflect the mean from $n = 3$–4 replicates with two-sided Welch's *t*-test used for all panels.

PYURF regulates CoQ and CI biology

As demonstrated above, our resource can be used to identify proteins involved in defined processes. Many diagnosed human mitochondrial disorders featuring CoQ[28] or complex I (CI)[29] deficiencies lack apparent mutations in established CoQ- and CI-related genes, suggesting that these pathways rely on proteins that have yet to be identified.

To search for MXPs involved in these processes, we first analysed the CoQ levels in each of our cell lines. As expected, the strongest losses of CoQ were observed in our sentinel lines lacking proteins required for CoQ biosynthesis (Extended Data Fig. 5a). Following these, the top hit among all MXPs was *PIGY* upstream open reading frame (*PYURF*), which was recently identified in a CRISPR-based screen for genes essential for mitochondrial respiration[30] but is otherwise uncharacterized (Fig. 3a). In line with these findings, our metabolomics data indicated that the PYURF^{KO} line had elevated levels of dihydroorotate, which requires CoQ to be converted to orotate (Extended Data Fig. 5c). Unexpectedly, our search for CI-related proteins based on NDUFS3 levels (our CI sentinel protein) again yielded PYURF as the top MXP hit (Fig. 3a and Extended Data Fig. 5b), suggesting that PYURF may somehow bridge these essential and interrelated pathways.

To further explore the connection between PYURF and these pathways, we analysed the full proteomic profile of the PYURF^{KO} line. The abundance of three CoQ-related proteins (COQ3, COQ5 and COQ7) and three CI assembly factors (AFs) (NDUFAF3 (AF3), NDUFAF5 (AF5) and NDUFAF8 (AF8))[31] was markedly decreased in this line (Fig. 3b). The PYURF^{KO} line stood out among all our cell lines for the coordinated loss of these proteins (Fig. 3c, d). Other CI subunits also had decreased abundance in this line, particularly those that make up the CI Q module where CoQ binding occurs (Fig. 3b). We validated these large-scale analyses with targeted measurements of CoQ- and CI-related proteins and CoQ precursors (Extended Data Fig. 5d–f). To ensure that these effects were not unique to HAP1 cells, we silenced *PYURF* expression in HEK293 cells and observed comparable loss of CoQ- and CI-related proteins without significant effects on their corresponding mRNA levels (Fig. 3e and Extended Data Fig. 5g). Notably, in humans, but not in mice, *PYURF* is bicistronic with *PIGY*[32]. To ensure that the observed effects were driven by PYURF, we silenced *Pyurf* in mouse C2C12 cells and found similar depletion of Coq5 and AF5 (Extended Data Fig. 5h).

In reanalysing our recent systematic analyses of mitochondrial protein–protein interactions[26], we found that PYURF interacted with only two proteins: AF5 and COQ5, which themselves interacted with AF8 and other COQ proteins, respectively (Fig. 3f). AF5, COQ5 and COQ3 are all members of the *S*-adenosylmethionine-dependent methyltransferase (SAM-MT) family (Fig. 3f), suggesting that PYURF selectively binds members of this protein class probably through its Trm112-like domain[33]. We confirmed these interactions by immunoprecipitation of tagged PYURF from HEK293 cells (Fig. 3g) and HAP1 mitochondria (Extended Data Fig. 5i and Supplementary Table 4), and we demonstrated direct binding between purified recombinant PYURF and AF5 using differential scanning fluorimetry (Fig. 3h and Extended Data Figs. 5j and 6). PYURF–AF5 binding was markedly reduced by introducing any of four mutations to conserved residues in the PYURF Trm112-like domain (Fig. 3h). PYURF did not induce SAM binding or catalytic activity of AF5, which is consistent with the documented role for Trm112 in stabilizing SAM-MT proteins without contributing to substrate binding or catalysis[33] (Extended Data Fig. 5k). To test the effect of PYURF on AF5 stability, we measured AF5 levels in HEK293 cells and found a marked reduction in its half-life after loss of *PYURF* expression (Extended Data Fig. 7a, b). Consistent with these observations, PYURF^{KO} cells exhibited a major loss of assembled CI as determined by blue native PAGE as well as diminished basal and uncoupled oxygen consumption rates (Fig. 3i, j and Extended Data Fig. 7c, d). Notably, other mitochondrial SAM-MTs in our study were not significantly affected by loss of *PYURF*. This selective targeting suggests a particular need to regulate these SAM-MTs, which is consistent with earlier findings that calibration of Coq5p levels in yeast is vital for cell health[34,35]. Collectively, these data support a model in which PYURF binds and stabilizes SAM-MTs in the interconnected CI and CoQ pathways (Fig. 3k). On the basis of these observations, we propose renaming PYURF as NDUFAFQ.

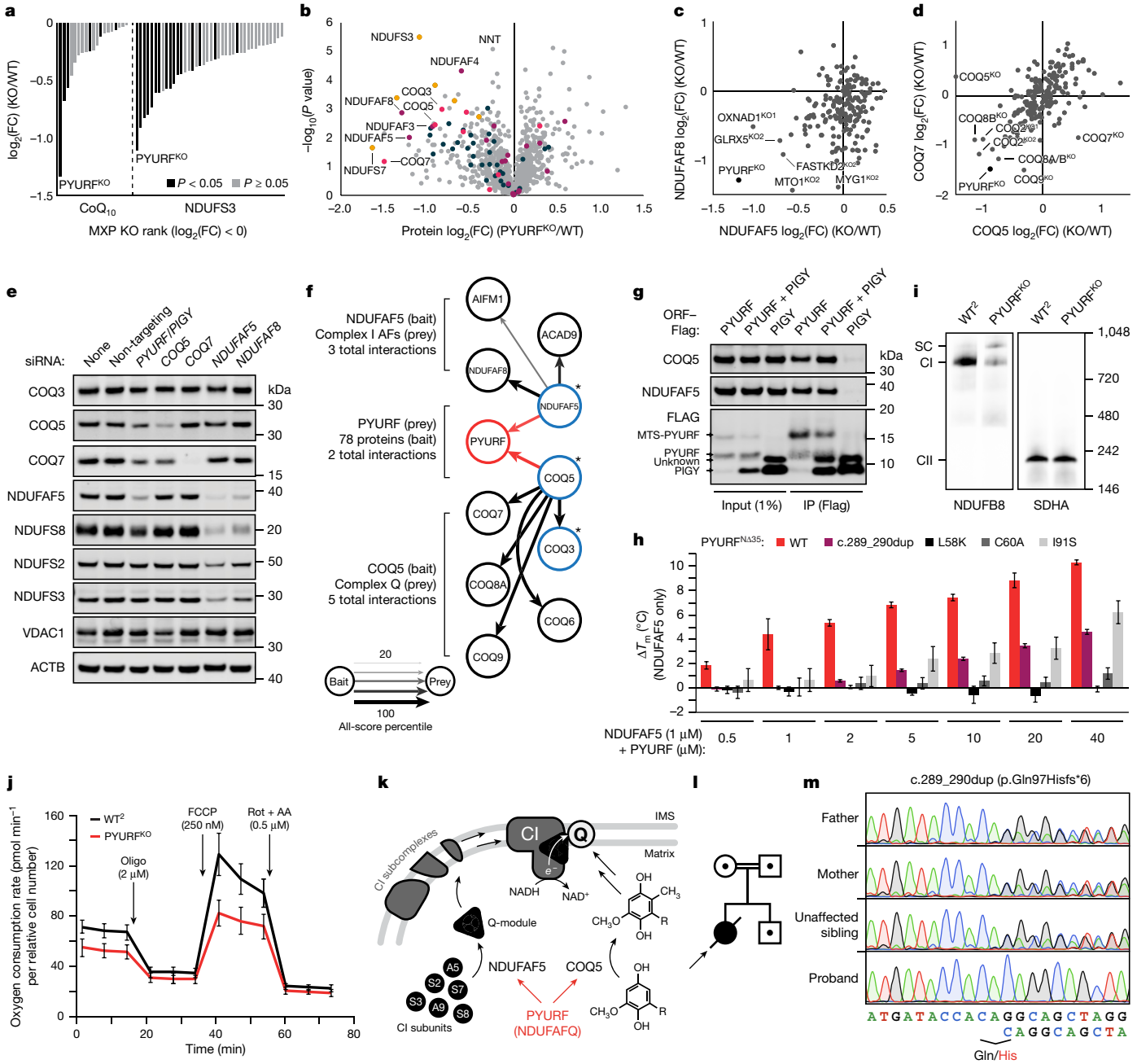


Fig. 3 | PYURF (NDUFAFQ) is a CoQ- and CI-related chaperone disrupted in human disease. a, Relative abundance of CoQ_{10} and NDUFS3 across MXP KO cell lines with $\log_2(FC) < 0$. **b**, Relative protein abundance in PYURF^{KO} cells compared with WT versus statistical significance with CoQ-related (COQ3–COQ9), CI, CI Q module and CI-AF proteins highlighted. **c**, **d**, Relative abundance of NDUFAF5 versus NDUFAF8 (**c**) and COQ5 versus COQ7 (**d**) compared with WT across all KO lines. In **a**–**d**, data reflect the mean from $n = 3$–4 replicates with two-sided Welch's *t*-test. **e**, CI, CI-AF and CoQ biosynthetic protein levels in HEK293 cells treated for 5 d with the indicated siRNAs, assessed by immunoblotting. **f**, Meta-analysis of protein–protein interaction data for NDUFAF5 and COQ5 (2 of the 78 bait proteins that interact with PYURF) and the related CI-AF and COQ proteins. Asterisks indicate CI SAM-MTs. **g**, COQ5 and NDUFAF5 levels in immunoprecipitates from HEK293 cells transfected to express PYURF–Flag, PYURF–Flag–PIGY–Flag and PIGY–Flag. IP, immunoprecipitation. **h**, Change in melting temperature of NDUFAF5 with increasing concentrations of WT PYURF, the c.289_290dup patient variant or point mutants relative to NDUFAF5 only, measured by differential scanning fluorimetry (mean ± s.d., $n = 3$). **i**, Levels of assembled CI (left) and CII (right) in HAP1 WT and PYURF^{KO} cells assessed by blue native PAGE and immunoblotting. SC, respiratory supercomplex. See Methods and Supplementary Table 7 for information about WT^2. **j**, Mitochondrial stress test profile showing cellular oxygen consumption rate normalized to relative cell number versus time for WT and PYURF^{KO} cells (mean ± s.d., $n = 10$–14). Oligo, oligomycin; FCCP, carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone; Rot + AA, rotenone + antimycin A. **k**, Model of PYURF function in coordinating the CI assembly and CoQ biosynthesis pathways. **l**, Pedigree of a consanguineous kindred and their offspring. An arrow indicates the deceased index case. Dots indicate carriers of the c.289_290dup *PYURF* variant. **m**, Sanger sequencing traces showing heterozygous carrier status in the father, mother and unaffected sibling as well as the homozygous *PYURF* variant in the proband. For western source data, see Supplementary Fig. 1.

Disruption of PYURF causes disease

Using the MITOMICS resource to establish functional connections between MXPs and known pathways can accelerate the molecular diagnosis of orphan diseases, which remains a prominent challenge in mitochondrial medicine. Given the strong connection between PYURF and CoQ/CI processes, we explored whether such unresolved cases of mitochondrial disease might harbour pathogenic *PYURF* variants. We identified a child born to consanguineous, first-cousin parents who presented at birth with profound metabolic acidosis (Fig. 3l). Clinical features included muscle hypotonia, failure to thrive, developmental delay, optic atrophy, persistently elevated lactate levels in the blood and cerebrospinal fluid, and abnormal cerebellar white matter and cerebellar atrophy on magnetic resonance imaging (Extended Data Fig. 7e and Supplementary Methods).

To determine the precise genetic aetiology, we performed unbiased whole-exome sequencing. We prioritized exonic (coding) or splice-site variants with a minor allele frequency (MAF) ≤ 0.01 (refs. [36–38]) that were predicted to be pathogenic according to in silico tools[39–41]. Our analysis of autosomal recessive variants in nuclear genes encoding mitochondrial[21] proteins (Extended Data Fig. 7f and Supplementary Methods) indicated a homozygous frameshift variant in the second exon of *PYURF* (GenBank, NM_032906.4) as the only conceivable candidate genetic cause: c.289_290dup (p.Gln97Hisfs*6); this variant was present in one allele (MAF = 0.000006287) in the Genome Aggregation Database (gnomAD)[38]. Sanger sequencing clearly demonstrated segregation with clinical disease in the family (Fig. 3m).

In the absence of cells from the patient, we engineered this mutation into HAP1 cells to test whether the encoded PYURF variant is sufficient for disruption of CoQ- and CI-related processes. Cells harbouring the mutation closely mirrored the PYURF^{KO} line, with substantial loss of CoQ- and CI-related proteins (Extended Data Fig. 7g). Finally, we purified the mutant version of PYURF, which exhibited markedly reduced binding affinity for AF5 (Fig. 3h). Although the requisite patient cell line to definitively assign pathogenicity was unavailable, impairment of CI and CoQ processes in cells harbouring this specific loss-of-function *PYURF* variant provides a molecular basis for the observed cellular phenotypes. This demonstrates that the MITOMICS resource provides assistance for de-orphanizing mitochondrial proteins and resolving the genetic causes for rare diseases.

Composite analyses link MXPs to function

Beyond performing targeted searches for proteins that fill specific gaps in knowledge using select molecular queries, as with PYURF above, the MITOMICS data enable systematic, unbiased analyses that can link MXPs to proteins of related function using the full complement of our molecular measurements.

To begin, we applied *t*-stochastic neighbour embedding (*t*-SNE), a technique well suited for the visualization of high-dimensional datasets. In this analysis, each molecule was represented by a vector of its *q*-adjusted relative differences across all KO lines (Supplementary Methods). We analysed the full *t*-SNE plot (Fig. 4a) using HDBSCAN[42], a spatial cluster detection method that identifies dense clusters against the background distribution of points; we further identified the closest HDBSCAN cluster for each background point. This analysis indicated tight clusters for OxPhos- and mitoribosome-related proteins (Fig. 4a, b and Extended Data Fig. 8a–c). Beyond the established proteins in these pathways, these clusters included aspartic acid, which is consistent with the established role of cellular respiration in supporting aspartate biosynthesis[43,44], a phosphatidylcholine species and a number of poorly characterized proteins. The latter included MXP C16orf91 and uncharacterized proteins C18orf21 and GTPBP8, among others, suggesting that these proteins are linked to these pathways in ways not yet appreciated.

We extended this guilt-by-association analysis across the full *t*-SNE plot by defining a conservative plot radius of one unit (Supplementary Methods) and recording the molecules that fell within this radius for each MXP. These data suggested a range of MXP functions, including roles for DHRS4 in bridging mitochondrial and peroxisomal lipid metabolism, C14orf159 in propionyl-CoA metabolism and various others (Extended Data Fig. 8d–k and Supplementary Table 5).

As a complementary approach, we systematically surveyed the MITOMICS dataset for outlier molecular changes for each gene KO in the study (Fig. 4c, d and Supplementary Table 6). Molecules (proteins, lipids or metabolites) whose abundance is changed substantially more in one gene KO than in all others probably have a functional relationship with that gene[12]. This analysis suggested multiple such gene–molecule relationships across our dataset (Fig. 4e, f), including a particularly strong connection between MXP RAB5IF (C20orf24) and TMCO1. Our proteomics data indicated that the abundance of TMCO1 was markedly and specifically diminished in the RAB5IF^{KO} lines, which we confirmed via western blots (Fig. 4g and Extended Data Fig. 9a, b, d). We further tested whether this relationship persisted in a distinct cell line (HEK293) by using siRNA to silence either *RAB5IF* or *TMCO1* expression. Again, loss of *RAB5IF* led to nearly complete loss of TMCO1 protein without affecting *TMCO1* mRNA levels (Fig. 4h, i). Moreover, the effect was reciprocal: silencing of *TMCO1* had similar effects on RAB5IF protein without affecting its mRNA levels (Fig. 4h, i). These results establish a clear mutual dependence for these two poorly characterized proteins, which holds true across hundreds of cell lines within the DepMap project[25] (Extended Data Fig. 9c) and further shows how different analytical tools can be applied to the MITOMICS resource to explore protein function.

RAB5IF pathogenic variants cause CFSMR

The biological functions of TMCO1 and RAB5IF remain nebulous. Recent studies have described TMCO1 as a mitochondrial protein[43], but also as a member of an endoplasmic reticulum (ER) translocon[44] or an ER channel that prevents Ca^{2+} stores from overfilling[45]. However, we observed no changes to Ca^{2+} release in response to thapsigargin in RAB5IF^{KO} HAP1 cells (Extended Data Fig. 9e) or in HeLa cells (Extended Data Fig. 9f, g). Although RAB5IF was recently annotated as a mitochondrial respiratory chain assembly factor[46], we did not observe a notable loss of respiratory chain proteins in its absence. Despite these unclear functions, it is well established that mutations in *TMCO1* cause cerebrofaciothoracic dysplasia, also referred to as craniofacial dysmorphism, skeletal anomalies and mental retardation syndrome (CFSMR; MIM 213980)[47]. Given the strong connection between TMCO1 and RAB5IF in our data, we theorized that mutations in *RAB5IF* might underlie unresolved cases of CFSMR.

In a previous study[47], the clinical features of five families with CFSMR were described; among these, disease in four families was associated with biallelic mutations in *TMCO1*. In the remaining family, the *TMCO1* locus was excluded by homozygosity mapping despite the clinical CFSMR diagnosis being fully consistent with mutation of this locus. We revisited these data and identified 11 distinct homozygosity regions, including the critical region on chromosome 20 where *RAB5IF* is located (Extended Data Fig. 9h). Sanger sequencing indicated a homozygous loss-of-function variant in the first exon, c.75G>A (p.Trp25*), in the affected individual (Fig. 4j and Extended Data Fig. 9i), which was poorly covered by previous exome sequencing data. This variant, which presumably leads to early truncation of the protein, was not found in gnomAD. Five unaffected individuals from the same family lacked this mutation in homozygosity; however, two of the individuals with cleft lip and/or palate were heterozygous for this mutation (Fig. 4j and Extended Data Fig. 9j). This suggests that heterozygosity for this *RAB5IF* variant may represent low-penetrant variation leading to cleft lip or palate, which is a component of CFSMR. Of the 11 reported families with *TMCO1*-related CFSMR, only one heterozygous individual had a

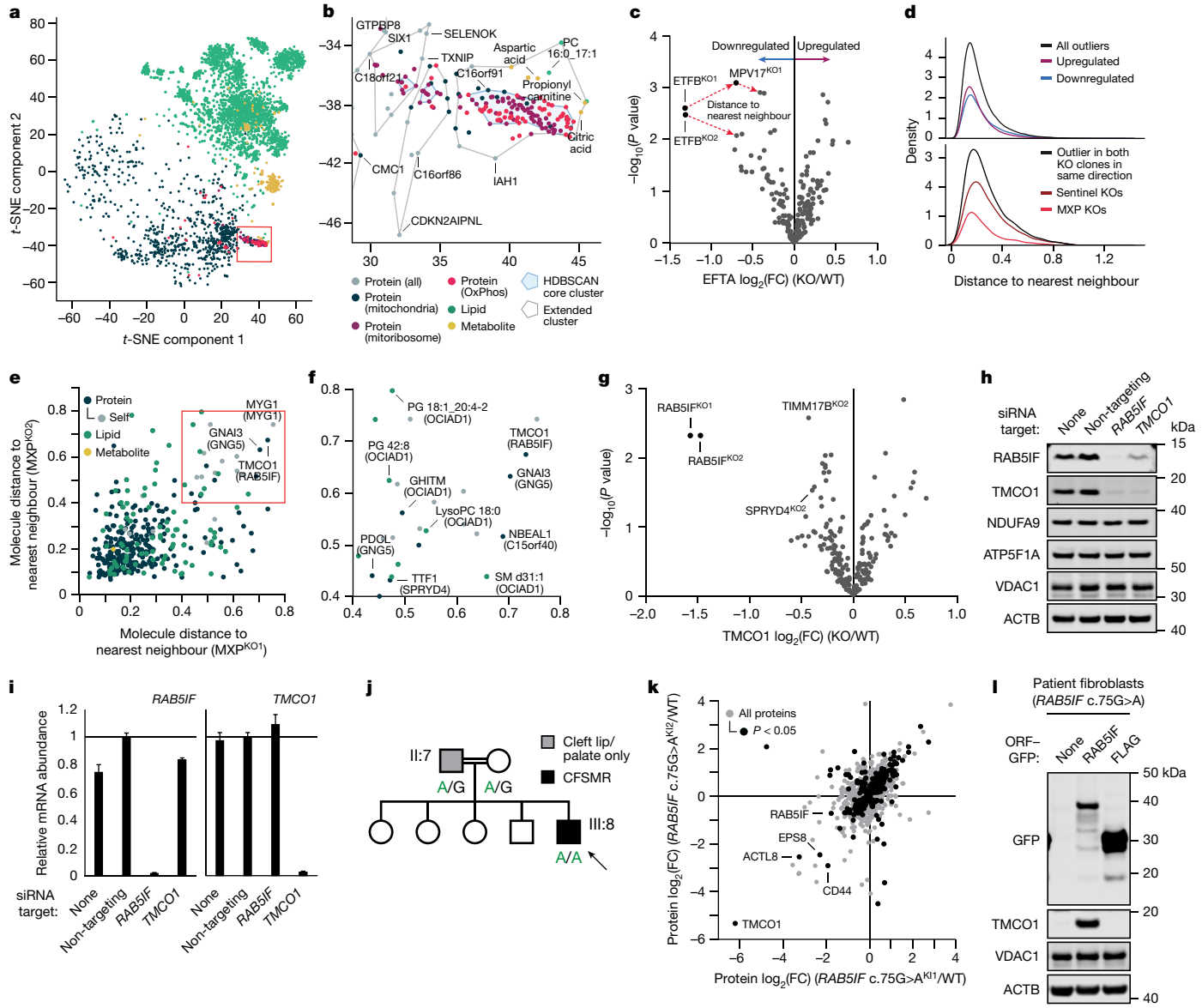


Fig. 4 | *t*-SNE and KO-specific phenotype analyses connect MXPs to mitochondrial functions. a, *t*-SNE analysis of the MITOMICS data including mean $\log_2$ fold change and associated multiomic *q* values from 191 conditions. **b**, Expanded view of the box in **a** showing clusters of known OxPhos and mitoribosome proteins and other biomolecules clustering with these pathways. **c**, Example of the KO-specific phenotype detection approach using ETFB^{KO1} and ETFB^{KO2} as outlier KOs for ETFA showing relative abundance versus statistical significance across all KO lines (mean from $n = 3–4$ replicates; two-sided Welch's *t*-test). **d**, Normalized density plots of nearest neighbour distances showing the number of KO-specific phenotypes identified as upregulated or downregulated and those identified in both KO clones and in the same direction of regulation for sentinel KOs and MXP KOs. **e**, Molecule distance to nearest neighbour in MXP KO clone 1 versus KO clone 2. **f**, Expanded view of the box in **e** showing molecules with large nearest neighbour distances detected for both MXP KO clones. The KO target is indicated in brackets. **g**, Relative abundance of TMCO1 versus statistical significance across all KO lines (mean from $n = 3–4$ replicates; two-sided Welch's *t*-test). **h**, **i**, Levels of the indicated proteins (**h**) and transcripts (**i**) in HEK293 cells treated with siRNA for 2 d as assessed by immunoblotting and qPCR (mean ± s.d., $n = 3$), respectively. **j**, Abbreviated pedigree of an affected family and associated phenotypes. An arrow indicates the proband. **k**, Relative protein abundance in HAP1 *RAB5IF* c.75G>A knock-in cells (clone 1 versus clone 2) compared with WT cells (mean from $n = 3$ replicates; two-sided Student's *t*-test). **l**, Level of TMCO1 in primary patient fibroblasts with the *RAB5IF* c.75G>A mutation transfected to express RAB5IF–GFP or Flag–GFP as assessed by immunoblotting. For western source data, see Supplementary Fig. 1.

cleft lip[47]. It is reasonable to speculate that cleft lip would lead to some selection pressure in individuals heterozygous for *RAB5IF* mutation, for whom cleft lip seems to be more common than in individuals heterozygous for *TMCO1* mutation. This concept is supported by gnomAD loss-of-function observed/expected upper bound fraction scores, which were low for *RAB5IF* (0.602) and high for *TMCO1* (1.362). Notably, the loss of TMCO1 observed in the RAB5IF^{KO} cell line was recapitulated in HAP1 cells engineered to express the mutation identified in the patient (Extended Data Fig. 9k). We performed full proteomic analysis of two independent clones harbouring this mutation, which showed high overall correlation between the lines; TMCO1 was the most affected protein in each case (Fig. 4k).

To further investigate the pathogenicity of this variant, we established a fibroblast cell culture from a patient biopsy and confirmed that these cells exhibited loss of RAB5IF and TMCO1 (Extended Data Fig. 9l). Reintroduction of wild-type (WT) RAB5IF–GFP into these cells by transfection resulted in a substantial increase in TMCO1 levels (Fig. 4l), unequivocally demonstrating that defective *RAB5IF* is the cause of this

patient's clinical phenotype. Overall, these data inextricably link two poorly understood proteins, RAB5IF and TMCO1, thereby providing a new route to explore the function of each protein and understand the underlying pathophysiology of a debilitating disorder. The partial localization of TMCO1 to the ER suggests that the connection between these proteins, and the aetiology of CFSMR, may involve inter-organelle interaction.

Systematic analyses of MITOMICS data

In this study, we demonstrate the power of the MITOMICS resource by identifying new proteins central to core mitochondrial pathways, presenting a range of new molecular hypotheses that will motivate further mechanistic investigations and providing molecular diagnoses for two unresolved human diseases. Beyond what is leveraged here, our dataset immediately enables a vast array of additional analyses. The MITOMICS website is equipped with built-in tools including outlier analysis, volcano plot profiles, molecule ranking across cell lines, scatter correlations, principal-component analysis, gene ontology analyses and *t*-SNE. Our full datasets can also be easily downloaded for custom analyses.

Moving forward, our deep, multidimensional dataset and analysis platform, particularly when integrated with other large-scale biological and patient registry data, promises to help advance multiple important areas of mitochondrial biology and medicine[48]. Specifically, our resource can accelerate the functional characterization of orphan proteins, facilitating the discovery of new disease-associated genes and improving understanding of mitochondrial disease pathomechanisms and genotype–phenotype correlations. Furthermore, it can be used to devise more robust diagnostics and therapeutics for the extensive array of human disorders underpinned by mitochondrial dysfunction.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04765-3.

1. Calvo, S. E. & Mootha, V. K. The mitochondrial proteome and human disease. *Annu. Rev. Genomics Hum. Genet.* **11**, 25–44 (2010).
2. Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* **491**, 374–383 (2012).
3. Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).
4. Pagliarini, D. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
5. Frazier, A. E., Thorburn, D. R. & Compton, A. G. Mitochondrial energy generation disorders: genes, mechanisms, and clues to pathology. *J. Biol. Chem.* **294**, 5386–5395 (2019).
6. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
7. Weissig, V. Drug development for the therapy of mitochondrial diseases. *Trends Mol. Med.* **26**, 40–57 (2020).
8. Sung, A. Y., Floyd, B. J. & Pagliarini, D. J. Systems biochemistry approaches to defining mitochondrial protein function. *Cell Metab.* **31**, 669–678 (2020).
9. Hebert, A. S. et al. The one hour yeast proteome. *Mol. Cell. Proteomics* **13**, 339–347 (2014).
10. Hutchins, P. D., Russell, J. D. & Coon, J. J. LipiDex: an integrated software package for high-confidence lipid identification. *Cell Syst.* **6**, 621–625 (2018).
11. Peterson, A. C. et al. Development of a GC/quadrupole-Orbitrap mass spectrometer, part I: design and characterization. *Anal. Chem.* **86**, 10036–10043 (2014).
12. Stefely, J. A. et al. Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nat. Biotechnol.* **34**, 1191–1197 (2016).
13. Phang, J. M., Liu, W., Hancock, C. & Christian, K. J. The proline regulatory axis and cancer. *Front. Oncol.* **2**, 60 (2012).
14. Tran, D. H. et al. Mitochondrial $NADP^+$ is essential for proline biosynthesis during cell growth. *Nat. Metab.* **3**, 571–585 (2021).
15. Zhu, J. et al. Mitochondrial NADP(H) generation is essential for proline biosynthesis. *Science* **372**, 968–972 (2021).
16. Xu, Y., Malhotra, A., Ren, M. & Schlame, M. The enzymatic function of *tafazzin*. *J. Biol. Chem.* **281**, 39217–39224 (2006).
17. Kerner, J. & Hoppel, C. Fatty acid import into mitochondria. *Biochim. Biophys. Acta* **1486**, 1–17 (2000).
18. Santel, A. & Fuller, M. T. Control of mitochondrial morphology by a human mitofusin. *J. Cell Sci.* **114**, 867–874 (2001).
19. Eramo, M. J., Lisnyak, V., Formosa, L. E. & Ryan, M. T. The 'mitochondrial contact site and cristae organising system' (MICOS) in health and human disease. *J. Biochem.* **167**, 243–255 (2020).
20. Niemi, N. M. et al. Pptc7 is an essential phosphatase for promoting mammalian mitochondrial metabolism and biogenesis. *Nat. Commun.* **10**, 3197 (2019).
21. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
22. Rhee, H. W. et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
23. Greber, B. J. & Ban, N. Structure and function of the mitochondrial ribosome. *Annu. Rev. Biochem.* **85**, 103–132 (2016).
24. Perez, Y. et al. *SLC30A9* mutation affecting intracellular zinc homeostasis causes a novel cerebro-renal syndrome. *Brain* **140**, 928–939 (2017).
25. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
26. Floyd, B. J. et al. Mitochondrial protein interaction mapping identifies regulators of respiratory chain function. *Mol. Cell* **63**, 621–632 (2016).
27. Gopalakrishna, S. et al. C6orf203 is an RNA-binding protein involved in mitochondrial protein synthesis. *Nucleic Acids Res.* **47**, 9386–9399 (2019).
28. Stefely, J. A. & Pagliarini, D. J. Biochemistry of mitochondrial coenzyme Q biosynthesis. *Trends Biochem. Sci.* **42**, 824–843 (2017).
29. Thompson, K. et al. Recent advances in understanding the molecular genetic basis of mitochondrial disease. *J. Inherit. Metab. Dis.* **43**, 36–50 (2020).
30. Arroyo, J. D. et al. A genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.* **24**, 875–885 (2016).
31. Mimaki, M., Wang, X., McKenzie, M., Thorburn, D. R. & Ryan, M. T. Understanding mitochondrial complex I assembly in health and disease. *Biochim. Biophys. Acta* **1817**, 851–862 (2012).
32. Murakami, Y. et al. The initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-Y, a seventh component. *Mol. Biol. Cell* **16**, 5236–5246 (2005).
33. Li, W., Shi, Y., Zhang, T., Ye, J. & Ding, J. Structural insight into human N6AMT1–Trm112 complex functioning as a protein methyltransferase. *Cell Discov.* **5**, 51 (2019).
34. Lapointe, C. P. et al. Multi-omics reveal specific targets of the RNA-binding protein Puf3p and its orchestration of mitochondrial biogenesis. *Cell Syst.* **6**, 125–135 (2017).
35. Veling, M. T. et al. Multi-omic mitoprotease profiling defines a role for Oct1p in coenzyme Q production. *Mol. Cell* **68**, 970–977 (2017).
36. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
37. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
38. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
39. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
40. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
41. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
42. Campello, R. J. G. B., Moulavi, D. & Sander, J. in *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science* vol. 7819 (eds. Pei, J. et al.) 160–172 (Springer, 2013).
43. Zhang, Z. et al. Molecular cloning, expression patterns and subcellular localization of porcine *TMCO1* gene. *Mol. Biol. Rep.* **37**, 1611–1618 (2010).
44. McGilvray, P. T. et al. An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020).
45. Wang, Q. C. et al. TMCO1 is an ER Ca^{2+} load-activated Ca^{2+} channel. *Cell* **165**, 1454–1466 (2016).
46. Moutaoufik, M. T. et al. Rewiring of the human mitochondrial interactome during neuronal reprogramming reveals regulators of the respirasome and neurogenesis. *iScience* **19**, 1114–1132 (2019).
47. Alanay, Y. et al. TMCO1 deficiency causes autosomal recessive cerebrofaciothoracic dysplasia. *Am. J. Med. Genet. A* **164A**, 291–304 (2014).
48. Schlieben, L. D. & Prokisch, H. The dimensions of primary mitochondrial disorders. *Front. Cell Dev. Biol.* **8**, 600079 (2020).

Reporting summary

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

Data availability

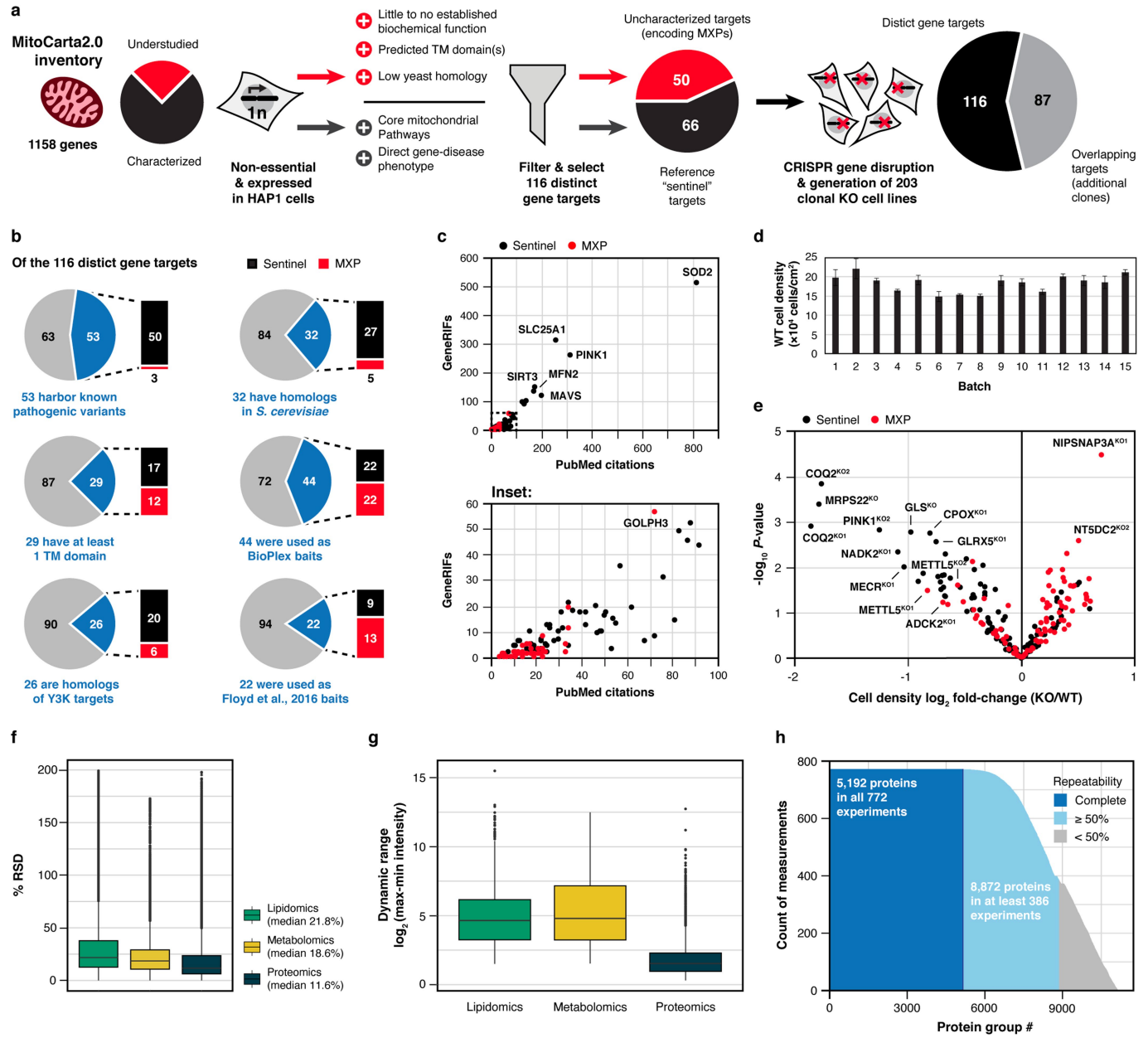
All associated mass spectrometry RAW files and search results were deposited into the MassIVE data repository under accession #MSV000086685. All associated code for data processing, analysis, and the companion webtool can be found in the GitHub repository https://github.com/coongroup/MITOMICS. Other relevant data are available from the corresponding authors upon reasonable request. HAP1 KO cell lines are available from Horizon Discovery.

Acknowledgements We thank J. Stefely for suggestions on experimental design, B. Floyd and N. Niemi for advice on knockout target selection, M. McDevitt and J. Stefely for guidance on lipid extraction, Z. Baker for data observations and A. Bartlett, M. Stefely, A. Sung and S. Hwang for graphic contributions. We thank Y. Murakami and T. Kinoshita (Research Institute for Microbial Diseases, Osaka University, Osaka, Japan) for providing the pRL-CMV-PreYF-PIG-YF expression constructs, J. Fan for advice on metabolite extraction and Y. Sancak for advice on cellular calcium analysis. This work was supported by NIH awards R35 GM131795 (D.J.P.), P41 GM108538 (J.J.C. and D.J.P.) and U54 AI117924 (Y.S. and M.C.); a UW2020 award (D.J.P. and J.J.C.); funds from the BJC Investigator Program (D.J.P.); and a grant from the Scientific and Technological Research Council of Turkey, 108S420 (N.A.A.) under the framework of ERA-NET for Research on Rare Disease, CRANIRARE Consortium (R07197KS). R.W.T. was supported by the Wellcome Centre for Mitochondrial Research (203105/Z/16/Z), the Medical Research Council International Centre for Genomic Medicine in Neuromuscular Disease (MR/S005021/1), the UK NIHR Biomedical Research Centre for Ageing and Age-related Disease award to the Newcastle upon Tyne Foundation Hospitals NHS Trust, the Mitochondrial Disease Patient Cohort (UK) (G0800674), the Lily Foundation, the Pathological Society and the NHS Highly Specialised Service for Rare Mitochondrial Disorders.

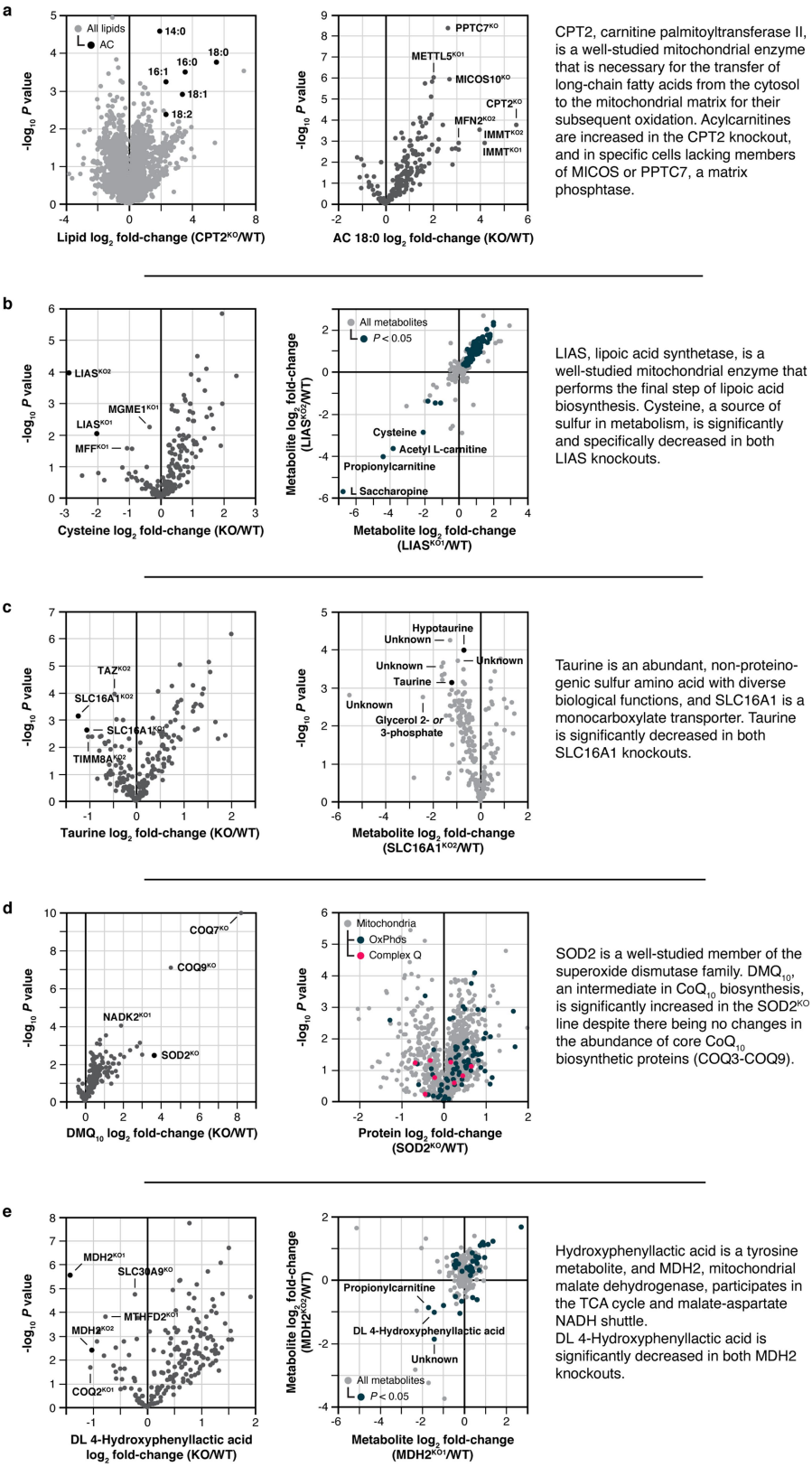
Author contributions J.W.R., E.S., J.J.C. and D.J.P. conceived the overall project and its design. J.W.R., M.M. and A.J. prepared samples and performed biochemical experiments. D.R.B. and N.W.K. designed and implemented the MITOMICS website. I.J.M., K.A.O., Y.S., P.D.H. and M.C. performed computational analyses including *t*-SNE. E.S., A.Z.S., P.D.H., S.R.P., V.L., A.S.H., C.E.V., M.J.P.R., M.S.W. and J.J.C. acquired and/or analysed the MS data. J.W.R., M.M., A.J. and D.J.P. analysed the biochemical data. A.C., A.P., J.R., Y.A., N.A.A. and R.W.T. were involved in the clinical care and molecular diagnosis of the patients described in this study. J.W.R. and E.S. led the preparation of final figures with assistance from most authors. D.J.P. led the preparation of the manuscript text with contributions from most authors. All authors critically reviewed and approved the final version of the manuscript.

Competing interests J.J.C. is a consultant for Thermo Fisher Scientific.

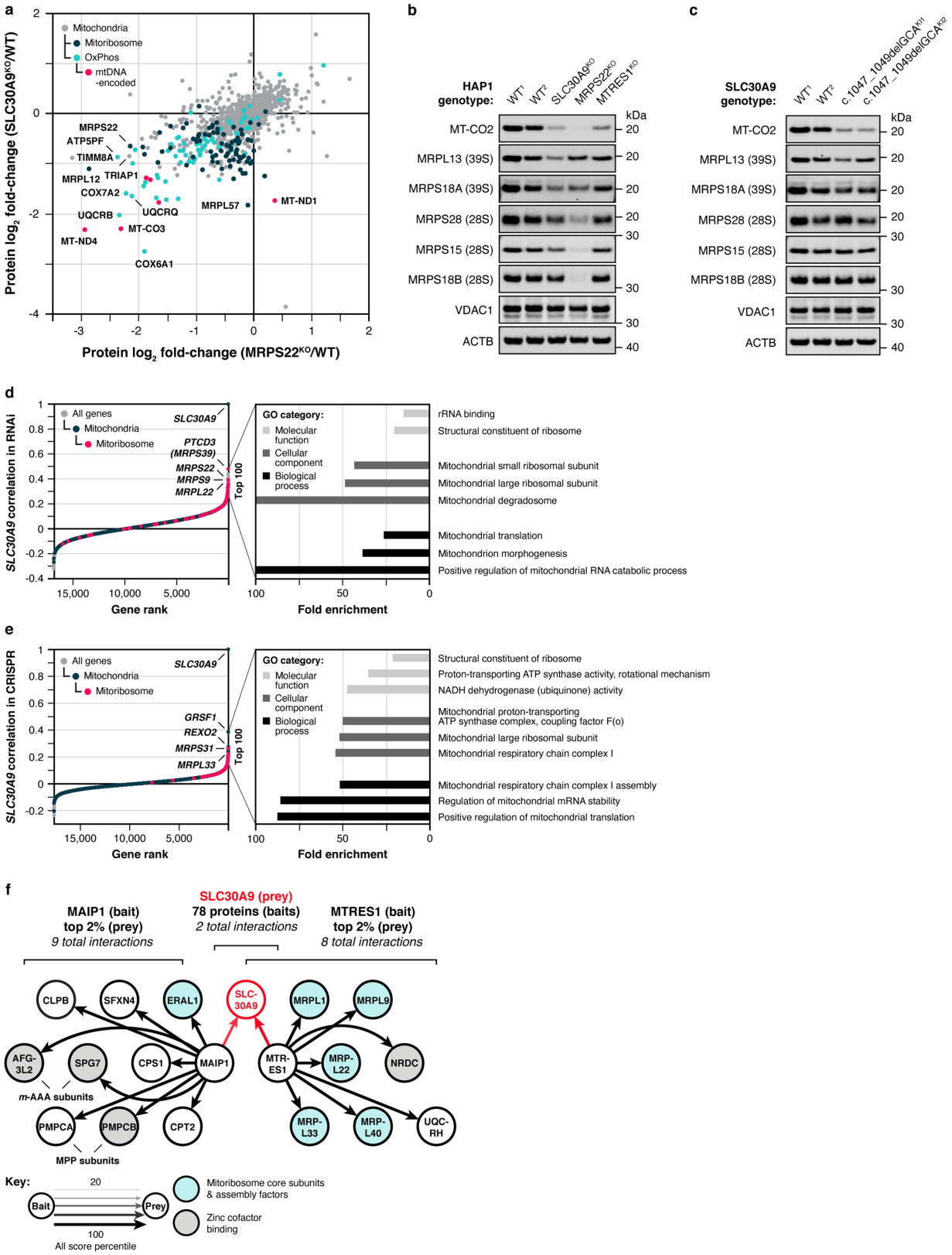
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04765-3.
Correspondence and requests for materials should be addressed to Joshua J. Coon or David J. Pagliarini.
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | MITOMICS design, target selection, and quality control. a, Criteria and filtering approach for knockout (KO) target selection. **b**, Features of each gene target and their representation in other select large-scale analyses at the time of selection. Metrics were taken from OMIM (omim.org), NCBI HomoloGene (ncbi.nlm.nih.gov/homologene), TMHMM (PMID: 11152613), The BioPlex Interactome (PMID: 28514442), The Y3K Project (PMID: 27669165), and Floyd et al., 2016 (PMID: 27499296). **c**, PubMed citations versus NCBI GeneRIFs (References Into Function) for each gene target at the time of selection. **d**, Cell density of wild-type (WT) reference cells across each analysis batch that were used to normalize cell growth measurements (mean ± s.d., n = 3-4). **e**, Relative cell density of each KO cell line compared to WT cells versus statistical significance (mean, n = 3-4, two-sided Welch's t-test). **f**, Distribution of % relative standard deviation (% RSD) of molecular abundance measurements made in 3-4 replicates of the KO cell lines. **g**, Distribution of $\log_2$ range in measured molecular abundances of all analytes calculated by subtracting the minimum observed intensity from the maximum observed intensity of each molecule across all cell lines. **h**, Histogram illustrating the count of quantitative measurements made per protein group across all analyzed samples.

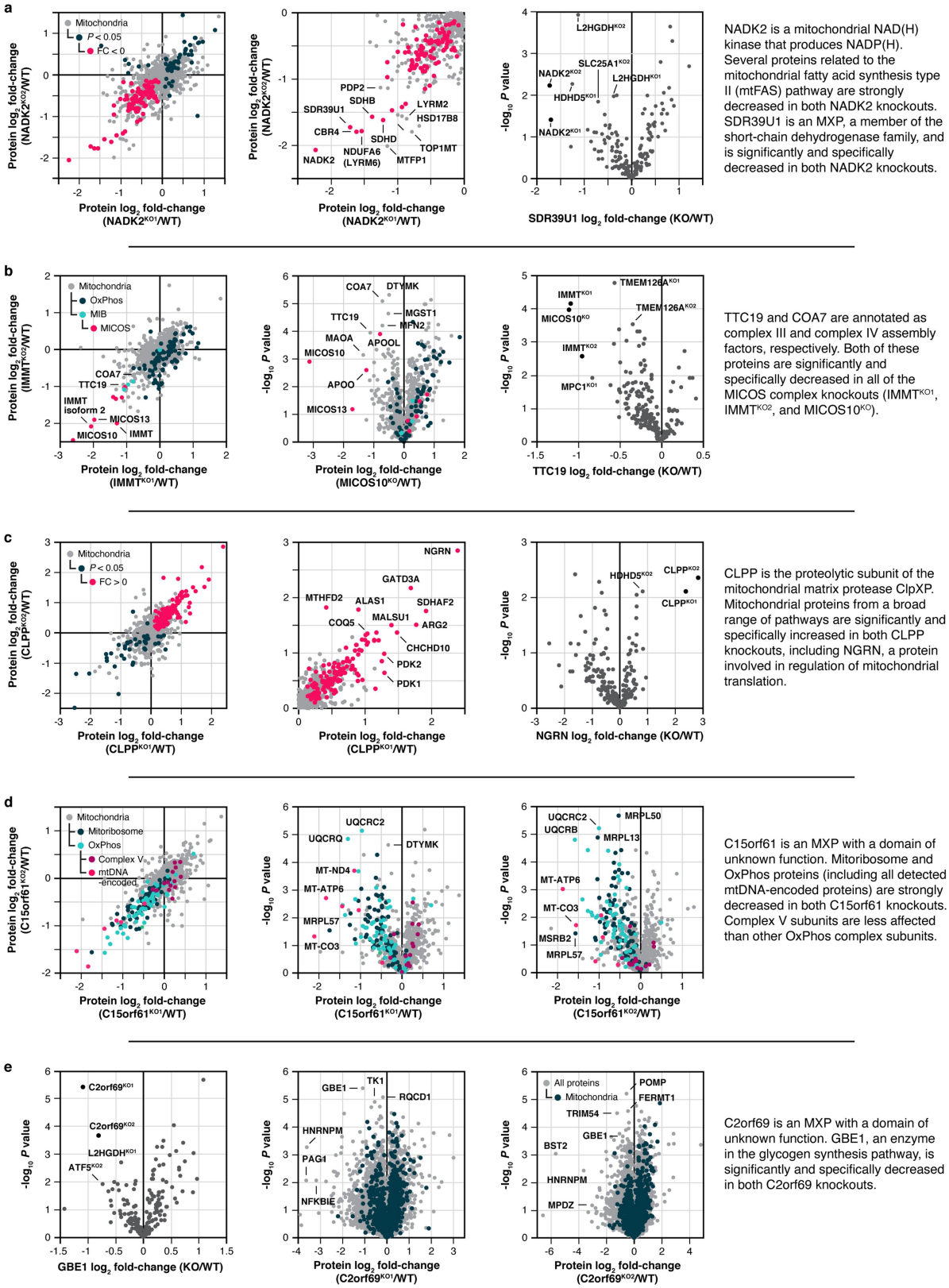


Extended Data Fig. 2 | MITOMICS profiles suggest new mitochondrial protein functions. a-e, Relative molecule abundance (protein, lipid, or metabolite) in the indicated KO compared to WT versus statistical significance, relative molecule abundance in KO versus KO compared to WT, or relative abundance of an individual molecule versus statistical significance across all KO lines with an accompanying summary of our observations. Data displayed as mean, $n = 3$-4, and two-sided Welch's t-test for all panels.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | SLC30A9 is necessary for mitoribosome and OxPhos protein integrity. a, Relative protein abundance in HAP1 MRPS22^{KO} cells versus SLC30A9^{KO} cells compared to WT cells with mitoribosome, OxPhos, and mtDNA-encoded proteins highlighted. Data displayed as mean, n = 3-4, and two-sided Welch's t-test. **b**, Level of mtDNA-encoded MT-CO2 and mitoribosome proteins in the indicated KO cell lines as assessed by immunoblotting. **c**, Level of the indicated proteins in HAP1 WT and *SLC30A9* c.1047_1049delGCA knock-in cells (two clones) as assessed by immunoblotting. **d**, **e**, Gene correlations with *SLC30A9* in DepMap project RNAi (**d**) and CRISPR (**e**) datasets with genes encoding mitochondrial and mitoribosome proteins highlighted and the top three GO annotations (most specific subclass term within a functional class) in each category for the 100 highest-ranking genes. **f**, Meta-analysis of protein-protein interaction data from Floyd et al., 2016 (PMID: 27499296) [26] displaying the two bait proteins (out of 78) that interacted with SLC30A9 and the top 2% of their interactors with mitoribosome core subunits, mitoribosome assembly factors, and zinc cofactor binding proteins (based on UniProt annotations) highlighted. For western source data, see Supplementary Figure 1.



NADK2 is a mitochondrial NAD(H) kinase that produces NADP(H). Several proteins related to the mitochondrial fatty acid synthesis type II (mtFAS) pathway are strongly decreased in both NADK2 knockouts. SDR39U1 is an MXP, a member of the short-chain dehydrogenase family, and is significantly and specifically decreased in both NADK2 knockouts.

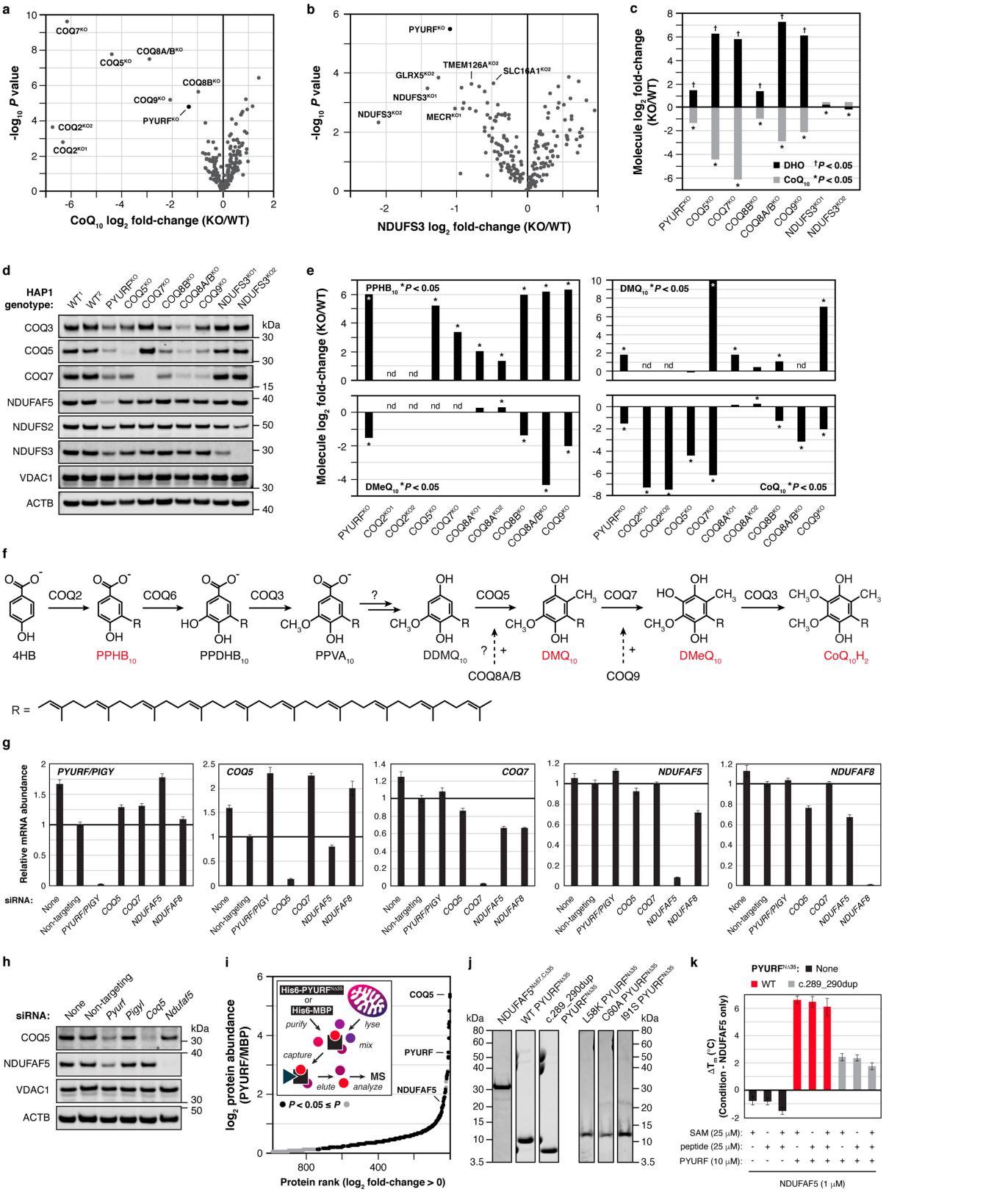
TTC19 and COA7 are annotated as complex III and complex IV assembly factors, respectively. Both of these proteins are significantly and specifically decreased in all of the MICOS complex knockouts ($IMMT^{KO1}$, $IMMT^{KO2}$, and $MICOS10^{KO}$).

CLPP is the proteolytic subunit of the mitochondrial matrix protease ClpXP. Mitochondrial proteins from a broad range of pathways are significantly and specifically increased in both CLPP knockouts, including NGRN, a protein involved in regulation of mitochondrial translation.

C15orf61 is an MXP with a domain of unknown function. Mitoribosome and OxPhos proteins (including all detected mtDNA-encoded proteins) are strongly decreased in both C15orf61 knockouts. Complex V subunits are less affected than other OxPhos complex subunits.

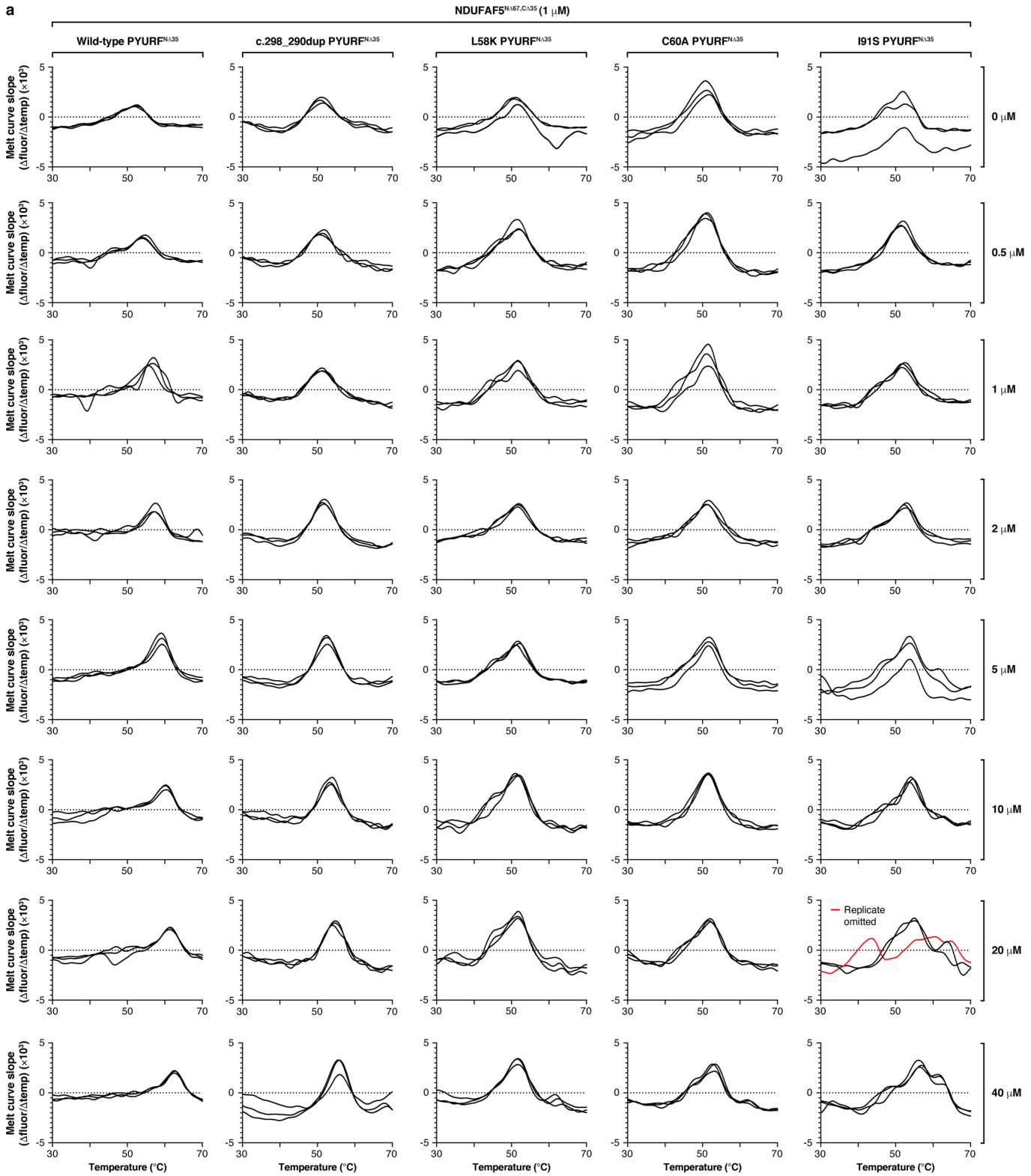
C2orf69 is an MXP with a domain of unknown function. GBE1, an enzyme in the glycogen synthesis pathway, is significantly and specifically decreased in both C2orf69 knockouts.

Extended Data Fig. 4 | Additional molecule-centric analyses suggest mitochondrial protein functions. a-e, Relative protein abundance in the indicated KO compared to WT versus statistical significance, relative protein abundance in KO versus KO compared to WT, or relative abundance of an individual protein versus statistical significance across all KO lines with an accompanying summary of our observations. Data displayed as mean, $n = 3$-4, and two-sided Welch's t-test for all panels.

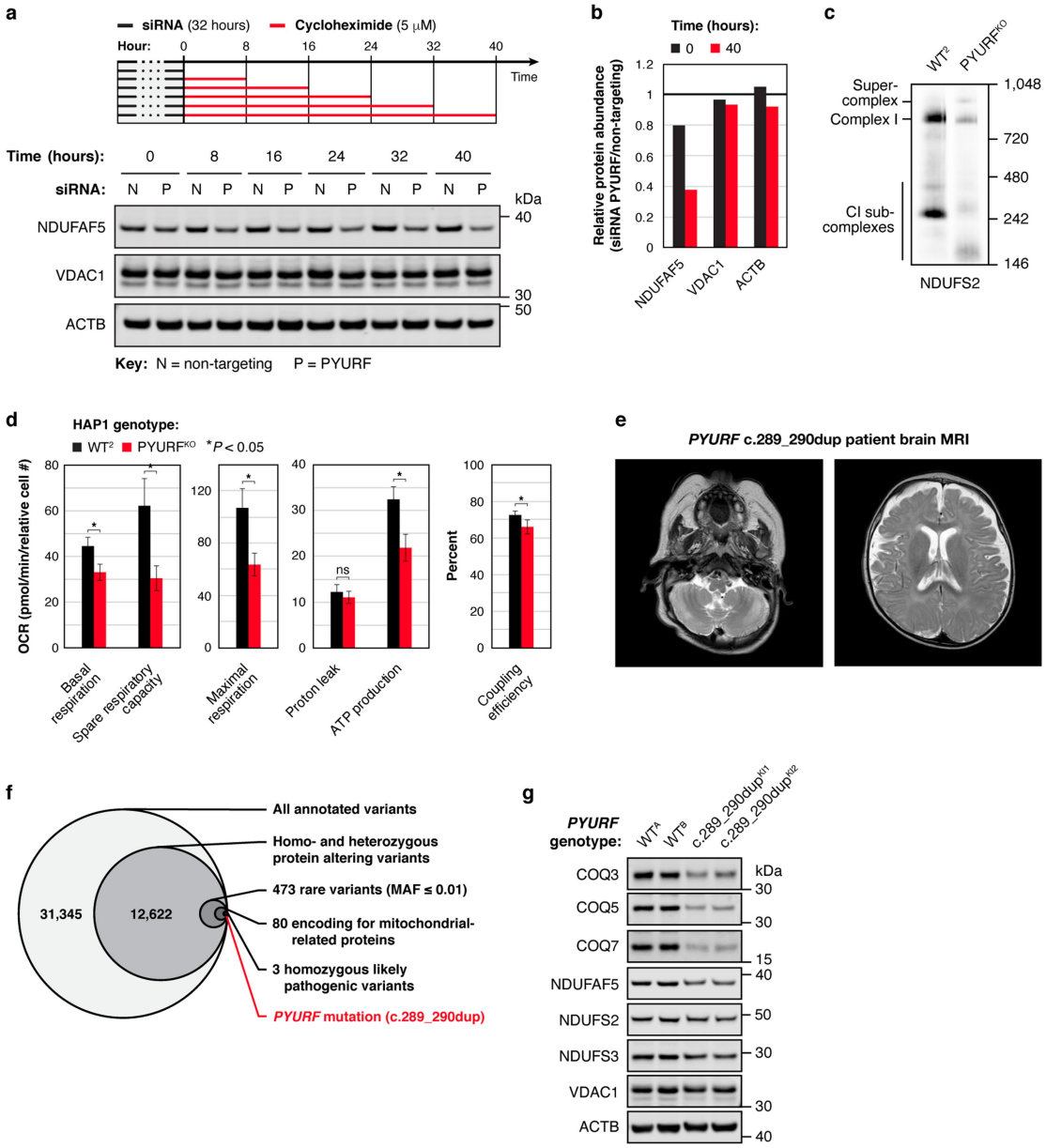


Extended Data Fig. 5 | See next page for caption.

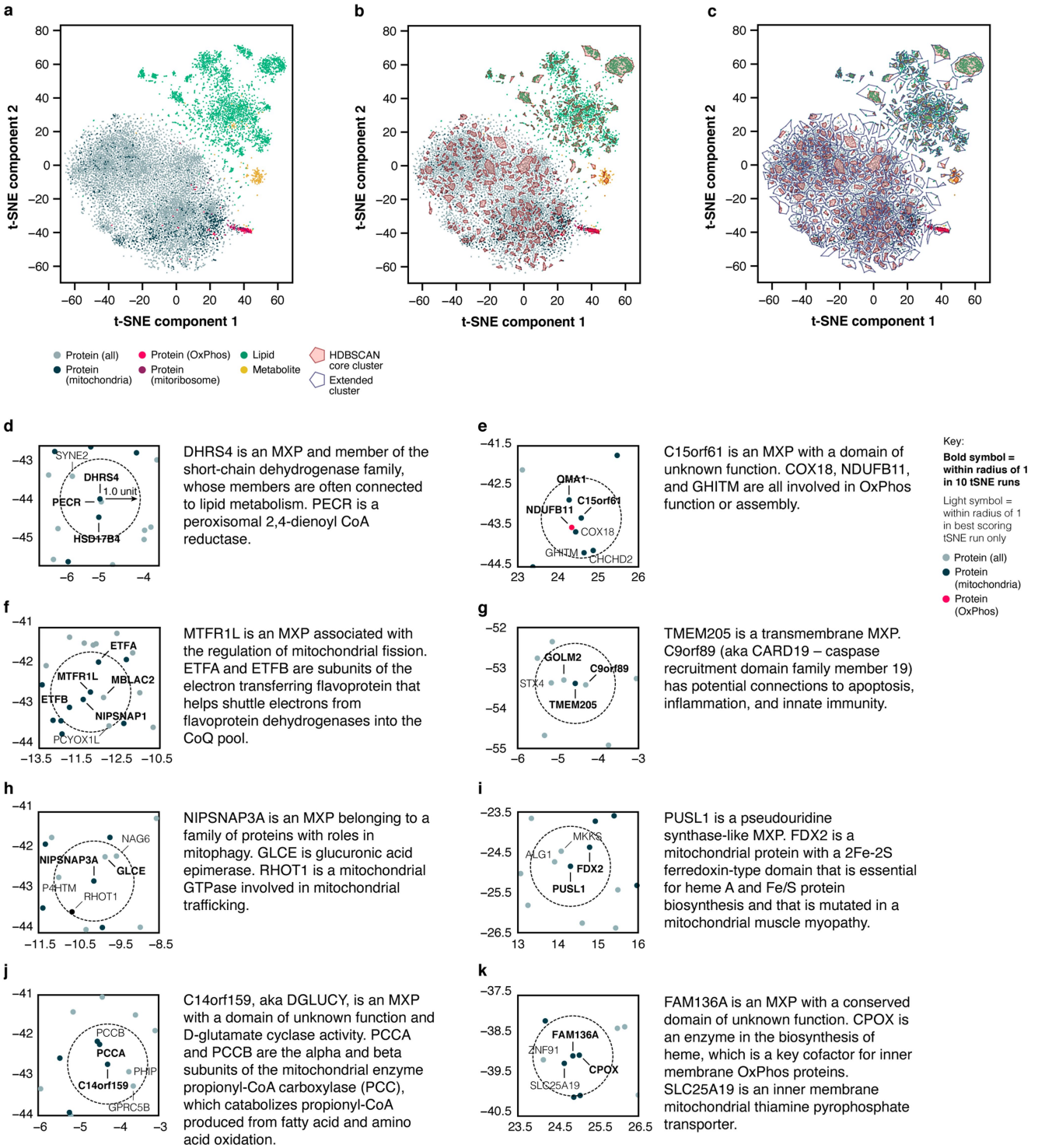
Extended Data Fig. 5 | PYURF (NDUFAFQ) is a CoQ- and CI-related chaperone. a, **b**, Relative abundance of CoQ_{10} (**a**) and NDUFS3 (**b**) versus statistical significance across all KO lines. **c**, Relative abundance of dihydroorotate (DHO) and CoQ_{10} in the indicated KO cell lines compared to WT cells. (**a**-**c**) Data displayed as mean, $n = 3$-4, and two-sided Welch's *t*-test. **d**, Level of complex I (CI), CI-assembly factor, and CoQ biosynthetic proteins in the indicated KO cell lines as assessed by immunoblotting. **e**, Relative abundance of CoQ_{10} and biosynthetic pathway intermediates analyzed via targeted LC-MS (mean, $n = 3$-4, two-sided Welch's *t*-test). PPHB, polyprenyl-hydroxybenzoate; DMQ, demethoxy-coenzyme Q; DMeQ, demethyl-coenzyme Q. **f**, CoQ biosynthesis pathway following polyisoprenoid tail attachment. Molecules quantified in (**e**) are indicated in red. 4-HB, 4-hydroxybenzoate; PPDHB, polyprenyl-dihydroxybenzoate; PPVA, polyprenyl-vanillic acid; DDMQ, demethoxy-demethyl-coenzyme Q. Supportive role for reactions is indicated by '+' symbol next to arrows. **g**, Level of the indicated transcripts in 293 cells treated with siRNA for five days as assessed by qPCR (mean ± s.d., $n = 3$). **h**, Level of COQ5 and NDUFAF5 in mouse C2C12 cells treated with the indicated siRNAs for five days as assessed by immunoblotting. **i**, Relative abundance of protein interactors for WT PYURF compared to maltose-binding protein (MBP) captured from a HAP1 mitochondrial lysate detected via immunoprecipitation (IP)-LC-MS/MS analysis (mean, $n = 3$, two-sided Student's *t*-test). **j**, Purity of NDUFAF5, WT PYURF, c.289_290dup patient variant, and point mutants analyzed via SDS-PAGE and Coomassie stain. **k**, Melting temperature of NDUFAF5 with combinations of WT PYURF or c.289_290dup mutant PYURF, peptide, and S-adenosylmethionine (SAM) compared to NDUFAF5 only as measured by differential scanning fluorimetry (mean ± s.d., $n = 3$). For western and gel source data, see Supplementary Figure 1.



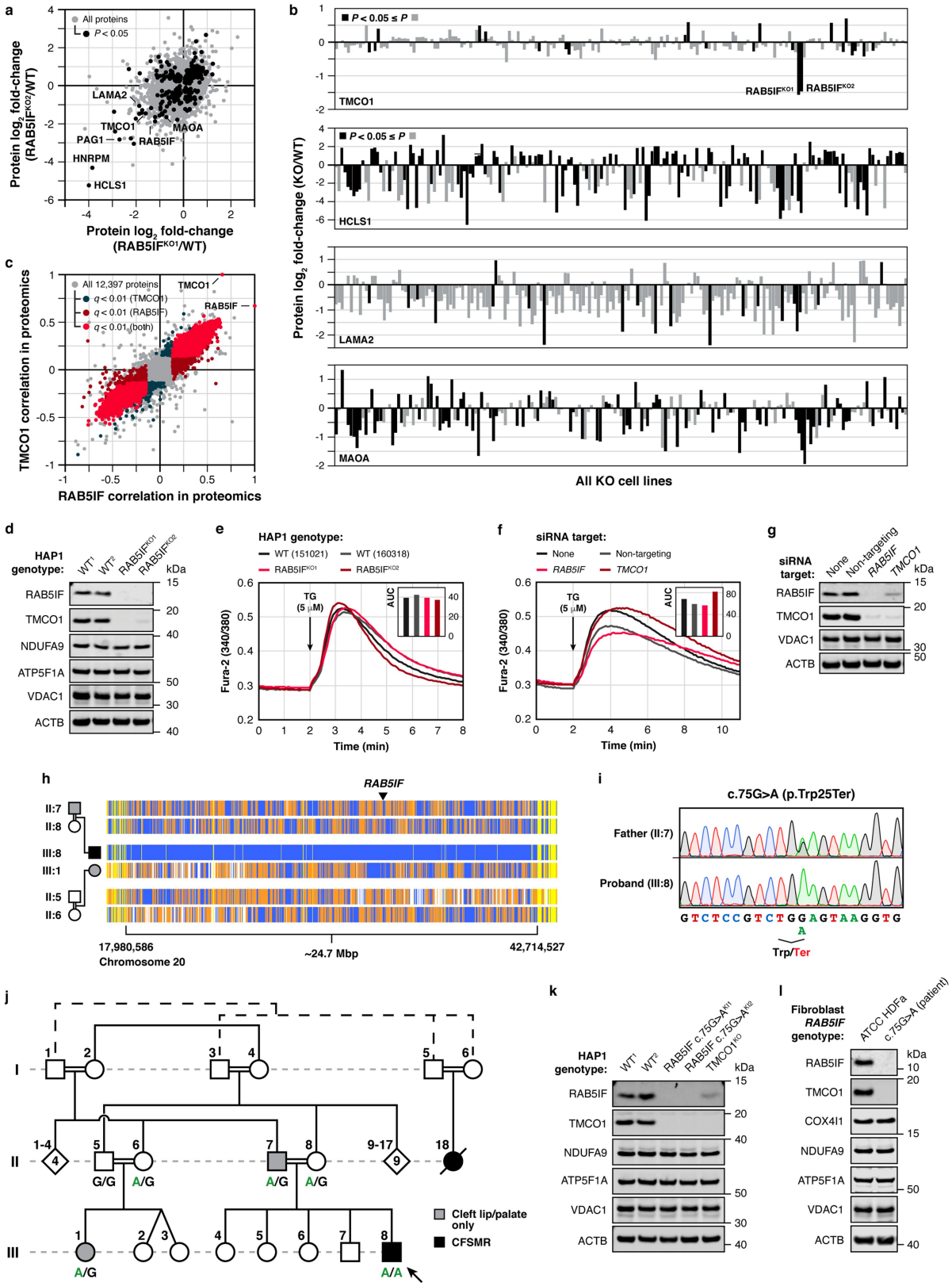
Extended Data Fig. 6 | Mutations to PYURF disrupt binding and stability of NDUFAF5. a, First-derivative plots of the differential scanning fluorimetry analysis in Fig. 3h ($n = 3$).



Extended Data Fig. 7 | PYURF (NDUFAFQ) is important for mitochondrial function and disrupted in human disease. a, **b**, Level of the indicated proteins in 293 cells during a cyclohexamide chase experiment following PYURF knockdown (**a**), and quantification of the immunoblot data (**b**). **c**, Level of assembled complex I in HAP1 WT and PYURF^{KO} cells as assessed by BN-PAGE and immunoblotting. **d**, Parameters of mitochondrial function for WT and PYURF^{KO} cells calculated from the mitochondrial stress test assay in Fig. 3j (mean ± s.d., n = 10-14, two-sided Student's t-test). **e**, Brian MRI of the *PYURF* case demonstrating increased extra axial CSF spaces, cystic high signal cerebellar white-matter, cerebellar atrophy, and decreased myelination in the internal capsule. **f**, Whole exome sequencing analysis and filtering for rare, autosomal recessive variants in nuclear genes encoding mitochondrial proteins. MAF, minor allele frequency. **g**, Level of the indicated proteins in HAP1 unedited *PYURF* WT cells and *PYURF* c.289_290dup knock-in cells (two clones each) as assessed by immunoblotting. For western source data, see Supplementary Figure 1.



Extended Data Fig. 8 | t-SNE analyses suggest functions for MXPs. a-c, t-SNE analysis of the MITOMICS data (mean $\log_2$ fold-changes and associated multi-ome q-values from 191 conditions) displaying all molecules (**a**), core clusters (**b**), and extended clusters (**c**). **d-k**, Analysis of the MXP KO targets in the t-SNE plot to identify proteins that fall within their close proximity (one unit radius) with accompanying summaries of our observations.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | RAB5IF is mutated in CFSMR. a, Relative protein abundance in RAB5IF^{KO1} cells versus RAB5IF^{KO2} cells compared to WT. **b**, Relative abundance of indicated proteins across all KO lines. (**a**, **b)** Data displayed as mean, n = 3-4, and two-sided Welch's t-test. **c**, Protein correlations with RAB5IF versus protein correlations with TMCO1 in the DepMap proteomics dataset. **d**, Level of the indicated proteins in HAP1 WT and RAB5IF KO cells assessed by immunoblotting. **e**, **f**, Fura-2 fluorescence (mean, n = 3) following thapsigargin (TG) treatment in WT and RAB5IF KO cells (**e**), and HeLa cells treated with indicated siRNAs for three days (**f**) with area under the curve (AUC) measurements. **g**, RAB5IF and TMCO1 levels in HeLa cells treated with indicated siRNAs for three days assessed by immunoblotting. **h**, Homozygosity mapping of a 24.7 Mbp candidate region in chromosome 20p11.23-q13.12. Homozygous genotypes in the index (III:8) shown in blue. In other individuals, identical homozygous genotypes are also in blue, whereas contrasting homozygous genotypes are in white. Heterozygous genotypes are orange, while non-informative genotypes resulting from heterozygous SNPs in parent-child trios are yellow. Note that the index is homozygous for the candidate region, while the cousin (III:1) is heterozygous for the entire region. **i**, Sanger sequencing showing the c.75G > A (p.Trp25*) mutation as homozygous in the index (III:8) and heterozygous in his father (II:7). **j**, Pedigree of affected family with genotypes and associated phenotypes. Note that individuals II:7 and III:1 only have cleft lip and/or palate without other clinical features of CFSMR and are heterozygous for the *RAB5IF* variant. **k**, **l**, Indicated protein levels in HAP1 WT and *RAB5IF* c.75G > A knock-in cells (2 clones) (**k**), and normal adult human primary dermal fibroblasts (HDFa) and primary patient fibroblasts with the *RAB5IF* c.75G > A mutation (**l**) assessed by immunoblotting. For western source data, see Supplementary Figure 1.

nature research

Corresponding author(s): David J. Pagliarini
Joshua J. Coon

Last updated by author(s): Jan 19, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	To collect the mass spectrometric data the following commercial software was used: Xcalibur (Thermo Fisher Scientific, v.4.0 and above) and Thermo Scientific SII for Xcalibur (Thermo Fisher Scientific, v.1.3). Quantitative PCR and DSF data was collected using Applied Biosystems QuantStudio Real-Time PCR Software (version 1.2). Immunoblot data was collected using LI-COR Image Studio Software (version 5.2.5) or BioRad Image Lab Touch Software (version 2.4.0.03). Fura-2 fluorescence measurements were collected using BioTek Gen5 Software (version 1.11.5).
Data analysis	Mass spectrometric data were searched using MaxQuant (v.1.6.0.13 and 1.6.10.43), LipiDex (v1.0.2), Y3K-GC-Quantitation-Software (v.1.0), and TraceFinder (Thermo Fisher Scientific, v.4.0). The data were processed using ComBat R package sva (v3.30.1) and tsne R package (v. 0.1-3). All custom computer code could be found in the GitHub repository https://github.com/coongroup/MITOMICS. Original code used for creation of the webtool MITOMICS.app could be found at https://github.com/coongroup/Argonaut. Software used for data analyses include R version 4.0.2; R packages (version): drake (7.12.5), here (0.1), jsonlite (1.6.1), magrittr (1.5), qvalue (2.20.0), tidyverse (1.3.0), tsne (0.1.3). Code for computing mean log-fold-changes, p-values, q-values, and performing our tSNE analysis is available at https://github.com/coongroup/MITOMICS, licensed under the MIT License. Quantitative PCR data was analyzed using Applied Biosystems QuantStudio Real-Time PCR Software (version 1.2). DSF data was analyzed using Applied Biosystems Protein Thermal Shift Software (version 1.3). Immunoblot data was analyzed using LI-COR Image Studio Software (version 5.2.5) or BioRad Image Lab Touch Software (version 2.4.0.03).

Fura-2 fluorescence measurements were analyzed using GraphPad Prism 8 (version 8.4.3).

Cellular oxygen consumption measurements were analyzed using Agilent Seahorse Wave Desktop software (version 2.6.1).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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

Main Figures 1B, 2A-F, 3A-D, 4A-G, 4K, and Extended Data Figures 2A-E, 3A, 4A-E, 5A-C, 5E, 5I, 8A-K, and 9A-B have associated raw MS data.

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Identification of PYURF and RAB5IF as causative genes for a mitochondrial disorder and CFSMR, respectively, did not involve randomized controlled studies. Select Sanger sequencing data will be included in the published manuscript, and full Sanger sequencing data from all family members will be available from the authors upon request.
Data exclusions	Metabolomics samples analyzed between 1/3/2019 4:00pm and 1/4/2019 12:00 pm were excluded from analysis due to instrument failure.
Replication	All Sanger sequencing data are replicable. All other attempts at experiment replication were successful. Three or more replicates were used in each experiment.
Randomization	Randomization was not relevant to our study as our experiments were done quantitatively and data acquisition was performed by machines, mitigating investigator bias.
Blinding	As above, blinding was not relevant to our study as all measurements were derived from automated measurements or computational analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

COQ3 (Sigma, HPA031274-100UL); COQ5 (Thermo, PA5-61988); COQ7 (Proteintech, 15083-1-AP); NDUFAF5 (Abcam, ab192235); NDUFS8 (Abcam, ab170936); NDUFS2 (Abcam, ab192022); NDUFS3 (Abcam, ab14711); VDAC1 (Abcam, ab18988); ACTB (Abcam, ab8224); FLAG (Sigma, F1804-5MG); RAB5IF (Thermo, PA5-43332); TMCO1 (Proteintech, 27757-1-AP); NDUFA9 (Abcam, ab14713); ATP5F1A (Abcam, ab14748); MT-CO2 (Abcam, ab91317); MRPL13 (Proteintech, 16241-1-AP); MRPS18A (Thermo, PA5-113436); MRPS28 (Proteintech, 16378-1-AP); MRPS15 (Proteintech, 17006-1-AP); MRPS18B (Abcam, ab191891); NDUFB8 (Abcam, ab110242); SDHA (Abcam, ab14715); COX4I1 (Abcam, ab14744); GFP (Santa Cruz, sc-9996)

Validation

Antibodies for COQ3, COQ5, and COQ7 were validated for WB in this study through direct knockout and/or knockdown of CoQ-related proteins known to affect the abundance of the corresponding protein. NDUFAF5, NDUFS8, NDUFS2, NDUFS3, NDUFB8, COX4I1, MT-CO2, MRLP13, MRPS18A, MRPS28, MRPS15, MRPS18B were validated for WB in this study through positive control experiments including direct knockout and/or knockdown of complex I subunits, OxPhos assembly factors, and/or mitoribosome subunits known to affect the abundance of the protein. Antibodies for RAB5IF and TMCO1 were validated for WB in this study through knockout and/or knockdown of the corresponding protein. SDHA, ATP5F1A, VDAC1, and ACTB were chosen as loading controls based on the manufacturer's website, strong literature support, and protein products at expected molecular weights. The antibodies for FLAG and GFP were validated for immunoassays based on the manufacturer's website and strong literature support (e.g., PMID: 27499296).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

WT and KO HAP1 cells (Horizon Discovery), C2C12 (Sigma, 91031101-1VL), HeLa (ATCC, CCL-2), 293 cells (Pagliarini Lab, PMID: 27499296)

Authentication

HAP1, C2C12, and HeLa cells were authenticated by the commercial source. 293 cells were not authenticated.

Mycoplasma contamination

HAP1, C2C12, and HeLa cells were negative for mycoplasma contamination as tested by the commercial source. 293 cells were not tested.

Commonly misidentified lines
(See ICLAC register)

This study did not include the use of any commonly misidentified cell lines.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Our manuscript describes two individuals from unrelated families with candidate pathogenic variants in the PYURF gene and the RAB5IF gene, identified by clinical diagnostic investigations. A single family with CFSMR that was not associated with TMCO1 for which clinical features has already been published (Alanay et al., 2014 — see manuscript references) were included in this study.

Recruitment

The patient with the PYURF variant was investigated by the UK-based authors (Birmingham and Newcastle upon Tyne) by unbiased whole exome sequencing on account of the clinical presentation of the affected child which strongly suggested a mitochondrial or metabolic disorder. Recruitment of all six available members of the CFSMR family were recruited as described previously in Alanay et al., 2014.

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 (e.g., Hacettepe University Medical Faculty Clinical Research Local Ethics Board (Since 2012; Hacettepe University Non-Interventional Clinical Research Ethics Board, goetik@hacettepe.edu.tr). WES was performed following referral to the NHS Highly Specialised Mitochondrial Services, Newcastle upon Tyne.

Note that full information on the approval of the study protocol must also be provided in the manuscript.