Accessory subunits are integral for assembly and function of human mitochondrial complex I

David A. Stroud[1]*, Elliot E. Surgenor[1], Luke E. Formosa[1,2], Boris Reljic[2]†, Ann E. Frazier[3,4], Marris G. Dibley[1], Laura D. Osellame[1], Tegan Stait[3], Traude H. Beilharz[1], David R. Thorburn[3-5], Agus Salim[6], Michael T. Ryan[1]*

[1]Department of Biochemistry and Molecular Biology, Monash Biomedicine Discovery Institute, Monash University, 3800, Melbourne, Australia.

[2]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University 3086, Melbourne, Australia.

[3]Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne 3052, Australia

[4]Department of Pediatrics, University of Melbourne, Melbourne 3052, Australia.

[5]Victorian Clinical Genetics Services, Royal Children's Hospital 3052, Melbourne, Australia.

[6]Department of Mathematics and Statistics, La Trobe University 3086, Melbourne Australia.

*Correspondence to: d.stroud@monash.edu and michael.ryan@monash.edu

†Current address: Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria 3052, Australia

Complex I (NADH:ubiquinone oxidoreductase) is the first enzyme of the mitochondrial respiratory chain (RC) and is composed of 44 different subunits in humans, making it one of the largest known multi-subunit membrane protein complexes[1]. Complex I exists in supercomplex forms with RC complexes III and IV, which are together required for

the generation of a transmembrane proton gradient used for the synthesis of ATP[2]. Complex I is also a major source of damaging reactive oxygen species and its dysfunction is associated with mitochondrial disease, Parkinson's disease and aging[3-5]. Bacterial and human complex I share 14 core subunits essential for enzymatic function, however the role and requirement of the remaining 31 human accessory subunits is unclear[1,6]. The incorporation of accessory subunits into the complex increases the cellular energetic cost and has necessitated the involvement of numerous assembly factors for complex I biogenesis. We used gene-editing to generate human knockout cell lines for each accessory subunit. We found that 25 subunits are strictly required for assembly of a functional complex and one subunit is essential for cell viability. Quantitative proteomic analysis of cell lines revealed that loss of each subunit affects the stability of other subunits residing in the same structural module. Analysis of proteomic changes following loss of specific modules revealed ATP5SL and DMAC1 as novel factors required for assembly of the distal portion of the complex I membrane arm. Our results demonstrate the broad importance of accessory subunits in the structure and function of human complex I. Coupling gene-editing technology with proteomics represents a powerful tool for dissecting large multi-subunit complexes and enabling the study of complex dysfunction at a cellular level.

Mitochondrial complex I is a boot-shaped structure of ~1MDa with a hydrophilic matrix arm and a hydrophobic membrane arm[7-9]. These arms are assembled via intermediate modules through transient association with assembly factors[10]. The N-module at the tip of the matrix arm is involved in oxidation of NADH while the Q-module bridges the matrix and membrane arms and is involved in transfer of electrons along Fe-S clusters to ubiquinone. With the reduction of ubiquinone, four protons are pumped across the inner-membrane (IM) into the intermembrane space (IMS). The core structure of the membrane arm is defined by 7 subunits

encoded by mitochondrial DNA (mtDNA); ND1 at the base of the Q-module, followed by ND3, 6, and 4L, and the antiporter-like subunits ND2, ND4 and ND5[8,9]. The mechanisms of NADH oxidation and proton pumping are conserved from bacteria to humans, with 14 core (including the 7 mtDNA-encoded) subunits performing these roles[11].

To investigate the importance of the 31 accessory subunits, we used TALEN and CRISPR/Cas9 gene-editing tools to disrupt their genes in human HEK293T cells (Supplementary Table 1). Of the knockout (KO) lines generated, 24 were unable to grow on galactose-containing media indicating mitochondrial respiration defects (Fig. 1). Blue-native (BN)-PAGE and immunoblot analysis for subunits NDUFA9, NDUFA13 and NDUFB11, (located in different regions of the complex) revealed that loss of an individual accessory subunit often disrupted assembly of complex I (Fig. 1). Analysis of the supercomplex was also disrupted in the same cell lines (Extended Data Fig. 1) while assembly of complexes III and IV was not affected (Extended Data Fig. 2). For cell lines still capable of growth on galactose, a complex was present that did not markedly differ from the migration of mature complex I (Fig. 1, lanes 2-6). Other cell lines showed different subcomplexes including one that migrated slightly faster than complex I, consistent with loss of the N-module[12] (Fig. 1, lanes 5-8, marked with ‡).

In contrast to all other subunits, we found NDUFAB1 to be essential for cell viability (Extended Data Fig. 3a-c). NDUFAB1 is unique as it is the only subunit with a 2:1 stoichiometry within the complex, where it binds LYR motifs present in NDUFA6 and NDUFB9[7]. NDUFAB1 is also the mitochondrial acyl carrier protein[13] and associates with proteins involved in fatty-acid synthesis (LIPT2) and other proteins (Extended Data Fig. 3d; Supplementary Table 2) including LYRM7 that promotes biogenesis of the complex III Rieske subunit (UQCRFS1). An NDUFAB1^{KO} cell line was generated by complementing cells with the yeast mitochondrial acyl carrier protein (yACP1) (Extended Data Fig. 3e,

Supplementary Table 1). Since NDUFAB1^{KO} lacks assembled complex I and dies in galactose media (Fig. 1; Extended Data Fig. 3c, e), the essential role of NDUFAB1 is independent of complex I.

We selected a subset of representative KO lines for further analysis (Fig. 2a). Rescue of each line restored complex I assembly in all cases (Fig. 2b). Cell lines lacking NDUFV3, NDUFA12 or NDUFA7 that still grew on galactose, had negligible to moderate reductions in complex I activity and mitochondrial respiratory capacity (Fig. 2c). Knockout of NDUFS6 led to most of the N-module dissociating from complex I, however this did not severely impact complex I activity or respiratory capacity (Fig. 2c). This is consistent with previous patient cell studies[12] and suggests that the complex is less stable during BN-PAGE (Extended Data Fig. 4a). In NDUFA12^{KO} cells, the complex I assembly factor NDUFAF2, substituted for its paralog NDUFA12 leading to complex I appearing fully assembled (Extended Data Fig. 4b). In NDUFA2^{KO} cells, no N-module was present (Extended Data Fig. 4a) and these cells showed severe defects in complex I activity and respiration (Fig. 2c). We hypothesize that NDUFV3 may be the terminally assembled subunit of complex I due to its location[9] and lack of defects upon its loss. *In vitro* imported NDUFV3 also readily exchanged with the endogenous assembled protein (Extended Data Fig. 4c), while only *bona fide* subunits were enriched without assembly factors when complex I was isolated using NDUFV3 as bait (Extended Data Fig. 3d, Supplementary Tables 3-4). In contrast to most N-module subunits, knockout of membrane arm subunits resulted in severe mitochondrial respiration defects (Fig. 2c) and loss of assembled complex I with the concomitant accumulation of subcomplexes (Fig. 1).

The severity of knockouts observed for each accessory subunit on complex I assembly appeared to largely predict the impact of mutations in patients with mitochondrial disease (Extended Data Table 1)[14]. Almost all patients reported with mutations in genes encoding

three of the subunits that exhibit mild assembly defects (NDUFA12, NDUFS4, NDUFS6), have two nonsense mutations that block subunit expression. In contrast, patients having defects in 8 accessory subunits showing severe assembly defects carry missense mutations, suggesting that the complete loss of any of these subunits may be incompatible with human life (Extended Data Table 1).

Next we employed stable isotope labelling with amino acids in cell culture (SILAC) and quantitative mass-spectrometry to determine changes in levels of cellular proteins in the representative KO lines (Supplementary Table 5). Most of the >6,000 cellular proteins detected did not significantly differ from control except for complex I subunits themselves, which were consistently downregulated (Fig. 2e, Extended Data Fig. 5a). Of the other 20 mitochondrial proteins whose levels were changed >2-fold, 9 were similarly responsive in a cell line lacking functional complex IV[15] (Extended Data Fig. 5b; Supplementary Table 6) pointing to these gene products being related to general defects in oxidative phosphorylation. Besides affected gene sets related to complex I and oxidative phosphorylation, other affected pathways related to metabolism, transporter activity, translation and DNA replication (Fig. 2e; Extended Data Fig. 5d).

Hierarchical clustering of protein ratios in the representative KO cell lines (Extended Data Fig. 5a) identified clusters of complex I subunits that are similarly located in the structure (e.g. NDUFS4, NDUFA7 and NDUFA12). To increase the resolution of our clustering analysis, we measured the levels of mitochondrial proteins in the remaining 20 KO cell lines (Fig. 3a; Supplementary Table 5). Using the colour scheme from our heat maps, we mapped the levels of individual subunits in each KO onto the recently solved structure of bovine complex I[9] (Fig. 3b; Extended Data Fig. 6a). We uncovered clear structural correlations including the loss of subunits around the N-module upon KO of subunit NDUFA2 as well as loss of subunits from the distal membrane module in NDUFB11^{KO} cells (Fig. 3b).

Transcriptomic analysis of the representative KO cell lines revealed that the only genes with more than a two-fold difference in expression are those that were gene-edited (Extended Data Fig. 7). We conclude that the mutated target genes may be subject to nonsense-mediated mRNA decay while the other complex I subunits whose levels decrease are most likely proteolytically degraded[16].

Hierarchical clustering analysis of complex I subunits (Supplementary Table 7) identified five clusters containing subunits with similar stabilities across knockouts (Fig. 3a). Mapping of these clusters to the structure of bovine complex I[9] revealed distinct modules (Fig. 3c; Extended Data 6b). One cluster contains subunits encompassing the N-module while each other cluster partitions with mtDNA-encoded "ND" subunits. NDUFAB1 could not be assigned to any cluster with its level being almost unchanged, consistent with its separate functions. Subunits NDUFA9, NDUFB4, NDUFB6 and NDUFA11 were not clearly mapped to an individual module and may reside at module interfaces.

Complex I is assembled via a series of intermediate assembly modules and requires the involvement of >10 known assembly factors[17,18]. We generated KO cell lines of assembly factors known to function at different steps – NDUFAF1, NDUFAF2, NDUFAF4, NDUFAF6 and TIMMDC1 (Extended Data Fig. 8a). BN-PAGE analysis showed a reduction or loss of complex I assembly (Fig. 4a). Proteomic analysis indicated that complex I subunits belonging to different modules were affected to varying degrees (Extended Data Fig. 8b, c). The profile of changes in complex I subunits in assembly factor KO lines correlated with groups of complex I KO lines whose subunits belong to distinct modules (Fig. 4b) consistent with assembly models[18]. Since little is known about the assembly of the distal membrane module, we searched our proteomic dataset for proteins altered in KOs belonging to the ND4 and ND5 module relative to those belonging to ND1 and ND2 modules (Supplementary Table 8). ATP5SL, recently identified in a complex I subassembly[17], accumulated in ND4-

and ND5-module KO lines (Fig. 4c). In a separate analysis, the uncharacterized DMAC1 (*TMEM261*) was at elevated levels in membrane arm subunit KO lines when compared against matrix arm subunit KO lines (Fig. 4c; Supplementary Table 9). KO of either ATP5SL or DMAC1 led to specific and severe complex I assembly defects (Fig. 4d, Extended Data Fig. 9a, b) and turnover of N-module and distal membrane arm subunits (Extended Data Fig. 9c, d). Integration of the proteomic profiles in DMAC1^{KO} and ATP5SL^{KO} lines with those originating from our accessory subunit KO lines, indicated a strong correlation with the ND5 module (Fig. 4b).

While DMAC1 is absent from the MitoCarta2.0 database[19], we found it to be a mitochondrial inner-membrane protein (Extended Data Fig. 9e, f). Pulse-chase analysis revealed that mtDNA-encoded subunits formed a 600 kDa intermediate complex[20] in DMAC1^{KO} cells but then dissociated (Extended Data Fig. 9g) indicating a late-stage assembly defect similar to that seen upon loss of complex I assembly factor FOXRED1[21]. Proteins highly enriched with ATP5SL included complex I subunits of the ND4 module and FOXRED1 (Fig. 4e) while proteins enriched with DMAC1 included subunits ND4 and ND5, plus ATP5SL and FOXRED1 along with OXA1L, the membrane insertase for mtDNA-encoded subunits (Fig. 4e; Supplementary Table 10-11). ATP5SL and DMAC1 also interacted with newly translated ND5 (Fig. 4f). Since other complex I subunits, assembly factors and subunits of complexes III and IV were enriched in DMAC1 pull-downs, the integration of the ND4 and ND5 modules in the assembly pathway may intersect with supercomplex formation and occur concurrently with addition of the N-module, the final step in complex I assembly[22]. Due to the association of DMAC1 with the biogenesis of the distal region of complex I, we termed the protein Distal Membrane-arm Assembly Component 1.

In summary, we demonstrate that accessory subunits are integrally associated in modules, defined by the core structural and functional subunits of human complex I, assembly of

which require the concerted action of assembly factors. By defining the impact of individual subunit KOs, our data will facilitate validation of putative pathogenic variants found in complex I genes in patients while *DMAC1* and *ATP5SL* also represent new pathological gene targets. Our approach also serves as a powerful example of how coupling gene-editing and quantitative proteomics allows rapid insights into previously inaccessible aspects of human cellular function.

References

1 Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375-388 (2015).

2 Lapuente-Brun, E. *et al.* Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. *Science* **340**, 1567-1570 (2013).

3 Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* **491**, 374-383 (2012).

4 Morais, V. A. *et al.* PINK1 loss-of-function mutations affect mitochondrial complex I activity via NdufA10 ubiquinone uncoupling. *Science* **344**, 203-207 (2014).

5 Miwa, S. *et al.* Low abundance of the matrix arm of complex I in mitochondria predicts longevity in mice. *Nat. Commun.* **5**, 3837 (2014).

6 Hirst, J. Mitochondrial complex I. *Annu. Rev. Biochem.* **82**, 551-575 (2013).

7 Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* **515**, 80-84 (2014).

8 Zickermann, V. *et al.* Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347**, 44-49 (2015).

9 Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **xx**, xx (2016).

10 Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123 (2008).

11 Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494**, 443-448 (2013).

12 Lazarou, M., McKenzie, M., Ohtake, A., Thorburn, D. R. & Ryan, M. T. Analysis of the assembly profiles for mitochondrial- and nuclear-DNA-encoded subunits into complex I. *Mol. Cell Biol.* **27**, 4228-4237 (2007).

13 Runswick, M. J., Fearnley, I. M., Skehel, J. M. & Walker, J. E. Presence of an acyl carrier protein in NADH:ubiquinone oxidoreductase from bovine heart mitochondria. *FEBS Lett.* **286**, 121-124 (1991).

14 Rodenburg, R.J. Mitochondrial complex I-linked disease. *Biochim. Biophys. Acta* **1857**, 938-945 (2016).

15 Stroud, D. A. *et al.* COA6 is a mitochondrial complex IV assembly factor critical for biogenesis of mtDNA-encoded COX2. *Hum. Mol. Genet.* (2015).

16 Quiros, P. M., Langer, T. & Lopez-Otin, C. New roles for mitochondrial proteases in health, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **16**, 345-359 (2015).

17 Andrews, B., Carroll, J., Ding, S., Fearnley, I. M. & Walker, J. E. Assembly factors for the membrane arm of human complex I. *Proc Natl Acad Sci U S A* **110**, 18934-18939 (2013).

18 Sanchez-Caballero, L., Guerrero-Castillo, S. & Nijtmans, L. Unraveling the complexity of mitochondrial complex I assembly: A dynamic process. *Biochim. Biophys. Acta* **1857**, 980-990 (2016).

19 Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* (2015).

20 Stroud, D. A., Formosa, L. E., Wijeyeratne, X. W., Nguyen, T. N. & Ryan, M. T. Gene knockout using transcription activator-like effector nucleases (TALENs) reveals that human NDUFA9 protein is essential for stabilizing the junction between membrane and matrix arms of complex I. *J. Biol. Chem.* **288**, 1685-1690 (2013).

21 Formosa, L. E. *et al.* Characterization of mitochondrial FOXRED1 in the assembly of respiratory chain complex I. *Hum. Mol. Genet.* **24**, 2952-2965 (2015).

22 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).

Figure 1. Analysis of complex I assembly in KO cell lines. Mitochondria were solubilized in triton X-100 and analysed by BN-PAGE and immunoblotting (IB). CI, complex I; SDHA (complex II subunit), loading control; ‡, loss of N-module; #, subcomplexes. HEK293T*, control for NDUFAB1^{KO}. Galactose growth phenotypes and subunit positions indicated.

Figure 2. Metabolic and proteomic analysis of representative complex I accessory subunit KO lines. (a) Positions of subunits in complex I[9]. **(b)** Cell lines complemented with cDNA encoding the targeted gene. Analysis per Fig. 1. CI+III, supercomplex; ‡, loss of N-module; #, subcomplexes. **(c)** Upper panel, CI activity relative to citrate synthase (CS). $n = 3$ or $n = 4$ biological replicates (HEK293T, NDUFA7, NDUFV3). *P* values are from an unpaired *t*-test; *, $P<0.05$; **, $P<0.01$. Middle, mitochondrial basal and maximal oxygen consumption rates (OCR). Lower panel, glycolytic capacity. ECAR, extracellular

acidification rate. *n = 3* or *n* = 4 biological replicates (NDUFA7, NDUFA12, NDUFV3). Data are mean ± S.E.M. **(d)** Volcano plots showing relative levels of proteins in KO cells. *P* values are from an unpaired *t*-test; *n* = 3 biological replicates; red dots, CI subunits; black dots, *P*<0.05, >1.5-fold change; light grey dots, n.s. **(e)** Gene ontology (GO) enrichment map of pathways and functions altered in respiration deficient knockouts. Example GO terms are grouped according to general role.

Figure 3. Subunit stability correlates with structural modules. (a) Levels of CI subunits in KO lines. n.d., not detected. **(b)** Subunit levels for KOs mapped to the CI structure[9]. Grey, no data; yellow and arrow, KO subunit. Scale per (a). **(c)** Clusters defined in (a) mapped to the CI structure. Italics, core subunits; subunits not clustered removed for clarity.

Figure 4. Analysis of complex I assembly factors including DMAC1 and ATP5SL. (a) Complex I in assembly factor KO lines as per Fig.1. **(b)** KO lines compared via Pearson correlation and hierarchical clustering. **(c)** Volcano plots showing proteins with altered levels in KOs of subunits in specific modules. Light grey dots (n.s.), >5% False Discovery Rate (FDR), <1.5-fold change; module N/A, module not assigned. *P* values from an unpaired *t*-test employing permutation based FDR statistics; *n* = 27 biological replicates (ND4/ND5, ND1/ND2 modules), *n* = 56 (membrane arm), *n* = 23 (matrix arm). **(d)** Complementation of KOs per Fig. 2b. **(e)** Affinity enrichment of proteins from $DMAC1^{FLAG}$ or $ATP5SL^{FLAG}$ cells. *P* values are from an unpaired single-sided t-test. *n* = 3 biological replicates; light grey dots, n.s (*P*-value>0.05). **(f)** Radiolabelled mtDNA-encoded subunits in $ATP5SL^{FLAG}$ or $DMAC1^{FLAG}$ lines were immunoprecipitated with FLAG beads, analysed by SDS-PAGE and autoradiography.

Methods

Cell lines, gene-editing and screening

HEK293T cells[23], commonly used in complex I assembly studies[15,20,21,24,25], interactome[26] and mitochondrial complexome studies[24], were originally purchased from the ATCC and a clonal cell line was obtained after single cell sorting[20] and used as the parental line for all gene editing and proteomic work. Knockout cell lines were validated by sequencing of targeted alleles for insertions and deletions (indels), immunoblotting and subsequent proteomic analysis. Cell lines regularly undergo testing for mycoplasma contamination using PlasmaTest (InvivoGen). Gene-editing was performed using TALEN[27] pairs as described[15,28], or the pSpCas9(BB)-2A-GFP (PX458) CRISPR/Cas9 construct (a gift from Feng Zhang; Addgene plasmid # 48138[29]). Briefly, in the first round, TALEN constructs were designed using the ZiFiT Targeter[30]. For genes unsuccessfully targeted in the first round, CRISPR/Cas9 guide RNAs were designed for a second round of gene-disruption using CHOPCHOP[31]. Successful targeting strategies and constructs can be found in Supplementary Table 1. Gene edited and control HEK293T cells[15] were cultured in DMEM (ThermoFisher) supplemented with 10% (v/v) FBS and 50 µg/mL uridine. Transfection reagents used were: Lipofectamine 2000 and Lipofectamine LTX (ThermoFisher). During screening, glucose-free DMEM supplemented with 5mM galactose, 1mM sodium pyruvate, 10% (v/v) dialyzed FBS (ThermoFisher) and 50 µg/mL uridine was used to identify respiratory incompetent knockout clones. Respiratory competent knockout clones were identified by sequencing of a mixed PCR product covering the target region, where a loss of sequencing fidelity at the target indicates a candidate clone[28]. With the exception of the NDUFA9^{KO} and COA6^{KO} cell lines which were described previously[15,20], indels for individual alleles are summarized in Supplementary Table 1.

To generate NDUFAB1 knockout cells, clonal HEK293T cells were transduced with lentiviruses pLVX-TetOne-Puro-NDUFAB1*^{FLAG} or pLVX-TetOne-Puro-yACP1^{FLAG} (Clontech). NDUFAB1*^{FLAG} represents the C-terminally FLAG tagged human NDUFAB1 protein encoded by cDNA having undergone silent mutagenesis to remove the CRISPR/Cas9 target site. yACP1^{FLAG} indicates cDNA encoding the C-terminally FLAG tagged yeast (*S. cerevisiae*) ACP1. Transduced cells were grown in the presence of 2 μg/mL puromycin for 72 h, and expression of NDUFAB1*^{FLAG} or yACP1^{FLAG} was confirmed after a further 72 h of treatment with 1 μg/mL doxycycline (DOX; Sigma-Aldrich) followed by SDS-PAGE and immunoblotting with NDUFAB1 (Abcam) and FLAG (Sigma-Aldrich) antibodies. For subsequent gene-editing, cells cultured in the presence of 50 ng/mL DOX were transfected with pSpCas9(BB)-2A-GFP-NDUFAB1 and screened as described above.

For complementation, cDNAs encoding NDUFV3^{FLAG}, NDUFS6^{FLAG}, NDUFA8^{FLAG}, ATP5SL^{FLAG} and DMAC1^{FLAG} (TMEM261^{FLAG}) were cloned into pBABE-puro (Addgene #1764;[32]), whereas NDUFA1, NDUFA2, NDUFB7, NDUFB10, NDUFB11 and NDUFC1 cDNAs were cloned into pBMN-Z (Addgene #1734) in place of the LacZ insert. Retroviral constructs were used to transduce the corresponding main clone (Supplementary Table 1), following which expression was selected for through growth in galactose DMEM with the exception of NDUFS6^{KO} and NDUFV3^{KO} which were selected using 2 μg/mL puromycin. Transduction was verified by BN-PAGE or SDS-PAGE followed by immunoblotting with NDUFA9 or FLAG antibodies respectively.

Mitochondrial isolation, gel electrophoresis, immunoblotting and antibodies

Mitochondria were isolated as previously described[33]. Protein concentration was estimated by bicinchoninic acid assay (BCA; Pierce), and aliquots of crude mitochondria stored at -80˚C until use. SDS-PAGE was performed using samples solubilized in LDS sample buffer and separated on NuPAGE Novex Bis-Tris protein gels according to manufacturer's instructions

(ThermoFisher). Tris-Tricine SDS-PAGE, BN-PAGE and 2D-PAGE were performed as described previously[34-36]. Carbonate and swelling experiments were performed as described[37]. Immunoblotting onto PVDF membranes was performed using a Novex Semi-Dry Blotter (ThermoFisher) according to manufacturer's instructions. Horseradish peroxidase coupled secondary antibodies and ECL chemiluminescent substrate (BioRad) were used for detection on a BioRad ChemiDoc XRS+ imaging system. The following primary antibodies were used in this study: COX2 (ThermoFisher #A-6404), COX4 (Abcam, #ab110261), FLAG (Sigma-Aldrich, M2 clone), MIC10 (Aviva Systems Biology, #ARP44801_P050), NDUFA13 (Mitosciences #MS103-SP), NDUFAB1 (Abcam, #ab96230), NDUFB11 (Abcam, #ab183716), NDUFV1(Proteintech #11238-1-AP), NDUFS2 (Mitosciences, #MS114), anti-RC (Abcam, #ab110413 which contains antibodies against ATP5A, UQCRC2, COX1, SDHB and NDUFB8), SDHA (Abcam, #ab14715), TIMMDC1 (Sigma, # HPA053214), TOMM20 (Santa Cruz, #Sc11415) and UQCRC1 (ThermoFisher, #16D10AD9AH5), while rabbit polyclonal antibodies against NDUFA9[12], NDUFAF1 (CIA30)[38], NDUFAF2[21], NDUFAF4[21], NDUFB6[38] and HSP70[20] were raised in-house.

mRNA expression level analysis

For analysis of mRNA expression levels, total RNA was harvested from each cell-line in replicate with TRIzol™ (Thermo scientific). Total RNA was purified using Direct-zol™ columns according to the manufacturers specifications (Zymo Research). For cDNA synthesis 1 µg of total RNA was processed as the T12VN-PAT assay[39] adapted for multiplexing on the Illumina MiSeq instrument. We refer to this assay as mPAT for multiplexed PAT. The approach is based on a nested-PCR that sequentially incorporates the Illumina platform's flow-cell specific terminal extensions onto 3' RACE PCR amplicons. First, cDNA was generated using the anchor primer mPAT Reverse, next this primer and a pool of 50 gene-specific primers were used in 5 cycles of amplification. Each gene-specific

primer had a universal 5' extension (see Supplementary Table 12) for sequential addition of the 5' (P5) Illumina elements. These amplicons were then purified using NucleoSpin columns (Macherey-Nagel), and entered into second round of amplification using the universal Illumina Rd1 sequencing Primer and TruSeq indexed reverse primers from Illumina. Second round amplification was for 14 cycles. Note, that each experimental condition was amplified separately in the first round with identical primers. In the second round, a different indexing primer was used for each experimental condition. All PCR reactions were pooled and run using the MiSeq Reagent Kit v2 with 300 cycles (i.e. 300 bases of sequencing) according to the manufacturer's specifications. Data were analysed using established bioinformatics pipelines[40]. Figures were generated using the R framework.

Oxygen consumption and enzymatic activity measurements

Oxygen consumption (OCR) and extracellular acidification (ECAR) rates were measured in live cells using a Seahorse Bioscience XF24-3 Analyzer as described[15]. Briefly, 50,000 cells were plated per well in Seahorse Bioscience culture plates treated with 50 µg/mL poly-D-Lysine and grown overnight in standard culture conditions. The cellular OCR and ECAR were analyzed in non-buffered DMEM (Seahorse Biosciences) containing 5 mM glucose, 1 mM sodium pyruvate and 50 µg/mL uridine with the following inhibitors: 2 µM Oligomycin; 0.5 µM carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP); 0.5 µM Rotenone; and 0.3 µM Antimycin A. For each assay cycle, four measurement time points of 2 min mix, 2 min wait and 5 min measure were collected. For each cell line, 3-4 replicate wells were measured in multiple plates and CyQuant (ThermoFisher Scientific) was used to normalize measurements to cell number. Basal OCR and non-mitochondrial respiration (following rotenone and antimycin A injections) were calculated as a mean of the measurement points. Basal ECAR was calculated from the initial basal measurement cycle. To calculate proton leak and maximal respiration, the initial measurement following addition of oligomycin or

FCCP was used. Enzymatic activity measurements were performed as previously described[41] in three separate subcultures of each cell line. To accommodate unequal variance, statistical significance was determined through an unpaired two-sample, two-sided *t*-test using Welch's correction.

Radiolabeling of mtDNA-encoded translation products and protein import

Radiolabelling of mtDNA-encoded proteins was performed as previously described[15,34]. Isolated mitochondria were subjected to BN-PAGE or 2D-PAGE as described above, following which proteins were transferred to PVDF membranes and analysed by phosphorimager digital autoradiography (GE Healthcare Life Sciences). For immunoprecipitation of newly translated proteins, mitochondria were isolated from cells pulsed for 2 h and solubilized in 1% (w/v) digitonin, 20 mM Bis-Tris (pH 7.0), 50 mM NaCl, 0.1 mM EDTA, 10% (v/v) glycerol. After a brief clarification spin, complexes were incubated with anti-FLAG affinity gel (SigmaAldrich), the gel washed with 0.2 % (w/v) digitonin, 20 mM Bis-Tris (pH 7.0), 60 mM NaCl, 0.5 mM EDTA, 10% (v/v) glycerol, and enriched proteins eluted with the addition of 150 μg/mL FLAG peptide (SigmaAldrich). Samples were TCA precipitated to remove detergent and analysed by SDS-PAGE and phosphorimaging as above.

For protein import, NDUFA12, NDUFA7 and NDUFV3 cDNA were cloned into the pGEM-4Z plasmid (Promega). mRNA was transcribed using the mMESSAGE mMACHINE SP6 transcription kit (ThermoFisher Scientific) according to the manufacturer's instructions. Radiolabelled proteins were translated in the presence of [^{35}S]-methionine/cysteine using a Rabbit Reticulocyte Lysate System (Promega). Translated proteins were incubated with isolated mitochondria at 37˚C as previously described[12], following which mitochondria were analysed by SDS-PAGE or BN-PAGE as described above.

Quantitative mass-spectrometry using mitochondrial and whole-cell starting material, and data analysis

For NDUFV3^{KO}, NDUFS6^{KO}, NDUFA2^{KO}, NDUFA8^{KO}, NDUFA1^{KO}, NDUFS5^{KO}, NDUFC1^{KO}, NDUFB4^{KO}, NDUFB7^{KO}, NDUFB10^{KO}, NDUFB11^{KO} mass-spectrometry was performed from SILAC labelled whole-cell starting material as described by Kulak *et al.*[42] with modifications. Briefly, cells cultured in "heavy" $^{13}C_6{}^{15}N_4$-arginine, $^{13}C_6{}^{15}N_2$-lysine-containing or "light" SILAC DMEM[15] were harvested, washed in PBS and protein content determined by BCA assay. Measurements were performed in batches of 3-4 knockout cell lines in triplicate with a label switch. Each batch utilized a single pool of clonal HEK293T cells (1 sample grown in "heavy" DMEM, and 2 independent samples grown in "light" DMEM) and knockout cell lines were grown with the complementary label orientation (1 in "light" DMEM, and 2 in "heavy" DMEM). Equal amounts of "heavy" and "light" (typically 250 µg) control HEK293T and knockout cells were mixed, and cells were solubilized in 1% w/v sodium deoxycholate, 100 mM Tris-HCl (pH 8.1). Lysates were sonicated for 30 min at 60˚C in a sonicator waterbath, followed by denaturation and alkylation through the addition of 5 mM Tris(2-carboxyethy)phosphine (TCEP), 20 mM chloroacetamide and incubation for 5 min at 99˚C with vortexing. Samples were digested with trypsin overnight at 37˚C. Detergent was removed by ethyl acetate extraction in the presence of 2% formic acid (FA), following which the aqueous phase was concentrated by vacuum centrifugation. Peptides were reconstituted in 0.5% FA and loaded onto pre-equilibrated small cation exchange (Empore Cation Exchange-SR, Supelco Analytical) stage-tips made in-house. Tips were washed with 6 load volumes of 20% acetonitrile (ACN), 0.5% FA and eluted in 5 sequential fractions of increasing amounts (45-300 mM) of ammonium acetate, 20% ACN, 0.5% FA. A sixth elution was collected using 5% ammonium hydroxide, 80% ACN following which fractions were concentrated, desalted and reconstituted as previously described[15].

Peptides were reconstituted in 0.1% trifluoroacetic acid (TFA) and 2% ACN and fractions analysed sequentially by online nano-HPLC/electrospray ionization-MS/MS on a Q Exactive Plus connected to an Ultimate 3000 HPLC (Thermo-Fisher Scientific). Peptides were first loaded onto a trap column (Acclaim C_{18} PepMap nano Trap x 2 cm, 100 μm I.D, 5 μm particle size and 300 Å pore size; ThermoFisher Scientific) at 15 μL/min for 3 min before switching the pre-column in line with the analytical column (Acclaim RSLC C_{18} PepMap Acclaim RSLC nanocolumn 75 μm x 50 cm, PepMap100 C_{18}, 3 μm particle size 100 Å pore size; ThermoFisher Scientific). The separation of peptides was performed at 250 nL/min using a non-linear ACN gradient of buffer A (0.1% FA, 2% ACN) and buffer B (0.1% FA, 80% ACN), starting at 2.5% buffer B to 35.4% followed by ramp to 99% over 120 minutes (runs had a total acquisition time of 155 min to accommodate void and equilibration volumes). Data were collected in positive mode using Data Dependent Acquisition using m/z 375 - 1800 as MS scan range, HCD for MS/MS of the 12 most intense ions with $z \geq 2$. Other instrument parameters were: MS1 scan at 70,000 resolution (at 200 m/z), MS maximum injection time 50 ms, AGC target 3E6, Normalized collision energy was at 27% energy, Isolation window of 1.8 Da, MS/MS resolution 17,500, MS/MS AGC target of 1E5, MS/MS maximum injection time 100 ms, minimum intensity was set at 1E3 and dynamic exclusion was set to 15 sec.

For the remaining knockouts we utilized isolated mitochondria as starting material. Cells were cultured in SILAC DMEM as above and mitochondrial isolations performed in batches of 1-6 knockout cell lines in triplicate. Each batch contained a single set of clonal HEK293T mitochondria (2 independent isolations from “heavy” and 1 from “light” cells), with knockout mitochondria having the complementary label orientation (2 independent isolations from “light” DMEM and 1 from “heavy” cells). Mitochondria were isolated from cell pellets stored at -80˚C as previously described [43], but with modifications. Cells were resuspended in

20 mM HEPES-KOH (pH 7.6), 220 mM mannitol, 60 mM sucrose, 1 mM EDTA, 1 mM PMSF and homogenized as described above. The homogenate was centrifuged at 800 x *g* for 5 minutes at 4˚C, and the supernatant again centrifuged at 10,000 x *g* for 10 minutes at 4˚C. Crude mitochondria were resuspended in the above buffer and the two differential centrifugation steps repeated. The resuspended pellet was then layered onto a sucrose cushion consisting of 10 mM HEPES-KOH (pH 7.6), 500 mM sucrose, 1 mM EDTA. Samples were centrifuged at 10,000 x *g* for 10 minutes at 4˚C, following which the protein concentration was estimated by BCA assay. Equal amounts of "heavy" and "light" (typically 20 µg) control HEK293T and knockout mitochondria were mixed as described above, collected by centrifugation at 18,000 x *g* and solubilized in 8 M urea, 50 mM ammonium bicarbonate. Proteins were acetone precipitated, reduced and alkylated and desalted as previously described[15]. Peptides reconstituted in 0.1% TFA and 2% ACN were analysed on a Q Exactive Plus, or a LTQ-Orbitrap Elite Instrument. Instrument and method parameters for Q Exactive Plus were as described above however utilized a shorter gradient (90 min separation, 120 min total acquisition). For the Orbitrap Elite, instrument and method parameters were as previously described[15]. A single technical re-injection was collected for all mitochondrial samples.

All raw file names included identifiers for the batch, instrument and gradient used, knockout cell line being studied, and applicable label orientation. Raw files were analysed using the MaxQuant platform[44] version 1.5.4.1 searching against the Uniprot human database containing reviewed, canonical and isoform variants in FASTA format (June 2015) and a database containing common contaminants by the Andromeda search engine[45]. Default search parameters for a Arg10 and Lys8 labelled experiment were used with modifications. Briefly, cysteine carbamidomethylation was used as a fixed modification, and N-terminal acetylation and methionine oxidation were used as variable modifications. False discovery

rates of 1% for proteins and peptides were applied by searching a reverse database, and "Re-quantify" and "Match from and to", "Match between runs" options were enabled with a match time window of 2 minutes. Experimental groups based on data gathered using different instrumentation and/or acquisition parameters were given odd numbered fractions to avoid falsely matched identifications, whereas fractionated whole-cell samples were given sequential fraction numbers. Unique and razor peptides with a minimum ratio count of 2 were used for quantification.

Using the Perseus platform (version 1.5.4.1), identifications were matched to the MitoCarta2.0 database[19] using Ensembl ENSG id and gene name identifiers. Identifications labeled by MaxQuant as "Only identified by site", "Reverse" and "Potential Contaminant" were removed. Proteins having < 3 valid values in a single experimental group were removed. For mitochondrial samples, we found the correlation of $\log_2$-ratio data from biological replicates in the same experimental group to be moderate at best and as low as 0.3 in some cases. We surmised the main cause of this to be batch and labelling effect, the former due to differences in mitochondrial isolations between batches and latter due to one (of three) replicates within each experimental group always being subjected to a label switch. To account for these and potentially other factors, we adopted an approach that borrows principles from RUV-2[46] and SVA[47] methods for removing unwanted variations, with modifications in the algorithm for choosing the *control* proteins (*i.e.* those not found in MitoCarta 2.0[19]) and moderating the amount of adjustment for genes with small sample size due to missing values. Adjustments were performed in the R framework, following which the adjusted ratios were imported back into Perseus. The $\log_2$ ratio values for proteins in replicates were normally distributed and had equal variances. The mean $\log_2$-transformed ratios for each experimental group were calculated along with their standard deviation and *P*-value based on single sample two-sided *t*-test[15]. This statistical approach was consistent with

published quantitative SILAC analyses employing similar instrumentation and methods[15,48,49]. Groups having < 2 valid values were converted to "NaN" (not a number). A quality matrix was generated based on the standard deviation, and corresponding values having a standard deviation greater than 1 converted to "NaN". This threshold was determined empirically to remove outliers from the main distribution of standard deviations across all samples. These data can be found in Supplementary Data Table 5.

Figs 3b and Extended Data Fig. 6a, 8c and 9d were generated from a matrix containing log_2-transformed median SILAC ratios having a standard deviation <1 for complex I subunits (Supplementary Table 7) and data were mapped to homologous subunits in PDB XXXX[9].

For Fig. 3a, hierarchical clustering on rows (proteins) was performed using Pearson distance and average linkage. Data were pre-processed using k-means (clusters = 300). Images were generated using the PyMOL Molecular Graphics System, Version 1.7.2.1 (Schrödinger, LLC). Log_2 SILAC ratios for some proteins in their corresponding knockout cell line had very low (generally >4-fold reduction) ratios, whereas others were reported "NaN". This could be either due to the "Re-quantify" option being turned on for the MaxQuant search, which results in translation of peak shapes from an identified isotope pattern being translated to its unidentified label partner, or indels in some lines generating a non-functional (but still translated) protein as we have seen previously[15].

For the identification of proteins dysregulated between knockouts of discrete modules (Fig. 4c; Supplementary Tables 8 and 9), triplicate log_2–transformed SILAC ratios from Supplementary Table 5 were assigned to one of two groups based on the knockout being associated with the indicated module. Groups tested had comparable variance, and a modified Welch's two-sample *t*-test with permutation based FDR statistics[50,51] was used to determine significance. Parameters for the test were: 70% minimum valid values, 250 permutations and significance being an FDR of <0.05.

For the gene ontology enrichment analysis in Fig. 2c, proteins with a *P*-value <0.05 and with >1.5 fold change up or down were submitted to the DAVID online tool (david.abcc.ncifcrf.gov/home.jsp) for enriched biological processes (GOTERM_BP_FAT) and molecular function (GOTERM_MF_FAT). Functional annotation charts were exported and visualised using Cytoscape (version 3.4.0) and the Enrichment Map app[52] (version 2.1.0; *P*<0.005). Contents of enriched terms indicated in Fig. 2c are detailed in Extended Data Fig. 5d.

Affinity enrichment mass-spectrometry and data analysis

Affinity-enrichment experiments in Fig. 4e, Extended Data Figs 3d and 4d, and Supplementary Tables 2-4 and 10-11 were performed from HEK293T and knockout cells complemented with the FLAG-tagged protein cultured in "heavy" or "light" SILAC DMEM as previously described[15]. Mass-spectrometry was performed on a Q Exactive Plus as above but using a shorter gradient (25 min separation, 60 min total acquisition). For data analysis, raw files were analyzed using the MaxQuant platform as above using default search parameters for a Arg10 and Lys8 labelled experiment, with modifications. Briefly, cysteine carbamidomethylation was used as a fixed modification, and N-terminal acetylation and methionine oxidation were used as variable modifications. False discovery rates of 1% for proteins and peptides were applied by searching a reverse database, and "Re-quantify" and "Match from and to", "Match between runs" options were enabled with a match time window of 2 minutes. Unique and razor peptides with a minimum ratio count of 1 were used for quantification. Data analysis was performed using the Perseus framework. Identifications were matched to MitoCarta2.0 dataset[19] as above. Only proteins with a sequence coverage of 2 or more unique peptides were included in further analysis. Normalized SILAC ratios were inverted to achieve the orientation FLAG-tagged/HEK293T and proteins not present in > 2/3 replicates were removed. Log_{10}-transformed transformed values had a normal distribution

and comparable variance. For affinity-enrichment experiments, statistical method, sample size and analysis approaches were chosen based on published quantitative affinity-enrichment analyses employing similar instrumentation and methods[15,21,53,54]. *P*-values were calculated by a single (FLAG-tagged cell line enriched) sided t-test and the negative logarithmic *P*-value plotted against the mean of the three replicates.

Miscellaneous molecular biology

cDNA inserts were obtained from an in-house cDNA library generated from our clonal HEK293T line. Briefly, RNA was isolated using TRIzol Reagent (ThermoFisher) according to manufacturer's instructions. The Superscript III first strand synthesis kit (ThermoFisher Scientific) was used to generate cDNA primed with either Oligo(dT) or random hexamers. Inserts were amplified from the library using Q5 High Fidelity DNA Polymerase (NEB) and Gibson assembled into the relevant plasmid (see above) using the NEBuilder HiFi DNA Assembly Master Mix (NEB) according to manufacturer's instructions. Sanger sequencing was performed from PCR product or plasmid template DNA. DNA sequence assembly and alignment to sequencing reads was performed using SnapGene (GSL Biotech) and Geneious (Biomatters).

Immunofluorescence microscopy was performed as previously described[55] using primary antibodies (FLAG or TOMM20) at 1:500 dilutions. Primary antibodies were labelled with anti-mouse conjugated Alexa Fluor 488 and anti-rabbit conjugated Alexa Fluor 568 secondary antibodies (Molecular Probes). Hoechst (1µg/mL) was used to stain nuclei. Cells were visualized using either a Leica TCS SP8 equipped with HyD detectors. Images were processed using Image J[56]. All figures were prepared using Adobe Photoshop and Illustrator (CC2015.5).

Data Reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Additional References

23 DuBridge, R. B. *et al.* Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell Biol.* **7**, 379-387 (1987).

24 Heide, H. *et al.* Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab.* **16**, 538-549 (2012).

25 Vogel, R. O. *et al.* Identification of mitochondrial complex I assembly intermediates by tracing tagged NDUFS3 demonstrates the entry point of mitochondrial subunits. *J. Biol. Chem.* **282**, 7582-7590 (2007).

26 Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425-440 (2015).

27 Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460-465 (2012).

28 Reljic, B. & Stroud, D. A. Screening Strategies for TALEN-Mediated Gene Disruption. *Methods Mol Biol* **1419**, 231-252 (2016).

29 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281-2308 (2013).

30 Sander, J. D. *et al.* ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.* **38**, W462-468 (2010).

31 Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401-407 (2014).

32 Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587-3596 (1990).

33 Acin-Perez, R., Fernandez-Silva, P., Peleato, M. L., Perez-Martos, A. & Enriquez, J. A. Respiratory active mitochondrial supercomplexes. *Mol. Cell* **32**, 529-539 (2008).

34 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 (2007).

35 Wittig, I., Braun, H. P. & Schagger, H. Blue native PAGE. *Nat. Protoc.* **1**, 418-428 (2006).

36 Schagger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368-379 (1987).

37 Ryan, M. T., Voos, W. & Pfanner, N. Assaying protein import into mitochondria. *Methods Cell Biol.* **65**, 189-215 (2001).

38 Dunning, C. J. *et al.* Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *EMBO J.* **26**, 3227-3237 (2007).

39 Janicke, A., Vancuylenberg, J., Boag, P. R., Traven, A. & Beilharz, T. H. ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3' RACE applications. *RNA* **18**, 1289-1295 (2012).

40 Harrison, P. F. *et al.* PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **21**, 1502-1510 (2015).

41 Frazier, A. E. & Thorburn, D. R. Biochemical analyses of the electron transport chain complexes by spectrophotometry. *Methods. Mol. Biol.* **837**, 49-62 (2012).

42 Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319-324 (2014).

43 Johnston, A. J. *et al.* Insertion and assembly of human tom7 into the preprotein translocase complex of the outer mitochondrial membrane. *J. Biol. Chem.* **277**, 42197-42204 (2002).

44 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

45 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794-1805 (2011).

46 Gagnon-Bartsch, J. A. & Speed, T. P. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* **13**, 539-552 (2012).

47 Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724-1735 (2007).

48 Munch, C. & Harper, J. W. Mitochondrial unfolded protein response controls matrix pre-RNA processing and translation. *Nature* **534**, 710-713 (2016).

49 Wrobel, L. *et al.* Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485-488 (2015).

50 Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189**, 739-754 (2010).

51 Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5116-5121 (2001).

52 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).

53 Stroud, D. A. *et al.* Composition and topology of the endoplasmic reticulum-mitochondria encounter structure. *J. Mol. Biol.* **413**, 743-750 (2011).

54 Gebert, N. *et al.* Dual function of Sdh3 in the respiratory chain and TIM22 protein translocase of the mitochondrial inner membrane. *Mol. Cell* **44**, 811-818 (2011).

55 Richter, V. *et al.* Structural and functional analysis of MiD51, a dynamin receptor required for mitochondrial fission. *J. Cell Biol.* **204**, 477-486 (2014).

56 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).

57 Ostergaard, E. *et al.* Respiratory chain complex I deficiency due to NDUFA12 mutations as a new cause of Leigh syndrome. *J. Med. Genet.* **48**, 737-740 (2011).

58 Assouline, Z. *et al.* A constant and similar assembly defect of mitochondrial respiratory chain complex I allows rapid identification of NDUFS4 mutations in patients with Leigh syndrome. *Biochim. Biophys. Acta* **1822**, 1062-1069 (2012).

59 Haack, T. B. *et al.* Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9. *J. Med. Genet.* **49**, 83-89 (2012).

60 Kirby, D. M. *et al.* NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. *J. Clin. Invest.* **114**, 837-845 (2004).

61 van den Bosch, B. J. *et al.* Defective NDUFA9 as a novel cause of neonatally fatal complex I disease. *J. Med. Genet.* **49**, 10-15 (2012).

62 Hoefs, S. J. *et al.* NDUFA10 mutations cause complex I deficiency in a patient with Leigh disease. *Eur. J. Hum. Genet.* **19**, 270-274 (2011).

63 Angebault, C. *et al.* Mutation in NDUFA13/GRIM19 leads to early onset hypotonia, dyskinesia and sensorial deficiencies, and mitochondrial complex I instability. *Hum. Mol. Genet.* **24**, 3948-3955 (2015).

64 Haack, T. B. *et al.* Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J. Med. Genet.* **49**, 277-283 (2012).

65 Shehata, B. M. *et al.* Exome sequencing of patients with histiocytoid cardiomyopathy reveals a de novo NDUFB11 mutation that plays a role in the pathogenesis of histiocytoid cardiomyopathy. *Am. J. Med. Genet. A* **167A**, 2114-2121 (2015).

66 Fernandez-Moreira, D. *et al.* X-linked NDUFA1 gene mutations associated with mitochondrial encephalomyopathy. *Ann. Neurol.* **61**, 73-83 (2007).

67 Peralta, S. *et al.* Partial complex I deficiency due to the CNS conditional ablation of Ndufa5 results in a mild chronic encephalopathy but no increase in oxidative damage. *Hum. Mol. Genet.* **23**, 1399-1412 (2014).

Acknowledgments

We thank M. Curtis, P. Faou, M. Lazarou, B. Porebski, L. Twigg, R. Schittenhelm (Monash Biomedical Proteomics Platform), A. Barugahare and P. Harrison (Monash Bioinformatics Platform), Monash Micro Imaging and the Micromon NGS Facility for assistance. We acknowledge funding from NHMRC Project Grants (1068056, 1107094) and fellowships (1070916 to D.A.S., 541920 to A.E.F., 1022896 to D.R.T.), the Australian Mitochondrial Disease Foundation and the Victorian Government's Operational Infrastructure Support Program.

Contributions

D.A.S and M.T.R conceived the project and wrote the manuscript; D.A.S., D.R.T., and M.T.R. designed the experiments; D.A.S., E.E.S., L.E.F., B.R., M.G.D., L.D.O. and M.T.R. generated and analysed KO lines; D.A.S. performed proteomic experiments; A.E.F. and T.S. performed enzymology; T.H.B. undertook transcript analysis; A.S. developed normalization algorithms.

Author information

Data available via ProteomeXchange (PXD004666) and NCBI's Gene Expression Omnibus (GSE84913). The authors declare no competing financial interests. Correspondence to D.A.S. (d.stroud@monash.edu) or M.T.R (michael.ryan@monash.edu).

Extended Data Figures

Extended Data Figure 1. **Assembly analysis of the complex I/III/IV supercomplex (SC) in knockout cell lines.** Mitochondria were solubilized in digitonin and complexes separated by BN-PAGE followed by immunoblotting (IB) using the indicated antibodies. An antibody against complex V (CV) subunit ATP5A was used as loading control. #, subcomplexes; *, non-specific.

Extended Data Figure 2. Steady state levels of respiratory chain complexes I-IV and supercomplex forms in the 28 complex I accessory subunit knockout lines generated in this study. NDUFA9^{KO} has been analysed previously[20], whereas the NDUFAB1^{KO} is described in Extended Data Fig. 3. Mitochondria were solubilized in triton X-100 (TX100) or

digitonin (DIG) and analysed by BN-PAGE and immunoblotting with antibodies against NDUFA9 (complex I), SDHA (complex II), UQCRC1 (complex III) and COX4 (complex IV). In TX100 samples, some Complex III-IV supercomplex is retained. #, secondary clone later identified be an incomplete knockout.

Extended Data Figure 3. Generation and analysis of NDUFAB1^{KO} cell lines. (a) Scheme detailing knockout strategy of genomic *NDUFAB1* using doxycycline (DOX) -inducible expression of CRISPR/Cas9-resistant NDUFAB1 (NDUFAB1*^{FLAG}) or yACP1^{FLAG}. **(b)** *NDUFAB1* knockouts complemented with NDUFAB1^{FLAG} (NDUFAB1*-2) cells were cultured in media lacking DOX for the indicated times. Isolated mitochondria were analysed by BN-PAGE (triton X-100) or SDS-PAGE and immunoblotting (IB) with the indicated antibodies. **(c)** Brightfield images of cells grown +/- DOX, or + DOX in glucose or galactose cell culture medium. Scale bar, 25 μm. (**d**) SILAC labelled mitochondria from DOX-treated HEK293T or NDUFAB1^{FLAG} (NDUFAB1*-2) cells were solubilized in triton X-100 and incubated with anti-FLAG affinity gel. Elutions were mixed and analysed by LC-MS as described in the methods. Proteins enriched with NDUFAB1 include complex I subunits and LYRM proteins. *P* values are from an unpaired single-sided t-test. $n = 3$ biological replicates; light grey dots, *P*-value>0.05. (**e**) Mitochondria isolated from *NDUFAB1* knockouts complemented with yACP1^{FLAG} or NDUFAB1^{FLAG} were solubilized in triton X-100 and analysed by BN-PAGE and immunoblotting with the indicated antibodies.

Extended Data Figure 4. Analysis of N-module accessory subunits. (a) Mitochondria were isolated from cell lines, solubilized in triton X-100 and analysed by BN-PAGE and immunoblotting for N-module subunit NDUFV1 or non N-module subunit NDUFA9. ‡, complex lacking N-module; N*, subcomplex containing N-module. SDHA was used as a

loading control. **(b)** Mitochondria were solubilized in digitonin and analysed by BN-PAGE and immunoblotting for NDUFAF2. †, NDUFAF2 associated complex I. **(c)** [^{35}S]-methionine labelled proteins were imported into the indicated mitochondria, solubilised in digitonin and analysed by BN-PAGE and autoradiography. 10% of the input lysate was analysed by SDS-PAGE and autoradiography. CI^{SC}, complex I supercomplex *, non-specific band. **(d)** Mitochondria isolated from $NDUFV3^{KO}$ cells complemented with $NDUFV3^{FLAG}$ were solubilized in triton X-100 or digitonin and complexes bound to anti-FLAG affinity gel. Eluted proteins were analysed by LC-MS as described in the supplementary materials. *P* values are from an unpaired single-sided t-test. *n* = 3 biological replicates; light grey dots, n.s. *P*-value>0.05.

Extended Data Figure 5. Proteomic analysis of KO cell lines. (**a**) Relative levels of proteins in representative accessory subunit KO cell lines, clustered according to Euclidean distance. Column order is as per Fig. 2b. The inset shows complex I subunit specific clusters. (**b**) Volcano plot depicting proteins regulated in representative accessory subunit KO cell lines harbouring respiration defects ($NDUFA2^{KO}$, $NDUFA8^{KO}$, $NDUFS5^{KO}$, $NDUFC1^{KO}$, $NDUFB10^{KO}$, $NDUFB11^{KO}$, $NDUFB7^{KO}$). Proteins found to be regulated in a cell line with a severe complex IV defect[15] are shaded light blue (down) and green (up), suggesting their response is due to general defects in respiration. Inset, volcano plot depicting the relative level of proteins in a complex IV knockout cell line. *P* values are from an unpaired t-test; *n* = 8 independent means comprised each of 3 biological replicates (main panel), *n* = 3 (inset) biological replicates; light grey dots, *P*-value>0.05, <1.5-fold change. Data is reproduced in Supplementary Table 6. (**c**) Proteins affected >2-fold in levels in respiration-deficient subunit KO cell lines. Colour key according to (b). Bold, proteins listed in MitoCarta2.0. (**d**) Proteins associated with GO terms and groups outlined in Fig. 2d.

Extended Data Figure 6. Mapping of complex I subunit levels onto the structure. (a) Subunit levels in complex I accessory subunit KO lines were mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure of complex I[9] as in Fig. 3b. Both sides of complex I are shown. Median ratio data used in the preparation of this figure can be found in Supplementary Table 7. **(b)** Opposite side view of Fig. 3c. n.d., dark grey shading on the structures, subunits not quantified. Subunits not clustered to modules removed for clarity.

Extended Data Figure 7. mRNA expression levels in selected accessory subunit knockout lines. Transcripts were measured for nuclear-encoded complex I subunit genes along with control genes from complex II (SDHA), complex III (UQCRC1, UQCRFS1), complex IV (COX4L1, NDUFA4), complex V (ATP5B, ATP5H) and mt-ribosome (MRPS2, MRPL46) in KO lines (performed in duplicate).

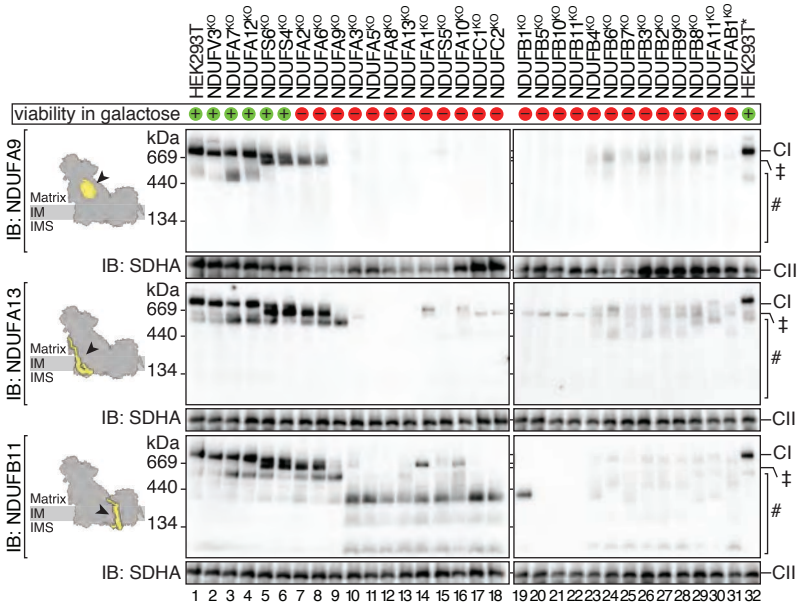
Extended Data Figure 8. Analysis of assembly factor knockout lines. (a) Mitochondrial proteins from the indicated cell lines were separated by SDS-PAGE and subjected to western blot analysis. **(b)** Volcano plots showing fold changes vs *P*-values for the mitochondrial proteins in assembly factor knockout cell lines. *P* values are from an unpaired *t*-test; $n = 3$ biological replicates; coloured dots are according to the key at bottom right; n.s., *P=>0.05*. **(c)** Subunit levels mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure as for Fig. 3b. n.d., dark grey shading on the structures, subunits not quantified. Both sides of complex I are shown.

Extended Data Figure 9. Characterization of DMAC1 and ATP5SL. (a) ATP5SL^{KO} mitochondria were solubilized in triton X-100 or digitonin and analysed by BN-PAGE and immunoblotting with the indicated antibodies. **(b)** As for (a) using DMAC1^{KO} mitochondria. **(c)** Volcano plots showing fold changes vs *P*-values for the mitochondrial proteins in ATP5SL and DMAC1 KO cell lines. *P* values are from an unpaired *t*-test; $n = 3$ biological replicates; coloured dots represent complex I subunits depicted in the key; n.s., *P>0.05.* **(d)** Subunit levels mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure as for Fig. 3b. n.d., dark grey shading on the structures, not quantified. Both sides of complex I are shown. **(e)** Mitochondria isolated from DMAC1 cells complemented with DMAC1^{FLAG} were resuspended in isotonic buffer, hypoosmotic swelling buffer, or triton X-100 (TX-100) followed by proteinase K (PK) incubation where indicated. Alternately, mitochondria were treated with 100 mM Na_2CO_3 and membrane-integral (pellet) and soluble or peripherally attached (supernatant, SN) proteins were separated by ultracentrifugation. Samples were analysed by SDS-PAGE and immunoblotting for TOMM20 (outer membrane); MIC10 (integral inner membrane protein exposed to intermembrane space); NDUFAF1 (matrix, soluble); NDUFS2 (matrix, peripheral). **(f)** DMAC1^{KO} cells complemented with DMAC1^{FLAG} were analysed by immunofluorescence microscopy with the indicated antibodies. TOMM20, mitochondrial protein. Scale bar, 20 μm. Representative result from 3 independent experiments. **(g)** Cells were pulsed with [^{35}S]-methionine for 1 h and chased for the indicated times as described in the supplementary materials. Isolated mitochondria were solubilized in triton X-100 and analysed by 2D-PAGE and autoradiography. ◊, 600 kDa complex; #, subcomplex containing ND1 and ND2.

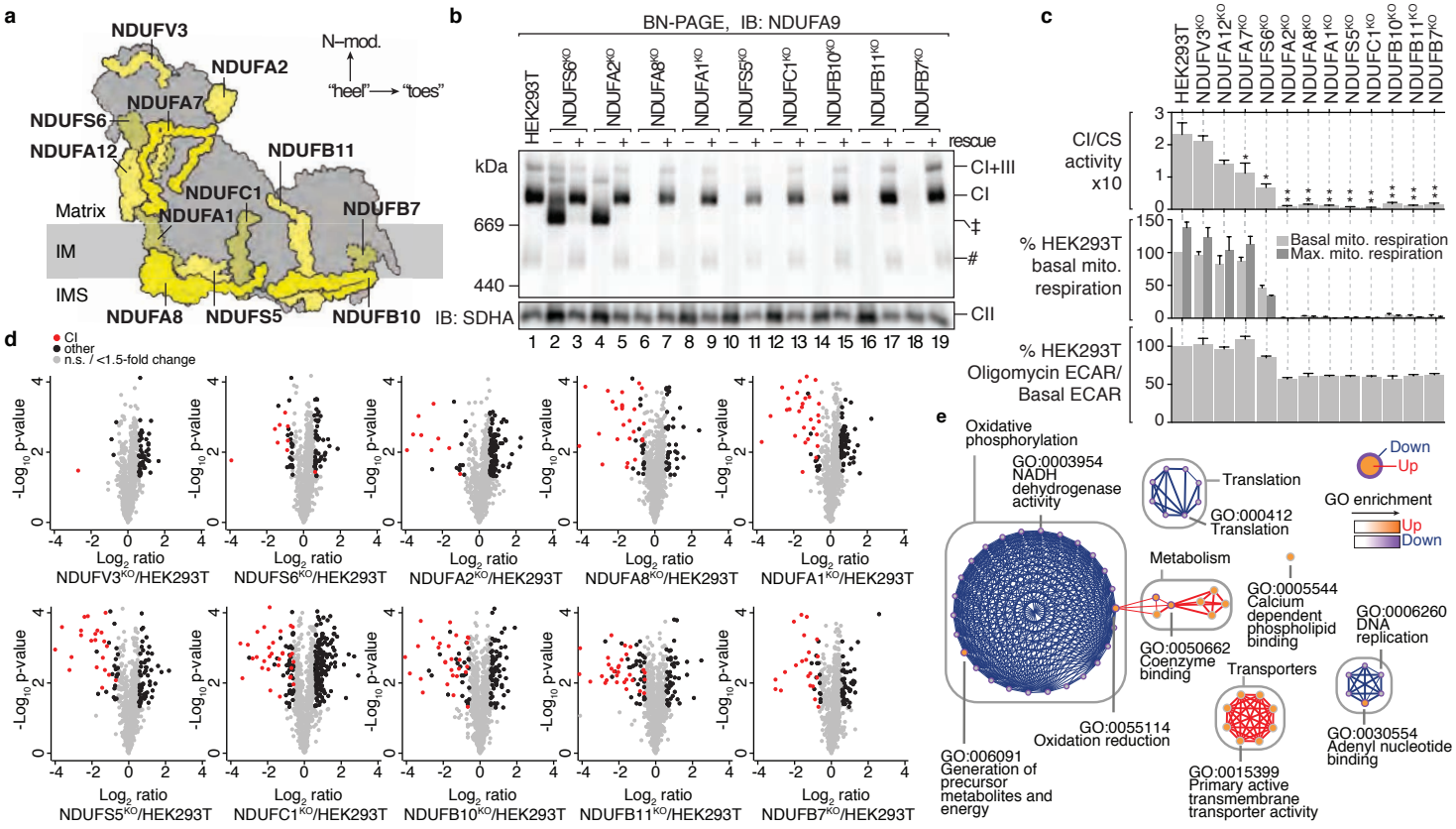
Extended Data Table 1: Pathogenic mutations in complex I accessory subunit genes in patients with mitochondrial disease

Gene Symbol	OMIM#	Mutations	Known/Predicted Impact on Subunit	Impact on Complex I Assembly
Mild Assembly Defects[a]				
NDUFA12	*614530	homozygous p.R60X	No detectable protein	Decreased CI assembly[57]
NDUFS4	*602694	homozygous p.K158fs homozygous p.W96X homozygous p.R106X homozygous p.W15X homozygous IVS1AS, G>A, -1 homozygous p.K154Kfs (and other similar homozygous or compound heterozygous mutations likely to be null mutations)	Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation	All mutations studied result in partially assembled CI lacking N-module[58]
NDUFS6	*603848	homozygous IVS2DS, T-A, +2 homozygous 4.175-KB DEL, EX3-4DEL homozygous p.C115Y homozygous p.Q118X[59]	Predicted Null mutation Predicted Null mutation Uncertain Predicted Null mutation	All mutations studied result in partially assembled CI lacking N-module[60]
Severe Assembly Defects				
NDUFA9	*603834	Homozygous p.R321P	Marked decrease in NDUFA9 protein amount	Decreased CI assembly[61]
NDUFA10	*603835	homozygous p.G99E compound heterozygous p.Met1?/p.Q142R[62]	Uncertain ~10% of normal levels of NDUFA13 protein	Decreased CI assembly[62]
NDUFA11	*612638	homozygous IVS1DS, G-A, +5	"Leaky" - 2:1 ratio of wildtype to normal transcript	Not assessed
NDUFA13	*609435	homozygous p.R57H[63]	30-40% of normal levels of NDUFA13 protein	Decreased CI assembly[63]
NDUFB3	*603839	homozygous p.W22R compound heterozygous p.W22R/p.G70X	Uncertain 1 predicted Null mutation & 1 uncertain	Decreased CI assembly[64]
NDUFB9	*601445	homozygous p.L64P	Some residual protein	Not assessed
NDUFB11	*300403	de novo heterozygous p.R88X (female) heterozygous p.R134SfsX3 (female) de novo heterozygous p.Y108X (female)[65] de novo heterozygous p.W85X (female)[65]	X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells	Not assessed
NDUFA1	*300078	hemizygous p.G8R (male) hemizygous p.R37S (male) hemizygous p.G32R (male) de novo heterozygous p.G32R (female)	X-linked gene: Uncertain X-linked gene: Uncertain X-linked gene: Uncertain X-linked gene: Uncertain	Decreased CI assembly[66]
NDUFA5	*601677	No patients reported but Ndufa5 knockout mice die around embryonic day-9		Decreased CI assembly[67]

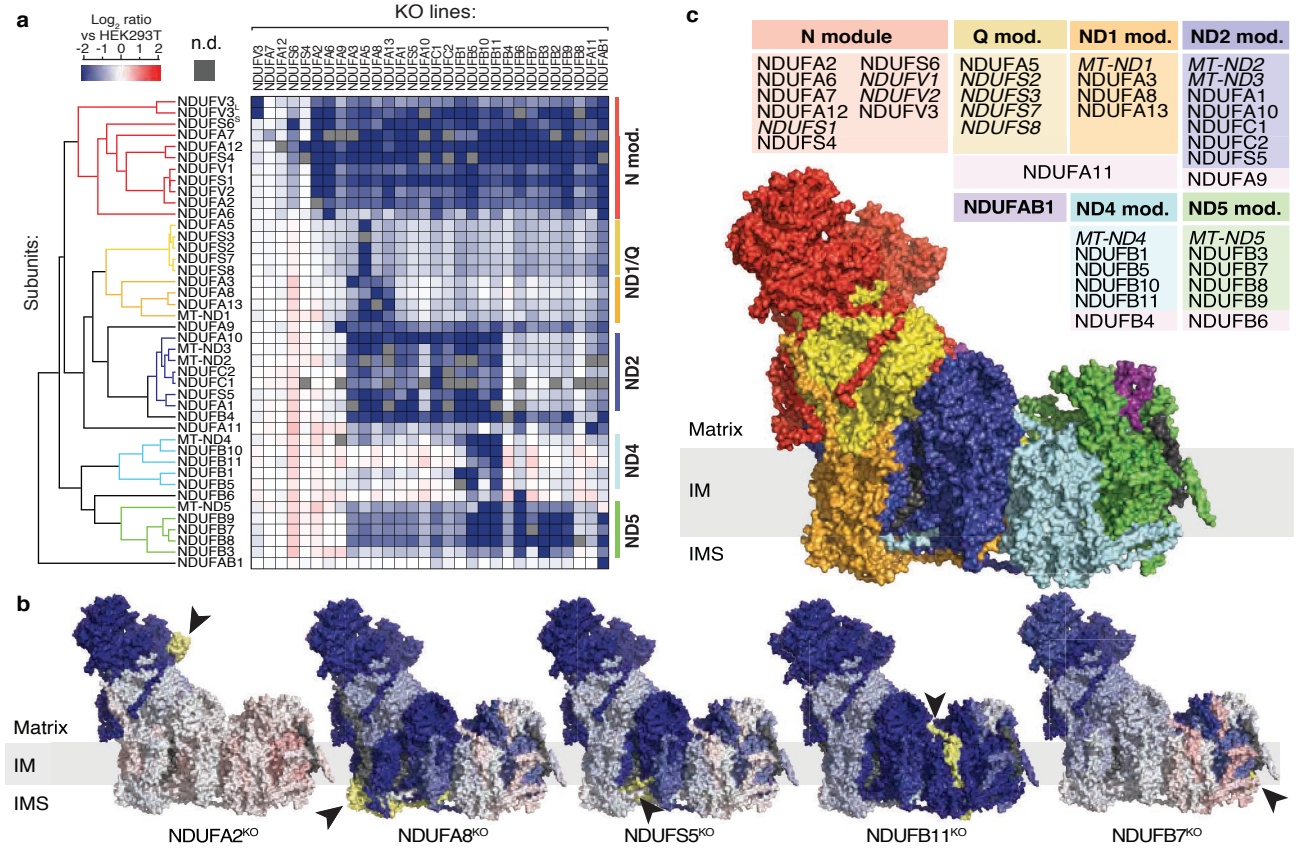
[a]Three accessory subunits in which knockouts cause mild complex I assembly defects have had patients reported with pathogenic mutations; in almost all cases the mutations are expected to cause two null alleles, suggesting that almost complete loss of function of these subunits may be required to cause human disease. Eight accessory subunits in which knockouts cause severe complex I assembly defects have had patients reported with pathogenic mutations; in almost all cases the patients have at least one missense mutation or some evidence that some residual subunit protein is present. This suggests that complete loss of function of these subunits may not be compatible with human life. Interpretation of the data for the NDUFB11 and NDUFA1 subunits is complicated by their being encoded on the X chromosome. Males thus have only one copy of these genes while females have 2 copies with some cells expressing the wildtype and some expressing the mutant allele. All reported NDUFB11 patients are female and all had stop codon or frameshift mutations expected to cause null alleles. Such patients often have skewed X-inactivation, with most cells expressing the wildtype allele, which may compensate partly for the severity of the defect. An additional subunit, NDUFA5, has not had patients with mutations identified but knockout of the mutation in mice results in embryonic lethality. This is consistent with the suggestion that human foetuses may not be viable if they have null-type mutations in both alleles of genes encoding accessory subunits linked to severe assembly defects. #, Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) 01/20/2016 World Wide Web URL: http://omim.org



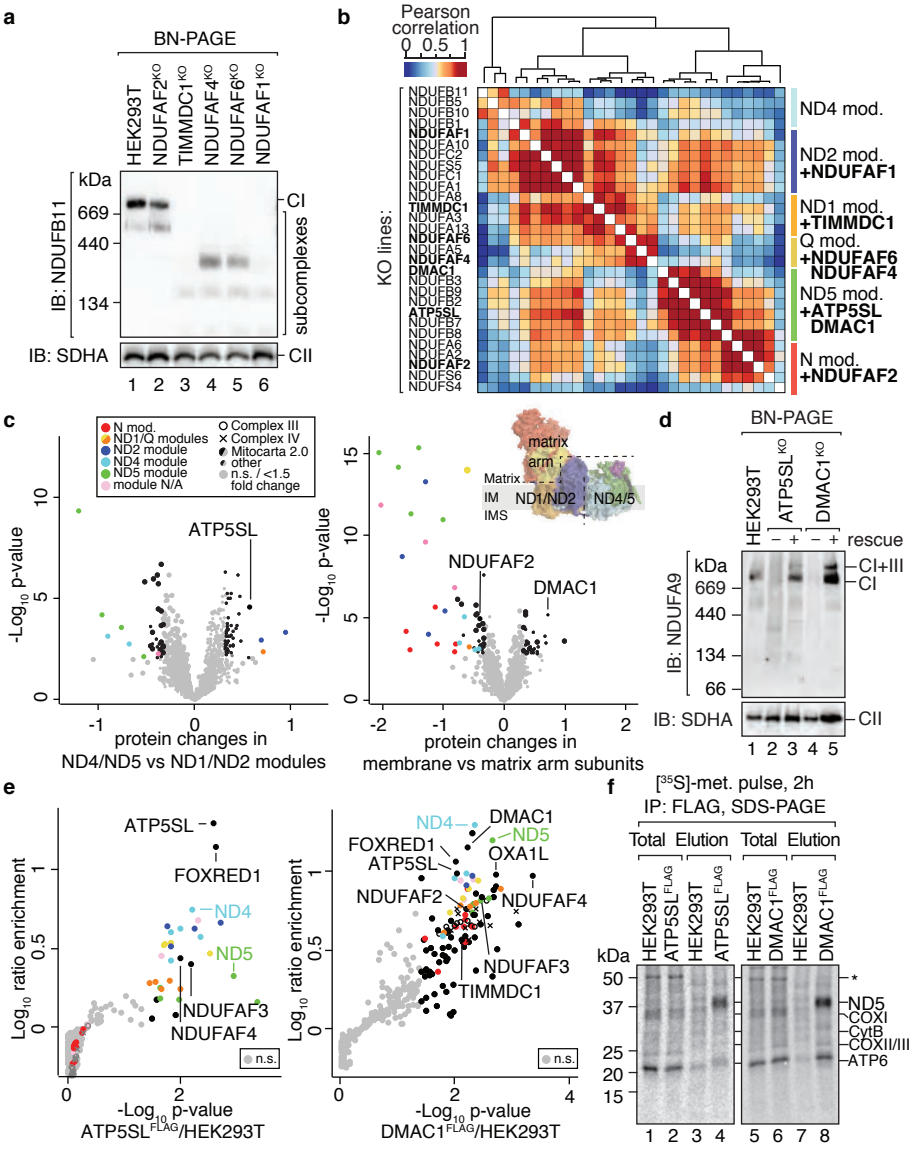
Stroud *et al*. Figure 1



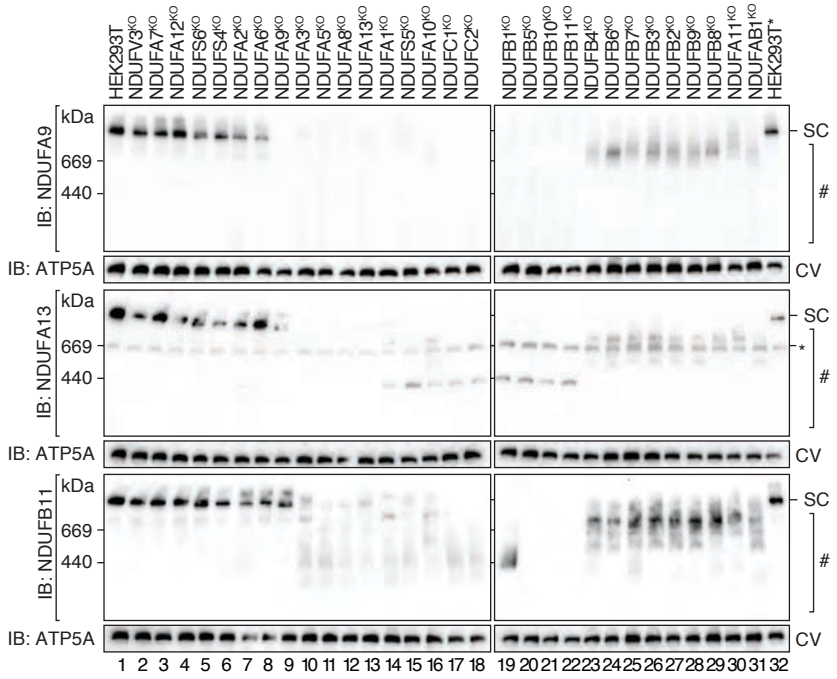
Stroud *et al.* Figure 2



Stroud *et al.* Figure 3

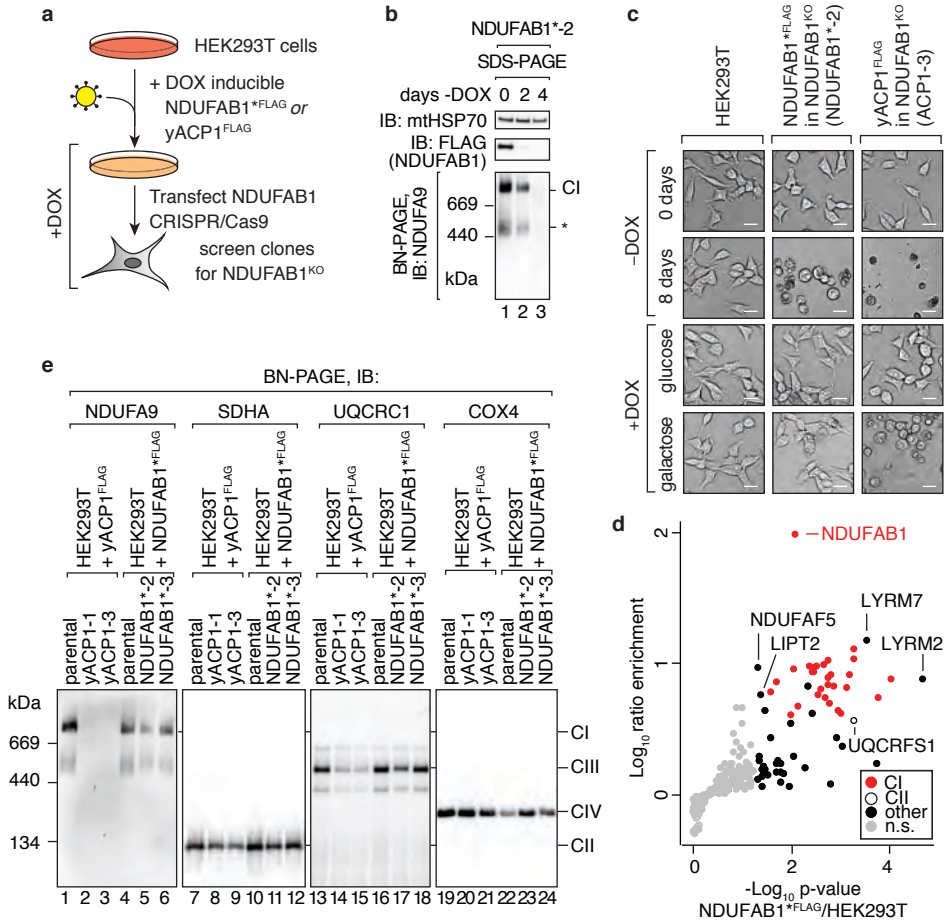


Stroud *et al.* Figure 4

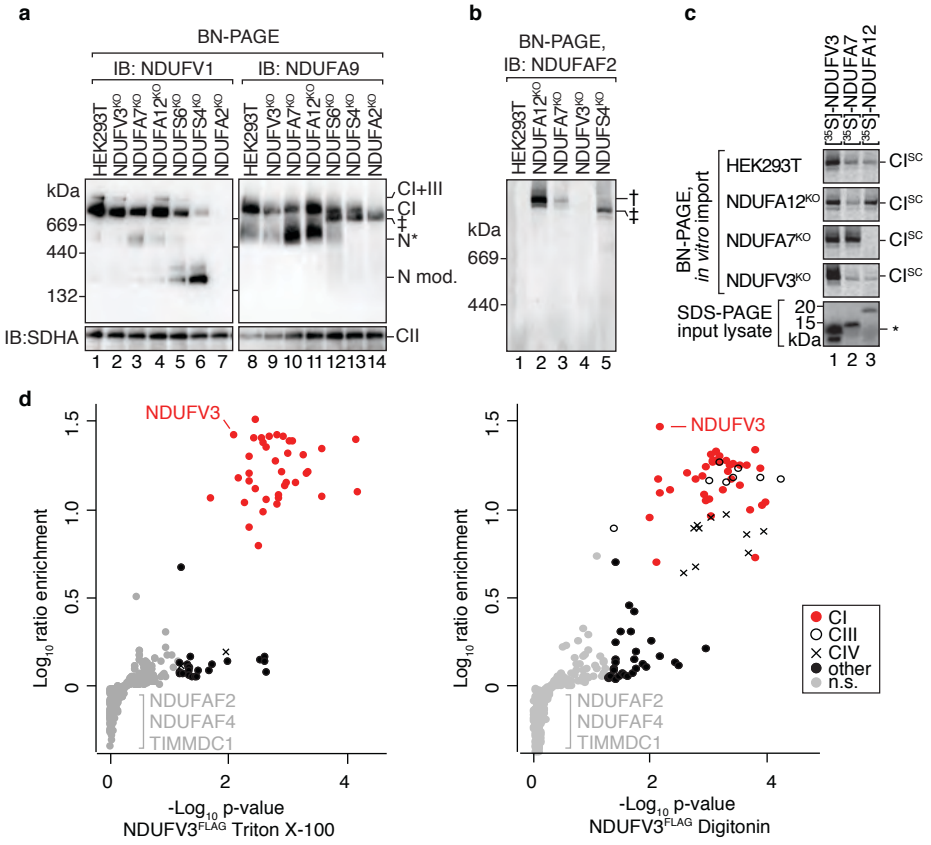


Stroud ED Fig. 1

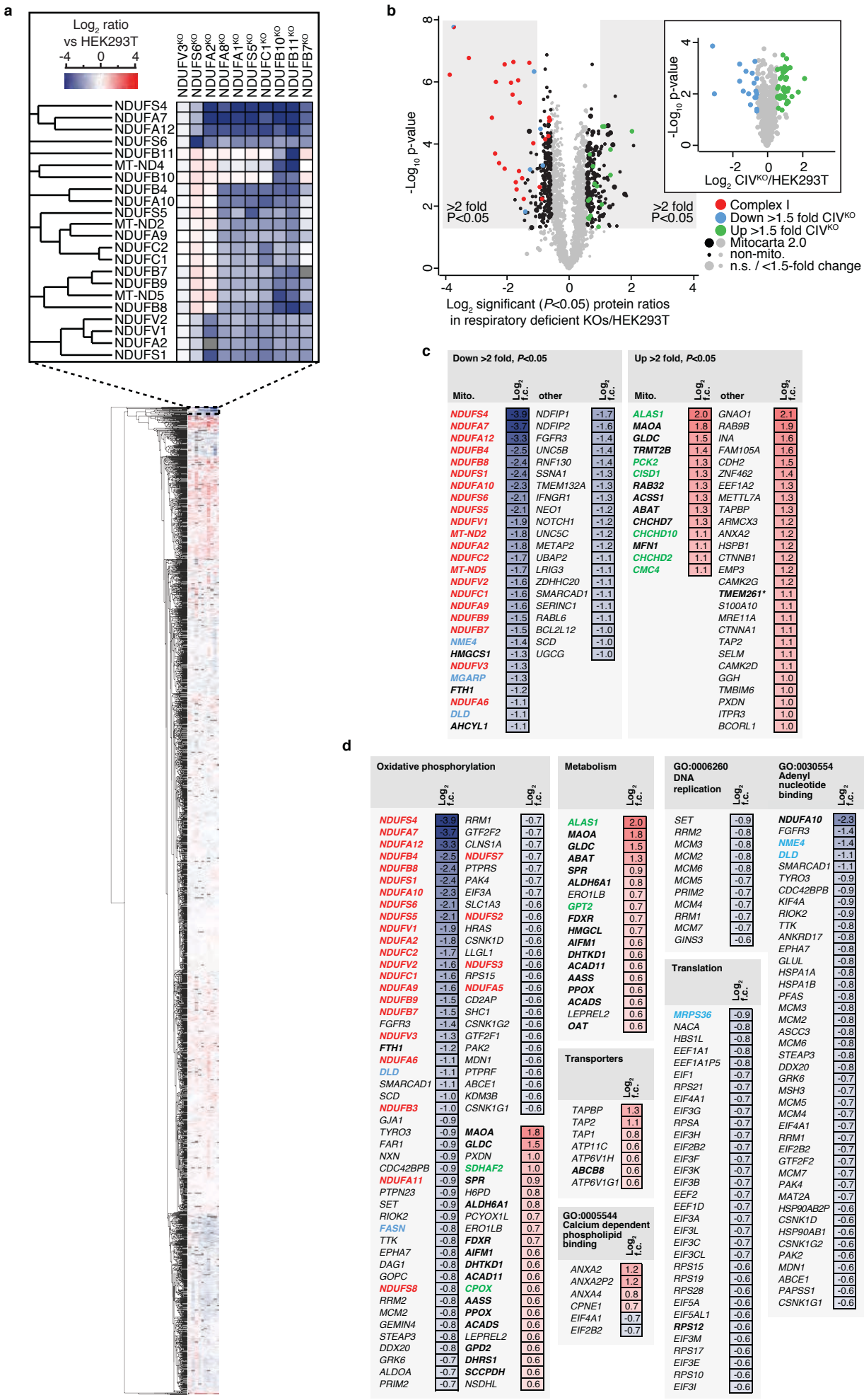
Stroud ED Fig. 2



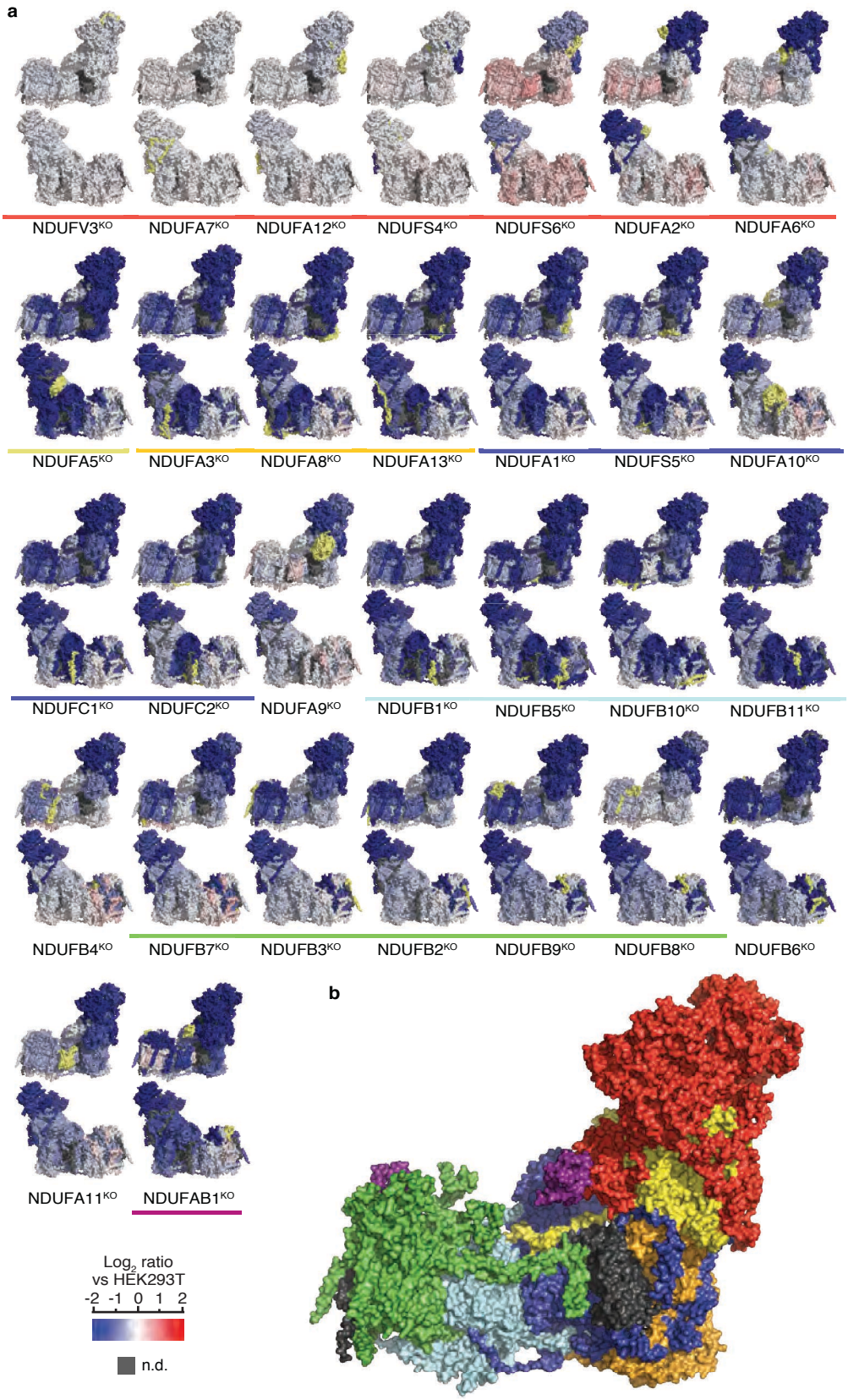
Stroud ED Fig. 3



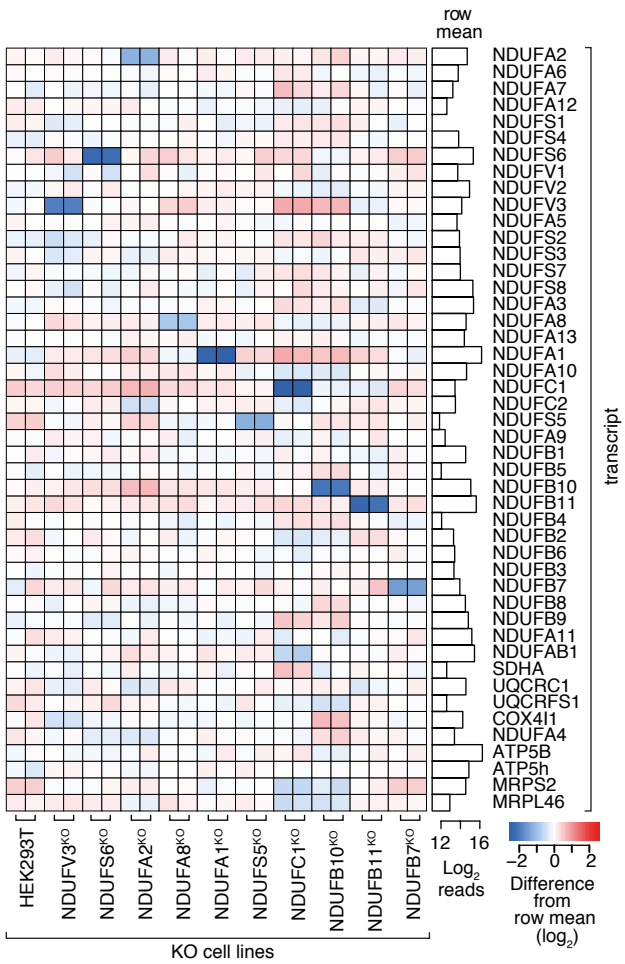
Stroud ED Fig. 4



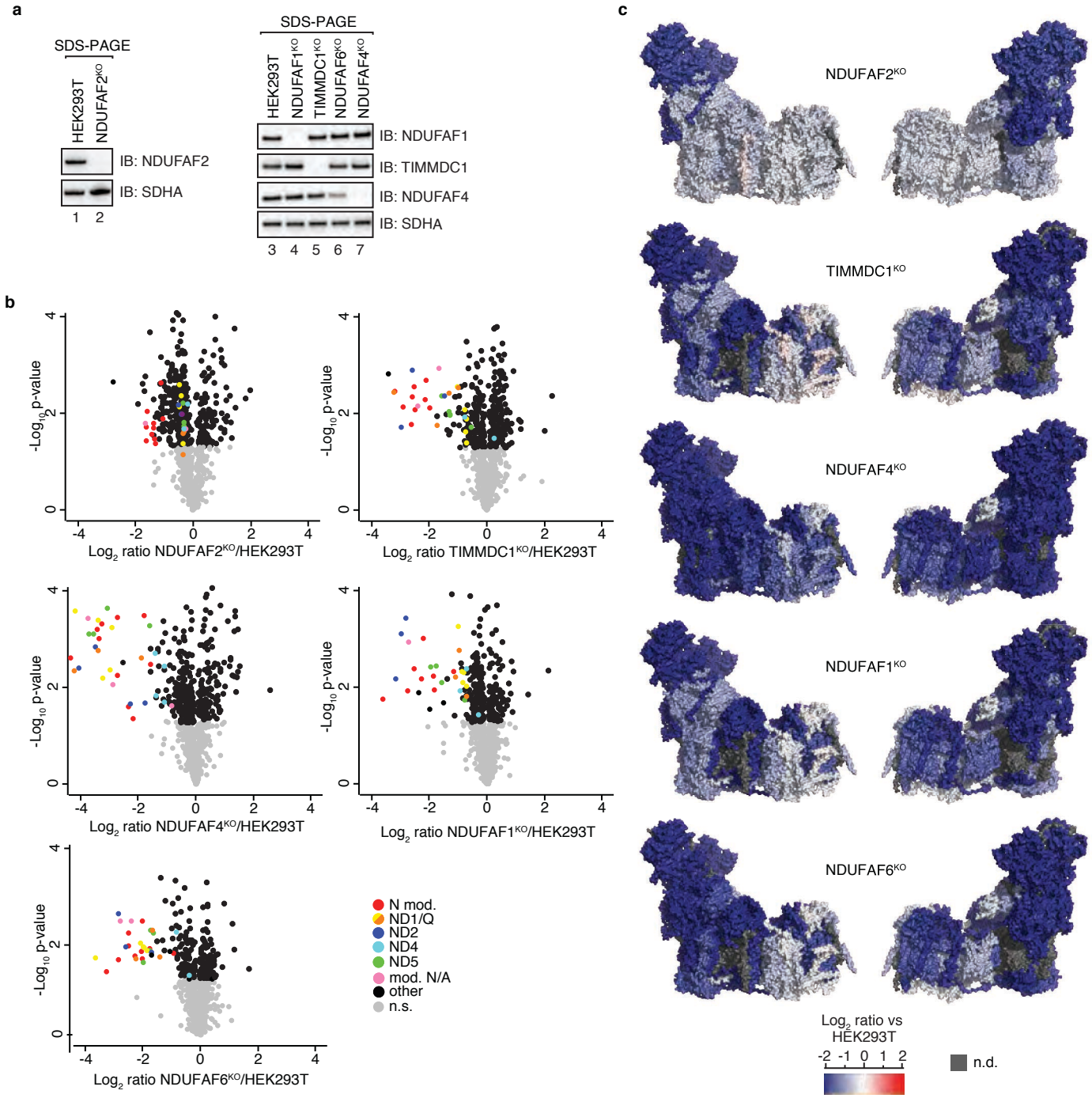
Stroud ED Fig. 5



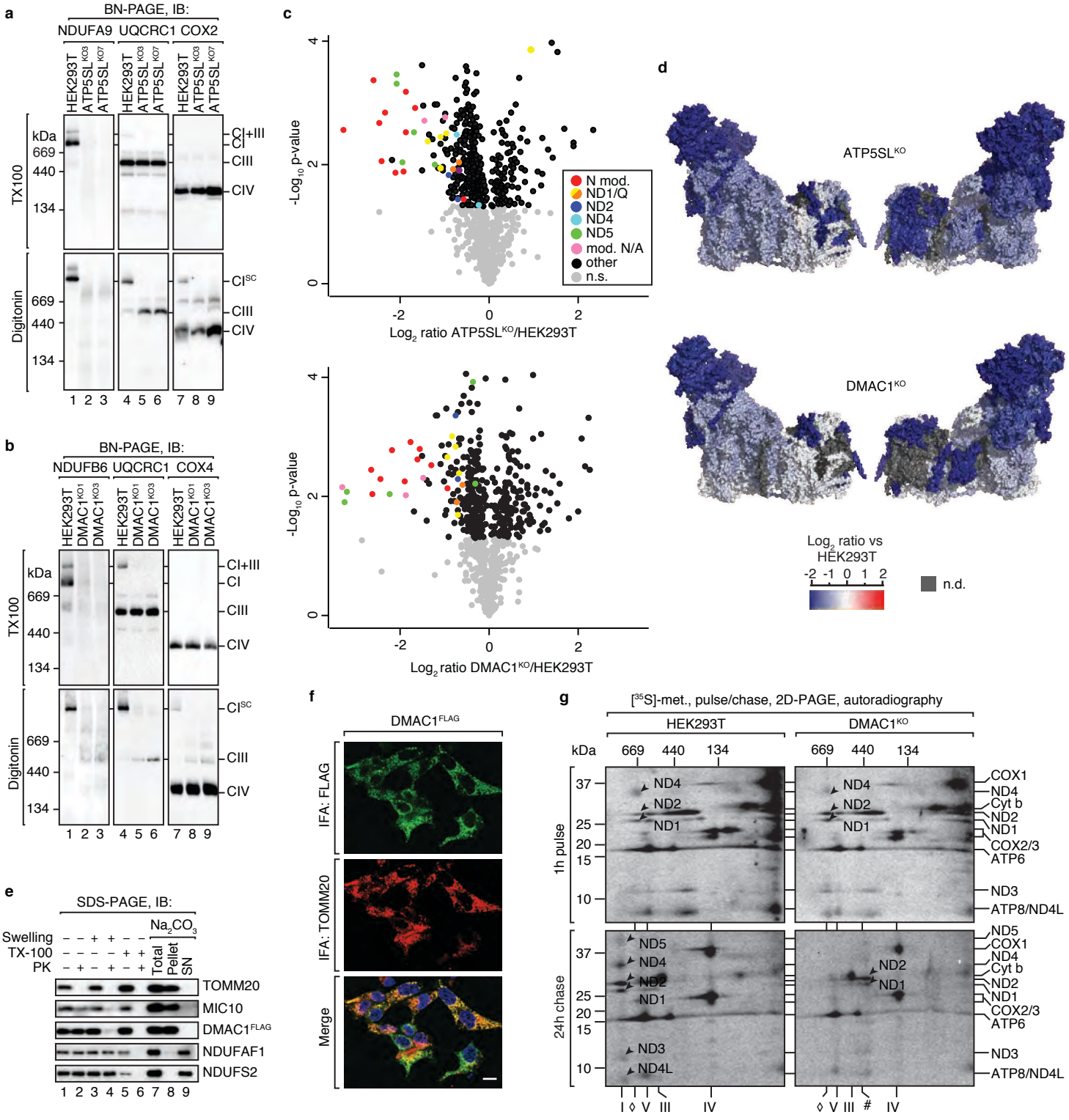
Stroud ED Fig. 6



Stroud ED Fig. 7



Stroud ED Fig. 8



Stroud ED Fig. 9

Author/s:
Stroud, DA; Surgenor, EE; Formosa, LE; Reljic, B; Frazier, AE; Dibley, MG; Osellame, LD; Stait, T; Beilharz, TH; Thorburn, DR; Salim, A; Ryan, MT

Title:
Accessory subunits are integral for assembly and function of human mitochondrial complex I

Date:
2016-10-06

Citation:
Stroud, D. A., Surgenor, E. E., Formosa, L. E., Reljic, B., Frazier, A. E., Dibley, M. G., Osellame, L. D., Stait, T., Beilharz, T. H., Thorburn, D. R., Salim, A. & Ryan, M. T. (2016). Accessory subunits are integral for assembly and function of human mitochondrial complex I. NATURE, 538 (7623), pp.123-+. https://doi.org/10.1038/nature19754.