Untargeted proteomics enables ultra-rapid variant prioritization in mitochondrial and other rare diseases

Daniella H. Hock[1,2,3,*], Nikeisha J. Caruana[1,4], Liana N. Semcesen[1], Nicole J. Lake[5], Luke E. Formosa[6], Sumudu S. C. Amarasekera[2], Tegan Stait[2], Simone Tregoning[2], Leah E. Frajman[2], David R. L. Robinson[1], Megan Ball[2,10,12], Boris Reljic[1,6], Bryony Ryder[7], Mathew J. Wallis[8,9], Anand Vasudevan[10], Cara Beck[3], Heidi Peters[11,12], Joy Lee[3,11], Natalie B. Tan[2,3,12], Mary-Louise Freckmann[13], MitoMDT Diagnostic Network for Genomics and Omics[**], Vasiliki Karlaftis[2,12], Chantal Attard[2,12], Paul Monagle[2,12,14,15], Amanda Samarasinghe[2], Rosie Brown[2,3], Weimin Bi[16,17], Monkol Lek[5], Robert McFarland[18,19], Robert W. Taylor[18,19], Michael T. Ryan[6], Zornitza Stark[2,3,12], John Christodoulou[2,3,12], Alison G. Compton[2,3,12], David R. Thorburn[2,3,12*], David A. Stroud[1,2,3*]

Affiliations

[1]Department of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Melbourne, VIC, 3010, Australia

[2]Murdoch Children's Research Institute, Melbourne, VIC 3052, Australia

[3]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, VIC, 3052, Australia

[4]Institute for Health and Sport (iHeS), Victoria University, Melbourne, VIC, 3011, Australia

[5]Department of Genetics, Yale School of Medicine, New Haven, CT 06510 USA

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

[7]Paediatric and Adult National Metabolic Service, Te Toka Tumai, Te Whatu Ora Health New Zealand, Auckland, New Zealand

[8]Tasmanian Clinical Genetics Service, Tasmanian Health Service, Hobart, TAS 7001, Australia

[9]School of Medicine and Menzies institute for Medical Research, University of Tasmania, Hobart, TAS 7001, Australia

[10]Royal Women's Hospital, Melbourne, VIC, 3052, Australia

[11]Department of Metabolic Medicine, Royal Children's Hospital, Melbourne, VIC 3052, Australia

[12]Department of Paediatrics, University of Melbourne, Melbourne, VIC 3010, Australia

[13]Department of Clinical Genetics, The Canberra Hospital, Canberra, Australian Capital Territory, Australia

[14]Department of Haematology, Royal Children's Hospital, Parkville VIC, Australia
[15]Kids Cancer Centre, Sydney Children's Hospital, Randwick, NSW, Australia
[16]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030, USA.
[17]Baylor Genetics, Houston, TX, 77021, USA.
[18]Mitochondrial Research Group, Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, NE2 4HH, United Kingdom
[19]NHS Highly Specialised Service for Rare Mitochondrial Disorders, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, NE1 4LP, UK

*To whom correspondence should be addressed. Tel: +61-3-8344-8317 Email: daniella.hock@unimelb.edu.au (Daniella H. Hock) Tel: +61-3-8341-6235 Email: david.thorburn@mcri.edu.au (David R. Thorburn); Tel: +61-3-8344-7316; Email: david.stroud@unimelb.edu.au (David A. Stroud).

**MitoMDT Diagnostic Network for Genomics and Omics members are listed at the end of the manuscript.

Abstract

Only half of individuals with suspected rare diseases receive a definitive genetic diagnosis following genomic testing. A genetic diagnosis allows access to appropriate patient care and reduces the number of potentially unnecessary interventions and related healthcare costs. Here, we demonstrate that an untargeted quantitative mass-spectrometry approach quantifying >6,000 proteins in primary fibroblasts representing >80% of known mitochondrial disease genes can provide functional evidence for 83% of individuals in a cohort of known primary mitochondrial diseases. We profiled >90 individuals, including 28 with confirmed disease and diagnosed 6 individuals with variants in both nuclear and mitochondrial genes. Lastly, we developed an ultra-rapid proteomics pipeline using minimally invasive peripheral blood mononuclear cells to support upgrade of variant pathogenicity in as little as 54 hours in critically ill infants with suspected mitochondrial disorders. This study supports the integration of a single untargeted proteomics test into routine diagnostic practice for the diagnosis of rare genetic disorders in clinically

actionable timelines, offering a paradigm shift for the functional validation of genetic variants.

Introduction

Despite advances in genomic sequencing approaches, only 35-70% of individuals with suspected rare disease receive a molecular diagnosis following whole exome sequencing (WES) or whole genome sequencing (WGS)[1-5]. Negative cases generally remain undiagnosed due to variants being refractory to short-read sequencing approaches or when detected variants of uncertain significance (VUS, Class 3) are detected, requiring functional evidence to be upgraded to likely pathogenic (Class 4) or pathogenic (Class 5). Mitochondrial disease is an umbrella term for a group of over 300 rare monogenic disorders affecting mitochondrial energy production in the form of ATP[6]. These disorders can arise from sporadic or inherited variants in either nuclear or mitochondrial DNA (mtDNA), presenting at any stage of life with a myriad of symptoms affecting either a single organ or in a multisystemic manner[7]. Mitochondrial diseases affect approximately 1 in 5,000 live births[8] with limited treatments available, reinforcing the importance of a genomic diagnosis for early intervention in affected individuals. A genomic diagnosis also facilitates informed reproductive options such as prenatal diagnosis (PND) and assisted reproductive technologies such as preimplantation genetic testing (PGT) and mitochondrial replacement therapy (MRT)[9].

Functional approaches to assess VUS pathogenicity in mitochondrial disease have historically relied on targeted and low throughput tests such as respiratory chain enzymology (RCE), SDS-polyacrylamide gel electrophoresis (PAGE) or Blue Native (BN)-PAGE and immunoblotting. RCE assesses the activity of the mitochondrial respiratory chain complexes I-IV and is typically normalised to the activity of a single enzyme, citrate synthase (CS), to account for variability in mitochondrial content determined by sample quality, amount and storage[10]. Western blotting has also been used for the confirmation of VUS pathogenicity based on protein abundance or differential size, although it relies on having a strong genetic lead for disease causation from genomic data and the commercial availability of antibodies for the desired protein.

99 Other non-targeted functional approaches such as transcriptomics and proteomics have
100 also been applied to the diagnosis of mitochondrial diseases. In selected cohorts,
101 transcriptomics has been shown to increase the diagnosis of mitochondrial disorders by
102 10-16%[11, 12] and other rare Mendelian disorders by 7-17%[13-15]. Although transcriptomic
103 analysis can be a powerful tool in assessing pathogenicity of intronic and splice variants,
104 this approach has limited power in offering functional information on missense variants,
105 one of the most common and challenging classes of variants to assess[16, 17]. Splice
106 variants are predicted to account for only 10% of pathogenic variants associated with
107 autosomal recessive disorders[18], in line with current diagnostic rates achieved by
108 transcriptomics. Missense variants, on the other hand, account for 60% of pathogenic
109 variants[18] and approximately 40% of them are expected to result in reduced protein levels
110 due to protein instability and turnover[19]. Proteomics can detect such changes and has
111 been demonstrated to contribute to the diagnosis of Mendelian disorders by providing
112 functional evidence for not only missense variants but also splice, deep intronic and copy
113 number variants [11, 20-25]. Despite this, there is a lack of data to inform the general utility of
114 proteomics in rare disease diagnoses as an untargeted approach like genomics and
115 transcriptomics.
116 Here, we present a systematic analysis demonstrating the utility of proteomics in the
117 detection of rare mitochondrial disorders and provide the first validated pipeline for ultra-
118 rapid functional testing using peripheral blood mononuclear cells (PBMCs). Reference
119 proteomics data and associated bioinformatics tools for performing the proteome analysis
120 approaches developed for these investigations can be applied to user-uploaded data
121 through our interactive web tool (https://rdmassspec.shinyapps.io/RDMSExplorer/).
122
123 ***Results***
124 To validate our untargeted proteomic approach in the detection of mitochondrial disorders,
125 we first analysed primary fibroblast cell lines from a cohort of 24 patients with confirmed
126 genetic diagnoses based on known pathogenic or likely pathogenic nuclear or mtDNA-
127 encoded variants. This validation cohort (VC) included cell lines with defects in subunits
128 or assembly factors for each of the five individual OXPHOS complexes as well as defects
129 of mtDNA translation that impact multiple complexes, causing a combined OXPHOS

130 biochemical defect (**Fig 1A, Supplementary Table 1**). Principal component analysis
131 showed that proteomes segregate according to their primary defect (**Fig. 1B**) and
132 hierarchical clustering of OXPHOS subunits and assembly factors identified clear profiles
133 (**Supplementary Fig. 1A, B**). We have previously shown similar profiles to be present in
134 gene-edited HEK293T knockout of Complex I subunits[26]. Comparison of mitochondrial
135 proteomes from our validation cohort of fibroblast cell lines and corresponding gene-
136 edited HEK293T knockout cell lines (knockout cohort, KC) showed good correlation of
137 dysregulated proteins (**Supplementary Fig. 2A, B**).
138
139 The maximum theoretical coverage of known mitochondrial disease gene products from
140 the fractionated library was 80%, with a routine coverage of 64% in a standard 95-minute
141 library-supported DIA run from fibroblast samples (**Fig. 1C**). Our approach could also
142 detect proteins corresponding to 37% of the Mendeliome (i.e., encoded by known disease
143 genes) in a standard 95-minute run, supporting the potential utility of proteomics as a
144 disease agnostic test in the resolution of many rare monogenic disorders. In terms of
145 variant type in the validation cohort, the most common type was nuclear single nucleotide
146 variants (48%) followed by mtDNA single nucleotide variants (16%), indel (16%), splice
147 site (13%) and deep intronic (6%) (**Fig. 1D**). We quantified the abundance of the protein
148 of interest in 67% of our investigations with the most common outcome being the protein
149 abundance readily quantified in controls and absent in proband (42%) followed by a
150 decrease in the patient relative to controls (21%) (**Fig. 1E**). In 29% of cases the protein
151 of interest was not detected in either proband or control fibroblasts. For the cases where
152 the protein was identified in both proband and control fibroblasts, analysis of the standard
153 deviation show that ND1 abundance in VC6 (*MT-ND1*) was 97% relative to control median
154 and within the control distribution, while ND5 abundance in VC12 (*MT-ND5*) was 118%
155 relative to control median and approximately 3 standard deviations above the control
156 median (**Fig. 1F**). Other identified proteins were outside the control range and >9 standard
157 deviations below control median (**Fig. 1F**).
158
159

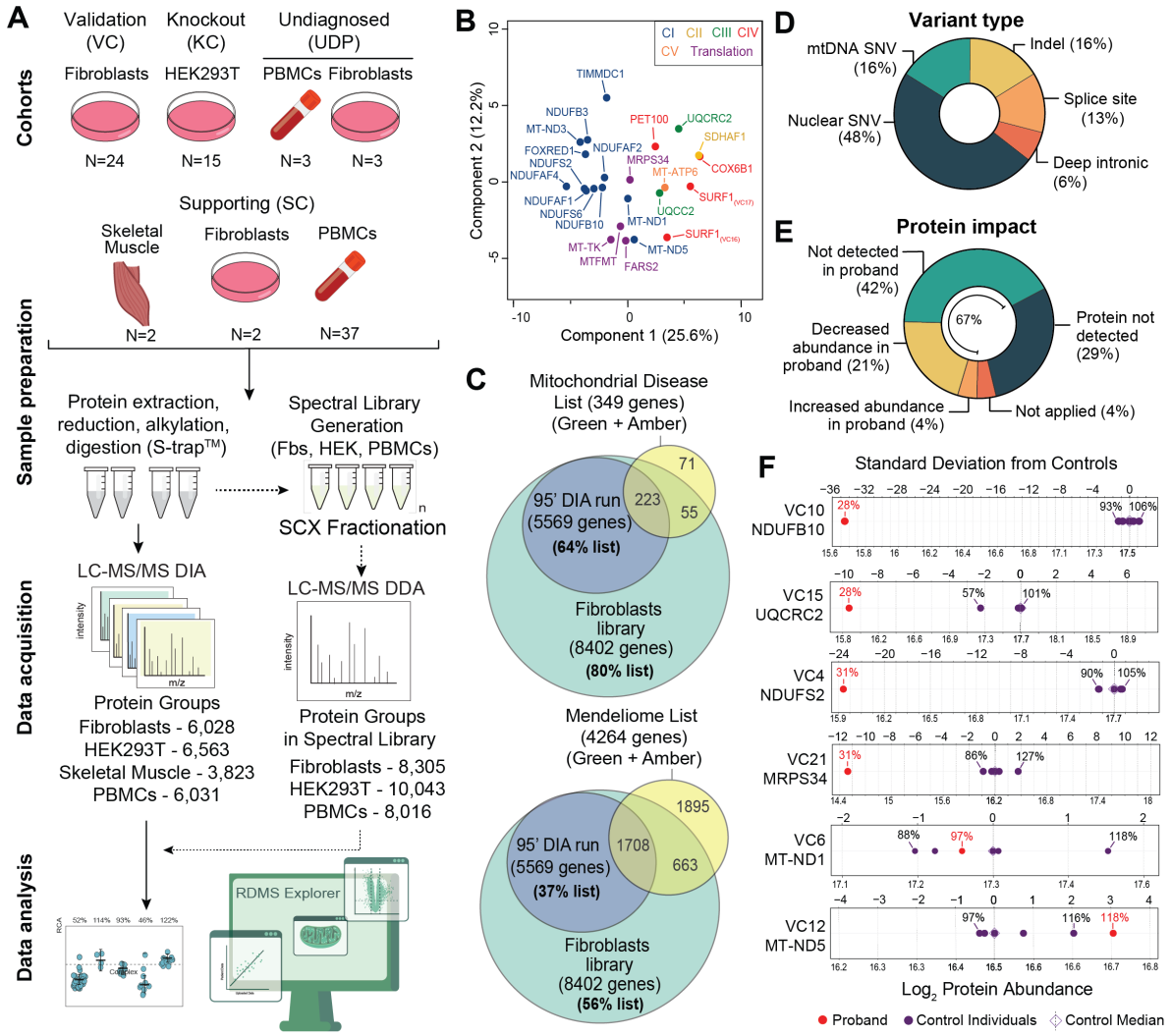

Figure 1. Study design and validation cohort analysis A – Quantitative proteomics experimental design overview. A fibroblast validation cohort (VC), HEK293T knockout cohort (KC), undiagnosed patient cohort (UDP) and supporting cohort (SC) prepared using S-Trap™ columns. Digested peptides were subjected to Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) Data Independent Acquisition (DIA). Pooled control peptides were fractionated using strong cation-exchange (SCX) chromatography and data acquired as Data Dependent Acquisition (DDA) to generate spectral libraries for fibroblasts (Fbs), HEK293T (HEK) and PBMCs. Raw data were searched using Spectronaut® software and data analyses were performed with a combination of Perseus software, Python and R. **B –** Principal Component Analysis (PCA) of the fibroblast validation cohort (N=24) relative to controls based on the differential abundance of whole cell proteins calculated from t-test. **C –** Venn diagram showing the coverage of Mendeliome (v. 0.12869, top panel) and Mitochondrial disease (v. 0.787, bottom panel) lists including green (diagnostic-grade) and amber (borderline diagnostic-grade) genes retrieved from PanelApp Australia[27]. **D –** Summary of the genetic variant types analysed in the validation cohort (VC). **E –** Summary of the findings in the VC based on protein identification and abundance. In 42% (10/24) of the cases the protein expected to be affected by the genetic variant is not detected in the patient while it is detected in the controls. In 29% (7/24) of the cases, the protein is not detected via quantitative proteomics, in 21% (5/24) of the cases, the protein is decreased in abundance in the patient compared to the 5 controls, and in 4% (1/24) of the cases, the protein

is increased compared to the controls analysed. SNV = single nucleotide variant. **F –** Protein standard deviation from control median of the respective affected gene in controls (purple) and probands (red) in the validation cohort (VC). Standard deviation was calculated from the median control variance.

Mitochondrial content has previously been calculated in proteomics data using the mean abundance of experimentally validated mitochondrial proteins in the MitoCarta 3.0 database[28, 29]. We noticed that the mean abundance of mitochondrial proteins as a group varied between cell lines in our validation cohort (**Supplementary Fig. 3A**) but not control individuals (**Supplementary Fig. 3B)**. This phenomenon is typically taken into consideration in enzymology analyses by expressing enzyme rates as citrate synthase (CS) ratios to account for mitochondrial proliferation or cell/tissue variability[10, 30]. CS enzyme activity in our validation cohort varied greatly (**Supplementary Fig. 3A, upper panel**), consistent with altered mitochondrial content. Surprisingly, while CS activity did have a weak positive correlation with the relative abundance of CS protein as detected by proteomics (**Supplementary Fig. 3C, left panel**), it did not correlate with mitochondrial content calculated from the abundance of mitochondrial proteins (**Supplementary Fig. 3C, middle panel**). In line with this, CS protein abundance also did not correlate with mitochondrial content calculated from proteomics data (**Supplementary Fig. 3C, right panel**). Since the nature of proteomics-based RCA is relative protein abundance, we concluded that mean mitochondrial protein abundance is the appropriate metric to normalise proteomics data. We built a differential mitochondrial abundance correction step into our RCA calculations (see **Materials and Methods** for detailed information). This correction is of benefit for cell lines that may have different mitochondrial content to controls and corrects for mitochondrial content before calculating the relative abundance of each complex.

We next sought to systematically benchmark our RCA analysis against clinical RCE. RCE was performed on 23 of the 24 fibroblast lines in the validation cohort, and was also performed on skeletal muscle (SKM) or lymphoblastoid cell lines (LCLs) for 17 of the 24 patients. Results are presented relative to CS (**Supplementary Table 2**) and the Bernier criteria[31] were used to classify results. Fibroblast RCE detected a definite defect (<30%, major criterion) in the expected complex(es) for 19 of 24 fibroblast lines (79%) while three

210 [VC4 *(NDUFS2),* VC7 *(NDUFAF1)* and VC10 *(NDUFB10*)] had a probable defect (minor
211 criterion) and RCE did not detect a defect in VC19 (*COX6B1*) fibroblasts (**Fig. 2A left**
212 **panel**, **Supplementary Table 2**). When skeletal muscle or LCLs were also available,
213 RCE revealed a major defect in 9 of 17 of the analyses (53%), and a minor defect in four
214 (23%) while three were not deficient (18%) and one was not available (6%)(**Fig. 2A, right**
215 **panel)**.
216
217 Relative complex abundance (RCA) analysis was performed incorporating the correction
218 for mitochondrial content (**Supplementary Fig. 4**) as described above. We classified a
219 major defect as an RCA abundance ≤65% relative to controls or ≤75% with absent
220 detection of the protein of interest in the proband with >2 peptides detected in controls
221 while minor defects were classified by an RCA abundance of ≤75% relative to controls
222 (**Supplementary Table 2**). Using these criteria, a major RCA defect was detected in 83%
223 of the cell lines in the validation cohort, a minor defect in a single case, VC19 (*COX6B1*),
224 and no defect was detected in three cell lines with mtDNA variants (**Fig. 2B**). Overall,
225 there is no strong correlation between RCE activity and an RCA value relative to control
226 (R=0.072, p=0.78) (**Fig. 2C**).
227
228 For isolated Complex I disorders, 83% (10/12 patients) had a complex I abundance lower
229 than 75% relative to controls analysed (p<0.0001). Two of the three Complex I-deficient
230 fibroblast lines harbouring missense mtDNA-encoded variants were refractory to RCA
231 analysis but diagnostic by RCE. mtDNA variant heteroplasmy in fibroblasts was 92% for
232 VC6 *MT-ND1* and 71% for VC12 *MT-ND5*, highlighting a limitation of our approach in the
233 detection of some mtDNA-driven disorders. Importantly, in all 10 fibroblast lines where
234 RCA was diagnostic for a Complex I disorder, lower abundance (<65% relative to
235 controls) of Complexes III, IV and V were not noted (**Supplementary Fig. 4**). While
236 Complex II was reduced to <75% in four of these, RCE similarly identified a probable
237 Complex II defect for one (*FOXRED1*), supporting the accuracy and specificity of RCA
238 analysis. In 7/7 individuals with isolated Complex II, III or IV disorders, the abundance of
239 the relevant complex was reduced to <75% controls (**Fig. 2D**) however t-test significance
240 varied, especially in CII and IV defects where the number of detected subunits was as

241 low as two. All were specific for the relevant isolated defect except for *UQCRC2* where a
242 Complex I abundance of 67% control (p<0.0001) was noted, as expected for Complex III
243 disorders[32].
244
245 For isolated disorders involving Complexes I-IV due to nuclear defects, we determine an
246 RCA value in fibroblasts of <65% controls to be diagnostic. The same threshold may also
247 be applied to isolated disorders caused by pathogenic variants in mtDNA including those
248 leading to Complex V deficiency, however smaller reductions may not exclude
249 pathogenicity as seen in VC6 (*MT-ND1*), VC12 (*MT-ND5*) and VC20 (*MT-ATP6*) (**Fig.**
250 **2D**). Interestingly, in the case of VC6 this was likely not due to heteroplasmy, which was
251 92% NC_012920.1(*MT-ND1*):m.3949C>T in fibroblasts. On the other hand, in VC12 the
252 NC_012920.1(*MT-ND5*):m.13513G>A variant found at 71% in fibroblasts, did not show
253 the expected RCA defect while it satisfied the major criteria in RCE (**Supplementary**
254 **Table 2**), suggesting impaired enzyme kinetics in the absence of a structural Complex I
255 defect. Finally, we found that defects directly impacting mitochondrial translation such as
256 *MRPS34* fibroblasts presenting with a destabilised mitochondrial ribosome small
257 subunit[25], *MT-TK* which encodes the mitochondrial lysine transfer RNA[33], *MTFMT* which
258 encodes the mitochondrial methionyl-tRNA formyltransferase[34], and *FARS2* encoding
259 mitochondrial phenylalanyl-tRNA synthetase, led to combined Complex I and IV defects
260 that each met criteria of an RCA value <65% control (**Fig. 2D**), consistent with the major
261 criterion observed in RCE (**Supplementary Table 2**). Taken together, fibroblast RCA
262 detected an OXPHOS defect at <65% relative to controls in the expected complex(es) for
263 20 of 24 fibroblast lines tested (83%) (**Fig. 2B**). One fibroblast line, VC19 (*COX6B1*), had
264 an RCA of 71% for Complex IV relative to controls, which is in line with a previous report
265 showing remaining assembled complex[35] and the peripheral position of the late
266 assembled COX6B1 subunit in the structure of Complex IV[36, 37]. Three fibroblast lines
267 where a definitive defect was not observed by RCA harboured variants in mtDNA,
268 suggesting that this class of variant may be more refractory to RCA analysis. Despite this,
269 proteomics-based RCA outperformed RCE in the detection of primary mitochondrial
270 disease. It is important to note here that the nature of proteomics as an untargeted
271 approach means that additional analyses can be performed from the same proteomic

data, such as reduced abundance of specific proteins encoded by the gene of interest (**Fig. 1E, F**) and co-dependent proteins in structural modules (**Supplementary Fig. 1B**). This is exemplified in the VC2 (*NDUFS6*) case where despite showing a Complex I abundance of 74%, topographical mapping of Complex I subunit abundances against the cryo-EM structure show specific reduction of the N-module of Complex I, providing strong evidence in supporting disease causation.

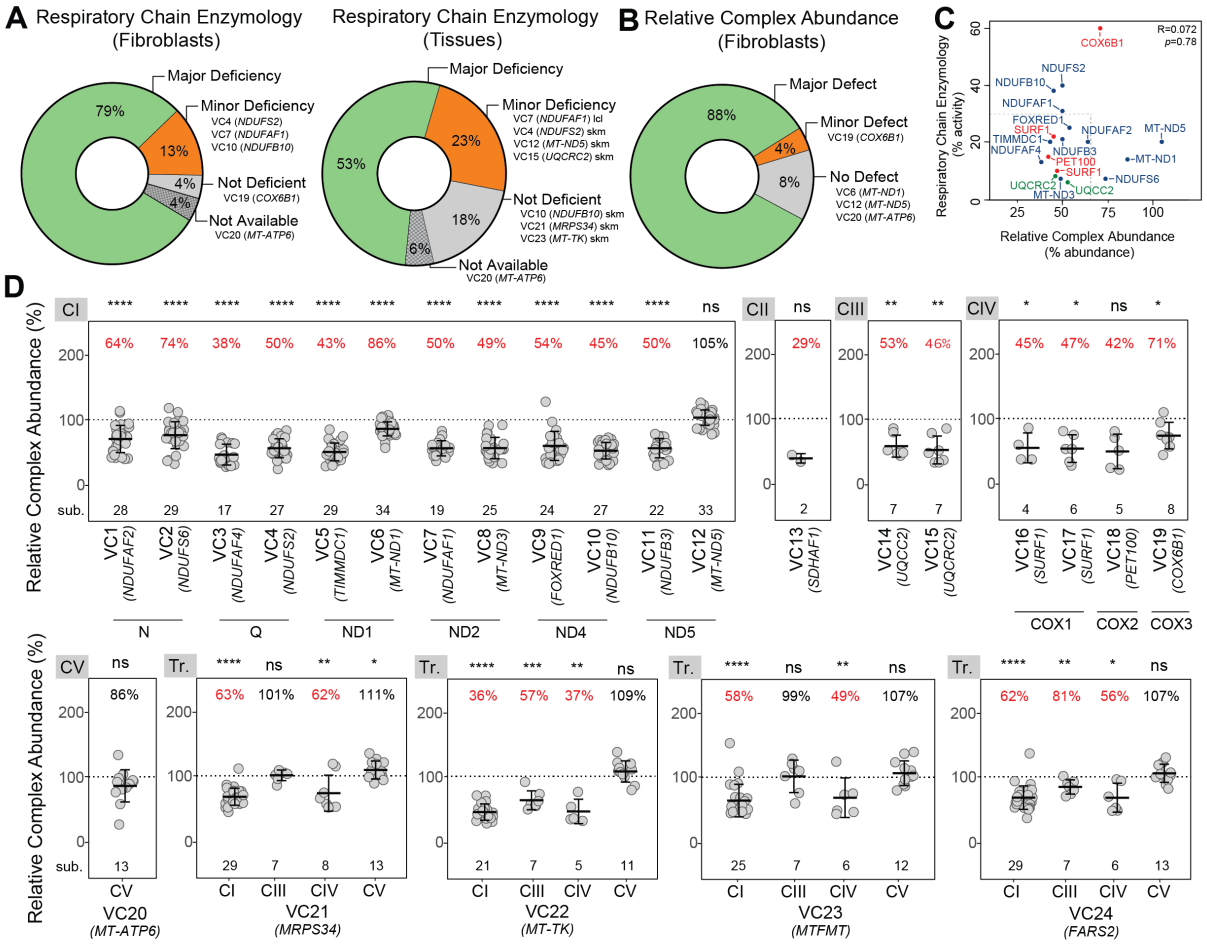

Figure 2. Proteomics outperforms clinical respiratory chain enzymology in detection of mitochondrial disorders A – (left) Summary of the RCE results in fibroblasts according to the Bernier criteria[31]. (right) Summary of the RCE results in tissues (skeletal muscle, skm or lymphoblastoid cell line, LCL) according to the Bernier criteria[31]. **B –** Summary of the RCA results from quantitative proteomic data. **C –** Pearson correlation between RCA results and Respiratory Chain Enzymology results for Complexes I, III and IV. **D –** RCA results of OXPHOS complexes from the validation cohort showing the predicted affected complexes for each cell line. CI-V = Complex I-V. Tr. = translation. Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = p<0.0001, *** = p<0.001, ** = p<0.01, * = p<0.05, ns = not significant, p>0.05.

290 After benchmarking quantitative proteomics and RCA analysis against the validation
291 cohort, we applied our label-free DIA method to six undiagnosed cases where primary
292 fibroblast or skeletal muscle were available for proteomic analysis.
293
294 **UDP1 (*MT-ATP6*)**
295 UDP1 (*MT-ATP6*) was a four-year-old child who had intrauterine growth retardation,
296 truncal hypotonia, microcephaly, global developmental delay, left ventricular
297 noncompaction and Wolff-Parkinson-White syndrome. Biochemically there was a
298 marginal elevation of blood lactate (2.8 mmol/L; normal range 0.7 – 2.0) and 3-
299 methylglutaconic aciduria. Missense variants were identified in *MT-ATP6*
300 [NC_012920.1(*MT-ATP6*):m.8672T>C;p.(Leu49Pro)], a *de novo* variant with 81%
301 heteroplasmy in blood, and *ATAD3A* (heterozygous c.683C>T;p.Thr228Met) from WES
302 plus mtDNA sequencing from blood and both classified as a VUS. Clinical RCE assays
303 typically do not include Complex V activity and proteomics was sought to provide
304 functional evidence. Despite proteomic RCA analysis showing no reduction in abundance
305 of Complex V (104% relative to controls), there was a reduction in the number of Complex
306 V subunits identified from 13 to 10 compared to the diagnosed *MT-ATP6* patient as well
307 as other cases in the same batch (**Fig. 3A** and **Supplementary Fig. 4**). The Complex V
308 subunits that are not detected in UDP1 (*MT-ATP6*) but are detected in VC20 (*MT-ATP6*)
309 are ATP5IF1, the ATPase inhibitor, and ATP5MD (ATP5MK/DAPIT) and ATP5MPL
310 (ATP5MJ/MP68/6.8PL), which like ATP6, are incorporated at a late stage in the assembly
311 of Complex V[36, 38]. The latter suggests a specific turnover of late stage assembled
312 subunits of Complex V below the limit of detection of proteomics in the proband, which
313 contributes to the unchanged Complex V levels seen in the RCA of UDP1 (*MT-ATP6*).
314 We processed the same data using q-value sparse run-wise imputation in the
315 Spectronaut® that allows for peptides that were identified in one sample to be imputed at
316 the limit of detection in samples where they were not detected. This allowed us to estimate
317 the fold changes of ATP6 (*MT-ATP6*) and ATP5MPL proteins in UDP1 (*MT-ATP6*) and
318 correlate their reduced levels with VC20 (*MT-ATP6*) as well as confirming unchanged
319 levels of ATAD3A (**Fig. 3B**). Another data analysis strategy we explored for UDP1 (*MT-*
320 *ATP6*) was the use of an *in-silico* library-free analysis including the protein sequence of

321 the ATAD3A variant p.Thr228Met. As expected, the peptide containing the methionine at
322 position 228 was only detected in the patient (**Fig. 3C**) and present at ~43% of the
323 abundance relative to the canonical peptide expressed from the other allele
324 (**Supplementary Fig. 5A**). Taken together, these results suggest that the *MT-ATP6*
325 p.Leu49Pro variant is the likely diagnosis for this case.
326
327 The lack of a Complex V defect in proteomics RCA analysis of both *MT-ATP6* cases
328 (VC20 and UDP1) prompted us to analyse additional patients with known pathogenic
329 variants affecting Complex V. Supporting cohort probands SC1 and SC2 were two
330 unrelated infants presenting with lactic acidosis, cardiac abnormalities (persistent patent
331 ductus arteriosus, ventriculo-septal defect and cardiomyopathy), mild dysmorphic facial
332 features, hypospadias, hyperammonaemia and 3-methyl-glutaconic aciduria. Targeted
333 gene sequencing based on this phenotype identified in both individuals a homozygous
334 intronic founder variant NM_017866.6(*TMEM70*):c.317-2A>G classified as Class
335 5/pathogenic, with several patients previously reported[39-43]. Skeletal muscle samples
336 were available for both cases and compared against three unrelated individuals using the
337 library-free DIA approach. Whole muscle proteomic analyses showed reduction of several
338 Complex V subunits in SC1 and SC2 respectively (**Fig. 3D** and **E**). RCA analyses
339 quantified the relative abundance of Complex V at 19% and 20% (**Fig. 3F and G**),
340 demonstrating that defects in Complex V can be detected using RCA analysis. TMEM70
341 has also been implicated in Complex I assembly[44, 45], although no defects in Complex I
342 abundance were noted in these muscle samples.
343

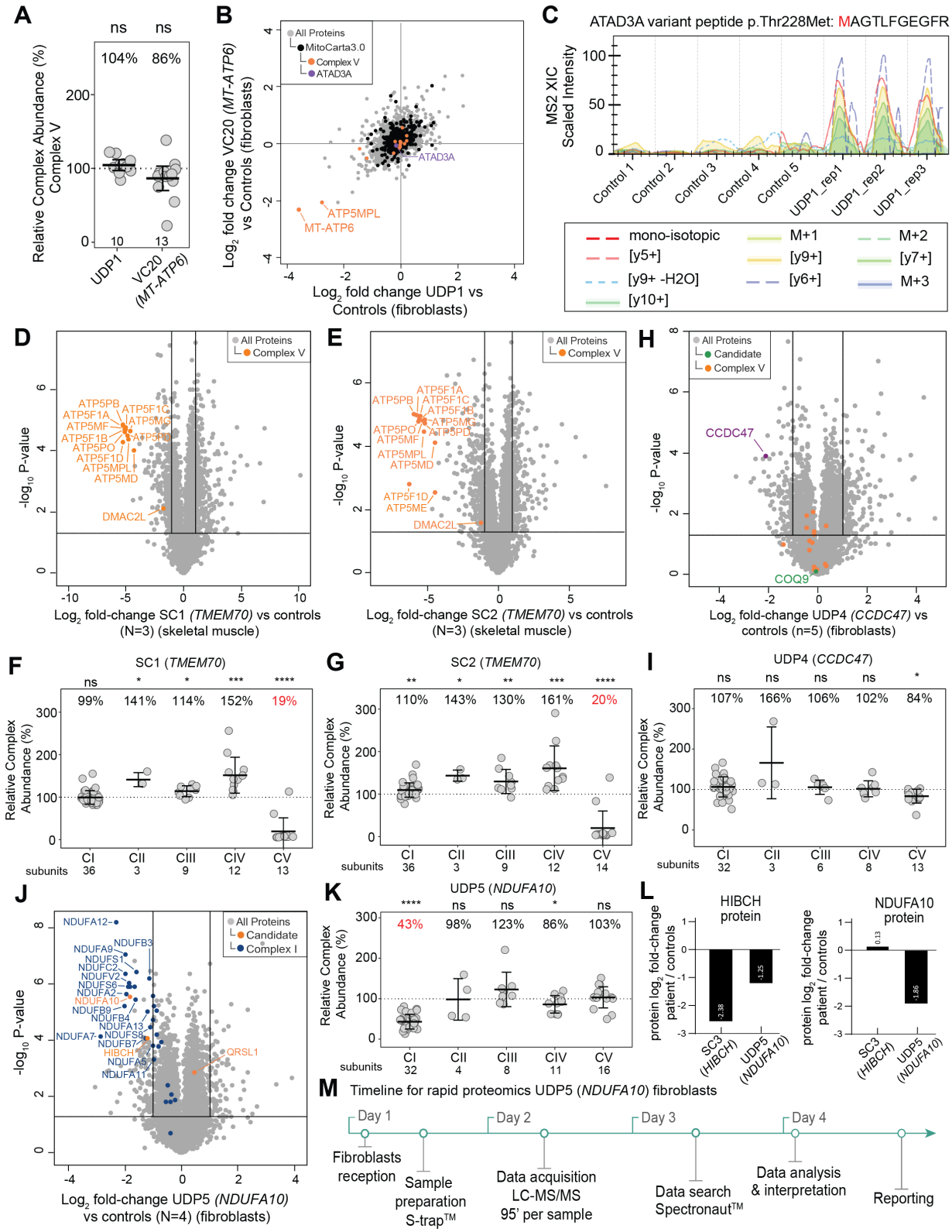

Figure 3. Proteomics supports the diagnosis of patients with suspected mitochondrial disorders. A – Relative Complex Abundance (RCA) of Complex V subunits in undiagnosed patient UDP1 (*MT-ATP6*) and diagnosed VC20 (*MT-ATP6*) patient. Middle bar represents mean complex abundance. Upper and

lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. ns = not significant, $p>0.05$. **B –** Correlation between $\log_2$ fold-changes from whole-cell proteins from run-wise imputed data between undiagnosed patient UDP1 (*MT-ATP6*) and diagnosed patient VC20 (*MT-ATP6*) harbouring mutations in *MT-ATP6* and controls showing reduced abundance in late assembly proteins, MT-ATP6 and ATP5MPL, for both patients. **C –** Spectral intensity of the peptide containing the ATAD3A p.Thr228Met variant detected in the UDP1 patient but not in controls. **D –** Volcano plot of whole cell proteins from SC1 (*TMEM70*) whole-cell skeletal muscle compared to controls (N=3) showing reduced abundance of Complex V subunits (orange dots). Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent from a two-sample t-test. **E –** Volcano plot of whole cell proteins from SC2 (*TMEM70*) whole-cell skeletal muscle compared to controls (N=3) showing reduced abundance of Complex V subunits (orange dots). Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent from a two-sample t-test. **F –** Relative Complex Abundance (RCA) of OXPHOS subunits in undiagnosed patient SC1 (*TMEM70*). Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = $p<0.0001$, *** = $p<0.001$, * = $p<0.05$, ns = not significant, $p>0.05$. **G –** Relative Complex Abundance (RCA) of OXPHOS subunits in undiagnosed patient SC2 (*TMEM70*). Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = $p<0.0001$, *** = $p<0.001$, ** = $p<0.01$, * = $p<0.05$. **H –** Volcano plot of whole cell proteins from UDP4 (*CCDC47*) fibroblasts compared to controls showing reduced abundance of CCDC47. No significant changes to Complex V proteins (orange). Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent from a two-sample t-test. **I –** Relative Complex Abundance (RCA) of OXPHOS complexes from UDP4 (*CCDC47*) fibroblasts compared to controls. * = $p<0.05$, ns = not significant, $p>0.05$. **J –** Volcano plot of whole cell proteins from UDP5 (*NDUFA10*) fibroblasts compared to controls showing reduced abundance of NDUFA10 and other structural subunits of Complex I. Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent. Blue = Complex I subunits. **K –** Relative Complex Abundance (RCA) of OXPHOS complexes from UDP5 (*NDUFA10*) fibroblasts compared to controls (N=4) showing isolated Complex I defect. Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = $p<0.0001$, * = $p<0.05$, ns = not significant, $p>0.05$. **L –** NDUFA10 and HIBCH protein abundance from two-sample t-test in SC3, a confirmed *HIBCH* proband and UDP5 (*NDUFA10*) whole-cell fibroblasts compared to controls showing an approximate half reduction in HIBCH levels in UDP5 (*NDUFA10, HIBCH* carrier) compared to SC3 and unchanged NDUFA10 levels in SC3. **M –** Timeline for the rapid proteomics from live fibroblast sample receipt to results for UDP5 (*NDUFA10*) case achieved in less than five days.

UDP4 (*CCDC47*)

UDP4 (*CCDC47*) was born to consanguineous parents, presenting with intrauterine growth restriction, severe failure to thrive, short stature, hypotonia, marked developmental delay, mildly dysmorphic features, severe pruritus woolly hair and increased plasma bile acids (224 μmol/L; normal range <7), Ricketts and decreased mitochondrial CIII activity in liver and muscle[46]. WES performed on DNA extracted from fibroblasts and analysed using a mitochondrial disease gene panel identified single heterozygous missense

392 variants NM_001358921.2(*COQ2*):c.460G>C; p.(Gly154Arg) and
393 NM_020312.4(*COQ9*):c.826C>T p.(Arg276Trp). Suspecting a compound oligogenic
394 mechanism impacting Coenzyme Q biosynthesis, RCA analysis was performed on
395 fibroblast cells. No abundance changes to COQ9 or respiratory chain complexes I-IV
396 were identified, although an abundance of 84% in Complex V was noted (**Fig. 3H** and **I**).
397 This reduction was regarded as not biologically significant as no rare variants in Complex
398 V related genes were flagged from sequencing data. Transcriptomic analysis performed
399 in RNA extracted from skeletal muscle showed a reduced abundance of *CCDC47*
400 transcript **(Supplementary Fig. 5B).** Re-analysis of proteomics data was sought to
401 provide functional evidence to support the transcriptomic findings, which confirmed a
402 reduction in the abundance of CCDC47 in primary fibroblasts from imputed data (**Fig. 3H**).
403 Re-analysis of sequencing data identified a homozygous variant
404 NM_020198.3(*CCDC47*):c.431C>G;p.(Thr144Asn). SpliceAI predicted (δ score 0.964)
405 the activation of a donor-splice site five nucleotides upstream of this variant, and PCR
406 analysis of the muscle cDNA showed the presence of both the missense transcript and a
407 second shorter one, missing 121 bp from the 3' end of exon 4 resulting in a loss-of-
408 function type transcript p.(Val143Glyfs*15) **(Supplementary Fig. 5C)**. This variant is not
409 observed in the gnomAD population database nor any clinical cases previously but is
410 predicted to be classified as likely pathogenic/Class 4, supported by the observed splicing
411 defect, as well as reduced abundance in *CCDC47* transcript and CCDC47 protein levels.
412 CCDC47 has been previously linked to Trichohepatoneurodevelopmental Syndrome[47]
413 (MIM 618268) which presents an overlapping phenotype to UDP4 (*CCDC47*). The
414 observed reduction in *CCDC47* transcripts due to incorrect splicing is in keeping with
415 previously reported patients, all of whom harbour loss-of-function variants[47, 48].
416
417 **UDP5 (*NDUFA10*)**
418 UDP5 presented with nystagmus, hypotonia, vomiting and weight loss at 20 weeks of age.
419 She had lactic acidosis, concentric left ventricular hypertrophy with 48% ejection fraction,
420 T2 hyperintense lesions in the substantia nigra and passed away at 24 weeks of age.
421 Targeted exome sequencing analysis flagged candidate variants
422 NM_004544.4(*NDUFA10*):c.914T>C; p.(Leu305Pro) homozygous VUS/Class 3,

423 NM_014362.4(*HIBCH*):c.891+1G>A, heterozygous likely pathogenic/Class 4 and
424 NM_018292.5(*QRSL1*):c.22G>C, p.(Glu8Gln) heterozygous VUS/Class 3. Fibroblast
425 proteomics detected NDUFA10 and HIBCH below 2-fold-change (representing >50%
426 reduction) in the patient relative to controls (**Fig. 3J and Supplementary Figure 5D**).
427 QRSL1 level was slightly increased but not significant, which led us to further investigate
428 *NDUFA10* and *HIBCH* leads. NDUFA10 is a structural subunit of Complex I and additional
429 functional evidence for pathogenicity could be drawn from reduced abundance of other
430 Complex I subunits, with RCA analysis quantifying Complex I abundance at 43% in UDP5
431 (*NDUFA10*) compared to controls, meeting the criteria for a major defect (**Fig. 3K**). We
432 further assessed whether the levels of HIBCH protein were compatible with a carrier
433 (heterozygous) state by analyzing fibroblasts from an unrelated patient (SC3) diagnosed
434 with bi-allelic pathogenic variants: NM_014362.4(*HIBCH*):c.891+1G>A &
435 NM_014362.4(*HIBCH*):c.470G>A; p.(Arg157Gln). Proteomic results show that HIBCH
436 protein in SC3 (*HIBCH*) is almost two times less abundant than in UDP5 (*NDUFA10*) (**Fig.
437 3L, left panel**) while the levels of NDUFA10 protein were unchanged in SC3 (*HIBCH*),
438 suggesting that a reduction in NDUFA10 is not a secondary defect from a primary *HIBCH*
439 defect (**Fig. 3L right panel**). The lack of any decrease in Complex I in SC3 (*HIBCH*)
440 suggests that the phenotype in UDP5 (*NDUFA10*) is likely to have arisen from the
441 *NDUFA10* variant while the *HIBCH* variant is expressed as a carrier and further
442 investigation for a second missed variant in *HIBCH* was not undertaken. UDP5 was the
443 first child of non-consanguineous parents of Māori ethnicity. The mother was 6 weeks
444 pregnant with a second child at the time of analysis and the family sought prenatal genetic
445 testing (PGT). With this in mind, we prioritised sample processing and analysis, and were
446 able to return results in less than 4 days from receipt of fibroblasts (**Fig. 3M**) with the
447 results sufficient to support the upgrade of the *NDUFA10* variant c.914T>C;
448 p.(Leu305Pro) to likely pathogenic/Class 4. This facilitated PGT, which confirmed
449 absence of the *NDUFA10* variant, allowing continuation of the pregnancy and delivery of
450 a healthy child who, at the time of writing, is 15-months old.
451
452 The rapid turnaround time (TAT) in this case was only possible as fibroblasts from UDP5
453 were available for analysis at the time of proteomics testing. Establishing and culturing a

454 new fibroblast line typically takes up to ~2 months and requires a skin biopsy[49] which is
455 increasingly considered invasive and can potentially delay a diagnosis. We sought to
456 explore the feasibility of rapid proteomics-based testing for suspected mitochondrial
457 disorders using peripheral blood mononuclear cells (PBMCs), which can be readily
458 obtained from whole blood and have been used in various diagnostic functional testing
459 approaches previously. To test the applicability of PBMCs in proteomic testing we first
460 generated a fractionated spectral library and compared it against the fibroblast spectral
461 library and coverage of Mendeliome proteins (**Fig. 4A**). The PBMC library was generated
462 from peptides of 39 normal individuals aged 0-17 years of age and can be used to
463 confidently identify 8,016 gene products, including 52% of known Mendeliome genes[27].
464 In comparison, our fibroblast spectral library covers 8,305 genes and 55% of the
465 Mendeliome. We performed a Principal Component Analysis (PCA) of whole-cell proteins
466 in the 37 PBMC samples showing that over 99% of the proteins are tightly clustered at
467 the whole-cell level and over 97% at the mitochondrial level (**Fig. 4B**).
468
469 Next, we validated the use of PBMCs against fibroblasts by analysing both sample types
470 from a diagnosed patient with pathogenic variants in *MRPL50* (Twin 2 in [21], SC4 and
471 SC5). MRPL50 is a structural subunit of the mitoribosome[50] and defects in this subunit
472 lead to impaired assembly of the large mitoribosomal subunit (mtLSU)[21]. We performed
473 a correlation of the $\log_2$ fold-change levels of mitochondrial proteins identified in both
474 samples relative to controls which showed reduced abundance of multiple proteins
475 belonging to the mtLSU in both sample types (**Fig. 4C**). RCA quantified the abundance
476 of the mtLSU in 63% and 48% in PBMCs and fibroblasts respectively **(Supplementary**
477 **Fig. 5E)**.

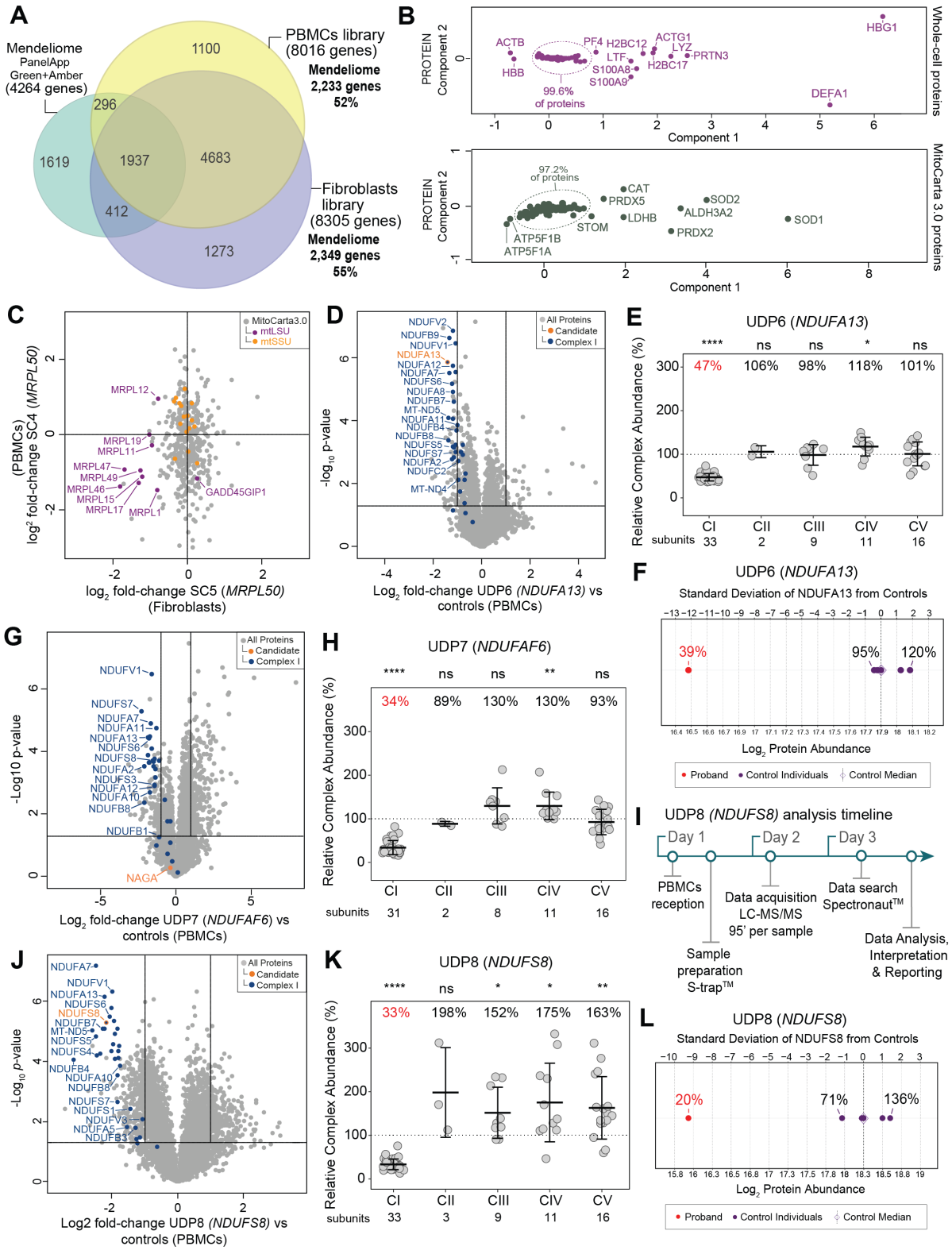

Figure 4. Utility of ultra-rapid proteomics supports the diagnosis of critically ill infants with suspected mitochondrial disorders A – Venn diagram showing the coverage of Mendeliome genes (PanelApp Australia Green and Amber entries, 4,264 genes) in the fibroblast spectral library (55% of

Mendeliome list of genes) and PBMC library (52% of Mendeliome list of genes). **B –** Principal Component Analysis (PCA) of the pilot PBMC normative data (N=36) for whole-cell proteins (top panel) and MitoCarta3.0 proteins (lower panel). **C –** Scatter plot showing the correlation between PBMC (SC4) and fibroblast (SC5) samples based on $\log_2$ fold-changes from whole-cell proteins in a diagnosed patient with variants in *MRPL50* relative to controls[21] showing reduced abundance of the proteins belonging to the large subunits of the mitochondrial ribosome (mtLSU; purple), mtSSU = mitoribosome small subunit (orange). **D –** Volcano plot of whole cell proteins from UDP6 (*NDUFA13*) PBMCs compared to controls (N=5) showing reduced abundance of NDUFA13 (orange dot) and other structural subunits of Complex I. Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent. Blue = Complex I subunits**. E –** Relative Complex Abundance (RCA) of OXPHOS complexes from UDP6 (*NDUFA13*) PBMCs compared to controls (N=5) showing an isolated Complex I defect. Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = p<0.0001, * = p<0.05, ns = not significant, *p*>0.05. **F –** Protein range for NDUFA13 in PBMCs in UDP6 (*NDUFA13,* red dot) and controls (N=5, purple dots) showing standard deviation of -12.2 from the control median**. G –** Volcano plot of whole cell proteins from UDP7 (*NDUFAF6*) PBMCs compared to controls (N=5) showing reduced abundance of NDUFS8 and other structural subunits of Complex I. Vertical lines represent ± 2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent. Blue = Complex I subunits. **H –** Relative Complex Abundance (RCA) of OXPHOS complexes from UDP7 (*NDUFAF6*) PBMCs compared to controls (N=5) showing an isolated Complex I defect. Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = p<0.0001, ** = p<0.01, ns = not significant, *p*>0.05. **I –** Timeline for ultra-rapid proteomics from PBMC sample receipt to results for UDP8 (*NDUFS8*) case achieved in less than 3 business days. **J –** Volcano plot of whole cell proteins from UDP7 (*NDUFS8*) PBMCs compared to controls (N=5) showing reduced abundance of NDUFS8 protein (orange dot) and other structural subunits of Complex I. Vertical lines represent ±2-fold-change equivalent and horizontal lines represent significance p-value ≤ 0.05 equivalent. Blue = Complex I subunits**. K –** Relative Complex Abundance (RCA) of OXPHOS complexes from UDP8 (*NDUFS8*) PBMCs compared to controls (N=5) showing isolated Complex I defect. Middle bar represents mean complex abundance. Upper and lower bars represent 95% confidence interval. Significance was calculated from a paired t-test between the individual protein means. **** = p<0.0001, ** = p<0.01, * = p<0.05, ns = not significant, *p*>0.05. **L –** Protein range for NDUFS8 in PBMCs in UDP8 (*NDUFS8,* red dot) and controls (N=5, purple dots) showing standard deviation of -9.3 from the control median.

UDP6 (*NDUFA13*)

UDP6 (*NDUFA13*) presented at 2 years of age with chronic ataxia. She was the first child to non-consanguineous healthy parents from Iran and had a healthy 2-month-old brother. There was no family history of any neurological issues. She presented with chronic, stable ataxia, having been unsteady and falling frequently since she started walking at 14 months of age. She had mild fine motor delay and a tremor. Her social, language, and cognitive development was age appropriate. On examination, UDP6 (*NDUFA13*) had normal growth parameters and was not dysmorphic. She had mild gait ataxia with a broad base. Her tone, power and reflexes were normal in upper and lower limbs bilaterally.

525 Ophthalmology examination was unremarkable. Magnetic resonance imaging (MRI) of
526 brain showed bilateral symmetrical T2/FLAIR hyperintensity of the cerebral peduncles of
527 the midbrain (and possibly the substantia nigra) and two symmetrical foci of high signal
528 in the dorsal medulla with diffusion-weighted imaging (DWI) restriction. There were no
529 basal ganglia or thalamic abnormalities. This was suggestive of a mitochondrial disorder.
530 Serum blood lactate was elevated 2.4 (RR 1.0-1.8). Other metabolic investigations were
531 normal. Audiology assessments and echocardiogram were normal. At 4 years of age, she
532 has ongoing mild ataxia that improved with physiotherapy and was no longer falling. She
533 has recently had deterioration in her vision and been diagnosed with optic neuropathy.
534 Trio exome sequencing identified compound heterozygous variants
535 NM_015965.7(*NDUFA13*):c.170G>A; p.(Arg57His) likely pathogenic/Class 4 and
536 NM_015965.7(*NDUFA13*):c.187G>A; p.(Glu63Lys) VUS/Class 3. Proteomics was sought
537 to provide functional evidence for the VUS/Class 3 variant and blood was collected for
538 PBMC isolation as fibroblasts were not available at the time of analysis. Five unrelated
539 PBMCs from age-matched normal donors were used as controls. A volcano plot showed
540 reduction of NDUFA13 protein and several other Complex I subunits in UDP6 (*NDUFA13*)
541 (**Fig. 4D**) and RCA quantified the isolated Complex I defect at 47% control satisfying the
542 criteria for a major defect (**Fig. 4E**). Analysis of the level of the NDUFA13 protein in the
543 proband demonstrated a residual abundance of 39%, which corresponds to greater than
544 12 standard deviations below the control median (**Fig. 4F**). Proteomics data were
545 subsequently used as functional evidence to upgrade the
546 NM_015965.7(*NDUFA13*):c.187G>A;p.(Glu63Lys) variant to likely pathogenic/Class 4.
547
548 **UDP7 (*NDUFAF6*)**
549 UDP7 (*NDUFAF6*) was referred to the study at five months of age with asymmetric early
550 onset growth restriction, central hypotonia, proximal renal tubular dysfunction, macrocytic
551 anaemia, severe exocrine pancreatic insufficiency, liver dysfunction with cholestasis and
552 mild persistent elevated lactate in CSF and blood. MRI showed symmetrical diffusion
553 abnormality involving corticospinal tract, areas of brainstem and medial cerebellar
554 hemisphere. He died at seven months of age due to progressive bulbar dysfunction with
555 a probable clinical diagnosis of Leigh syndrome. Clinical trio WES identified a single

556 pathogenic (Class 5) heterozygous NM_000262.3(*NAGA*):c.973G>A; p.(Glu325Lys)
557 variant, and two *in trans* variants in *NDUFAF6*, a frameshift
558 NM_152416.4(*NDUFAF6*):c.267delT, p.(Tyr323Ilefs*18) variant and a deep intronic
559 variant, NM_152416.4(*NDUFAF6*):c.298-768T>C, both classified as VUS/Class 3.
560 PBMCs were the sample of choice due to lower invasiveness and potential for fast
561 turnaround. Whole-cell proteomics identified no changes in the abundance of NAGA
562 protein relative to controls (**Fig 4G**). *NAGA* encodes the alpha-N-
563 acetylgalactosaminidase, a lysosomal enzyme whose deficiency causes a rare
564 autosomal recessive lysosomal storage disorder. Previous reports of NAGA-deficiency
565 show heterogenous presentations ranging from infantile-onset neuroaxonal dystrophy
566 (Schindler disease type I, MIM 609241) to an adult-onset disorder characterised by
567 angiokeratoma corporis diffusum and mild intellectual impairment (Schindler disease type
568 II, also known as Kanzaki disease MIM 609242)[51]. In addition to the proteomic studies,
569 UDP7 *(NDUFAF6)* did not appear to have clinical features consistent with NAGA-
570 deficiency, and infants with this condition typically develop normally until about a year old.
571 We then turned our investigations to *NDUFAF6* (*C8orf38*) which encodes an assembly
572 factor involved in biogenesis of the core peripheral arm subunit NDUFS8 of Complex I[52]
573 and has previously been associated with the stability of the ND1 module of Complex I[53,
574 54]. Previous reported cases of pathogenic variants in *NDUFAF6* present with a range of
575 symptoms, including Leigh syndrome and the Acadian variant of Fanconi syndrome 5
576 (proximal renal tubular dysfunction) (FRTS5, MIM 618913)[53-57]. UDP7 *(NDUFAF6)* had
577 clinical features consistent with Leigh syndrome, and while the aetiology of his liver
578 dysfunction and severe exocrine pancreatic insufficiency was somewhat less clear, both
579 would broadly fit with a mitochondrial cytopathy noting Complex I deficiencies often do
580 not correlate perfectly with a single clinical subdivision. The
581 NM_152416.4(*NDUFAF6*):c.298-768T>C variant has been reported as likely pathogenic
582 previously, specifically in relation to the FRTS5 phenotype[56]. A volcano plot from whole-
583 cell PBMC proteomics data showed reduction in the abundance of several subunits of
584 Complex I in UDP7 (*NDUFAF6*), including NDUFS8, (**Fig. 4G**) with RCA analysis
585 quantifying the isolated Complex I defect with a major defect of 34% relative to controls
586 (**Fig. 4H**). Functional proteomic data and phenotype match support *NDUFAF6* as being

587 the causative gene for this clinical case, and the VUS/Class 3 variants are in the process
588 of re-curation at the time of writing.
589
590 **UDP8 (*NDUFS8*)**
591 UDP8 (*NDUFS8*) was a neonate admitted to hospital presenting with persistent lactic
592 acidemia, hypertrophic cardiomyopathy, left pulmonary artery stenosis, thickened
593 pulmonary valve, hypotonia and microcephaly with progressive deterioration from 4 weeks
594 of age, succumbing at 5 weeks of age. Brain MRI showed abnormalities in the corpus
595 callosum and diffuse increased signal throughout the supratentorial white matter with
596 small foci of gliosis in the peritrigonal regions. Ultra-rapid trio WGS identified a
597 homozygous intronic variant NM_002496.4(*NDUFS8*):c.501+5G>A with biparental
598 inheritance. Due to the critical deteriorating condition, blood was collected from the
599 proband and subjected to ultra-rapid proteomics. Five unrelated age-matched controls
600 were used for comparison and the complete analysis was performed with a 54-hour
601 turnaround time (TAT) from PBMC sample reception to the reporting of results to the
602 clinical team (**Fig. 4I**). Whole-cell proteomic analysis of PBMCs showed significant
603 reduction of several Complex I subunits, including NDUFS8 (**Fig. 4J**) and RCA analysis
604 showed an isolated Complex I defect with a major defect of 22% relative to controls (**Fig.
605 4K**). Residual abundance of the NDUFS8 protein in the proband was quantified at 20%,
606 representing over 9 standard deviations below the control median (**Fig. 4L**). The strong
607 phenotype match together with functional evidence from proteomics support disease
608 causation due to variants in *NDUFS8* and leading to variant upgrade to likely
609 pathogenic/Class 4.
610
611 ***Discussion***
612 The estimated prevalence of rare disease in the population is in the range of 3.5-6%,
613 equivalent to 263-446 million individuals worldwide[58]. There are over 7,000 different rare
614 diseases known, with a current diagnostic yield of ~35-70% from WES or WGS[1-5].
615 Mitochondrial disease is a group of rare diseases caused by variants in over 300 known
616 genes[6, 16], where RCE has been historically performed to confirm a specific OXPHOS
617 defect in the functional validation of variant pathogenicity[16].

618

619 We have previously used an approach we term relative complex abundance (RCA) where
620 proteomics data are used to quantify the residual abundance of OXPHOS complexes in
621 clinical samples from individuals suspected of mitochondrial disease. We have shown
622 that RCA acts as an effective proxy for clinically accredited RCE, including when
623 appropriate samples have not been available[20-22, 24, 25, 59-61]. In an RCA analysis, the
624 abundance of a complex in a sample is calculated from the mean of each individual
625 subunit abundance detected by more than two peptides across each complex. The power
626 of RCA analysis relies on the quantification of multiple peptides per protein in a complex
627 with a high degree of subunit stability co-dependence, as we and others have previously
628 shown for Complex I, III, IV, the mitoribosome[11, 17, 20-22, 25, 59, 60, 62, 63] and non-mitochondrial
629 complexes such as the exocyst and nuclear pore complexes[64, 65]. RCE, on the other hand,
630 typically relies on enzyme rate estimates determined in several replicates of patient
631 samples and compared against age-matched controls. In some instances, we have found
632 RCA demonstrated higher sensitivity by detecting a specific OXPHOS defect where RCE
633 was inconclusive[20, 22]. RCA can also show higher specificity for use in functional validation
634 of variant pathogenicity by detecting reduction of specific parts of the complex (e.g. loss
635 of subunits in the catalytic Complex I N-module in VC1-*NDUFAF2* and VC2-*NDUFS6*),
636 implicating only the 10 N-module subunits as candidate genes compared with loss of
637 Complex I RCE potentially being caused by variants in any of more than 50 genes
638 encoding Complex I subunits and assembly factors[26, 66]. In other cases, some variants
639 encoded by the mtDNA resulted in no defect on RCA analysis despite over 70% and 90%
640 heteroplasmy levels in fibroblasts. This suggests that some mtDNA variants can be
641 refractive to RCA analysis due to an enzymatic and not a structural defect, where RCE
642 might be an appropriate test to provide functional evidence. In terms of variant type,
643 mitochondrial DNA variants account for only 0.54% of rare diseases while autosomal
644 recessive variants account for over 41%[67] and are more likely to result in a loss-of-
645 function protein[68], which can be detected via quantitative proteomics.

646

647 RCE is typically performed on a tissue biopsy sample (e.g. skeletal muscle, liver) or
648 primary cultured cells (e.g. fibroblasts). Establishing and growing primary fibroblasts from

649 skin biopsies to a sufficient volume of cells for RCE analysis involves weeks to months of
650 cell culture[49], and, in some cases, tissue specificity results in only mild or undetectable
651 defects as measured by RCE[10]. Here we have demonstrated the feasibility of using
652 PBMCs in diagnostic proteomics testing. In our hands, sufficient PBMCs for these
653 analyses can be readily obtained from as a little as 1 ml of whole blood and the process
654 from PBMC isolation to data acquisition performed in less than 48 hours. Together these
655 properties and the results presented in this study demonstrate that PBMCs can be
656 combined with our ultra-rapid proteomics pipeline for functional validation of variants
657 causing defects in the mitochondrial OXPHOS system and the mitoribosome.
658
659 The untargeted nature of quantitative proteomics also allows this technique to be applied
660 beyond mitochondrial disease, as we have shown here for a mitochondrial disease
661 phenocopy (UDP4; *CCDC47*) and elsewhere for nuclear pore, neuronal exocytic vesicle
662 trafficking and rigid spine disorders[64, 65, 69]. Over 50% of known disease-associated
663 genes[27] are routinely detected by our quantitative proteomics pipeline in PBMCs or
664 fibroblasts, and over 45% of all disease genes listed in OMIM are part of a protein
665 complex[16]. This means that pathogenic variants in one of these genes can lead to
666 reduction of the protein of interest, which in some cases will result in downstream complex
667 reduction as an additional layer of functional evidence to support variant upgrade. In
668 terms of variant type, we show that quantitative proteomics can provide functional
669 evidence for a wide range of genetic variants including CNV[62], splice site and deep
670 intronic[20, 22], and missense variants[21]. Missense variants are the most common type of
671 variant accounting for 60% of pathogenic variants associated with autosomal recessive
672 disorders[18] and are usually refractory to transcriptomic analysis. For some cases,
673 proteomics can also provide supportive functional evidence supportive when the protein
674 encoded by the candidate gene is not detectable, either due to protein abundance being
675 below the limit of detection or for genes encoding tRNAs, as seen in VC22 (*MT-TK*),
676 further expanding the potential application of proteomics in rare disease diagnosis.
677
678 A genomics-first approach to rare disease diagnosis has markedly increased diagnostic
679 rates and, in the case of rapid and ultra-rapid genomic testing, shortened diagnostic

odysseys to days instead of months[65, 70-72]. In the case of mitochondrial and many other rare diseases, diagnoses are often now achieved using just a blood sample, sparing many patients from an invasive biopsy of muscle or other tissues. As a result, diagnostic centres such as ours now receive a third or less of the number of samples for RCE testing than received 10 years ago. This provides an incentive for untargeted functional tests, such as RNA sequencing and proteomics, that could be financially viable and widely available in a pathology-certified context.

This study also highlights the limitations of quantitative proteomics in detecting abundance changes in variants leading to catalytic defects and some mtDNA-encoded variants in fibroblasts. Moreover, the detection of proteins is related to the limit of detection as determined by the liquid chromatography, mass-spectrometry instrumentation, and data acquisition methods used. However, recent advances in mass-spectrometry instrumentation including development of the asymmetric track lossless (Astral) analyzer offer further improvements in sensitivity and greatly reduced run times[73]. Quantitative proteomics thus offers a further paradigm shift by providing functional evidence for variants in thousands of genes in a single test. This driver will facilitate the translation of proteomics testing for rare diseases into certified pathology laboratories. Clinically delivered proteomics can potentially replace hundreds of tests targeted to specific diseases, which are usually restricted to research settings and carry less weight in the upgrade of variants. Our study also demonstrates that proteomics can be delivered alongside ultra-rapid genomic sequencing approaches to provide functional data in a clinically relevant timeline.

Materials and Methods

Ethics statement

This study was conducted in accordance with the revised Declaration of Helsinki and following the Australian National Health and Medical Research Council statement of ethical conduct in research involving humans. Samples were obtained after receiving written, informed consent for diagnostic or research investigations from the respective responsible human ethics institutional review boards. HREC/RCH/34228, HREC/RCH/34183, HREC/89419/RCHM-2022 and HREC/82160/RCHM-2022 were

712 approved by the Royal Children's Hospital, Melbourne, Ethics in Human Research
713 Committee. HREC/16/MH/251 was approved by the Melbourne Hospital Ethics in Human
714 Research Committee. The REC reference 2002/205 by the Newcastle and North
715 Tyneside Local Research Ethics Committee.
716
717 *Genomic investigations*
718 Genomic investigations resulting in diagnosis of patients within the validation cohort (VC)
719 are described in the publications indicated by PMID within **Supplementary Table 1**, with
720 the exceptions of VC13 (*SDHAF1*), VC15 (*UQCRC2*) and VC24 (*FARS2*) which are as
721 follows; Individual VC13 (*SDHAF1)* underwent clinical singleton WES through Baylor
722 Genetics as previously described[74]. For individual VC15 (*UQCRC2*), homozygosity
723 mapping from Illumina HumanCytoSNP-12 v2.1 array data identified substantial long
724 contiguous stretches of homozygosity (LCSH) accounting for ~3.7% of the genome. Due
725 to the observed complex III RCE defect in this patient, candidate gene sequencing of
726 exons using genomic DNA through PCR and sanger sequencing was performed on
727 complex III-related genes *UQCC1* and *UQCRC2* which lay within these regions of LCSH
728 (primers available upon request). Blood DNA for individual VC24 (*FARS2*) underwent
729 clinical singleton WES testing through the Victorian Clinical Genetics Service (VCGS) as
730 previously described[75]. mtDNAseq was also used to quantify mtDNA variant
731 heteroplasmy levels for individuals VC6 (*MT-ND1*) and VC12 (*MT-ND5*) as previously
732 described[75]. Individuals within the undiagnosed patient (UDP) cohort were investigated
733 as follow: blood DNA from UDP1 (*MT-ATP6*) underwent WES with mitochondrial DNA
734 sequencing (mtDNAseq) as part of the Australian Genomics Mitochondrial Flagship as
735 described[75]. For individual UDP4 (*CCDC47*), sequencing and analysis was performed as
736 previously described[25]. Individual UDP5 (*NDUFA10*) had the Comprehensive Metabolism
737 Panel performed by Blueprint Genetics. Individuals UDP6 (*NDUFA13*) and UDP7
738 (*NDUFAF6*) both underwent clinical trio WES through the Victorian Clinical Genetics
739 Service (VCGS)[75]. UDP8 (*NDUFS8*) had acute-care rapid trio WGS and analysis at the
740 VCGS as previously published[65].
741
742 *Respiratory chain enzymology*
743 Respiratory chain enzyme activities in fibroblasts and skeletal muscle were measured by
744 spectrophotometry as described[10]. Complex I (CI) was measured as rotenone-sensitive
745 NADH:coenzyme Q_1 oxidoreductase, Complex II (CII) as succinate:coenzyme Q_1
746 oxidoreductase, Complex III (CIII) as decylbenzylquinol:cytochrome *c* reductase, and
747 Complex IV (CIV) as cytochrome *c* oxidase. Citrate synthase (CS) was measured as

748 production of coenzyme A (CoA.SH) from oxaloacetate using the thiol reagent 5,5'-dithio-
749 bis-(2-nitobenzoic acid). Enzyme activities were calculated as initial rates (CI, CII, and
750 CS) or as first-order rate constants (CIII and CIV)[10].
751
752 *Cell Culture Conditions*
753 Fibroblast and HEK293T cell lines were cultured in Dulbecco's Modified Eagle Medium
754 (DMEM) High Glucose, Sodium Pyruvate and Glutamine (Sigma-Aldrich) supplemented
755 with 10% (v/v) Fetal Calf Serum (FCS; CellSera), 100 U/mL Penicillin-Streptomycin
756 (Gibco) and 50 µg/mL Uridine. Cells were maintained at 37°C with 5% CO2.
757
758 *Generation of knockout lines*
759 The NDUFS2^{KO} cell line was generated using a CRISPR/Cas9 plasmid encoding the
760 guide RNA 5'-TGAGGGCTTTGTGCGGCTTCCGG-3' cloned into the pSpCas9-2A-GFP
761 vector [26]. The CRISPR was transfected into HEK293T cells and single cells were obtained
762 by Fluorescence-activated cell sorting at the Monash Flow Cytometry Platform (Flowcore).
763 Singe cell populations were screened by loss of cell viability on galactose-containing
764 DMEM and subsequently validated by the loss of protein by western blotting and indel
765 sequencing, which identified two deletions c.[18_36del] and c.[17_42del] Supplementary
766 table 1). The UQCRC2^{KO} cell line was generated using a CRISPR/Cas9 plasmid encoding
767 the guide RNA 5'-CACCGGTACTTACACATCACCCCGC designed using the
768 CHOPCHOP tool and subcloned into pSpCas9(BB)-2A-GFP vector as previously
769 described[26]. The plasmids were validated by sequencing, and then transfected into target
770 cell lines using Lipofectamine™ 3000 Transfection Reagent (Thermo Fischer Scientific)
771 according to manufacturer's guidelines. Single cells expressing GFP were isolated on a
772 FACS Aria Fusion (BD Biosciences) cell sorter and clonal populations were subsequently
773 screened for relevant gene knockouts using SDS-PAGE and immunoblotting and indel
774 sequencing, which identified two deletions c.[9141_9199del] and [9146_9167del]
775 (Supplementary table 1). The SURF1^{KO} cell line was generated using a CRISPIR/Cas9
776 system with the HEK293T parental cell line. Two guide RNAs targeting the 5'UTR and
777 exon 2 regions of *SURF1* were cloned into individual pSpCas9-2A-GFP vectors (Addgene,
778 plasmid 48138 [18]). Both constructs were transfected simultaneously into HEK293T cells
779 using Lipofectamine 3000 (Thermo Fisher Scientific), according to the manufacturer's
780 instructions. Verification of successful HEK293T SURF1^{KO} was performed via three
781 primer screening PCR were used to amplify incorporated indels via PCR and by SDS-
782 PAGE immunoblotting. The product of external Primers (outside the double CRISPR
783 target sites) underwent M13 primer Sanger Sequencing of alleles, using the Topo™ TA

784 Cloning Kit (Thermo Fisher Scientific) as per the manufacturer's instructions. Sequencing
785 identified two deletions [c.18_219del] and [c.18_221delinCG] (Supplementary table 1).
786 The ATP5PD^{KO} cell line was generated using a CRISPR/Cas9 plasmid encoding the
787 guide RNA 5'-CACCGCCTTTCCTTGTGGGCAGGT designed using the CHOPCHOP
788 tool and subcloned into pSpCas9(BB)-2A-GFP vector as previously described[26]. The
789 plasmids were validated by sequencing, and then transfected into target cell lines using
790 Lipofectamine™ 3000 Transfection Reagent (Thermo Fischer Scientific) according to
791 manufacturer's guidelines. Single cells expressing GFP were isolated on a FACS Aria
792 Fusion (BD Biosciences) cell sorter and clonal populations were subsequently screened
793 for relevant gene knockouts using indel sequencing, which identified a single deletion
794 c.[33481insT] (Supplementary table 1).
795
796 *PBMC isolation from whole blood*
797 Neonatal blood samples were collected from healthy neonates following routine
798 intramuscular administration of Vitamin K. For older infants and children ages 30 days to
799 18 years, venous blood samples were obtained from hospital outpatients undergoing
800 minor elective day surgery. An outline of the study protocol and complete detail of
801 participant recruitment, inclusion and exclusion criteria and sample collection are
802 described in our previously published HAPPI Kids study protocol[76]. PBMCs were isolated
803 from fresh blood samples of the 37 HAPPI Kids controls using Ficoll®-Paque Plus (GE
804 Healthcare Life Sciences) according to the instructions of the manufacturer with manual
805 removal of the PBMC layer using a Pasteur pipette. For the UDP6 (*NDUFA13*), UDP7
806 (*NDUFAF6*) and UDP8 (*NDUFS8*) probands, PBMC isolation using Ficoll®-Paque Plus
807 was performed in SepMate™ tubes (STEMCELL Technologies), based on
808 manufacturer's protocols with moderate braking applied for both washes, and a reduction
809 in centrifugation speed and duration to 200 *g* for five minutes for the second wash.
810
811 *Data-Independent Acquisition Mass Spectrometry of Fibroblast, HEK293T and PBMCs*
812 *Cohorts and Data Analysis*
813
814 *Spectral Library Sample Preparation*
815 A HEK293T spectral library was generated from both whole cell and mitochondrial
816 HEK293T samples to search data-independent acquisition (DIA) HEK293T knockout cell
817 lines. A total of 100 µg of mitochondrial sample isolated from wildtype HEK293T cells
818 underwent carbonate extraction at pH 11.5 to separate soluble and integral mitochondrial
819 proteins, as previously described[77]. The resulting pellet and supernatant samples, as well

820 as 25 µg whole cell HEK293T sample, were solubilized in lysis buffer for a final
821 composition of 5% SDS and 50 mM Triethylammonium bicarbonate (TEAB) pH 8.5. The
822 lysed whole cell, mitochondrial pellet and supernatant samples were processed using S-
823 trap™ Micro Spin Columns (ProtiFi) according to the manufacturer's instructions, where
824 samples were reduced with 40 mM 2-chloroacetamide (CAA, Sigma) and alkylated with
825 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP; BondBreaker, Thermo
826 Fischer Scientific). Isolated proteins were digested at 1:25 trypsin to protein ratio at 37°C
827 overnight and eluted peptides were dried down using a CentriVap Benchtop Vacuum
828 Concentrator (Labconco). Fibroblasts and PBMCs spectral libraries were generating from
829 whole cell samples of either five healthy fibroblast controls or PBMC pellets isolated from
830 39 healthy individuals ranging from 0-17 years old. Fibroblast or PBMC pellets were
831 solubilized 5% SDS and 50 mM TEAB pH 8.5, and total protein determined by Pierce
832 bicinchoninic acid (BCA) Protein Assay Kit (Thermo Fisher Scientific). A total of 25 µg for
833 each control was processed using the S-Trap™ Micro Spin Columns (ProtiFi), as
834 described above.
835 Peptides from HEK293T, fibroblast and PBMC samples were reconstituted in 0.5% formic
836 acid (FA) for fractionation. Fibroblast and PBMC control samples were pooled upon
837 reconstitution to create a single fibroblast, HEK293T and PBMC peptide mixture and
838 loaded onto a strong cation exchange (Empore Cation Exchange-SR, Supelco Analytical)
839 stage-tips made as described[78]. Stage tips were washed with 20% acetonitrile (ACN) and
840 0.5% FA and eluted over seven fractions of increasing concentrations of ammonium
841 acetate (45-300 mM), 20% ACN and 0.5% FA for whole cell fibroblasts, PBMC and
842 HEK293T samples. HEK293T mitochondrial and supernatant samples, were collected
843 over five fractions. All samples underwent a final elution with 5% ammonium hydroxide
844 and 80% ACN, followed by concentration using a CentriVap Benchtop Vacuum
845 Concentrator (Labconoco). Fractions were desalted on SDB-XC (poly(styrene-divinyl-
846 benzene); Supelco Analytical) stage tips made in-house as previously described[78].
847
848 *Spectral Library Mass Spectrometry Data Dependent Acquisition*
849 Fractions were reconstituted in 0.1% trifluoracetic acid (TFA) and 2% ACN and each
850 library was analyzed by liquid chromatography (LC)-tandem mass spectrometry (MS/MS)
851 on an Orbitrap Eclipse Mass Spectrometer (Thermo Fischer Scientific) operating on DDA
852 mode over a 125-minute gradient. Tryptic peptides were loaded onto an Acclaim Pepmap
853 nano-trap column (Dinoex-C18, 100 Å, 75 µm × 2 cm) at an isocratic flow of 5 µl/min of
854 2% ACN and 0.1% FA for six minutes before switching with an Acclaim Pepmap RSLC
855 analytical column (Dinoex-C18, 100 Å, 75 µm × 50 cm). The separation of peptides was

performing using a nonlinear 125-minute gradient of solvent A (5% dimethyl sulphoxide (DMSO), 0.1% FA) and solvent B (5% DMSO, 100% ACN, 0.1% FA). The flow gradient was (i) 0-6 min at 3% solvent B, (ii) 6-95 min at 3-23% solvent B, (iii) 95-105 min at 23-40% solvent B, (iv) 105-110 min at 40-80% solvent B, (v) 110-115 min at 80-80% solvent B, (vi) 115-116 min at 80-3% solvent B and equilibrated at 3% solvent B for 10 minutes before the next injection. Briefly, the data was collected using positive polarity with a MS1 scan range of 375-1500 m/z and resolution set to 120,000. Other MS1 instrument parameters include: ACG target of 4e5, maximum injection time of 50 ms and Isolation window of 1.6. MS2 instrument parameters include: scan range of 150-2000 m/z, resolution of 15,000, HCD collision energy of 30%, ACG target of 5e3, maximum injection time of 22 ns and dynamic exclusion of 30 s.

Sample Preparation of HEK293T, PBMC and Fibroblast Cell Lines

Whole cell pellets of HEK293T cells, PBMCs and fibroblasts were collected and washed twice with PBS before resuspension in 5% SDS, 50 mM tetraethylammonium bromide (TEAB) pH 8.5 buffer and 125 U of benzonase per ml (Sigma-Aldrich). Skeletal muscle samples were solubilized with a probe sonicator with 30% amplitude in 5% SDS, 50 mM TEAB pH 8.5 buffer on ice with a cycle of 10 seconds on 10 seconds off for 1 minute and clarified at 16,000 x g for 5 minutes. Protein concentration was performed with Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) and 25 µg of each sample was aliquoted in triplicates for patients and singlicate for individual controls (N=3-5). Samples were processed using S-trap™ micro spin columns (ProtiFi) according to the manufacturer's instructions. Proteins were digested with trypsin (Thermo Fisher Scientific) at 1:10 trypsin to protein ratio. Peptides were dried down using a CentriVap Benchtop Vacuum Concentrator (Labconco) and reconstituted in 45 µL 2% ACN, 0.1% trifluoroacetic acid (TFA) and 2 µL injected for liquid chromatography tandem mass spectrometry (LC-MS/MS). Samples were analyzed for 125 minutes (HEK293T) or 95 minutes (Fibroblasts, skeletal muscle, PBMCs) on an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) operating in DIA mode.

Mass Spectrometry Data Independent Acquisition

Peptides were reconstituted in 45 µL of 2% ACN and 0.1% TFA and 2 µl was injected into an Orbitrap Eclipse mass spectrometer (Thermo Fischer Scientific) for LC-MS/MS analysis equipped with trap and analytical columns described above. All samples were analysed with the mass spectrometer operating in DIA mode. The separation of peptides was performed using a nonlinear gradient of solvent A and solvent B across 125-minutes

892 for HEK293T knockout cell lines, and 95-minutes for all other sample types (fibroblasts,
893 PBMCs and skeletal muscle). Briefly, for the 125-minute method, the flow gradient was
894 (i) 0-6 min at 3% solvent B, (ii) 6-95 min at 3-23% solvent B, (iii) 95-105 min at 23-40%
895 solvent B, (iv) 105-110 min at 40-80% solvent B, (v) 110-115 min at 80-80% solvent B,
896 (vi) 115-116 min at 80-3% solvent B and equilibrated at 3% solvent B for 10 minutes
897 before the next injection. The data was collected using positive polarity with a MS1 scan
898 range of 350-1200 m/z at a resolution of 120,000. Other MS1 instrument parameters
899 include: ACG target of 1e6, maximum injection time of 50 ms, Isolation window of 24 with
900 1 m/z overlap. For MS2 parameters, the scan range was 200-2000 m/z at 15,000
901 resolution, HCD collision energy of 30%, AGC target of 1e6 and maximum injection time
902 of 22 ms. For the 95-minute method, the flow gradient was (i) 0-6 min at 3% solvent B,
903 (ii) 6-7 min at 3-4% solvent B, (iii) 7-82 min at 4-25% solvent B, (iv) 82-86 min at 25-40%
904 solvent B, (v) 86-87 min at 40-80% solvent B, (vi) 87-90 min at 80-80% solvent B, (vii) 90-
905 91 min at 80-3% solvent B and equilibrated at 3% solvent B for 5 minutes before the next
906 injection. Instrument parameters were consistent with those used in the 125-minute
907 method described above, with changes made to the following parameters: MS1 scan
908 range of 350-1400 m/z, MS1 maximum injection time of 45 ms, Isolation window of 13.7
909 and MS2 ACG target of 5e5.
910
911 *Proteomic data search*
912 For spectral library, raw files were imported into Spectronaut®[79] (v.15.2.210819.50606 for
913 HEK293T library, v.14.8.201029.4778 for fibroblasts library, v. 16.0.220606.53000 for
914 PBMCs library) and the three libraries were generated using the 'Pulsar' option with
915 default BGS Factory settings, searching against Uniprot human database containing
916 reviewed canonical and isoforms sequences (42,386 entries). The resulting HEK293T
917 library contained 194,270 precursors, the fibroblast library contained 131,627 precursors,
918 and the PBMC library contained 148,522 precursors. For DIA samples, raw files were
919 imported into Spectronaut® and searched against the libraries generated above. Default
920 BGS Factory search parameters were used with changes made to exclude single hit
921 proteins, no selection of 'Major Group Top N' and 'Minor Group Top N' and data filtering
922 setting set to 'Q-value' or 'Q-value sparse' with 'run-wise imputation' as an imputing
923 strategy. 'Q-value' datasets (unimputed) were used for generation of RCA values while
924 'Q-value sparse' with 'run-wise imputation' datasets were used to generate heatmaps and
925 correlation plots to reduce missing values for visualization, and in volcano plots for VC
926 and KC cohorts or when the protein of interest was not well detected. UDP1 (*MT-ATP6*)
927 data was also searched using BGS Factory direct DIA settings (library free) with changes

928 as above with inclusion of the p.Met228 variant peptide fasta file in the protein database
929 search. SC1 and SC2 (*TMEM70*) skeletal muscle were searched with default BGS
930 Factory direct DIA settings with changes as above.
931
932 *Protein filtering and clean up*
933 Proteins were filtered in each cohort to ensure that the data was of high quality and with
934 a minimal amount of missing data for analysis. Proteins within the validation cohort (VC)
935 were filtered out if identified by a single peptide in all controls or proband samples,
936 maintaining proteins that were identified by >2 peptides in both groups. Proteins were
937 then filtered to least three or more valid values for the controls and at least two or more
938 valid values for the proband samples. For the knockout cohort (KC), proteins were filtered
939 out if identified by a single peptide in all controls and proband samples. Proteins were
940 then filtered to at least two valid values for the controls and at least two valid values for
941 the knockout sample. For the undiagnosed proband cohort (UDP), proteins were filtered
942 out if identified by a single peptide in all controls or proband samples. Proteins were then
943 filtered for 70% valid values for the controls due to the different number of controls per
944 batch and 70% valid values for the proband samples. For the supporting cohort (SC),
945 samples SC1, SC2, SC3 and SC5 proteins were filtered out if identified by a single peptide
946 in all controls or proband samples. Proteins were then filtered for 70% valid values for the
947 controls and 70% valid values for the proband samples. Sample SC4 had no filtering
948 completed as this was a previously published dataset[21] and the data was unlogged and
949 complex subunit gene names manually updated to match the current MitoCarta3.0
950 naming convention for Complex V subunits. For the 36 PBMC control samples (SC6-41),
951 proteins were filtered out if they were identified by a single peptide across all 36 samples
952 and further filtered for at least 70% valid values.
953
954 *Relative Complex Abundance Plots*
955 Relative complex abundance (RCA) plots were generated for all VC, KC, SC and UDP
956 cohorts and plots can be viewed within the RDMS Explorer platform. All samples
957 underwent mitochondrial normalization where the mean of the mitochondrial proteins for
958 each specific sample was determined using the average of all control mitochondrial
959 proteins or proband mitochondrial proteins and then used as a normalization factor for
960 each group and applied to each sample. To develop the RCA plots, the average of the
961 proband and controls was determined and then transformed to log10. The difference
962 between the log10 of the proband and controls was determined and then unlogged.
963 Statistical difference between the control and proband complex subunits was determined

using a paired t-test and summary statistics including mean and the error bars indicating standard deviation were determined for each complex and the graph plotted within R (v.4.3.2). Complex subunits were determined using curated annotations from MitoCarta3.0[28]. For Supplementary Fig. A, relative mitochondrial abundance from MitoCarta3.0 '+' entries were generated as above without the mitochondrial normalization step. For Supplementary Fig. B, control mitochondrial abundance plots were generated using log2 abundances from MitoCarta3.0 '+' entries for each batch and significance performed with an ANOVA test.

Protein range graphs

Protein range graphs were developed using an in-house R script. The median of the protein of interest in control samples is calculated as well as the median of the proband replicates. The log2 fold-change is calculated between the proband and control and converted to percentage for the proband as the abundance relative to the control median for plotting. The range of the controls is also displayed as the abundance relative to the control median. The log2 abundance per standard deviation (SD) is rounded to one decimal place with the top axis SD present to determine how many standard deviations the protein is from the control median.

Volcano and Correlation plots

The peptide and valid value cleaned matrices for all samples were annotated with MitoCarta3.0[28] entries in Perseus[80]. All analyses were performed in Perseus v 1.6.14.0. Volcano plots were visualized using scatter plot function from a two-sample t-test using p-value threshold of 0.05 in Perseus. For the correlation plots (Fig 3B, Supplementary Fig.2B), proteomic data searched with q-value sparse with run-wise imputation settings were correlated via gene name using gene name entries and the logFCs in R and plotted using scatter plot function in Perseus. For Fig. 4C, previously published data from MRPL50 Fibroblasts[21] were correlated with MRPL50 PBMCs using gene name entries and the logFCs in R and plotted using scatter plot function in Perseus.

RDMS Explorer Website.

All the code used to develop the RDMS Explorer website (https://rdmassspec.shinyapps.io/RDMSExplorer/) and in-house analysis performed in R can be found at (https://doi.org/10.5281/zenodo.12883602) and (https://github.com/njcaruana/rdms_oxphos).

1000 *Heatmaps and Venn diagrams*
1001 For heatmaps output matrices searched with q-value sparse run-wise imputation settings
1002 Spectronaut runs were log_2 transformed and a two-sided t-test using p-value for
1003 truncation was performed in Perseus for each cell line against their respective batch
1004 controls. The Morpheus software was used for heatmap visualization
1005 (https://software.broadinstitute.org/morpheus). For Supplementary Fig. 1A, proteins were
1006 manually annotated based on MitoCarta3.0 as 'assembly factor' or 'complex I-V' and
1007 subset for visualization in Morpheus using the following parameters metric 'one minus
1008 Pearson correlation', linkage method 'average', cluster 'rows' and group columns by
1009 'complex/assembly factor'. P-value results from t-test statistical analysis were used to
1010 visualize the significance of each data point by overlaying the p-value matrix onto the t-
1011 test matrix in Morpheus and using the option size by 'p-value matrix' size minimum '0'
1012 and maximum '1.301' (p-value<0.05). For the correlation heatmap plot (Supplementary
1013 Fig. 2A), hierarchical clustering was used with the following parameters in Morpheus:
1014 metric 'one minus Pearson correlation', linkage method 'average' and cluster 'columns'.
1015 Topographical heatmaps (Supplementary Fig. 1B) were generated from log_2-transformed
1016 fold-change from t-test in Pymol using the PDB structures (CI: 5LDW, CII: 1ZOY, CIII:
1017 1BGY, CIV: 5Z62, CV: 7AJD) as previously described[26]. Venn diagrams were generated
1018 with DeepVenn[81] from gene names. Green (diagnostic-grade) and amber (borderline
1019 diagnostic-grade) lists for mitochondrial disease (v.0.787) and Mendeliome gene lists
1020 (v.0.12869) were retrieved from PanelApp Australia[27].
1021
1022 *Principal Component Analysis and Donut plots*
1023 PCA plots were generated in Perseus v 1.6.14.0 from t-test fold-change data for VC
1024 cohort (Fig. 1B) or protein abundances for PBMCs controls (Fig. 4B). Donut plots were
1025 generated in GraphPad Prism (v.10.2.0).
1026
1027 *RNA sequencing and transcriptomic analysis*
1028 RNA sequencing and analysis was performed as described[20]. In brief, RNA was extracted
1029 from muscle using miRNeasy Mini Kit (Qiagen) and treated with RNase-Free DNase
1030 (Qiagen) and RNA quality and quantity measured using TapeStation RNA ScreenTape
1031 analysis (Agilent) and Qubit RNA HS (Thermo Fisher Scientific). UDP4 underwent
1032 RNAseq within a group of 10 undiagnosed patients with available muscle samples, all
1033 with an RNA integrity number of 7.2 - 8.9. The Kapa Biosystems mRNA Hyper Prep Kit
1034 was used to generate sequencing libraries from 500ng RNA, and paired-end sequencing

1035 on an Illumina instrument was performed at the Yale Center of Genome Analysis to
1036 achieve 50-100 million read coverage.
1037
1038 For outlier detection, 100 control muscle samples were retrieved from GTEx using the
1039 Sequence Read Archive (SRA) toolkit and converted into FASTQ using Fastq-dump.
1040 Patient and control FASTQ files were aligned to hg38 human reference genome with
1041 GENCODE v26 annotations using STAR (v2.5.3a), with twopassMode = 'Basic' to enable
1042 detection of novel splice junctions, and duplicates marked using Picard (v2.9). RNA-
1043 SeQC (v2.3.4) was used to generate quality metrics from STAR-aligned bam files.
1044 Outliers were removed per the following criteria: (a) <45 million QC-passed reads, (b)
1045 <18,000 genes detected and (c) manual identification via Principal Component Analysis.
1046 This resulted in the removal of 11 samples, leaving 89 control samples from muscle.
1047 Outlier expression was assessed using the DROP (Detection of RNA Outliers Pipeline)
1048 pipeline v0.9.0 (https://github.com/gagneurlab/drop), with default settings[82]. Analysis was
1049 restricted to 6106 genes, which included Mendeliome and 'MitoExome' genes, as
1050 described previously[20].
1051
1052 *cDNA studies*
1053 The total RNA extracted from skeletal muscle for RNA sequencing was also synthesized
1054 into cDNA using the SuperScript III First-Strand Synthesis System (Thermo Fisher
1055 Scientific) as per both manufacturer's protocols and as described previously[83]. To
1056 examine mRNA splicing in *CCDC47*, PCR primers (5' GCGTGACTGAGCTACGGTT 3';
1057 5' CTCTGGGATGGCTTTACATGG 3') were designed to amplify the entire ORF from
1058 cDNA. PCR products were analyzed on 1% agarose (Bioline) gels using a 1-kb Plus DNA
1059 Ladder (Thermo Fisher Scientific) prior to cloning into a pCR2.1-TOPO vector using the
1060 TOPO TA cloning kit (Thermo Fisher Scientific) and transformation into TOP10 competent
1061 cells (Thermo Fisher Scientific). Individual colonies were examined and then underwent
1062 sanger sequencing.
1063
1064 **Conflict of Interest**
1065 The authors declare that the research was conducted in the absence of any commercial
1066 or financial relationships that could be construed as a potential conflict of interest.
1067
1068 **MitoMDT Diagnostic Network for Genomics and Omics**
1069 David R Thorburn, Aleksandra Filipovska, Michael T Ryan, David A Stroud, Diana
1070 Stojanovski, David Coman, Sean Murray, Ryan L Davis, John Christodoulou, Roula

Ghaoui, Suzanne CEH Sallevelt, Cas Simons, Stefan J Siira, Shanti Balasubramaniam, Alison G Compton, Daniel G MacArthur, Nicole J Lake, Amanda Samarasinghe, Yoni Elbaum, Catherine Atthow, Pauline McGrath, Ellenore M Martin, Madeleine Harris, Tegan Stait, Leah E Frajman, Simone Tregoning

Author contribution

Conceptualization: DHH, DRT, DAS
Methodology: DHH, NJC, LNS, NJL, LuEF, SSCA, TS, ST, LeEF, DRLR, MB, BoR, AGC, VK, CA
Investigation: DHH, MB, BrR, MJW, AV, CB, HP, JL, ZS, RM, RWT, AGC, WB, NBT, MLF
Visualization: DHH, NJC, LNS, NJL, SSCA
Website design and coding: NJC
Funding acquisition: DHH, JC, AGC, DRT, DAS
Project administration: AS, RB, DRT, DAS
Supervision: PM, ML, RM, RWT, MTR, ZS, JC, AGC, DRT, DAS
Writing – original draft: DHH, DRT, DAS
Writing – review & editing: all authors reviewed and edited the manuscript

Funding

This research was supported by Australian National Health and Medical Research Council (NHMRC) Project and Ideas grants (1140906 to DAS; 1164479 to DRT; 2010939 to MTR), Investigator Fellowships (2009732 to DAS, 2010149 to LuEF and 1155244 to DRT) and a Principal Research Fellowship (1155244 to DRT) along with funding by Australian Genomics Health Alliance (Australian Genomics) NHMRC Targeted Call for Research grant GNT1113531. Additional support came from the Australian Medical Research Future Fund Genomics Health Futures Mission (2007959 to DRT) and Acute Care Genomics (GHFM76747). The US Department of Defense Congressionally Directed Medical Research Programs (PR170396 to DRT). We thank the Mito Foundation for the provision of instrumentation through research equipment grants to DAS and DHH. Additionally, LuEF acknowledges support from the Mito Foundation. This work was also supported by grants from Royal Children's Hospital Foundation [2021-1377]. Work at the MCRI is supported through the Victorian Government's Operational Infrastructure Support Program. RWT is funded by the Wellcome Centre for Mitochondrial Research (203105/Z/16/Z), the Mitochondrial Disease Patient Cohort (UK) (G0800674), the Medical Research Council (MR/W019027/1), the Lily Foundation, Mito Foundation, the Pathological Society, the UK NIHR Biomedical Research Centre for Ageing and Age-related disease award to the Newcastle upon Tyne Foundation Hospitals NHS Trust,

LifeArc and the UK NHS Highly Specialised Service for Rare Mitochondrial Disorders of Adults and Children.

Acknowledgments

We acknowledge and thank all the families who contributed to this study. We thank all members of the Stroud and Stojanovski lab for input into experimental design and interpretation of data. We thank the Bio21 Mass Spectrometry and Proteomics Facility (MMSPF) for the provision of instrumentation, training, and technical support. The Chair in Genomic Medicine awarded to JC is generously supported by The Royal Children's Hospital Foundation. The authors thank staff of the Pathology Collection Department at The Royal Children's Hospital for obtaining the consent of participants and the collection of samples. The authors thank staff of the Anaesthetic and Surgical Departments at the Royal Children's Hospital.

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[84] partner repository with the dataset identifier PXD055473. De-identified genomic data from this study are available for ethically approved research. For queries about the genomic data sets please contact: david.thorburn@mcri.edu.au.

References

1. Haack, T.B. *et al.* Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J Med Genet* **49**, 277-283 (2012).
2. Taylor, R.W. *et al.* Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *JAMA* **312**, 68-77 (2014).
3. Ohtake, A. *et al.* Diagnosis and molecular basis of mitochondrial respiratory chain disorders: exome sequencing for disease gene identification. *Biochim Biophys Acta* **1840**, 1355-1359 (2014).
4. Wortmann, S.B., Koolen, D.A., Smeitink, J.A., van den Heuvel, L. & Rodenburg, R.J. Whole exome sequencing of suspected mitochondrial patients in clinical practice. *J Inherit Metab Dis* **38**, 437-443 (2015).
5. Davis, R.L. *et al.* Use of Whole-Genome Sequencing for Mitochondrial Disease Diagnosis. *Neurology* **99**, e730-e742 (2022).
6. 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).
7. Gorman, G.S. *et al.* Mitochondrial diseases. *Nat Rev Dis Primers* **2**, 16080 (2016).
8. Skladal, D., Halliday, J. & Thorburn, D.R. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain* **126**, 1905-1912 (2003).
9. Smeets, H.J.M., Sallevelt, S. & Herbert, M. Reproductive options in mitochondrial disease. *Handb Clin Neurol* **194**, 207-228 (2023).

1150 10. Frazier, A.E., Vincent, A.E., Turnbull, D.M., Thorburn, D.R. & Taylor, R.W.
1151 Assessment of mitochondrial respiratory chain enzymes in cells and tissues.
1152 *Methods Cell Biol* **155**, 121-156 (2020).
1153 11. Kremer, L.S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing.
1154 *Nat Commun* **8**, 15824 (2017).
1155 12. Yepez, V.A. *et al.* Clinical implementation of RNA sequencing for Mendelian disease
1156 diagnostics. *Genome Med* **14**, 38 (2022).
1157 13. Fresard, L. *et al.* Identification of rare-disease genes using blood transcriptome
1158 sequencing and large control cohorts. *Nat Med* **25**, 911-919 (2019).
1159 14. Murdock, D.R. *et al.* Transcriptome-directed analysis for Mendelian disease
1160 diagnosis overcomes limitations of conventional genomic testing. *J Clin Invest* **131**
1161 (2021).
1162 15. Chen, E. *et al.* Rates and Classification of Variants of Uncertain Significance in
1163 Hereditary Disease Genetic Testing. *JAMA Netw Open* **6**, e2339571 (2023).
1164 16. Alston, C.L., Stenton, S.L., Hudson, G., Prokisch, H. & Taylor, R.W. The genetics of
1165 mitochondrial disease: dissecting mitochondrial pathology using multi-omic
1166 pipelines. *J Pathol* **254**, 430-442 (2021).
1167 17. Kopajtich, R. *et al.* Integration of proteomics with genomics and transcriptomics
1168 increases the diagnostic rate of Mendelian disorders. *medRxiv*,
1169 2021.2003.2009.21253187 (2021).
1170 18. Xiao, Q. & Lauschke, V.M. The prevalence, genetic complexity and population-
1171 specific founder effects of human autosomal recessive disorders. *NPJ Genom Med*
1172 **6**, 41 (2021).
1173 19. Sahni, N. *et al.* Widespread macromolecular interaction perturbations in human
1174 genetic disorders. *Cell* **161**, 647-660 (2015).
1175 20. Amarasekera, S.S.C. *et al.* Multi-omics identifies large mitoribosomal subunit
1176 instability caused by pathogenic MRPL39 variants as a cause of pediatric onset
1177 mitochondrial disease. *Hum Mol Genet* **32**, 2441-2454 (2023).
1178 21. Bakhshalizadeh, S. *et al.* Deficiency of the mitochondrial ribosomal subunit,
1179 MRPL50, causes autosomal recessive syndromic premature ovarian insufficiency.
1180 *Hum Genet* **142**, 879-907 (2023).
1181 22. Helman, G. *et al.* Multiomic analysis elucidates Complex I deficiency caused by a
1182 deep intronic variant in NDUFB10. *Hum Mutat* **42**, 19-24 (2021).
1183 23. Horvath, R. *et al.* Childhood onset mitochondrial myopathy and lactic acidosis
1184 caused by a stop mutation in the mitochondrial cytochrome c oxidase III gene. *J*
1185 *Med Genet* **39**, 812-816 (2002).
1186 24. Van Haute, L. *et al.* TEFM variants impair mitochondrial transcription causing
1187 childhood-onset neurological disease. *Nat Commun* **14**, 1009 (2023).
1188 25. Lake, N.J. *et al.* Biallelic Mutations in MRPS34 Lead to Instability of the Small
1189 Mitoribosomal Subunit and Leigh Syndrome. *Am J Hum Genet* **101**, 239-254 (2017).
1190 26. Stroud, D.A. *et al.* Accessory subunits are integral for assembly and function of
1191 human mitochondrial complex I. *Nature* **538**, 123-126 (2016).
1192 27. Martin, A.R. *et al.* PanelApp crowdsources expert knowledge to establish
1193 consensus diagnostic gene panels. *Nat Genet* **51**, 1560-1565 (2019).

1194 28. Rath, S. *et al.* MitoCarta3.0: an updated mitochondrial proteome now with sub-
1195 organelle localization and pathway annotations. *Nucleic Acids Res* **49**, D1541-
1196 D1547 (2021).
1197 29. Granata, C. *et al.* High-intensity training induces non-stoichiometric changes in the
1198 mitochondrial proteome of human skeletal muscle without reorganisation of
1199 respiratory chain content. *Nat Commun* **12**, 7056 (2021).
1200 30. Thompson, K., Stroud, D.A., Thorburn, D.R. & Taylor, R.W. Investigation of oxidative
1201 phosphorylation activity and complex composition in mitochondrial disease. *Handb*
1202 *Clin Neurol* **194**, 127-139 (2023).
1203 31. Bernier, F.P. *et al.* Diagnostic criteria for respiratory chain disorders in adults and
1204 children. *Neurology* **59**, 1406-1411 (2002).
1205 32. Acin-Perez, R. *et al.* Respiratory complex III is required to maintain complex I in
1206 mammalian mitochondria. *Mol Cell* **13**, 805-815 (2004).
1207 33. Rahman, S. *et al.* Leigh syndrome: clinical features and biochemical and DNA
1208 abnormalities. *Ann Neurol* **39**, 343-351 (1996).
1209 34. Tucker, E.J. *et al.* Mutations in MTFMT underlie a human disorder of formylation
1210 causing impaired mitochondrial translation. *Cell Metab* **14**, 428-434 (2011).
1211 35. Massa, V. *et al.* Severe infantile encephalomyopathy caused by a mutation in
1212 COX6B1, a nucleus-encoded subunit of cytochrome c oxidase. *Am J Hum Genet* **82**,
1213 1281-1289 (2008).
1214 36. Hock, D.H., Robinson, D.R.L. & Stroud, D.A. Blackout in the powerhouse: clinical
1215 phenotypes associated with defects in the assembly of OXPHOS complexes and
1216 the mitoribosome. *Biochem J* **477**, 4085-4132 (2020).
1217 37. Zong, S. *et al.* Structure of the intact 14-subunit human cytochrome c oxidase. *Cell*
1218 *Res* (2018).
1219 38. Signes, A. & Fernandez-Vizarra, E. Assembly of mammalian oxidative
1220 phosphorylation complexes I-V and supercomplexes. *Essays Biochem* **62**, 255-270
1221 (2018).
1222 39. Cizkova, A. *et al.* TMEM70 mutations cause isolated ATP synthase deficiency and
1223 neonatal mitochondrial encephalocardiomyopathy. *Nat Genet* **40**, 1288-1290
1224 (2008).
1225 40. Catteruccia, M. *et al.* Persistent pulmonary arterial hypertension in the newborn
1226 (PPHN): a frequent manifestation of TMEM70 defective patients. *Mol Genet Metab*
1227 **111**, 353-359 (2014).
1228 41. Torraco, A. *et al.* TMEM70: a mutational hot spot in nuclear ATP synthase deficiency
1229 with a pivotal role in complex V biogenesis. *Neurogenetics* **13**, 375-386 (2012).
1230 42. Braczynski, A.K. *et al.* ATP synthase deficiency due to TMEM70 mutation leads to
1231 ultrastructural mitochondrial degeneration and is amenable to treatment. *Biomed*
1232 *Res Int* **2015**, 462592 (2015).
1233 43. Magner, M. *et al.* TMEM70 deficiency: long-term outcome of 48 patients. *J Inherit*
1234 *Metab Dis* **38**, 417-426 (2015).
1235 44. Sanchez-Caballero, L. *et al.* TMEM70 functions in the assembly of complexes I and
1236 V. *Biochim Biophys Acta Bioenerg* **1861**, 148202 (2020).

1237 45. Guerrero-Castillo, S. *et al.* The Assembly Pathway of Mitochondrial Respiratory
1238 Chain Complex I. *Cell Metab* **25**, 128-139 (2017).
1239 46. Mowat, D. *et al.* Respiratory chain complex III [correction of complex] in deficiency
1240 with pruritus: a novel vitamin responsive clinical feature. *J Pediatr* **134**, 352-354
1241 (1999).
1242 47. Morimoto, M. *et al.* Bi-allelic CCDC47 Variants Cause a Disorder Characterized by
1243 Woolly Hair, Liver Dysfunction, Dysmorphic Features, and Global Developmental
1244 Delay. *Am J Hum Genet* **103**, 794-807 (2018).
1245 48. Yang, Q. *et al.* Clinical and genetic analysis of trichohepatoneurodevelopmental
1246 syndrome caused by a CCDC47 variant. *Heliyon* **10**, e27955 (2024).
1247 49. Iannello, G. *et al.* Simple, Fast, and Efficient Method for Derivation of Dermal
1248 Fibroblasts From Skin Biopsies. *Curr Protoc* **3**, e714 (2023).
1249 50. Amunts, A., Brown, A., Toots, J., Scheres, S.H.W. & Ramakrishnan, V. Ribosome.
1250 The structure of the human mitochondrial ribosome. *Science* **348**, 95-98 (2015).
1251 51. Bakker, H.D. *et al.* Human alpha-N-acetylgalactosaminidase (alpha-NAGA)
1252 deficiency: no association with neuroaxonal dystrophy? *Eur J Hum Genet* **9**, 91-96
1253 (2001).
1254 52. Sung, A.Y. *et al.* Systematic analysis of NDUFAF6 in complex I assembly and
1255 mitochondrial disease. *Nat Metab* **6**, 1128-1142 (2024).
1256 53. Pagliarini, D.J. *et al.* A mitochondrial protein compendium elucidates complex I
1257 disease biology. *Cell* **134**, 112-123 (2008).
1258 54. McKenzie, M. *et al.* Mutations in the gene encoding C8orf38 block complex I
1259 assembly by inhibiting production of the mitochondria-encoded subunit ND1. *J Mol*
1260 *Biol* **414**, 413-426 (2011).
1261 55. Baide-Mairena, H. *et al.* Mutations in the mitochondrial complex I assembly factor
1262 NDUFAF6 cause isolated bilateral striatal necrosis and progressive dystonia in
1263 childhood. *Mol Genet Metab* **126**, 250-258 (2019).
1264 56. Hartmannova, H. *et al.* Acadian variant of Fanconi syndrome is caused by
1265 mitochondrial respiratory chain complex I deficiency due to a non-coding mutation
1266 in complex I assembly factor NDUFAF6. *Hum Mol Genet* **25**, 4062-4079 (2016).
1267 57. Kim, J., Lee, J. & Jang, D.H. NDUFAF6-Related Leigh Syndrome Caused by Rare
1268 Pathogenic Variants: A Case Report and the Focused Review of Literature. *Front*
1269 *Pediatr* **10**, 812408 (2022).
1270 58. Nguengang Wakap, S. *et al.* Estimating cumulative point prevalence of rare
1271 diseases: analysis of the Orphanet database. *Eur J Hum Genet* **28**, 165-173 (2020).
1272 59. Van Bergen, N.J. *et al.* Severe NAD(P)HX Dehydratase (NAXD) Neurometabolic
1273 Syndrome May Present in Adulthood after Mild Head Trauma. *Int J Mol Sci* **24** (2023).
1274 60. Tucker, E.J. *et al.* Premature Ovarian Insufficiency in CLPB Deficiency:
1275 Transcriptomic, Proteomic and Phenotypic Insights. *J Clin Endocrinol Metab* **107**,
1276 3328-3340 (2022).
1277 61. Van Bergen, N.J. *et al.* Biallelic Variants in PYROXD2 Cause a Severe Infantile
1278 Metabolic Disorder Affecting Mitochondrial Function. *Int J Mol Sci* **23** (2022).
1279 62. Frazier, A.E. *et al.* Fatal perinatal mitochondrial cardiac failure caused by recurrent
1280 de novo duplications in the ATAD3 locus. *Med (N Y)* **2**, 49-73 (2021).

63. Borna, N.N. *et al.* Mitochondrial ribosomal protein PTCD3 mutations cause oxidative phosphorylation defects with Leigh syndrome. *Neurogenetics* **20**, 9-25 (2019).
64. Van Bergen, N.J. *et al.* Mutations in the exocyst component EXOC2 cause severe defects in human brain development. *J Exp Med* **217** (2020).
65. Lunke, S. *et al.* Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat Med* **29**, 1681-1691 (2023).
66. Formosa, L.E., Dibley, M.G., Stroud, D.A. & Ryan, M.T. Building a complex complex: Assembly of mitochondrial respiratory chain complex I. *Semin Cell Dev Biol* **76**, 154-162 (2018).
67. Frederiksen, S.D. *et al.* Rare disorders have many faces: in silico characterization of rare disorder spectrum. *Orphanet J Rare Dis* **17**, 76 (2022).
68. Turner, T.N. *et al.* Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. *Hum Mol Genet* **24**, 5995-6002 (2015).
69. Dofash, L.N. *et al.* Biallelic variants in HMGCS1 are a novel cause of rare rigid spine syndrome. *medRxiv*, 2023.2010.2025.23297129 (2023).
70. Gorzynski, J.E. *et al.* Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting. *N Engl J Med* **386**, 700-702 (2022).
71. Australian Genomics Health Alliance Acute Care, F. *et al.* Feasibility of Ultra-Rapid Exome Sequencing in Critically Ill Infants and Children With Suspected Monogenic Conditions in the Australian Public Health Care System. *JAMA* **323**, 2503-2511 (2020).
72. Dimmock, D. *et al.* Project Baby Bear: Rapid precision care incorporating rWGS in 5 California children's hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am J Hum Genet* **108**, 1231-1238 (2021).
73. Guzman, U.H. *et al.* Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat Biotechnol* (2024).
74. Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* **369**, 1502-1511 (2013).
75. Akesson, L.S. *et al.* Early diagnosis of Pearson syndrome in neonatal intensive care following rapid mitochondrial genome sequencing in tandem with exome sequencing. *Eur J Hum Genet* **27**, 1821-1826 (2019).
76. Hoq, M. *et al.* A prospective, cross-sectional study to establish age-specific reference intervals for neonates and children in the setting of clinical biochemistry, immunology and haematology: the HAPPI Kids study protocol. *BMJ Open* **9**, e025897 (2019).
77. Robinson, D.R.L. *et al.* Applying Sodium Carbonate Extraction Mass Spectrometry to Investigate Defects in the Mitochondrial Respiratory Chain. *Front Cell Dev Biol* **10**, 786268 (2022).
78. 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).

79. Bruderer, R. *et al*. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410 (2015).
80. Tyanova, S. & Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol Biol* **1711**, 133-148 (2018).
81. Hulsen, T. arXiv:2210.04597 (2022).
82. Yepez, V.A. *et al*. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc* **16**, 1276-1296 (2021).
83. Calvo, S.E. *et al*. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* **4**, 118ra110 (2012).
84. Perez-Riverol, Y. *et al*. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552 (2022).