Integrated multi-omics reveals anaplerotic rewiring in methylmalonyl-CoA mutase deficiency

Patrick Forny[1a], Ximena Bonilla[2a], David Lamparter[3,8a], Wenguang Shao[4,8a], Tanja Plessl[1], Caroline Frei[1], Anna Bingisser[1], Sandra Goetze[4,8], Audrey van Drogen[4,8], Keith Harshman[3,8], Patrick G. A. Pedrioli[4,8], Cedric Howald[3], Florian Traversi[1], Sarah Cherkaoui[1], Raphael J. Morscher[1], Luke Simmons[5], Merima Forny[1], Ioannis Xenarios[6,8], Ruedi Aebersold[7,8], Nicola Zamboni[7,8], Gunnar Rätsch[2], Emmanouil Dermitzakis[3,8], Bernd Wollscheid[4,8*], Matthias R. Baumgartner[1*], D. Sean Froese[1*]

[1] Division of Metabolism, University Children's Hospital Zurich, University of Zurich, Switzerland

[2] Biomedical Informatics, Department of Computer Science, Swiss Federal Institute of Technology / ETH Zürich, Switzerland

[3] Health 2030 Genome Center, Geneva, Switzerland

[4] Institute of Translational Medicine, Department of Health Science and Technology, Swiss Federal institute of Technology / ETH Zürich, Switzerland

[5] Division of Child Neurology, University Children's Hospital Zurich, University of Zurich, Switzerland

[6] Agora Center, Rue du Bugnon 21, Lausanne, Switzerland

[7] Institute of Molecular Systems Biology, Department of Biology, Swiss Federal Institute of Technology / ETH Zürich, Switzerland

[8] PHRT Swiss Multi-OMICS Center / smoc.ethz.ch, Switzerland

[a] These authors contributed equally to this study.

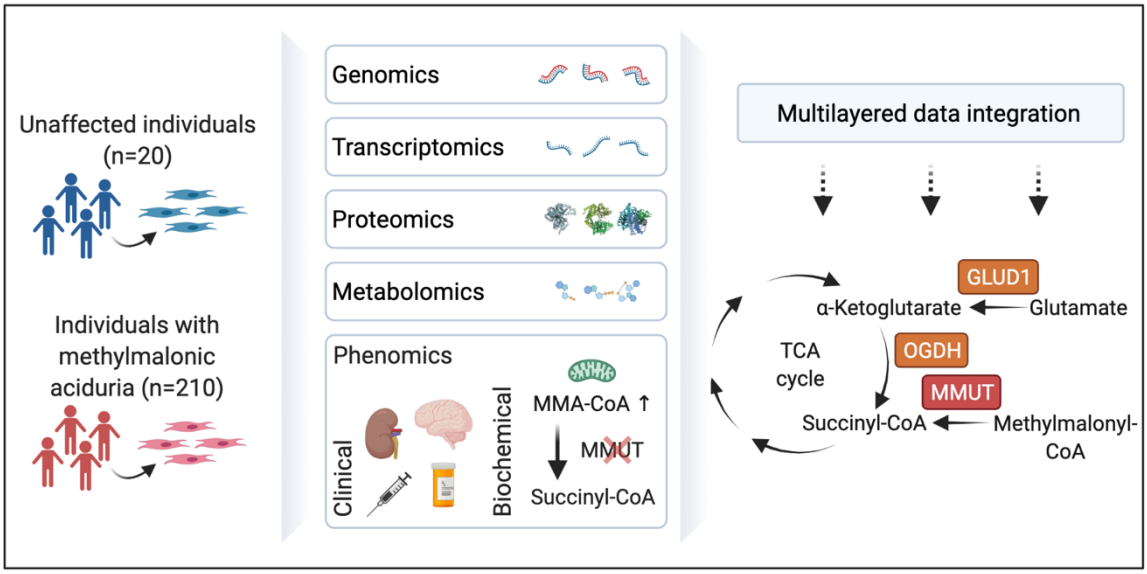
[*] Correspondence: sean.froese@kispi.uzh.ch (Lead Contact), matthias.baumgartner@kispi.uzh.ch, bernd.wollscheid@hest.ethz.ch

Abstract

Multi-layered omics technologies can help define relationships between genetic factors, biochemical processes and phenotypes thus extending research of monogenic diseases beyond identifying their cause. We implemented a multi-layered omics approach for the inherited metabolic disorder methylmalonic aciduria. We performed whole genome sequencing, transcriptomic sequencing, and mass spectrometry-based proteotyping from matched primary fibroblast samples of 230 individuals (210 affected, 20 controls) and related the molecular data to 105 phenotypic features. Integrative analysis identified a molecular diagnosis for 84% (179/210) of affected individuals, the majority (150) of whom had pathogenic variants in methylmalonyl-CoA mutase (*MMUT*). Untargeted integration of all three omics layers revealed dysregulation of TCA cycle and surrounding metabolic pathways, a finding that was further supported by multi-organ metabolomics of a hemizygous *Mmut* mouse model. Stratification by phenotypic severity indicated downregulation of oxoglutarate dehydrogenase and upregulation of glutamate dehydrogenase in disease. This was supported by metabolomics and isotope tracing studies which showed increased glutamine-derived anaplerosis. We further identified MMUT to physically interact with both, oxoglutarate dehydrogenase and glutamate dehydrogenase providing a mechanistic link. This study emphasizes the utility of a multi-modal omics approach to investigate metabolic diseases and highlights glutamine anaplerosis as a potential therapeutic intervention point in methylmalonic aciduria.

Take home message: Forny et al. have combined integrative multi-omics with clinical and biochemical features and demonstrate an increased diagnostic rate compared to genome sequencing alone and the identification of anaplerotic rewiring as a potentially therapeutically targetable feature of the rare inborn error of metabolism methylmalonic aciduria.

Graphical abstract



Keywords

Rare disease, inborn error of metabolism, multi-omics, RNA-seq, WGS, proteomics, methylmalonic aciduria

Main

Inborn errors of metabolism (IEMs), first described by Archibold Garrod (Garrod, 1908), are inherited diseases resulting in inadequate function of metabolic proteins. IEMs represent a group of nearly 1500 diseases with a combined incidence of approximately 1:800 births. They present a clinically and genetically heterogeneous picture making them inherently difficult to diagnose (Ferreira et al., 2021; Rahman, 2020). Beyond their diagnostic challenges, the pathomechanisms of many IEMs are not well understood, hence most IEMs lack rationalized treatment approaches (Hirano et al., 2018).

Technological advances in genomics and mass spectrometry, leveraging datasets of whole molecule classes (omics), have recently led to a paradigm shift in their use as diagnostic tools. For example, single-layer whole genome sequencing has achieved diagnostic rates of 30-50% (Palmer et al., 2021; Schon et al., 2021), while dual-layer combination with RNA sequencing can improve this by 10-35% (Cummings et al., 2017; Frésard et al., 2019; Gonorazky et al., 2019; Kremer et al., 2017). In spite of these advances a significant number of patients remain undiagnosed and course prediction also remains poor, mainly due to a lack of pathomechanistic understanding and often unclear genotype/phenotype relationships.

Multi-layered omics data have the potential to not only increase diagnosis rates of inborn errors of metabolism, but also to uncover novel mechanistic insights into disease pathophysiology (Argmann et al., 2016), thus potentially indicating new therapeutic targets. Such an approach is key to move beyond the traditional “one gene – one disease” view of these disorders which fails to explain phenotypic heterogeneity based on genetic variation only.

Here, we have generated multi-layered omics information from individuals suspected with the prototypic IEM methylmalonic aciduria (MMA). MMA has approximately 20 potential monogenic causes most of which result in deficiency of the methylmalonyl-CoA mutase (MMUT) enzyme due to loss of function or proper cofactor assembly. Even though the position of MMUT as an anaplerotic tricarboxylic acid (TCA) cycle enzyme producing succinyl-CoA from methylmalonyl-CoA is well known, the main metabolic disturbances in MMA remain an open question.

By combining WGS, RNA-seq, and DIA-MS derived proteotype information with phenotypic features we identified disease causative and pathogenic features in a cohort of MMA affected individuals. These revealed differentially expressed transcripts and proteins directly related to anaplerosis of the TCA cycle. Moreover, follow-up studies utilizing untargeted metabolomics and [^{13}C]glutamine tracing identified a shift from reductive to oxidative TCA cycling for glutamine derived metabolites due to dysregulation of glutamate dehydrogenase and oxoglutarate dehydrogenase enzymes, which we found to physically interact with MMUT. These may be compensation and potential intervention points in MMA.

Results

Monogenic disease variant detection through multi-omics

To extend the understanding of MMA from the causative genomic lesions to the affected biochemical processes, we preformed WGS, RNA-seq and DIA-MS based proteotyping on fibroblasts taken from 230 individuals (210 affected by MMA, 20 unaffected), representing a mainly European cohort collected over a 25-year period (**Fig. 1a** and **Extended data Fig. 1a**). Overall quality assurance tests revealed a mean of high quality aligned genomic reads of 8.7×10^8 at a median genomic coverage of >38-fold (**Extended data Fig. 1b**). A median of 3.74 million SNVs were called using the Genome Analysis Toolkit (McKenna et al., 2010) and DeepVariant (Poplin et al., 2018). RNA-seq data showed a median Phred score of >36.3 at three and more cycles (**Extended data Fig. 1c**), while proteomics data showed a high reproducibility with 2218 proteins detected in at least 75% of samples (**Extended data Fig. 1d**). For nine of the 230 samples RNA extraction yielded insufficient nucleic acid amounts to proceed with transcriptome sequencing; hence, these datasets were excluded from all further analysis (transcriptomics data of sample IDs 22, 54, 59, 78, 89, 109, 123, 207, 221).

MMA is caused by defective processing of methylmalonyl-CoA – a product of propionyl-CoA catabolism and an anaplerotic substrate of the TCA cycle – resulting from deficiency of the enzyme methylmalonyl-CoA mutase (MMUT) or proteins related to production of its cofactor (adenosylcobalamin) or substrate (**Fig. 1a**). Biochemical assay of propionyl-CoA catabolism (PI) (Froese and Baumgartner, 2020), and MMUT enzyme activity (Forny et al., 2014) strongly correlated across all samples ($r=0.73$, $p<0.0001$) (**Extended data Fig. 2a**). Fibroblasts from 150 individuals with MMA had reduced MMUT activity, including 123 which did not increase upon cofactor supplementation (**Extended data Fig. 2b**, **c**), while those of 60 individuals had MMUT activity similar to controls (**Fig. 1b**, **Extended data Table 2**). Those with reduced MMUT activity were designated as MMUT-deficient, the others were designated as MMA affected individuals with unknown cause.

In the MMUT-deficient samples we searched for disease causing variations in the *MMUT* gene and identified a bi-allelic molecular diagnosis in 148/150 individuals (**Extended data Table 1**). Pathogenic variants constituted 165 missense alleles, 105 truncating alleles, 21 splicing alleles, two alleles with in-frame deletions and three alleles containing copy number variants (**Fig. 1c, d** and **Extended data Fig. 2d**), of which 41 variants were novel (**Extended data Table 1**). RNA-seq identified reduced *MMUT* RNA expression in cells from MMUT-deficient individuals compared to the other groups (**Fig. 1e**). Individuals with strongly reduced RNA expression were enriched for splicing and/or truncating variants, consistent with nonsense mediated decay (**Extended data Fig. 2e**). DIA-MS based proteome measurements revealed reduced MMUT protein levels in MMUT-deficient cell lines (**Fig. 1f**) which was distributed across all variant types (**Extended data Fig. 2e**). Consistent with its disease-causing role, MMUT RNA and protein levels were positively and significantly associated with PI and MMUT activity (**Extended data Fig. 2f**) and MMUT represented the most significantly dysregulated RNA and protein of MMUT-deficient samples when compared against all other samples (**Fig. 1g** and **Extended data Fig 2g**).

In the 60 unknown samples, we identified bi-allelic disease-causing variants in *ACSF3* (17 individuals) (**Fig. 1h**), *TCN2* (3 individuals), *SUCLA2* (1 individual) and MMAB (1 individual) (**Extended data Table 1**). By searching RNA-seq for aberrantly expressed genes using OUTRIDER (Brechtmann et al., 2018) (**Extended data Fig 3**), we identified two individuals with very low *ACSF3* expression; two with aberrant *SUCLA2* expression, in whom we confirmed predicted splicing and copy number variants at the genomic level (**Extended data Table 1**); two with very low *MMAA* expression, confirmed by complementation analysis; and one with low *MMAB* transcript, also confirmed by complementation analysis. Therefore, we identified a molecular cause for 29/60 remaining individuals, including 18 pathogenic mutations, of which 10 were novel (**Extended data Table 1**). In sum, we found a diagnosis for 177/210 (84%) affected individuals (**Fig. 1i**), including 150 with deficiency of MMUT and 19 of *ACSF3*, the largest cohort of *ACSF3* individuals so far described.

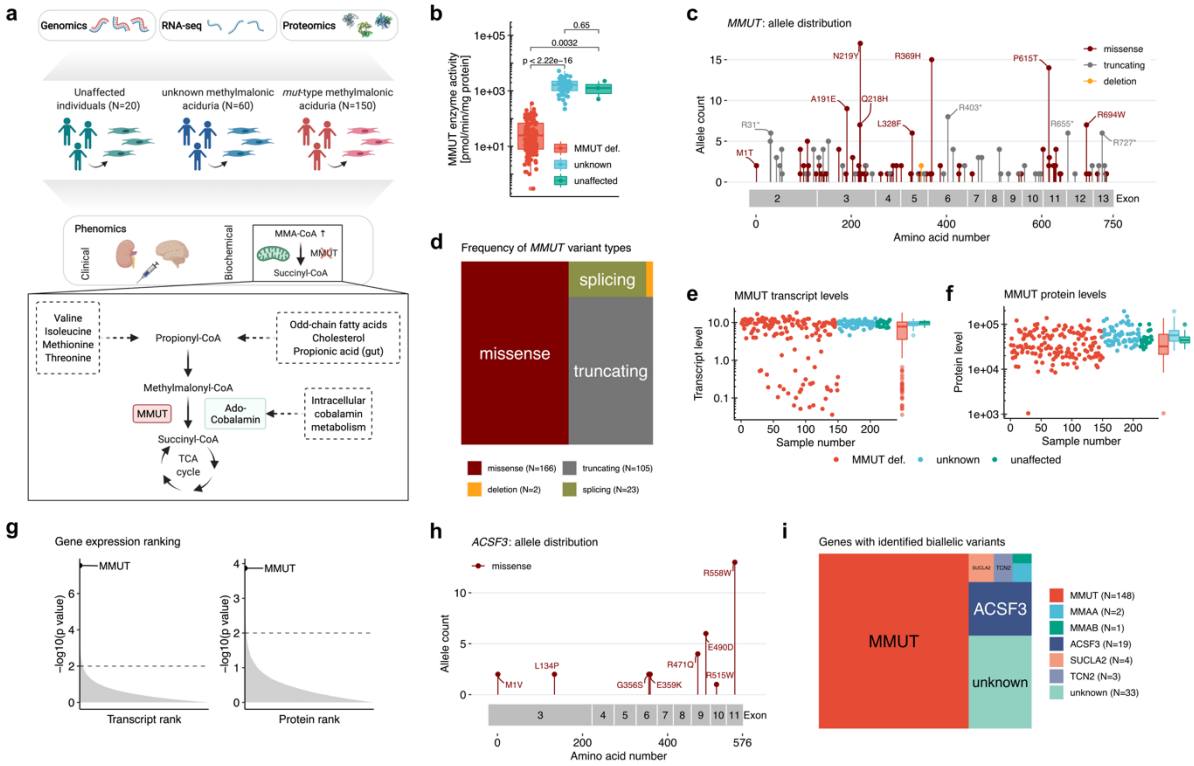


Fig. 1: Multi-facetted omics view enables a molecular diagnosis in 84% of individuals. a, Study overview with depiction of the propionate pathway including its precursors and the pathways catalyzed by MMUT. **b**, MMUT enzyme activity per study sub cohort. **c**, Lollipop plot of all pathogenic variants found on the *MMUT* gene. **d**, Proportions of variant types as identified on the *MMUT* gene. **e**, Transcript levels of MMUT divided in study sub cohorts. **f**, MMUT protein levels divided in study sub cohorts. **g**, Gene ranks according to p-values as calculated by gene-wise Welch's *t*-test in the proteomics and transcriptomics data. **h**, Lollipop plot of pathogenic variants identified in *ACSF3*. **i**, Proportions of affected genes identified in the whole cohort.

Phenotypic description and association to disease severity

A correlation matrix of all continuous numeric variables as well as discrete phenotypic variables (n = 105), spanning clinical symptoms at presentation and during disease course (n = 53), clinical treatments required and response (n = 23), clinical chemistry of blood or tissues including metabolite measurements (n = 13), and *in vitro* biochemical parameters (n = 13), revealed a cluster of

164 biochemical parameters (MMUT activity, PI) that showed strong correlation across many variables
165 (**Fig. 2a**, **b**).

166 Since these few variables strongly associated to many clinical features, we postulated that most
167 disease characteristics might be well predicted by one or a select few variables. As proof-of-principle,
168 we established a clinical severity score (CSS), which incorporated the outcome of five typical clinical
169 features ((Forny et al., 2021), for composition see **Methods**), whereby a score of 0 represented
170 asymptomatic individuals, 1 mild disease, and score 2 or higher moderate to severe disease.
171 Comparison of the CSS against all phenotypic parameters demonstrated significant correlation with
172 54 individual variables (**Fig. 2c**), including many classical phenotypic symptoms of MMAuria, such
173 as acidosis, hyperammonemia and muscular hypotonia as well as the requirement for dietary and
174 pharmacological interventions (**Extended data Fig. 4a**). Importantly, the CSS also inversely
175 correlated with age of onset (**Extended data Fig. 4b**), a parameter that on its own has been used as an
176 indication of clinical severity (Hörster et al., 2009).

177 Multiple correlation analysis identified PI in the presence of hydroxocobalamin (PI+) to significantly
178 correlate to 41 phenotypic features, the most of any individual continuous variable (**Fig. 2c**). This
179 contrasts, for example, to age at onset which significantly correlated with only 17 parameters (**Fig.**
180 **2c**). Closer inspection revealed PI+ to be inversely related to disease severity, including significant
181 correlation with e.g. glomerular filtration rate, age at disease onset, methylmalonic acid concentration
182 in plasma, presence of clinical interventions such as protein restriction and the CSS (**Fig. 2d** and
183 **Extended data Fig. 4c**,**d)**. Therefore, we used PI+ as an approximation of clinical disease severity in
184 this study.

185

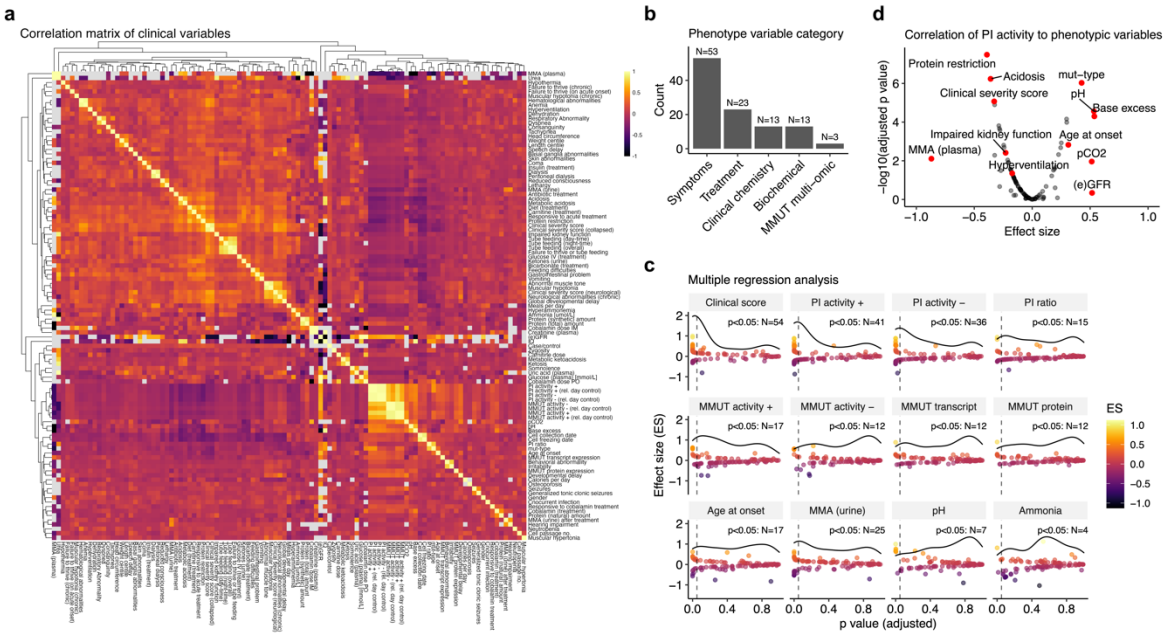


Fig. 2: Phenomics analysis reveals two main surrogate markers of disease severity (clinical severity score and PI+ activity). a, Correlation matrix of all continuous numeric and discrete phenotype variables. **b**, Number of phenotypic traits according to five phenotype subcategories. **c**, Panel of selected phenotypic traits and their overall strength of representing the entirety of the phenomics dataset (here termed clinical disease severity) as assessed by linear modelling. Each point represents the result of linear regression against one other phenotypic variable with the effect size on the y-axis and the resulting p-value on the x-axis. The horizontal curved line indicates the density of data points as distributed along the x-axis. Vertical dashed line indicates threshold of significance (p-value <0.05). **d**, Linear regression results of the PI+ activity variable compared against the rest of the phenotypic variables.

Multi-layered biology reveals disruption of TCA cycle and associated pathways

To identify disease-associated expression alterations of genes, proteins, and pathways, we attempted global assessment of transcript and protein expression. Since patients with TCN2, SUCLA2 and ACSF3 deficiency do not show typical signs and symptoms of classical MMA, we compared MMUT-deficient (150 samples) with all non-MMUT-deficient samples (control, 80 samples).

Investigation of transcripts and proteins using differential correlation patterns (Pearson correlation method), dimensionality reduction via Principal Component Analysis and DESeq2 did not immediately yield clear grouping of the data, nor obvious expression pattern differences between the groups (**Extended data Fig. 5a**, **b**, **c**). However, multi-omics factor analysis (Argelaguet et al., 2018), integrating both genetic data layers and proteotyping data, identified mitochondrial metabolic pathways and in particular the electron transport chain and the TCA cycle to be enriched in MMUT-deficient samples (**Fig. 3a**). In more detail, the proteins SLC16A3, CS, MDH2, and OGDH were found to be the main drivers of this particular factor variance in the proteotyping data within the TCA-associated gene sets (**Fig. 3b**). Linear discriminant analysis of genes shared between transcriptomics and proteotyping indicated MMUT as the strongest and SUCLA2, OGDH and PDHB to be top drivers of separation between MMUT-deficient and control samples (**Fig. 3c**).

Further, gene set enrichment analysis utilizing sample stratification by disease severity, both by CSS and PI+, also identified oxidative phosphorylation and the TCA cycle as over-represented pathways in the proteomics (CSS and PI+) and transcriptomics (CSS) datasets (**Fig. 3d**). Within the underlying differential expression analysis, performed using a linear mixed modelling approach (Lamparter et al., 2017), genes in the proteomics dataset that had an effect size $|>0.2|$ and were significantly (p-value $<$ 0.01) enriched for mitochondrial localization, as listed in MitoCarta 3.0 (Rath et al., 2021) (**Fig. 3e**). Closer examination identified MMUT to be significantly downregulated in disease at both the RNA and protein level, while ALDH2, which catalyzes the interchange between methylmalonate and methylmalonate semialdehyde, was upregulated in both (**Fig. 3f**, **g**). A further upregulated transcript was PDK4 (**Fig. 3g**), which is responsible for the phosphorylation and as a consequence inactivation of the pyruvate dehydrogenase complex. However, the proteins with the overall largest effect size were OGDH (downregulated in disease) and GLUD1 (upregulated in disease) (**Fig. 3f**).

The biological relevance of these changes were confirmed in a hemizygous mouse model of MMUT-deficiency (Forny et al., 2016) (**Fig. 3h**). Untargeted metabolomics of brain, heart, kidney, liver, plasma and urine confirmed elevated levels of the eponymous metabolite methylmalonic acid in mutant animals, while pathway analysis identified significantly dysregulated TCA cycle pathways in

all tissues and urine (**Fig. 3i**). Transcriptomics of brain tissue further confirmed the expected 50% reduction in *Mmut* transcript of mutant mice, along with enrichment of electron transport chain and oxidative phosphorylation pathways (**Fig. 3j** and **Extended data Fig. 6**).

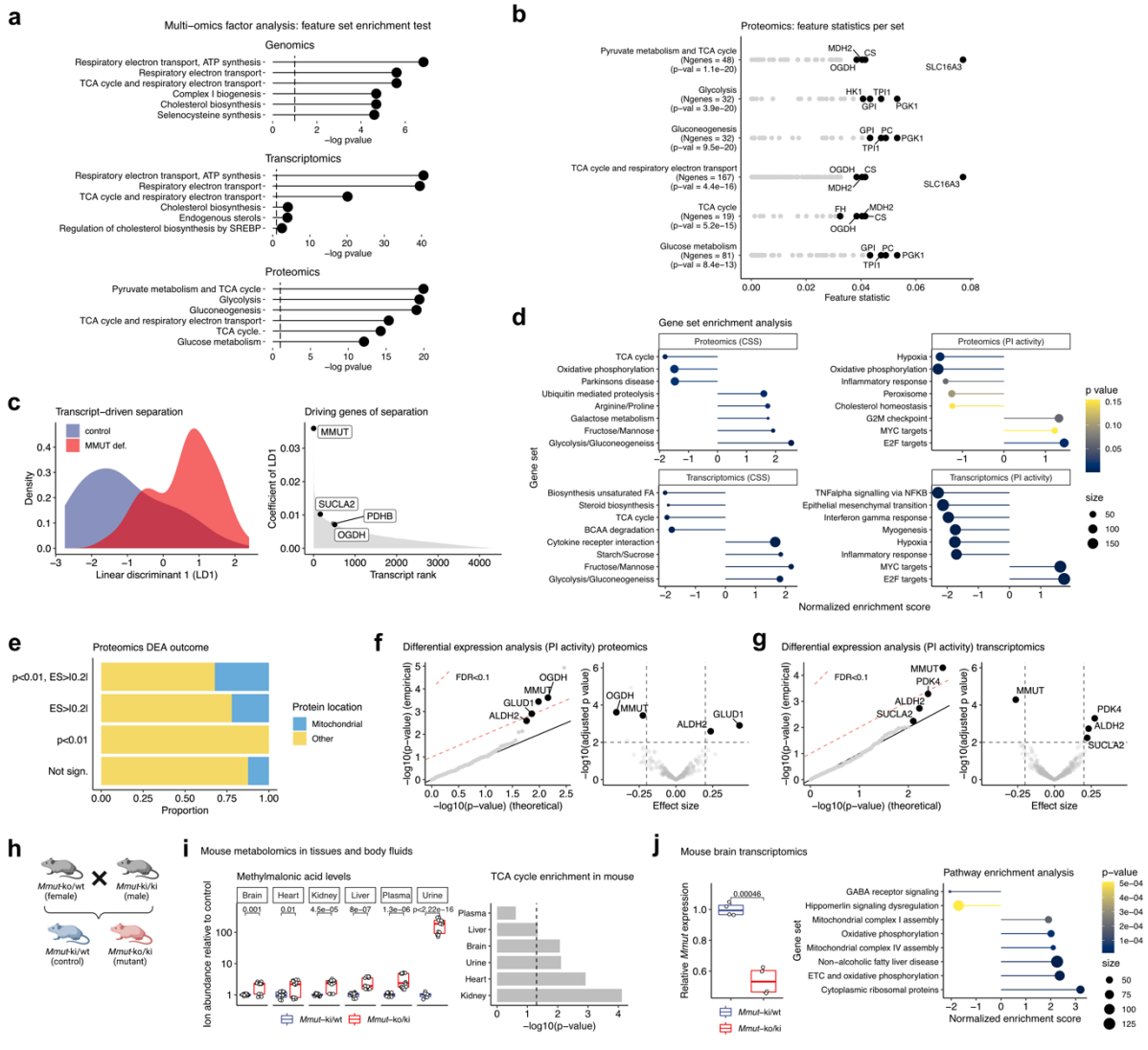


Fig. 3: Untargeted inspection of omics data layers highlights TCA cycle and associated pathways as well as oxidative phosphorylation gene sets to be dysregulated in MMA. a, Gene set enrichment test using the multi-omics factor analysis as previously published (Argelaguet et al., 2018). **b**, Detailed feature statistics of the top enriched gene sets following multi-omics factor analysis in the proteomics data. **c**, Linear discriminant model (split to assign training and test data: 0.5) of transcripts separates MMUT-deficient from control driven by MMUT and other genes related to the

TCA cycle. **d**, Gene set enrichment analysis based on effect size ranking derived from differential expression analysis using sample stratification based on the clinical severity score or PI activity. **e**, Proportions of proteins localized to the mitochondria (according to MitoCarta 3.0) split according to the result of the differential expression analysis. The splits are indicated by dashed horizontal and vertical lines in (f) and (g). **f**, Q-Q and volcano plots illustrating the results of the differential expression analysis based on a linear mixed modelling approach according to (Lamparter et al., 2017) applied to the proteomics data, restricted to enzymes (or their encoding genes) localized in the mitochondria. **g**, Same as (F) but for transcriptomics data. **h**, Breeding scheme of *Mmut* deficient mice (Forny et al., 2016). **i**, Untargeted metabolomics in mouse tissues and body fluids, depicting boxplots for methylmalonic acid and gene set enrichment analysis results. **j**, RNA-seq on mouse brain tissue. Boxplots of the relative *Mmut* transcript abundance and gene set enrichment analysis following DESeq2 analysis, dot size represents number of genes per set.

MMUT deficiency leads to transcript and protein alterations in proximal enzymes

Since both, data-driven and clinically stratified analyses indicated TCA and associated pathways to be disrupted in disease, we performed concerted investigation of TCA cycle anaplerotic and cataplerotic enzymes from which we had both RNA and protein information. As controls we included isoforms of TCA enzymes which are not involved in these pathways (**Fig. 4a**). Direct comparison of RNA and protein expression between MMUT-deficient and control cells revealed only MMUT to be significantly dysregulated (**Fig. 4a outer band**). However, differential expression analysis, performed analogous to **Fig. 3f**, **g**, indicated RNA expression of SUCLA2 to be upregulated with increased disease severity, while protein expression of OGDH and SUCLA2 were reduced and GLUD1 increased in disease; MMUT RNA and proteins levels were downregulated in disease as expected (**Fig. 4a middle band**).

Examination of RNA-protein expression correlation in all samples revealed a limited Spearman correlation of 0.14 at the gene level (4318 transcript-protein pairs) and 0.40 at the sample level (**Fig. 4b** and **Extended data Fig. 7a**), similar to findings by others (Nusinow et al., 2020). Comparison of RNA-protein correlation in MMUT-deficient cells compared to controls revealed that, while 1158 pairs (26.8%) correlated significantly (p-value < 0.05) in both genotypes (**Fig. 4c all colored points** and **Extended data Fig. 7b**) in accordance with previous studies (Williams et al., 2016), the correlation of some genes segregated depending on the genotype (MMUT-deficient vs control) (**Fig. 4c**). In particular, OGDH, GLUD1, CS and GLS showed higher RNA-protein correlation in MMUT-deficient samples than controls, while SUCLA2 had reduced RNA-protein correlation (**Fig. 4a**, **c** and **Extended data Fig. 7c**). OGDH and SUCLA2 were among the genes with the strongest genotype-dependent RNA-protein correlation changes (**Fig. 4c**). Interestingly, we found poor RNA-protein correlation for MMUT in both control and MMUT-deficient cells (**Fig. 4a**, **c** and **Extended data Fig. 7c**).

Finally, MMUT protein levels positively correlated to protein levels of many TCA and anaplerotic enzymes in control but not MMUT-deficient cells**,** while there was little or no protein expression correlation between MMUT and non-TCA protein isoforms in either genotype (**Fig. 4a center** and **Extended data Fig. 7d**). Such a relationship is exemplified by MMUT:ACO2 and MMUT:ACO1 (**Fig. 4a center** and **Extended data Fig. 7e**) and provides the first indication that MMUT may be part of an interaction network with these mitochondrial TCA cycle and anaplerotic enzymes (Heusel et al., 2019). Examination of pair-wise correlation between all proteins (**Fig. 4d**) and transcripts (**Extended data Fig. 7f**) in these pathways suggests that TCA cycle and anaplerotic enzymes have a positive correlation with each other, which is not altered in MMUT deficiency, unless MMUT is included in the comparison. Overall, the above findings suggest that disruption of MMUT RNA and protein expression drives regulatory changes in certain TCA and anaplerotic enzymes.

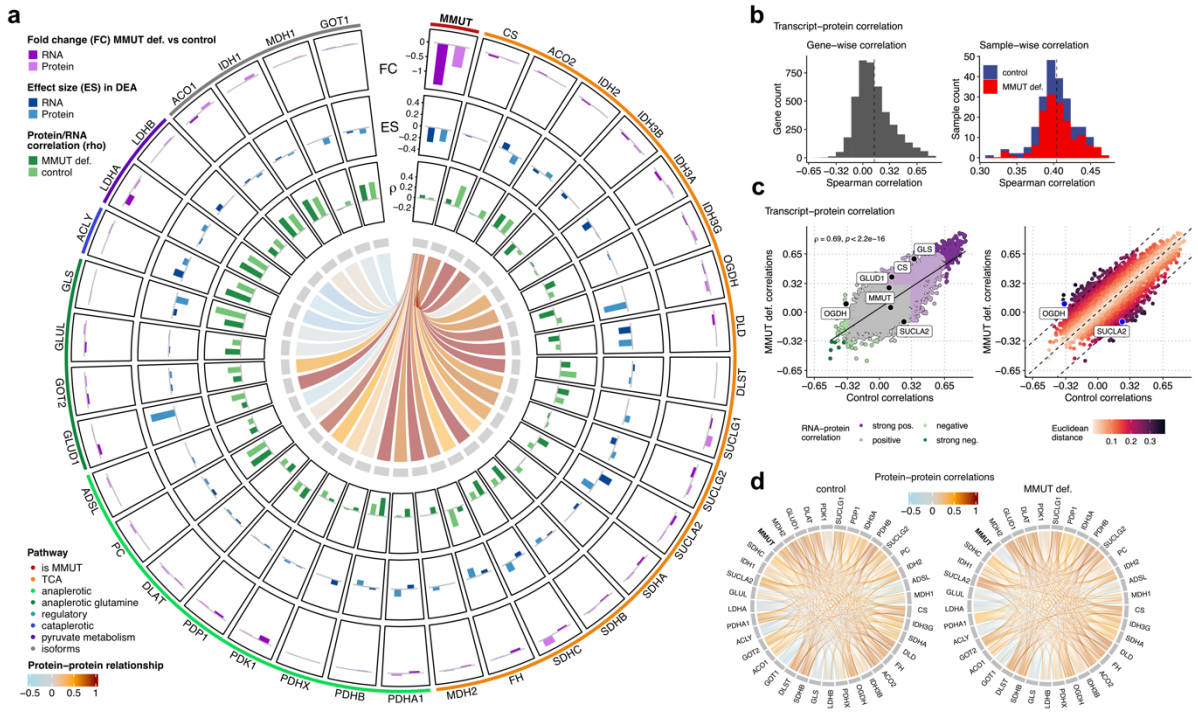


Fig. 4: Transcript-protein and protein-protein correlation analysis reveals coordinated relationships between MMUT and TCA genes and proteins but not their isoforms. a, Circos plot depicting raw fold changes of transcripts and proteins, effects sizes derived from differential expression analysis, transcript-protein correlations, and correlative relationships of the MMUT protein to TCA proteins and their corresponding isoforms. **b**, Histograms of Spearman correlations across 4318 transcript-protein pairs (left), and 221 samples (right). **c**, Scatter plot of Spearman correlations in MMUT-deficient against control. Euclidean distance from the diagonal is calculated based on the formula |(MMUT def. correlation – control correlation)|/sqrt(2). **d**, Chord plot depicting all protein-protein correlations of TCA and related proteins; color code as in (a) indicates the correlation coefficient.

Metabolomics suggest rewiring of TCA cycle anaplerosis

To examine the functional consequences of the above RNA and protein expression changes, we performed untargeted polar metabolomic analysis on a selection of 6 MMUT-deficient and 6 control fibroblast cell lines (for selection criteria, see **Extended data Fig. 8a** and **Methods**). Among the most

307 strongly changed metabolites we found decreased glutamine and alanine as well as increased hexoses,
308 2-methylcitrate, oxo-adipate and amino-adipate. Pyruvate was also significantly accumulated in
309 MMUT-deficient samples (**Fig. 5a**). Combining these alterations with observed changes in RNA and
310 protein expression in the same samples provides insight into how TCA cycle anaplerosis is disrupted
311 in MMUT-deficiency (**Fig. 5b**).

312 The increased pyruvate level (**Fig. 5a, c**), increased expression of the PDH complex inhibitor *PDK4*
313 (**Fig. 3g**), and strongly elevated 2-methylcitrate (**Fig. 5c**), a biomarker of MMA formed by citrate
314 synthase mediated mis-condensation of oxaloacetate with propionyl-CoA instead of acetyl-CoA,
315 together suggest reduced incorporation of pyruvate into the TCA cycle. This is further supported by
316 the analysis of our cross-dimensional data which identified stronger pyruvate to *PDHB* transcript and
317 pyruvate to PDHX protein correlations (**Fig. 5d**), implying reduced pyruvate turnover per transcript
318 and protein respectively. Consecutive metabolites in a pathway often correlate strongly (Williams et
319 al., 2016), here exemplified by malate and fumarate (**Fig. 5d**). Although we found such a correlation
320 between pyruvate and its downstream metabolite (iso)citrate in both control and MMUT-deficient
321 cells, the right shift of MMUT-deficient cells suggests that a higher concentration of pyruvate is
322 required to create the same amount of (iso)citrate as controls (**Fig. 5d**), again supporting reduced
323 incorporation of pyruvate into the TCA cycle.

324 Despite this reduced pyruvate mediated anaplerosis, 2-oxoglutarate, succinate, fumarate and malate
325 levels remained unchanged in MMUT-deficient cells (**Fig. 5c** and **Extended data Fig. 8b**),
326 suggesting overall TCA cycling is stably maintained. We reasoned this may be due to an increase
327 from alternate anaplerotic sources, particularly glutaminolysis. Glutamine strongly correlated with
328 glutamate in both MMUT-deficient and control samples (**Fig. 5d**). However, in MMUT-deficient
329 cells, glutamine and glutamate were decreased (**Fig. 5c**), and there was a negative correlation of
330 glutamate with 2-oxoglutarate, suggestive of more tightly regulated glutamine flux under disease
331 conditions (**Fig. 5d**). Abrogated positive correlation of glutamate and aspartate potentially points to an
332 increased dependance on glutamate for TCA cycling but not production of cataplerotic metabolites
333 (**Fig. 5d**). In line with this, we found an inverse relationship between *MMUT* transcript and 2-

oxoglutarate in MMUT-deficient cells (**Fig. 5d**). As a further consequence, we found increased oxoadipate and aminoadipate (**Fig. 5a**), upstream metabolites of 2-oxoadipate dehydrogenase which shares its E2 (DLST) and E3 (DLD) components with 2-oxoglutarate dehydrogenase (**Fig. 5e**), suggestive of a preference for 2-oxoglutarate over 2-oxoadipate metabolism.

These results point to an increased reliance on the glutamine anaplerotic pathway in disease; a hypothesis we tested by assessing relative anaplerotic glutamine flux into the TCA cycle using MS-based stable isotope tracing. For this experiment, control, MMUT-KO and DLST-KO 293T cells, validated by Western blotting and enzyme activity measurements (**Extended data Fig. 9**) were cultured in glutamine-free glucose-DMEM supplemented with 4 mM [$^{13}C_5$]glutamine for 4 hours. It is expected that (M+5) [$^{13}C_5$]glutamine entering the TCA cycle through (M+5) glutamate and 2-oxoglutarate undergoing oxidative metabolism would generate (M+4) succinate, fumarate, malate, aspartate, citrate and aconitate; while reductive carboxylation would generate in (M+5) aconitate and citrate, as well as (M+3) aspartate, malate, fumarate and succinate (**Fig. 5f**) (Chen et al., 2018). Therefore, by determining the abundance of ^{13}C enrichment and isotopologue distribution we can quantify the glutamine-glutamate-oxoglutarate flux into the TCA cycle and the relative contribution to oxidative metabolism versus reductive carboxylation. Labeled glutamine was equally incorporated into glutamate (**Extended data Fig. 8c**) and subsequently 2-oxoglutarate (**Fig. 5g**) in all three cell lines, suggesting a glutamine-derived anaplerotic flux independent of the genetic defect. Consistent with loss of oxoglutarate dehydrogenase activity, glutamine-derived flux was almost exclusively directed towards the reductive TCA pathway in the DLST-KO cells, indicated most prominently by increased fractional incorporation of M+3 fumarate and decreased M+4 fumarate, aspartate, and citrate (**Fig. 5g, h**); culminating in reduced M+3/M+4 fumarate and aspartate and M+5/M+4 citrate. MMUT deficiency, by contrast, led to slight increase of oxidative TCA cycle flux compared to control, signified by increased fractional incorporation of M+4 citrate (**Fig. 5h**) resulting in a clearly increased M+4/M+5 citrate ratio (**Fig. 5i**). In sum, we found preferred oxidative utilization of anaplerotic glutamine potentially compensating for the lack of succinyl-CoA in MMUT deficiency.

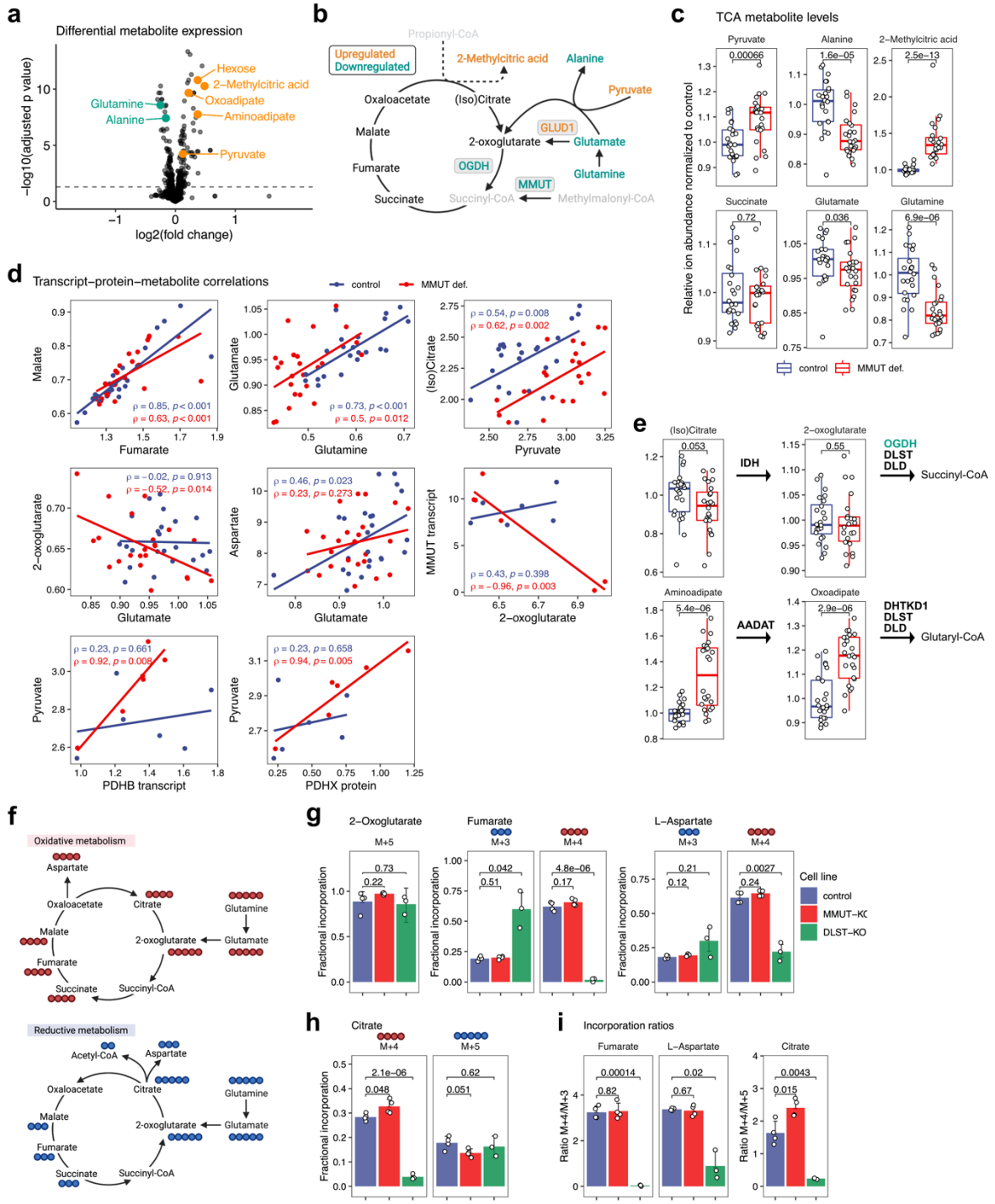


Fig. 5: Polar metabolomics in selected patient fibroblasts and glutamine flux studies in CRISPR-KO HEK293 cells highlight increased glutamine anaplerosis. a, Volcano plot depicting differentially expressed metabolites. Highlighted are the ones particularly relevant to this study. **b**, Schematic depiction of the TCA cycle and relevant anaplerotic reactions. The color code indicates dysregulations on metabolite and protein level; grey metabolites were not detected. **c**, Boxplots to

compare relevant TCA cycle metabolites, including the pathologically relevant biomarker 2-methylcitrate formed upon condensation of oxaloacetate with propionyl-CoA; p-values calculated by Wilcoxon rank test. **d**, Pearson correlation of a certain key metabolite against a metabolite, transcript, or protein variable. **e**, Levels of metabolites involved in the two enzymatic steps catalyzed by two oxoacid dehydrogenase complexes (OGDC and OADC) and their proximal reactions. **f**, Schematic representation of labeling of TCA cycle and associated metabolites derived from glutamine anaplerosis. **g**, Enrichment of specific isotopologues of interest. **h**, All detected isotopologues of citrate. **i**, Ratios of isotopologue as indicated by the y-axis label of specific metabolites of interest; p-values are calculated by pairwise t-test.

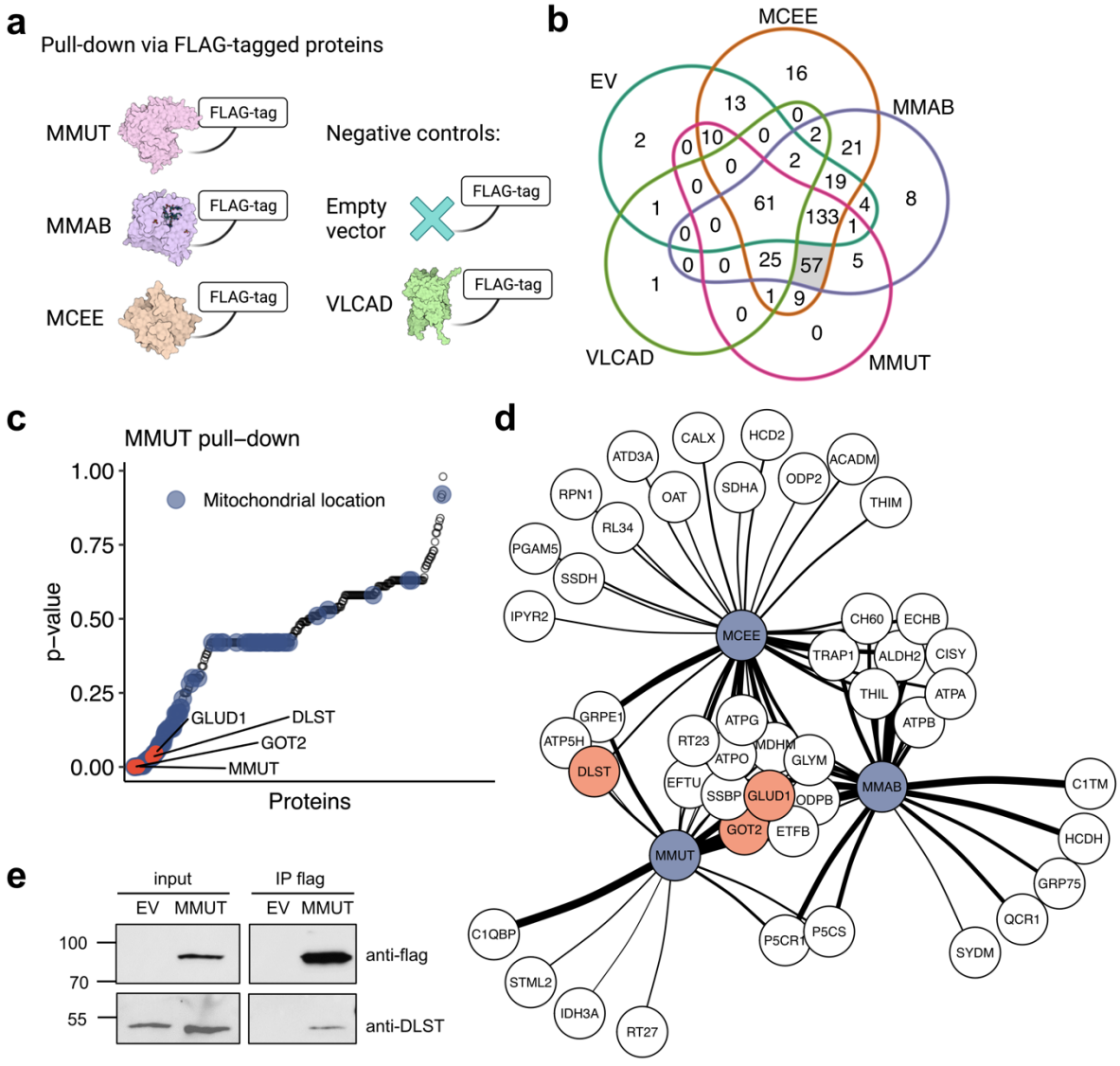
MMUT physically interacts with 2-oxoglutarate dehydrogenase and other anaplerotic enzymes

The strong protein expression correlation between MMUT and proximal TCA cycle and anaplerotic enzymes (**Fig. 4** and **Extended data Fig. 7)** suggested they may be part of a shared protein complex (Heusel et al., 2019). To gain insights into this potential physical relationship, we took advantage of over-expression of C-terminally flag-tagged versions of MMUT, two pathway members (MMAB and MCEE) expected to participate in any multi-protein complex containing MMUT, and two negative controls (empty vector and VLCAD) in 293T cells (**Fig. 6a**). Using cross-linking affinity purification coupled to mass spectrometry, at 1.0% FDR using 2 peptides minimum at 95% threshold, each of these 'bait' proteins pulled-down a total of 100-350 different 'prey' proteins over 3 biological replicates (**Fig. 6b**). Within this intersection, we identified 57 prey proteins pulled-down by MMUT, MMAB and MCEE in at least one replicate, but not by empty vector or VLCAD in any replicate (**Fig. 6b**).

Analysis of variance (ANOVA) of the biological triplicates comparing MMUT with EV and VLCAD identified 22 proteins to be significantly enriched (nominal p-value < 0.05) in the MMUT sample (**Fig. 6c** and **Extended data Table 2**). All proteins were designated by Uniprot to have mitochondrial

393 localization, and included GLUD1, GOT2 and DLST (**Fig. 6c**). ANOVA comparing the intersection
394 of proteins confidently pulled-down by MMUT, MMAB and MCEE but not EV or VLCACD
395 identified 11 uniquely interacting proteins, including GLUD1 and GOT2; while the intersection of
396 two of MMUT, MMAB and MCEE against both negative controls identified 13 uniquely interacting
397 proteins, including DLST (**Fig. 6d** and **Extended data Table 2**). Interaction of MMUT and DLST
398 was further confirmed by Western blot analysis (**Fig. 6e**). These data indicate that MMUT physically
399 interacts with GLUD1 and the oxoglutarate dehydrogenase component DLST and suggests that
400 disruption of this interaction may underlie their altered regulation in disease.

401



402

Figure 6. MMUT interacts physically with GLUD1, DLST, and GOT2 as demonstrated by FLAG-tag pull-down. a, Outline of experimental and control groups indicating which protein was used with a FLAG-tag in a cross-linking affinity purification experiment coupled to subsequent analysis of the pull-down samples by mass spectrometry. **b**, Venn diagram of number of proteins pulled down by the different flag-tagged proteins. **c**, ANOVA p-values of all proteins pulled down by MMUT. Blue dots indicate proteins with mitochondrial location according to Uniprot. **d**, Interaction network of significantly enriched proteins (ANOVA p-value < 0.05). Thicker connector indicates lower p-value. **e**, Western blot of immunoprecipitation of flag-tagged MMUT probing for DLST.

Discussion

In this study, we used an integrated multi-omics approach to diagnose and uncover new pathomechanisms of the inborn error of metabolism MMA. Unique to this investigation was the relatively large (for a rare disease) set of patient samples along with corresponding phenotypes available, and the ability to coordinate aliquots from the same samples to generate molecular data at three data layers. Nevertheless, this approach can be generalizable in any setting of inborn monogenic diseases where a large enough biobank exists.

Our findings reinforce the value of using complementary datasets to increase diagnostic yield compared to WGS alone. Key to our total diagnostic yield of 84% was additional transcriptomic and biochemical analysis beyond WGS alone. Likewise, integrating multi-omics data layers with phenotypic information enabled sample stratification by disease severity and identification of multi-level alterations which were not apparent following examination of single omics layers. Such a move away from “data silos” into true integrative multi-layered analysis remains challenging, as it requires new analytical and statistical methods to combine these disparate data sets (Karczewski and Snyder, 2018). In this capacity, multi-omics factor analysis (Argelaguet et al., 2018) provided indication of disruption to RNA and proteins of TCA cycle and related pathways, a finding verified by the integration of phenotypic data utilizing both PI+ and a clinical severity score. Further overlaying metabolomic data provided indication of specific disturbance of glutamate dehydrogenase and oxoglutarate dehydrogenase, two proteins we found to physically interact with MMUT. Such orthogonal approaches, incorporating not only multi-omics layers, but also clinical information and complementary (experimental) information types, are invaluable for target validation.

Our findings highlight the importance of the loss of methylmalonyl-CoA as an anaplerotic source and point to oxoglutarate produced from glutamate as an alternative entry point into the TCA cycle. The rewiring of anaplerosis, potentially as a compensation mechanism, presents a novel therapeutic target for MMA. Such anaplerotic stimulating approaches have precedent in IEMs, including application of triheptanoin in long-chain fatty acid oxidation disorders (Vockley et al., 2021). Unfortunately, since fatty acid β-oxidation produces two acetyl-CoA and one propionyl-CoA per heptanoate molecule, use

of this particular compound in MMUT deficiency is not feasible. However, anaplerotic treatment of patients suffering from the MMA-related disorder propionic aciduria with the TCA cycle intermediate citrate has been attempted, with inconsistent results (Longo et al., 2017). Based on our findings, compounds such as dimethyl oxoglutarate, a membrane permeable alternative to 2-oxoglutarate, previously used in a model of OXPHOS dysfunction (Chen et al., 2018), may represent a more promising therapeutic strategy. Studies to delineate the efficacy of such approaches in preclinical and clinical models will be important for the ongoing development of new treatments.

Acknowledgements

This project was funded by the ETH domain strategic focus area "Personalized Health and Related Technology" (PHRT; https://www.sfa-phrt.ch). This work received financial support from the Swiss National Science Foundation [31003A_175779] to M.R.B. and [310030_192505] to D.S.F. and from the University Research Priority Program of the University of Zurich (URPP) ITINERARE—Innovative Therapies in Rare Diseases. P.F. was supported by the Filling the Gap grant awarded by the Medical Faculty, University of Zurich, Switzerland. We acknowledge the Functional Genomics Center Zurich of the University of Zurich for RNA sequencing of the mouse brain tissues and bioinformatics support as well as for support with the glutamine isotope studies.

Author Contributions

Conceptualization, I.X., D.V.M., R.A., G.R., E.D., B.W., M.R.B., D.S.F.; Methodology, P.F., X.B., D.L., W.S., T.P., M.F., L.S., R.J.M., S.C., S.G., A.v.D., K.H., P.P., N.Z.; Investigation, P.F., X.B., D.L., W.S., T.P., C.F., A.B., S.C., P.P., F.T., N.Z., G.R., D.S.F.; Writing – Original Draft, P.F., D.S.F.; Writing – Review & Editing, all co-authors; Visualization, P.F., X.B., D.L., W.S., C.F.; Supervision, P.F., K.H., P.P., F.T., I.X., G.R., E.D., B.W., D.S.F.; Project Administration, P.F., D.V.M., B.W.; Funding Acquisition, D.V.M., R.A., B.W., M.R.B., D.S.F.

Declaration of interests

The authors declare no competing interests.

Extended data

Supplementary figures and table in separate file.

Methods

Cohort and patient-derived fibroblast samples

Primary fibroblast samples and corresponding disease related information, including clinical and diagnostic data, was collected from 1989 to 2015. The information was obtained and used under the ethics approval granted by the Ethics Committee of the Canton of Zurich, Switzerland (no. KEK-2014-0211, amendment: PB_2020-00053). Upon collection, primary fibroblasts were cultured using Dulbecco's modified Eagle medium (Gibco, Life Technologies, Zug, Switzerland) with 10% fetal bovine serum (Gibco) and antibiotics (GE Healthcare, Little Chalfont, UK) and either used immediately, or exchanged to 90% fetal bovine serum and 10% dimyethlysulfoxide and stored in cryovials under liquid nitrogen. A frozen aliquot of each fibroblast line was sent for WGS, RNA-seq and DIA-MS analysis (**Fig. 1a**). RNA-seq and DIA-MS were always performed from matched aliquots.

Clinical disease severity score

The clinical disease severity score was based on five typical clinical signs/symptoms of MMA (Forny et al., 2021), including age at disease onset, as well as presence of neurological abnormalities, kidney impairment, hematological abnormalities, and failure to thrive. Each patient was assigned a score from 0-5 indicating increasing disease severity.

Biochemical assays

Propionate incorporation into acid precipitable material of primary fibroblasts was assessed according to a protocol described previously (Willard et al., 1976) with modification as described (Froese and

Baumgartner, 2020). Methylmalonyl-CoA mutase enzyme activity assay was performed in crude cell lysates as originally described (Baumgartner, 1983; Causey and Bartlett, 1984) using recent modifications (Forny et al., 2014).

Whole genome sequencing

Genomic DNA was isolated using the QIAmp DNA Mini Kit reagents (Qiagen; Hilden, Germany) following the protocol provided by the supplier. DNA was quantified by Qubit (Thermo Fisher; Waltham, USA). Whole genome sequencing libraries were prepared with the TruSeq DNA PCR-free library reagents (Illumina; San Diego, USA) using 1ug of genomic DNA following the protocol provided by the supplier. The genomic DNA libraries were quantified using the KAPA Library Quantification Complete Kit (Roche; Basel, Switzerland) according to the protocol supplied with the reagents. The quantified libraries were sequenced on the NovaSeq 6000 sequencer (Illumina) using a 150-nucleotide paired end run configuration following the protocol provided by the supplier.

RNA sequencing

Total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). The total RNA was quantified using the Qubit (Thermo Fisher) and quality-controlled using the Fragment Analyzer (Agilent, Santa Clara, USA). RNA-seq libraries were prepared using the TruSeq Stranded mRNA-seq reagents (Illumina) using 200ng of total RNA following the protocol provided by the supplier. The quality of the RNA-seq libraries was assessed on Fragment Analyzer (Agilent) and the libraries were quantified using the Qubit (Thermo Fisher). The libraries were sequenced on Illumina HiSeq 4000 using the 75-nucleotide paired end run configuration following the protocol provided by the supplier.

Sample preparation for mass spectrometry proteotyping measurements

Samples were processed in blocks of 8 taking into consideration a balance between disease types and control samples. All other factors within a block were randomized. 230 samples were processed in three batches. For sample processing, aliquots of primary fibroblast (~1e6 cells per vial, frozen in

either DMEM or FBS or a mix of DMEM/FBS, plus 10% DMSO) were washed twice in ice-cold PBS (Gibco). After centrifugation for 5 minutes at 300g, the cells were resuspended in lysis buffer (Preomics) at a ratio of 1:1 (vol pellet/vol lysis buffer) and incubated at 95°C for 10 mins. Samples were sonicated in a vial tweeter (Hielscher Ultrasound Technology) at 4°C for 3 cycles with an amplitude 100%, power 80% during 30 secs. For each sample, the total protein concentration was estimated by nanodrop. 100 ug of protein lysate were further processed with the iST kit (Preomics). The purified peptides were resuspended in LCLoad buffer containing iRT peptides (Biognosys) at a concentration of 1 ug/ul.

Spectral library generation

For spectral library generation, three times 24 samples (3x8 sample blocks) were pooled. Pooled sample batches were digested as described above. 100 µg of purified peptides were fractionated on a C18 column (YMC-Triart, C18, 3µm, 250 x 0.5 mm ID) according to pH on an Agilent HPLC 1260 system with a stepped 61 min gradient ranging from 95 % buffer A (20 mM ammonium formate acid/H2O) to 85% buffer B (20 mM ammonium formate/90 % ACN). 48 fractions were collected per sample and subsequently pooled to 24 fractions. Samples were resuspended in 5% ACN/0.1% FA and analyzed on a QExactive HF-X mass spectrometer (Thermo Fisher Scientific) in DDA mode. The same nLC 1200 configuration and mobile phase gradient elution conditions as for DIA were applied.

Full MS survey scans were acquired at a resolution of 60,000 with automatic gain control (AGC) target of 3e6 and a maximum injection time of 45 ms over a scan range of m/z 375-1500. A data-dependent top 12 method was used for HCD MS/MS with a normalized collision energy of 28 at a resolution of 15,000 and a fixed first mass of m/z 100. Precursor ions were isolated in a 1.4-Th window and accumulated to reach an AGC target value of 1e5 with a maximum injection time of 22 ms. Precursor ions with a charge state of 1 and 6 as well as isotopes were excluded for fragmentation. Dynamic exclusion was set to 15s.

DDA raw files were processed with Proteome Discoverer (v. 2.2) using a human UniProt database (release 201804) together with iRT peptides (Biognosys) and common contaminants. The processing

workflow consisted of SequestHT (Eng et al., 1994) and Amanda (Dorfer et al., 2014) nodes coupled with Percolator (Brosch et al., 2009). The following search parameters were used for protein identification: (i) a peptide mass tolerance of 10 ppm; (ii) an MS/MS mass tolerance of 0.02 Da; (iii) fully tryptic peptide search with up to two missed cleavages were allowed; (iv) carbamidomethylation of cysteine was set as fixed modification, methionine oxidation and protein N-term acetylation were set as variable modifications. Percolator was set at max delta Cn 0.05, with target FDR strict 0.01 and target FDR relaxed 0.05. The spectral library from Proteome Discoverer was imported into Spectronaut v12 (Biognosys, Schlieren, Switzerland) using standard parameters with 0.01 peptide spectrum match (PSM) FDR.

DIA-MS setup and data analysis

For DIA analysis samples were measured on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of HPLC-grade water with 0.1% (v/v) formic acid, and mobile phase B consisted of HPLC-grade ACN with 20% (v/v) HPLC-grade water and 0.1% (v/v) formic acid. Peptide separation was carried out on an ES806, 2 µm, 100 Å, 150 µm i.d. x 150 mm, C18 EASY-Spray column (Thermo Fisher Scientific) at a temperature of 50°C. For LC-MS/MS analyses, 2 μg of each sample were loaded onto the column via an Easy-nLC 1200 system (Thermo Fisher Scientific). Samples were loaded at 4 µL/min with 100% mobile phase A for 5 min. Peptide elution was performed using the following gradient: i) 2% to 8% mobile phase B in 4 min, ii) 8% to 32% mobile phase B in 49 min, iii) 32% to 60 % mobile phase B in 1 min, and iv) ramp to 98% mobile phase B in 1 min at 2 µL/min.

For DIA-Acquisition on a Q Exactive HF mass spectrometer, we applied a DIA method published elsewhere (Xuan et al., 2020). In short, we performed an MS1 scan over a mass range of m/z 400-1210 at a resolution of 120,000 with an AGC target value of 3e6 and with a maximum injection time of 50 ms. For MS/MS scans, resolution was at 0,000 with an AGC target value of 1e6 and with "Auto" maximum injection time. Precursor ions were isolated within a 15-Th window and fragmented

by HCD with normalized collision energy 28. 54 MS/MS scan windows were defined, interspersed every 18 scans with an MS1 scan.

DIA data Analysis was performed in Spectronaut v12 (Biognosys) using standard parameters. For identification, a Qvalue cut-off of 0.01 was applied on the precursor as well as on the protein level. The MS1 area was selected for quantification. Quantification parameters were set to mean peptide quantity for major group quantity, the top 3 peptides were selected for protein quantity calculation. Data filtering was set to Qvalue sparse, with no imputation. Cross-run normalization was set to local

Fibroblast cell line selection for polar metabolomics

We devised a model of OGDH and GLUD1 expression based on the differential expression method described earlier (Lamparter et al., 2017) and selected the top ten ranked MMUT deficient (MMA014, MMA92, MMA042, MMA67, MMA93, MMA104, MMA013, MMA030, MMA138, MMA036) and the last 10 ranked control cell lines (MMA219, MMA221, MMA227, MMA222, MMA213, MMA230, MMA226, MMA228, MMA225, MMA215). They were cultured as described above. Four cell lines in each group did not meet growth criteria, most likely due to the long freezing period, and were excluded. Six cell lines per group were selected for the polar metabolomics experiment.

Fibroblast sample preparation for polar metabolomics

100,000 cells per well were seeded in a 6 well plate and grown for 48 hours. Media was removed and cells washed with 2 mL per well of 150 mM ammonium hydrogen carbonate (NH_4HCO_3) at pH 7.4 twice. The whole plate was flash frozen in liquid nitrogen for 20 seconds and then stored at -80°C. Metabolites were extracted by putting the plate on dry ice and adding 400 ul of cold (-20°C) 40:40:20 acetonitrile:methanol:water and incubate at -20°C for 10 minutes. Supernatant was collected. Added another 400 ul of 40:40:20 acetonitrile:methanol:water and incubated at -20°C for 10 minutes. Plates were put on dry ice and cells were scraped mechanically. The supernatant including the scraped cells were collected in the same tube as during the first extraction step. Collection tubes were centrifuged at 13,000 rpm for 2 minutes at 4°C. Supernatants were stored at -20°C prior to metabolomics analysis.

Polar metabolomics in patient-derived fibroblasts

Metabolomics experiments were performed as previously published (Fuhrer et al., 2011).

Mouse care and handling

Ethics approval was obtained prior to the study from the Cantonal Veterinary Office Zurich under the license number 202/2014. Mice were kept under standard conditions, i.e. single-ventilated cages with controlled humidity and room temperature of 22°C. Generation of the *Mmut*-p.Met698Lys variant model and crossing with a *Mmut*-ko/wt model was done as previously described (Forny et al., 2016).

Harvesting of mouse tissues

Urine was collected in the morning after one night in a metabolic cage. The sediment was removed, and supernatant was flash frozen in liquid nitrogen. Tissue samples were harvested from mice aged 58 to 63 days. Animals were anesthetized by sevoflurane. Portal blood was taken and kept on ice to coagulate, centrifuged at 4°C and snap frozen in liquid nitrogen directly after. Liver, kidneys, heart, and brain were harvested and snap-frozen in liquid nitrogen. After the procedure, the mice were directly euthanized by cervical dislocation. All samples were stored at -80°C prior to analysis.

Metabolomics in mouse tissues

The mouse body fluid and tissue samples were prepared as previously published (Want et al., 2013). Sample analysis using liquid chromatography-mass spectrometry was performed as previously published (Abela et al., 2017). Ions were annotated to metabolites based on exact mass to the KEGG database (Kanehisa et al., 2017) considering [M-H+] and 0.01 Da mass accuracy. Significantly changing ions between mutant and control conditions were identified using a two-sample t-test. Pathway analysis was performed using annotated ion list ranked by p-value significance. Pathway enrichment were calculated using KEGG metabolic pathway definitions and a hypergeometric test.

Transcriptomics in mouse brains

Brain tissue samples were harvested as described above. Four females were used per genotype group, which were *Mmut*-ki/wt and *Mmut*-ko/ki. RNA was purified using DNAse kit (Qiagen, No. 79254) together with QIAmp RNA Blood Mini Kit (Qiagen, No. 52304). RNA-seq reads were aligned with the STAR-aligner (Dobin et al., 2013). As reference we used the Ensembl mouse genome build GRCm38. Gene expression values were computed with the function featureCounts from the R package Rsubread (Liao et al., 2013).

CRISPR gene-editing experiments

CRISPR-Cas9 editing was performed as described (Ran et al., 2013). Cas9 protein was provided as a plasmid (PX459-V2.0, Addgene, 62988) and guide RNA as gBLOCKS (Arbab et al., 2015) (IDT Technologies). HEK293T cells were transfected using the Neon transfection system (Thermo Fisher Scientific) containing 100,000 cells, 0.6 µg of Cas9 plasmid and 600 ng of guide RNA following manufacturer's instructions. 48 h after transfection, cells were collected, diluted to 1 cell/100 µl and transferred to a 96-well plate at 100 µl/well for clonal selection. Correct clones were confirmed by Sanger sequencing of genomic DNA.

Western blotting

Lysates of HEK293 cell lines were mixed with RIPA lysis buffer and 4X Laemmli buffer to obtain a concentration of 1 µg/µl protein. Samples were incubated at 96°C for 5 minutes. 20 µl of each sample was loaded onto a 10% SDS page gel. Proteins were separated by electrophoresis, transferred with a semi-dry method onto a nitrocellulose membrane (Whatman, GE Healthcare), blocked at room temperature for 2 hours with buffer A (5% skimmed milk, 1.2% w/v Tris-base, 9% w/v NaCl, 0.2% Tween 20, pH 7.6), incubated with primary antibodies dissolved in buffer A overnight at 4°C, and detected using secondary antibodies in buffer A. Primary antibodies used were probing for the following proteins: MMUT (Abcam, ab67869, 1:1000, host: mouse), OGDH (Atlas antibodies, HPA020347, 1:500, host: rabbit), GLUD (Abcam, ab166618, 1:2000, host: rabbit), Beta-actin (Sigma,

A1978, 1:5000, host: mouse). Secondary antibodies used were Anti-rabbit HRP (Santa Cruz, sc-2357, 1:5000, host: mouse), Anti-mouse HRP (Santa Cruz, sc-516102, 1:5000, host: goat).

OGDH and GLUD1 enzyme activity assays

Assays in all the HEK293 cell clones for assessment of enzymatic activity of glutamate dehydrogenase (Sigma-Aldrich, St. Louis, USA, catalogue number MAK099) and oxoglutarate dehydrogenase (Sigma-Aldrich, St. Louis, USA, catalogue number MAK189) using the plate reader Victor Nivo by PerkinElmer.

Glutamine tracing studies

HEK 293T cells were cultured on coverslips in DMEM (Gibco, catalog number: 11960044) without L-glutamine, sodium pyruvate and HEPES supplemented with 10% FBS, 1% antibiotic-antimycotic (Gibco) and [$^{13}C_5$]glutamine (final concentration 4 mM) (Sigma-Aldrich, catalog number: 605166) for 4 hours. Medium was then removed, coverslips were quickly dipped into sterile double-distilled water at 37°C and quenched in 80% methanol at -80°C. Cells were scrapped in methanol, collected, and centrifuged at 15'000 g for 15 minutes at 4°C. Supernatants were collected, snap frozen in liquid nitrogen, and stored at -80°C prior to LC/MS analysis. Samples were normalized based on protein content. Volume extracts equivalent to 4 µg protein were dried under a nitrogen stream, reconstituted in 20 µl water (MS grade) and incubated on shaker (800 rpm, 15°C, 10 min). Then, 80 µl injection buffer (90% acetonitrile, 8.8% methanol, 50 mM ammonium acetate) were added, vortexed and centrifuged (10,000 × g, 4°C, 10 min). A total of 50 µl of the supernatant was transferred to a glass vial with narrowed bottom (Total Recovery Vials, Waters, Milford, MA, USA) for LC-MS/MS injection. In addition, quality control (QC) standards, and pooled samples were prepared in the same way to serve as QC for the measurements. Metabolites were separated on a nanoAcquity UPLC (Waters, Milford, MA, USA) equipped with a BEH Amide capillary column (150 µm × 100 mm, 1.7 µm particle size, Waters, Milford, MA, USA). The injection volume was 1 µl. Bisolvent system were 5 mM ammonium acetate in water (A) and 5 mM ammonium acetate in 95% acetonitrile (B). The

flow rate was adjusted over the gradient from 3 to 2 µl/min. Gradient started with 5% A and increased to 50% A over 13 min. The following 5 min conditions were kept at 50% B, followed by 7 min reequilibration to 5% A. The UPLC was coupled to QExactive mass spectrometer (Thermo) by a Picoview nanoESI source (NewObjectives Inc.). MS data was acquired using negative polarization and all ion fragmentation (AIF). Full scan MS spectra were acquired in profile mode from 70-1050 m/z with an automatic gain control target of 1e6, an Orbitrap resolution of 70`000, and a maximum injection time of 200 ms. AIF spectra were acquired with the same setting and fragmented with a normalized collision energy of 35. Generated fragment ions were scanned in the linear trap. Data sets were evaluated with Skyline 21.2. Compounds were identified based mass error and retention time properties observed on references run along with the samples.

Affinity capture mass spectrometry

293T cells (ATCC #CRL-3216; Manassas, WV) were grown in Dulbecco's Modified Eagle Medium (Gibco, Carlsbad, CA) supplemented with 10% fetal bovine serum (Gibco) and antibiotics (GE Healthcare). Transient transfection of pCDNA3-C-Flag-LIC constructs was performed using Lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. 48 h after transfection, cells were crosslinked using 0.5% paraformaldehyde (PFA, Sigma-Aldrich) in PBS (Gibco) for 10 min at RT, the reaction was quenched with 1.25 M glycine/PBS (Sigma-Aldrich) for 10 min at 4°C, cells were centrifuged for 5 min at 2,000×*g* at 4°C, and the pellet resuspended in lysis buffer (1% Nonidet P-40, 0.5% deoxycholine, 150 mM NaCl, 50 mM Tris–HCl, pH7.5, all Sigma-Adrich). Pre-cleared cell extracts were immunoprecipitated with anti-flag M2 (F3165, Sigma–Aldrich) using Dynabeads Protein G (Thermo Fisher Scientific) according to manufacturer's instructions. Following washing, peptides were released by trypsin (100 ng/µl in 10 mM HCl) and supernatants collected, dried, dissolved in 0.1% formic acid and analysed using LC-MS/MS (Orbitrap, Thermo Fisher Scientific). Database searches were performed using Mascot (Matrix Science).

Data analysis

Data analysis was performed using R version 4.1.0. For the global data layer inspection we used the MOFA package version 1.3.1 (Argelaguet et al., 2018), the MASS package version 7.3-54 (Venables and Ripley, 2002), the fgsea package version 1.18.0 (Korotkevich et al., 2021). Gene enrichment analysis was perfomed using gene sets downloaded from http://www.gsea-msigdb.org/gsea/msigdb/index.jsp "MSigDB Collections" on 28 December 2020. Circos including chord plots were created using the circlize package version 0.4.13 (Gu et al., 2014). The Uniprot portal was accessed on 24 February 2021 at 4 PM to scrape protein localization data. The code for all the analyses and generation of figures is hosted on a repository on the GitHub platform under https://github.com/pforny/mma_pf.

References

Abela, L., Spiegel, R., Crowther, L.M., Klein, A., Steindl, K., Papuc, S.M., Joset, P., Zehavi, Y., Rauch, A., Plecko, B., et al. (2017). Plasma metabolomics reveals a diagnostic metabolic fingerprint for mitochondrial aconitase (ACO2) deficiency. PLoS One *12*, e0176363.

Arbab, M., Srinivasan, S., Hashimoto, T., Geijsen, N., and Sherwood, R.I. (2015). Cloning-free CRISPR. Stem Cell Reports *5*, 908–917.

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol *14*, e8124.

Argmann, C.A., Houten, S.M., Zhu, J., and Schadt, E.E. (2016). A Next Generation Multiscale View of Inborn Errors of Metabolism. Cell Metab *23*, 13–26.

Baumgartner, R. (1983). Activity of the cobalamin-dependent methylmalonyl-CoA mutase. The Cobalamins: Methods in Hematology. Churchill Livingstone, Edinburgh, New York 181–195.

Brechtmann, F., Mertes, C., Matusevičiūtė, A., Yépez, V.A., Avsec, Ž., Herzog, M., Bader, D.M., Prokisch, H., and Gagneur, J. (2018). OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. The American Journal of Human Genetics *103*, 907–917.

Brosch, M., Yu, L., Hubbard, T., and Choudhary, J. (2009). Accurate and sensitive peptide identification with Mascot Percolator. J Proteome Res *8*, 3176–3181.

Causey, A.G., and Bartlett, K. (1984). A radio-HPLC assay for the measurement of methylmalonyl-CoA mutase. Clinica Chimica Acta *139*, 179–186.

Chen, Q., Kirk, K., Shurubor, Y.I., Zhao, D., Arreguin, A.J., Shahi, I., Valsecchi, F., Primiano, G., Calder, E.L., Carelli, V., et al. (2018). Rewiring of Glutamine Metabolism Is a Bioenergetic Adaptation of Human Cells with Mitochondrial DNA Mutations. Cell Metabolism *27*, 1007-1025.e5.

Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., Bolduc, V., Waddell, L.B., Sandaradura, S.A., O'Grady, G.L., et al. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med *9*, eaal5209.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. J Proteome Res *13*, 3679–3684.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom *5*, 976–989.

Ferreira, C.R., Rahman, S., Keller, M., and Zschocke, J. (2021). An international classification of inherited metabolic disorders (ICIMD). Journal of Inherited Metabolic Disease *44*, 164–177.

Forny, P., Froese, D.S., Suormala, T., Yue, W.W., and Baumgartner, M.R. (2014). Functional characterization and categorization of missense mutations that cause methylmalonyl-CoA mutase (MUT) deficiency. Hum Mutat *35*, 1449–1458.

737 Forny, P., Schumann, A., Mustedanagic, M., Mathis, D., Wulf, M.-A., Nägele, N., Langhans, C.-D.,
738 Zhakupova, A., Heeren, J., Scheja, L., et al. (2016). Novel Mouse Models of Methylmalonic Aciduria
739 Recapitulate Phenotypic Traits with a Genetic Dosage Effect. J Biol Chem *291*, 20563–20573.

740 Forny, P., Hörster, F., Ballhausen, D., Chakrapani, A., Chapman, K.A., Dionisi-Vici, C., Dixon, M.,
741 Grünert, S.C., Grunewald, S., Haliloglu, G., et al. (2021). Guidelines for the diagnosis and
742 management of methylmalonic acidaemia and propionic acidaemia: First revision. Journal of
743 Inherited Metabolic Disease *44*, 566–592.

744 Frésard, L., Smail, C., Ferraro, N.M., Teran, N.A., Li, X., Smith, K.S., Bonner, D., Kernohan, K.D.,
745 Marwaha, S., Zappala, Z., et al. (2019). Identification of rare-disease genes using blood transcriptome
746 sequencing and large control cohorts. Nat Med *25*, 911–919.

747 Froese, S., and Baumgartner, M.R. (2020). Lysosomal Vitamin B12 Trafficking. In Ion and Molecule
748 Transport in Lysosomes, (CRC Press), p.

749 Fuhrer, T., Heer, D., Begemann, B., and Zamboni, N. (2011). High-throughput, accurate mass
750 metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. Anal
751 Chem *83*, 7074–7080.

752 Garrod, ArchibaldE. (1908). The Croonian Lectures ON INBORN ERRORS OF METABOLISM.
753 The Lancet *172*, 1–7.

754 Gonorazky, H.D., Naumenko, S., Ramani, A.K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D.,
755 Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M.A., et al. (2019). Expanding the Boundaries of RNA
756 Sequencing as a Diagnostic Tool for Rare Mendelian Disease. Am J Hum Genet *104*, 466–483.

757 Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances
758 circular visualization in R. Bioinformatics *30*, 2811–2812.

759 Heusel, M., Bludau, I., Rosenberger, G., Hafen, R., Frank, M., Banaei-Esfahani, A., van Drogen, A.,
760 Collins, B.C., Gstaiger, M., and Aebersold, R. (2019). Complex-centric proteome profiling by SEC-
761 SWATH-MS. Mol Syst Biol *15*, e8438.

762 Hirano, M., Emmanuele, V., and Quinzii, C.M. (2018). Emerging therapies for mitochondrial
763 diseases. Essays Biochem *62*, 467–481.

764 Hörster, F., Garbade, S.F., Zwickler, T., Aydin, H.I., Bodamer, O.A., Burlina, A.B., Das, A.M., De
765 Klerk, J.B.C., Dionisi-Vici, C., Geb, S., et al. (2009). Prediction of outcome in isolated methylmalonic
766 acidurias: combined use of clinical and biochemical parameters. Journal of Inherited Metabolic
767 Disease *32*, 630.

768 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new
769 perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res *45*, D353–D361.

770 Karczewski, K.J., and Snyder, M.P. (2018). Integrative omics for health and disease. Nat Rev Genet
771 *19*, 299–310.

772 Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021).
773 Fast gene set enrichment analysis.

774 Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T.B., Graf, E.,
775 Schwarzmayr, T., Terrile, C., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA
776 sequencing. Nat Commun *8*, 15824.

Lamparter, D., Marbach, D., Rueedi, R., Bergmann, S., and Kutalik, Z. (2017). Genome-Wide Association between Transcription Factor Expression and Chromatin Accessibility Reveals Regulators of Chromatin Accessibility. PLOS Computational Biology *13*, e1005311.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res *41*, e108.

Longo, N., Price, L.B., Gappmaier, E., Cantor, N.L., Ernst, S.L., Bailey, C., and Pasquali, M. (2017). Anaplerotic therapy in propionic acidemia. Mol Genet Metab *122*, 51–59.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297–1303.

Nusinow, D.P., Szpyt, J., Ghandi, M., Rose, C.M., McDonald, E.R., Kalocsay, M., Jané-Valbuena, J., Gelfand, E., Schweppe, D.K., Jedrychowski, M., et al. (2020). Quantitative Proteomics of the Cancer Cell Line Encyclopedia. Cell *180*, 387-402.e16.

Palmer, E.E., Sachdev, R., Macintosh, R., Melo, U.S., Mundlos, S., Righetti, S., Kandula, T., Minoche, A.E., Puttick, C., Gayevskiy, V., et al. (2021). Diagnostic Yield of Whole Genome Sequencing After Nondiagnostic Exome Sequencing or Gene Panel in Developmental and Epileptic Encephalopathies. Neurology *96*, e1770–e1782.

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol *36*, 983–987.

Rahman, S. (2020). Mitochondrial disease in children. J Intern Med *287*, 609–633.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281–2308.

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res *49*, D1541–D1547.

Schon, K.R., Horvath, R., Wei, W., Calabrese, C., Tucci, A., Ibañez, K., Ratnaike, T., Pitceathly, R.D.S., Bugiardini, E., Quinlivan, R., et al. (2021). Use of whole genome sequencing to determine genetic basis of suspected mitochondrial disorders: cohort study. BMJ e066288.

Venables, W.N., and Ripley, B.D. (2002). Modern applied statistics with S (New York: Springer).

Vockley, J., Burton, B., Berry, G., Longo, N., Phillips, J., Sanchez-Valle, A., Chapman, K., Tanpaiboon, P., Grunewald, S., Murphy, E., et al. (2021). Effects of triheptanoin (UX007) in patients with long-chain fatty acid oxidation disorders: Results from an open-label, long-term extension study. J Inherit Metab Dis *44*, 253–263.

Want, E.J., Masson, P., Michopoulos, F., Wilson, I.D., Theodoridis, G., Plumb, R.S., Shockcor, J., Loftus, N., Holmes, E., and Nicholson, J.K. (2013). Global metabolic profiling of animal and human tissues via UPLC-MS. Nat Protoc *8*, 17–32.

Willard, H.F., Ambani, L.M., Hart, A.C., Mahoney, M.J., and Rosenberg, L.E. (1976). Rapid prenatal and postnatal detection of inborn errors of propionate, methylmalonate, and cobalamin metabolism: a sensitive assay using cultured cells. Hum Genet *34*, 277–283.

817 Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta,
818 T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function.
819 Science *352*, aad0189–aad0189.

820 Xuan, Y., Bateman, N.W., Gallien, S., Goetze, S., Zhou, Y., Navarro, P., Hu, M., Parikh, N., Hood,
821 B.L., Conrads, K.A., et al. (2020). Standardization and harmonization of distributed multi-center
822 proteotype analysis supporting precision medicine studies. Nat Commun *11*, 5248.

823

824