

## **Supplementary Information:**

### **A new class of protein biomarkers based on subcellular distribution: application to a mouse liver cancer model**

Tatjana Sajic<sup>1</sup>, Rodolfo Ciuffa<sup>1</sup>, Vera Lemos<sup>2</sup>, Pan Xu<sup>2</sup>, Valentina Leone<sup>3,4</sup>, Chen Li<sup>1,5</sup>, Evan G. Williams<sup>1</sup>, Georgios Makris<sup>1</sup>, Amir Banaei-Esfahani<sup>1,6</sup>, Mathias Heikenwalder<sup>3</sup>, Kristina Schoonjans<sup>2</sup>, Ruedi Aebersold<sup>1,7\*</sup>

<sup>1</sup> Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

<sup>2</sup> Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>3</sup> Division of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>4</sup> Research Unit Radiation Cytogenetics, Helmholtz Zentrum München Research Center for Environmental Health (GmbH), Neuherberg, Germany

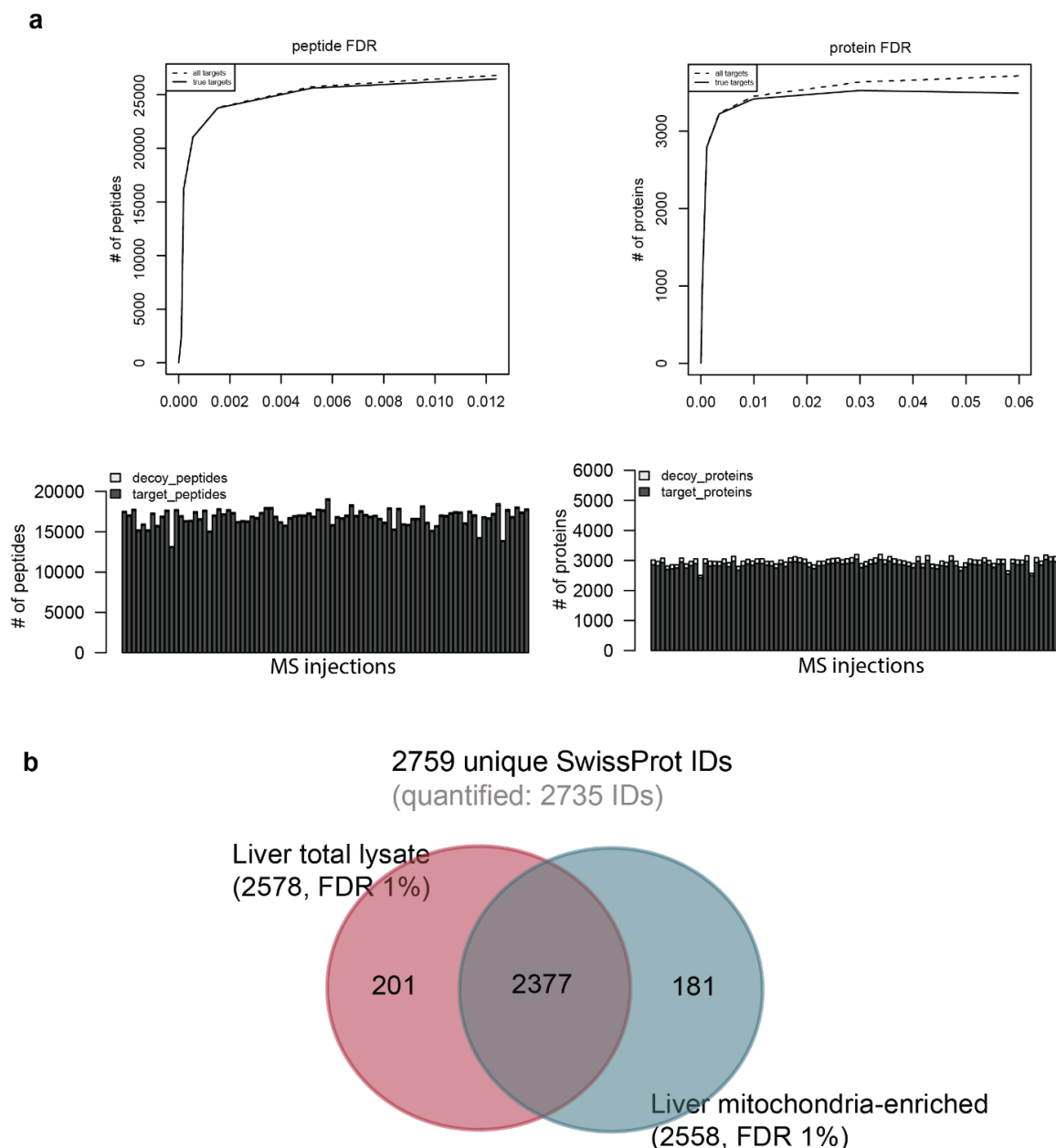
<sup>5</sup> Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC 3800, Australia

<sup>6</sup> PhD Program in Systems Biology, Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland

<sup>7</sup> Faculty of Science, University of Zurich, Zurich, Switzerland

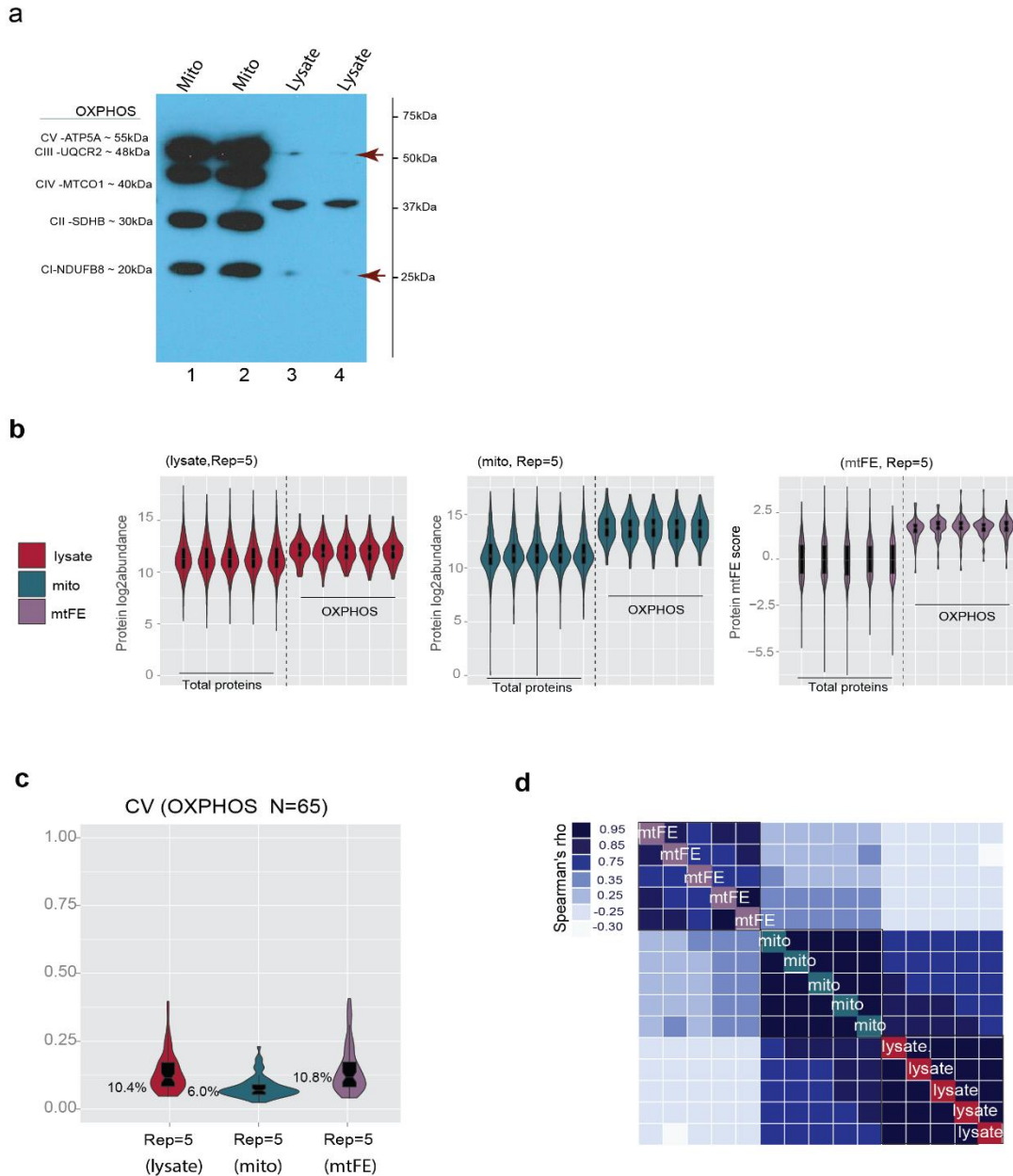
\*Corresponding author: [aebersold@imsb.biol.ethz.ch](mailto:aebersold@imsb.biol.ethz.ch)

**Supplementary Figure 1.**



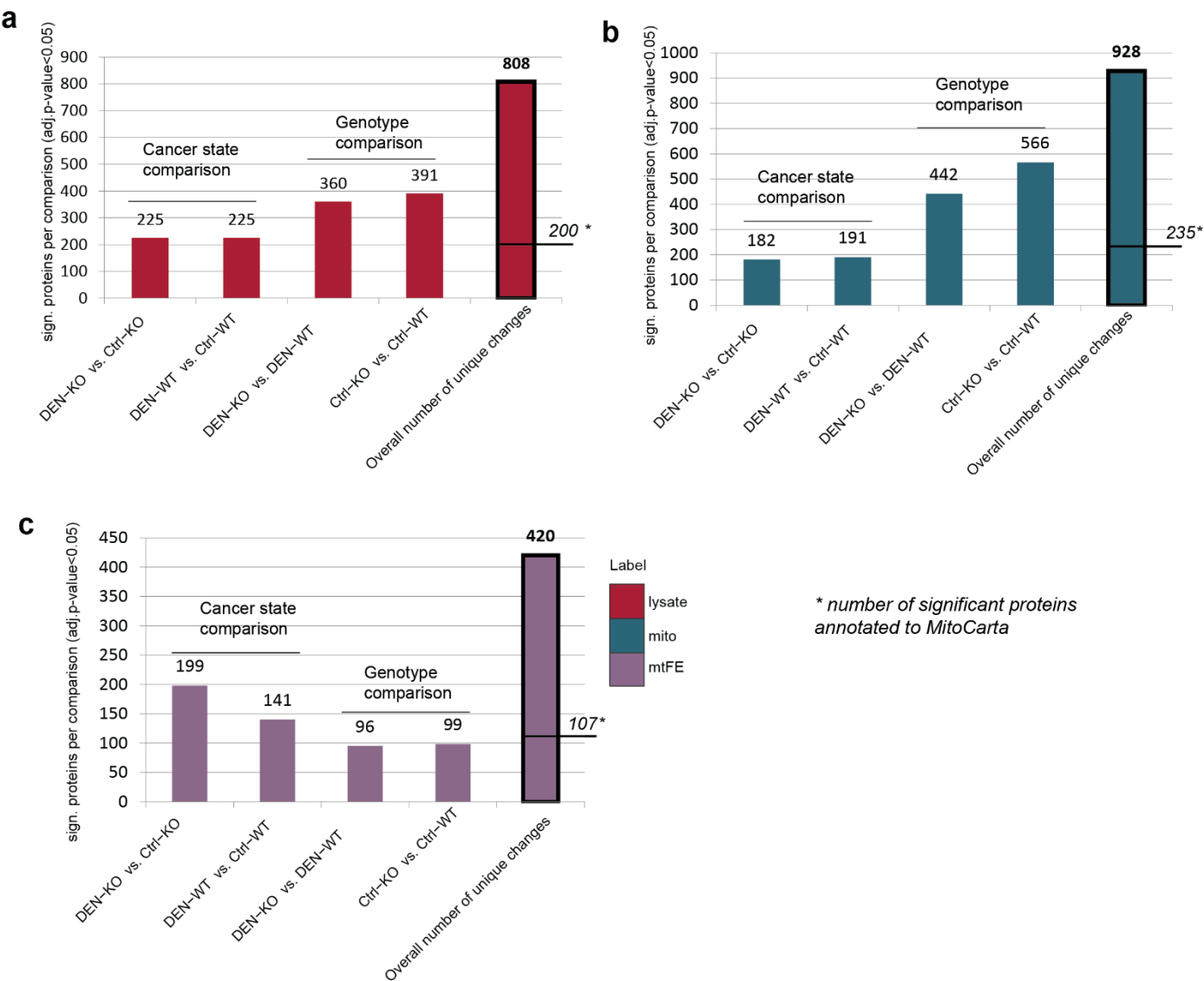
**Supplementary Figure 1: Reproducibility estimation at the peptide and protein level in the discovery cohort by SWATH2stats R-package.** (a) (Top) SWATH2stats global false discovery rate (FDR) report of OpenSWATH/pyProphet results. Receiver operating characteristics (ROC) curves of the estimation of FDR for number of identified peptides (left) and proteins (right). (Bottom) Barplot showing the stable and consistent peptides and proteins identification across all samples and replicates. (b) Venn diagram is showing the intersection of the proteins identified at 1% of protein FDR in total lysate and mitochondria-enriched fraction, respectively. The q-value threshold on level of OpenSWATH peptide query scores was used to assess the protein FDR.

**Supplementary Figure 2.**



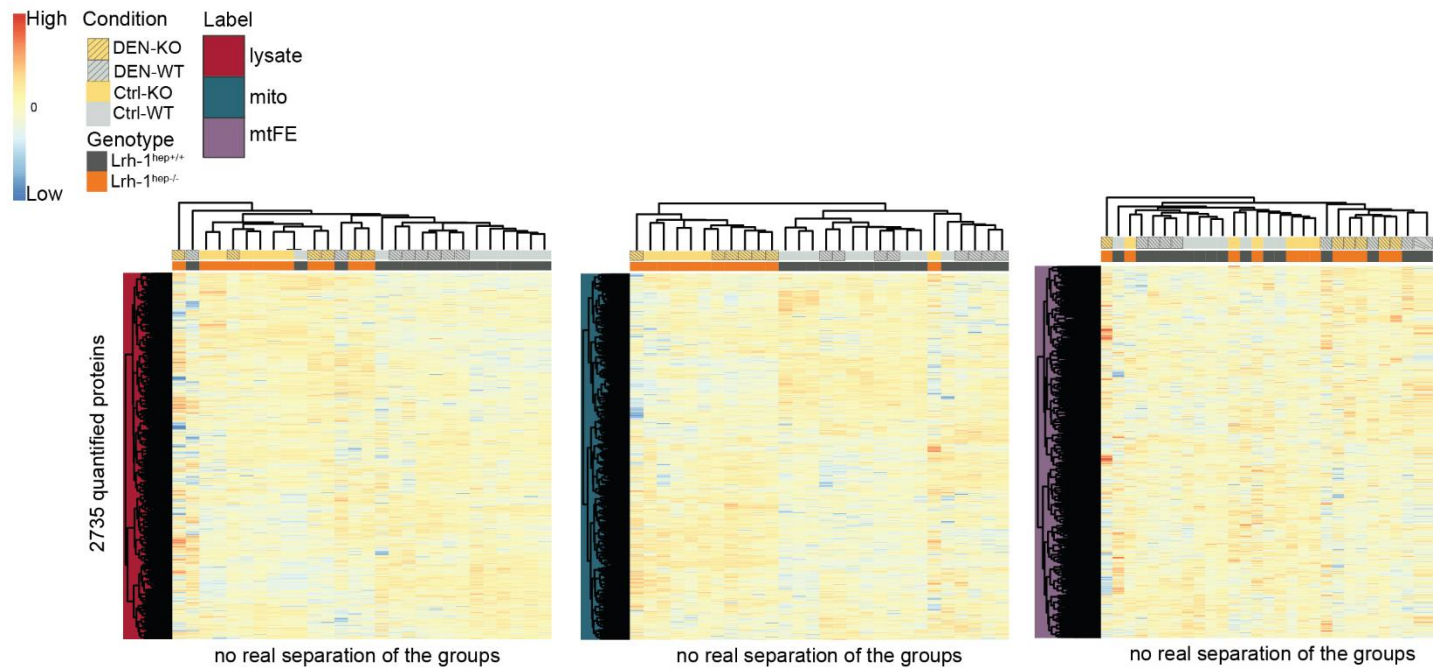
**Supplementary Figure 2: Control of data quality.** (a) Western Blot detection of five OXPHOS enzymes in mitochondrial fraction (lanes 1 and 2) and in cellular lysate (lanes 3 and 4). After long membrane exposure ( $t=1$  min) by chemiluminescence, the bands of CI subunit NDUFB8, CII-35KDa, and CV alpha subunit-53 KDa of OXPHOS enzymes appear on the lines corresponding to total cellular lysate that also contain mitochondria. (b) OXPHOS protein distribution across three different data readouts of replicated samples. (c) CVs (Coefficient of variation) calculated for statically localized OXPHOS enzymes in the enriched mitochondrial fraction, between five replicates of lysate, mito, and mtFE data. The replicates correspond to five repeated liver preparations of cellular lysate, mitochondria-enriched, and calculated mtFE data. (d) Spearman's rank correlation coefficients of protein log2 intensities in replicates of total lysate, enriched-mitochondria, and calculated mtES scores compared against each other. Black boxes highlight comparisons between five replicates ordered by strength of correlation. mtES protein scores, calculated from the lysate and enriched-mitochondrial fraction for each of the five respective replicates, present high range of Spearman's rho correlation of 0.75-0.90.

Supplementary Figure 3.



**Supplementary Figure 3: One-way ANOVA comparisons of liver conditions in lysate, mito, and mtFE scores data (a-c).** The number of ANOVA significant proteins of four distinct comparisons (i.e., comp1=Ctrl-WT vs Ctrl-KO, comp2= DEN-WT vs DEN-KO, comp3= Ctrl-WTvs DEN-WT, and comp4= Ctrl-KO vs DEN-KO) performed on each data set are presented by bar plots. The bar plots are shown by different color codes, corresponding to three respective datasets: lysate (dark red), mito (dark green) and mtFE (dark violet).

## Supplementary Figure 4.



**Supplementary Figure 4: Unsupervised hierarchical data clustering.** Unsupervised hierarchical clustering presents three heat-maps with no real separation of the groups among 28 samples. The based on the 2735 proteins, quantified as abundances in the total lysate and enriched-mitochondrion, and as calculated scores in mtFE data.

Supplementary Figure 5.

a

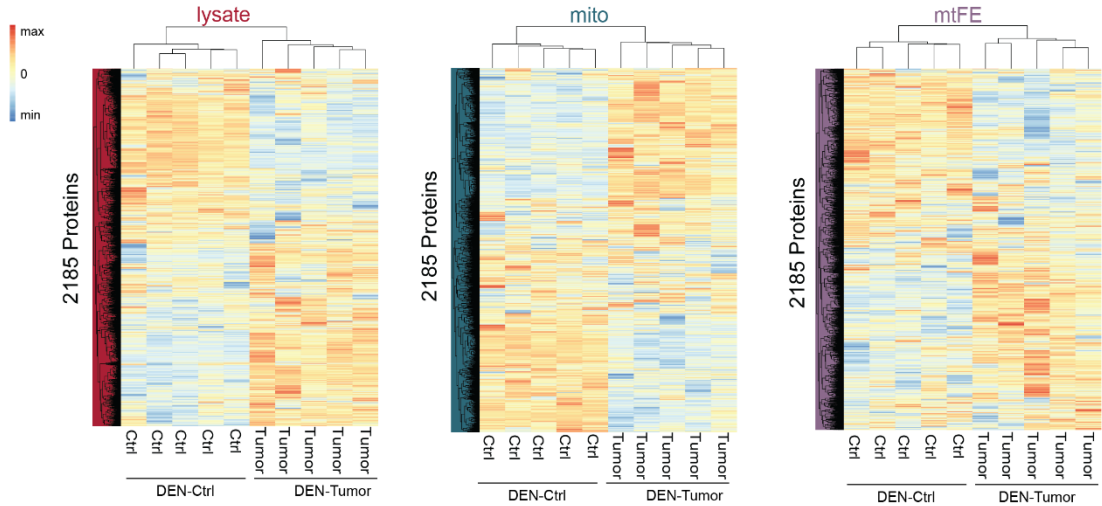
Naïve Bayes-Based Stratification (discovery cohort)

Cellular lysate (cohort 1)				Mitochondrial lysate (cohort 1)				mtES (cohort 1)			
UniProt	Gene Name	AUC	Accuracy	UniProt	Gene Name	AUC	Accuracy	UniProt	Gene Name	AUC	Accuracy
<b>Q8CGK3</b>	<b>Lonp1</b>	0.99	0.96	<b>Q8K354</b>	<b>Cbr3</b>	0.95	0.82	<b>Q8CGK3</b>	<b>Lonp1</b>	0.93	0.86
<b>P62270</b>	<b>Rps18</b>	0.93	0.96	<b>P47199</b>	<b>Cryz</b>	0.92	0.75	<b>P08228</b>	<b>Sod1</b>	0.93	0.93
<b>Q9CPQ8</b>	<b>Atp5l</b>	0.92	0.89	<b>P56393</b>	<b>Cox7b</b>	0.92	0.86	<b>Q9CR62</b>	<b>Slc25a11</b>	0.92	0.89
<b>P61922</b>	<b>Abat</b>	0.91	0.93	<b>Q9CXZ1</b>	<b>Ndufs4</b>	0.90	0.82	<b>P97493</b>	<b>Txn2</b>	0.91	0.86
<b>Q91VD9</b>	<b>Ndufs1</b>	0.91	0.82	<b>Q9CQZ6</b>	<b>Ndufb3</b>	0.89	0.79	<b>Q60597</b>	<b>Ogdh</b>	0.91	0.89
<b>Q9CR62</b>	<b>Slc25a11</b>	0.91	0.89	<b>Q9CQJ8</b>	<b>Ndufb9</b>	0.88	0.82	<b>P61922</b>	<b>Abat</b>	0.90	0.79
<b>Q9D0R2</b>	<b>Tars</b>	0.91	0.79	<b>Q9CQ75</b>	<b>Ndufa2</b>	0.87	0.82	<b>Q3U5Q7</b>	<b>Cmpk2</b>	0.89	0.75
<b>Q9DCT2</b>	<b>Ndufs3</b>	0.91	0.89	<b>Q9D0S9</b>	<b>Hint2</b>	0.87	0.82	<b>Q8VE22</b>	<b>Mrps23</b>	0.89	0.75
<b>P08228</b>	<b>Sod1</b>	0.90	0.93	<b>Q9D6J6</b>	<b>Ndufv2</b>	0.86	0.82	<b>Q9WVM8</b>	<b>Aadat</b>	0.88	0.82
<b>P62264</b>	<b>Rps14</b>	0.90	0.86	<b>Q8QZS1</b>	<b>Hibch</b>	0.86	0.75	<b>Q80XL6</b>	<b>Acad11</b>	0.87	0.86

\*UniProt name in bold overlap with predictors discovered by Logistic Regression-Based model

b

Validation cohort long term DEN, WT mice (n=10 liver samples)



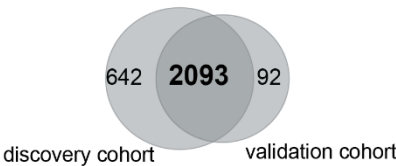
c.

Summary statistics of individual ROCs (validation cohort)

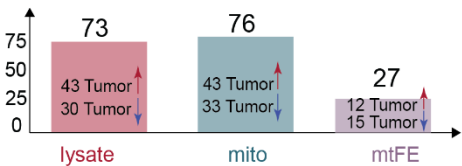
lysate	Rps18		Abat		Ndufs3		Ndufs1		Slc25a11	
	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)
	AUC	1.0 (1.0-1.0)	AUC	1.0 (1.0-1.0)	AUC	1.0 (1.0-1.0)	AUC	1.0 (1.0-1.0)	AUC	0.88 (0.62-1.0)
	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	0.80 (0.40-1.0)
	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)
mito	Ndufv2		Ndufs4		Ndufb3		Ndufb9			
	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)		
	AUC	1.0 (1.0-1.0)	AUC	1.0 (1.0-1.0)	AUC	0.96 (0.85-1.0)	AUC	1.0 (1.0-1.0)		
	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)		
	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	0.80 (0.40-1.0)	Specificity	1.0 (1.0-1.0)		
mtFE	Lonp1		Txn2		Ogdh		Cmpk2		Mrps23	
	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)	Cohort2	(CI 95%)
	AUC	0.88 (0.62-1.0)	AUC	1.0 (1.0-1.0)	AUC	0.88 (0.62-1.0)	AUC	1.0 (1.0-1.0)	AUC	1.0 (1.0-1.0)
	Sensitivity	0.80 (0.40-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	0.80 (0.40-1.0)	Sensitivity	1.0 (1.0-1.0)	Sensitivity	1.0 (1.0-1.0)
	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)	Specificity	1.0 (1.0-1.0)

d.

Common quantified proteins in discovery and validation cohort



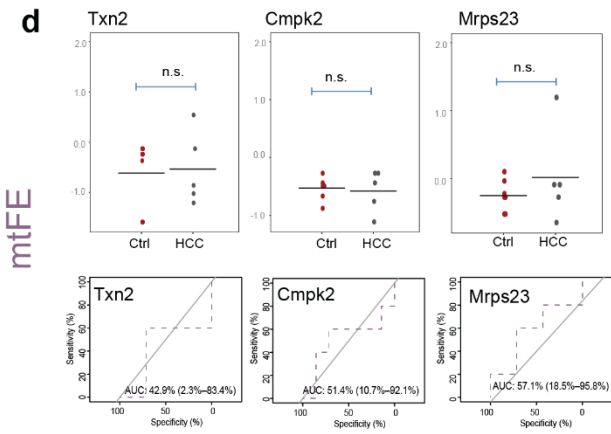
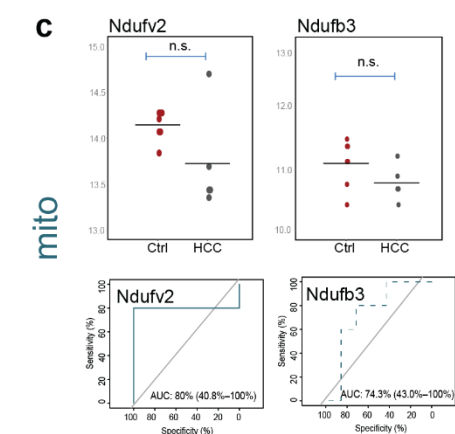
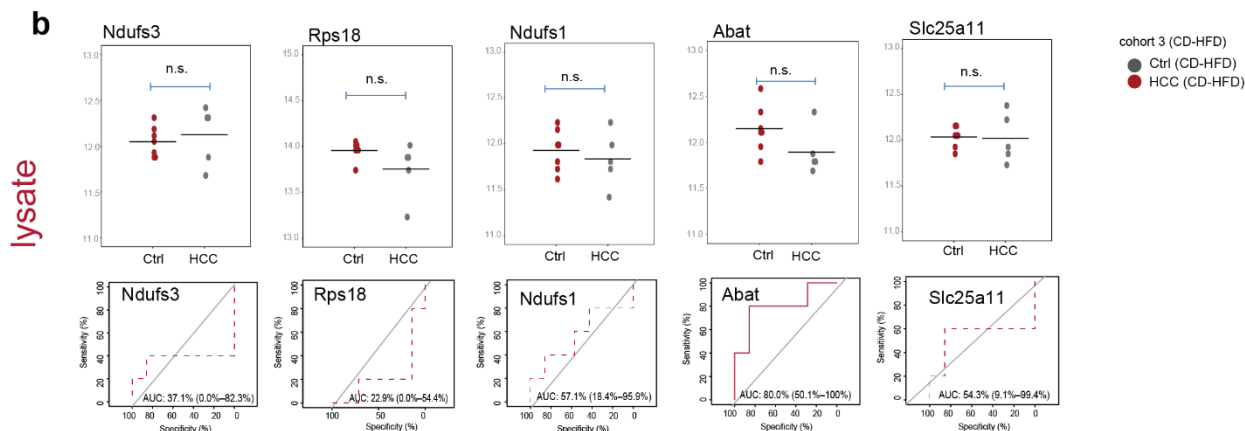
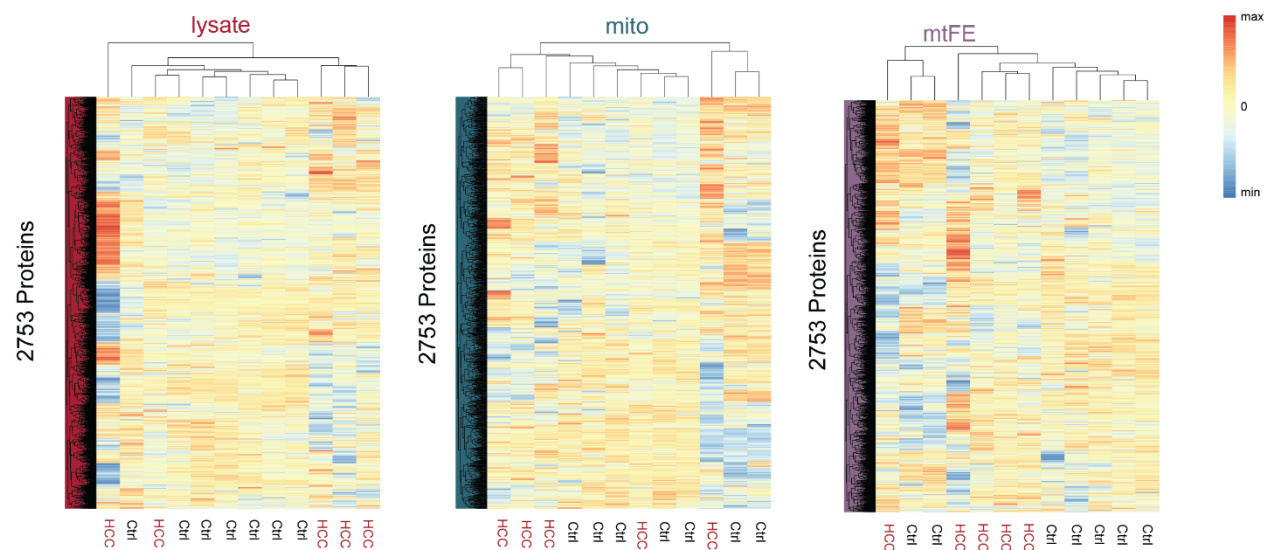
Common differentially expressed proteins in discovery and validation cohort



**Supplementary Figure 5: Protein biomarkers of three data types in DEN-induced liver carcinogenesis.** (a) Top selected tissue markers based on Naïve Bayes (NB) analysis. Three protein panels present top-selected markers ranked according to their average AUC value for three respective datasets: lysate (left panel), mito (mid panel) and mtFE data (right panel). (b) Heat map visualization of data clustering based on 2185 proteins quantified in the validation cohort of long-term DEN-treated WT liver. (c) Validation cohort summary statistics (i.e., specificity, sensitivity, and accuracy at 95% CI) of individual tissue-markers were calculated at an optimal cutoff point as determined by Youden's index (Youden, 1950) for each of the three respective datasets. Optimistic values for specificity, sensitivity, and accuracy at 95% CI in the small sample size available would be expected to be of inferior value in a cohort of large sample (d) Venn Diagram presents common quantified proteins across discovery and validation experiment. The bar plot shown by different color codes, corresponding to three respective datasets, (i.e. lysate (dark red), mito (dark green) and mtFE (dark violet)) present the number of common regulated proteins across discovery and validation cohort. Dark blue or dark red color of arrow corresponds to protein number with decreased or increased abundance in tumor respectively.

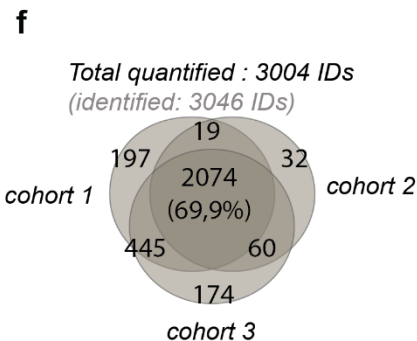
Supplementary Figure 6.

a NASH-induced HCC by CD-HFD in mice cohort 3 (n=12 liver samples)



e Summary statistics of individual ROCs (cohort 3)

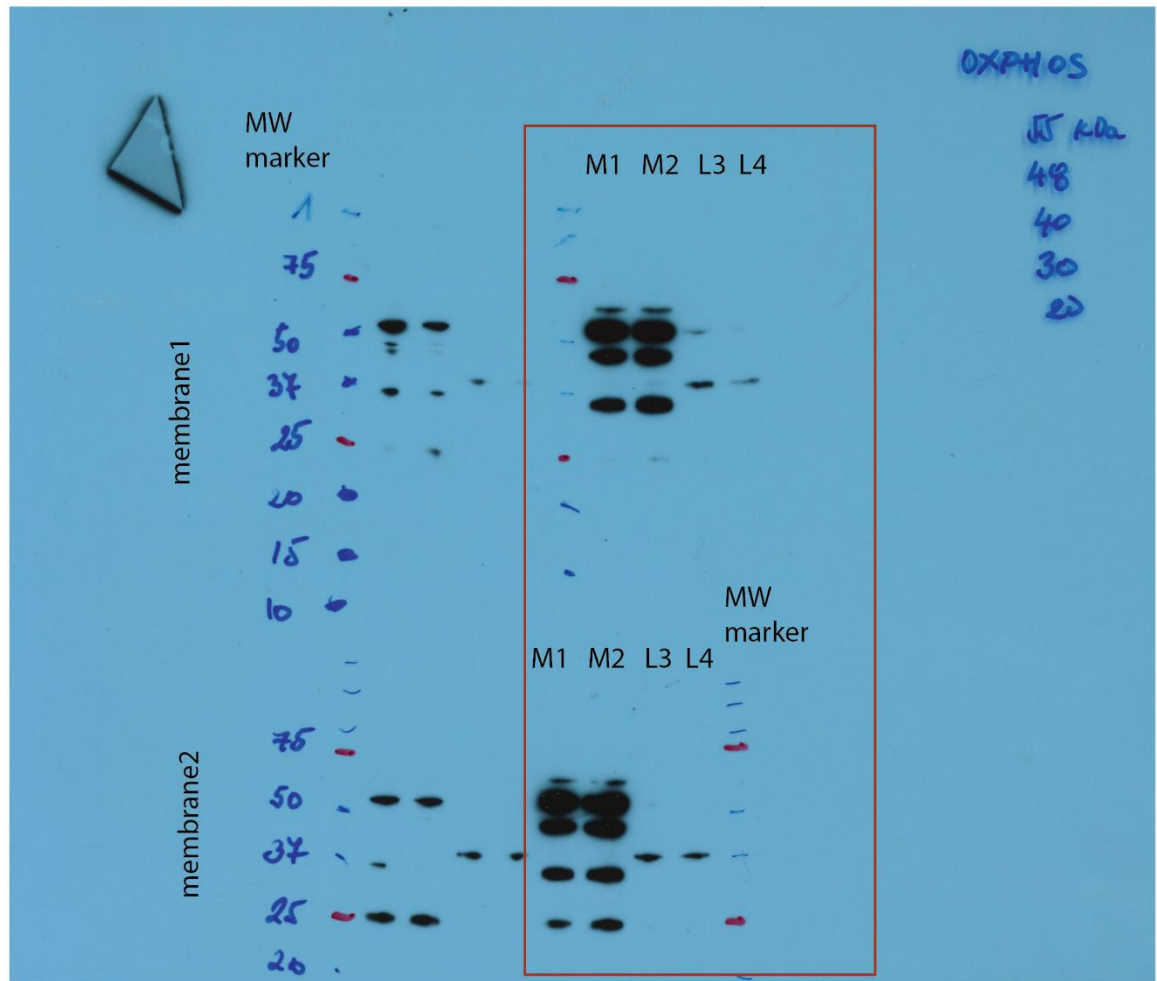
	Ndufs4		Ndufb9	
	Cohort3	(CI 95%)	Cohort3	(CI 95%)
mito	AUC	0.77 (0.39–1.0)	AUC	0.87 (0.65–1.0)
	Sensitivity	0.80 (0.40–1.0)	Sensitivity	0.80 (0.5–1.0)
	Specificity	0.86 (0.57–1.0)	Specificity	1.0 (1.0–1.0)
	Accuracy	0.83 (0.66–1.0)	Accuracy	0.92 (0.67–1.0)
mtFE	Lonp1		Ogdh	
	Cohort3	(CI 95%)	Cohort3	(CI 95%)
	AUC	0.86 (0.63–1.0)	AUC	0.83 (0.54–1.0)
	Sensitivity	0.60 (0.20–1.0)	Sensitivity	0.80 (0.40–1.0)
	Specificity	1.0 (0.43–1.0)	Specificity	0.86 (0.57–1.0)
	Accuracy	0.83 (0.66–1.0)	Accuracy	0.83 (0.66–1.0)





**Supplementary Figure 6: Lysate, mito and mtFE scores proteome data in the NASH-induced model of HCC.**

(a) Heat map visualization of data clustering based on proteins quantified in NASH-induced HCC mice cohort 3. Three heat maps correspond to respective datasets: lysate (dark red), mito (dark green) and mtFE (dark violet). (b-e) Dot plots and ROC curves of corresponding lysate, mito and mtFE markers that were not significantly dysregulated in the new mice model of HCC (b-d). (e) Summary statistics (i.e., specificity, sensitivity, and accuracy at 95% CI calculated with an optimal cutoff point as determined by Youden's index) of confirmed protein markers calculated in the mice cohort 3 for two respective datasets. (e) Venn diagram results represent the number of shared proteins that were quantified between mice cohorts. Of total 3004 quantified proteins, around 70 % were quantified and analyzed over different conditions/ mouse models of liver cancer.



**Supplementary Figure 7: Full membranes detected for Western blots shown in the main and supplementary figures.** Upper and lower membrane (i.e. membrane 1 and 2) are corresponding to experimental replicates of mitochondria-enriched and total lysate isolation by two different protocols. In the right red square, the samples were isolated by using a differential centrifugation steps in the sucrose density solution (200mM) as described in the main text. The representative picture in the main Figure 2 corresponds to the samples detected on the membrane 2 in the red square. The protein samples were in reduced state during gel electrophoresis. An antibody cocktail contains Anti-NDUFB8 (~20KD), Anti-SDHB (~30KD), Anti-UQCRC2 (~48Kgel D), Anti-MTCO1 (~40KD), AntiATP5A1 (~55KD) was used.