Data in Brief 33 (2020) 106453

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Dataset of a comparative proteomics experiment in a methylmalonyl-CoA mutase knockout HEK 293 cell model

Michele Costanzo[a,b,1], Marianna Caterino[a,b,1], Armando Cevenini[a,b], Vincent Jung[c], Cerina Chhuon[c], Joanna Lipecka[c], Roberta Fedele[b], Ida Chiara Guerrera[c,*], Margherita Ruoppolo[a,b,**]

[a] Department of Molecular Medicine and Medical Biotechnology, School of Medicine, University of Naples Federico II, 80131 Naples, Italy
[b] CEINGE – Biotecnologie Avanzate s.c.ar.l., 80145 Naples, Italy
[c] Proteomics Platform Necker, Université de Paris - Structure Fédérative de Recherche Necker, Inserm US24/CNRS UMS3633, 75015 Paris, France

ARTICLE INFO

*Article history:*
Received 7 September 2020
Revised 16 October 2020
Accepted 20 October 2020
Available online 22 October 2020

*Keywords:*
Proteomics
Label-free quantification
Methylmalonyl-CoA mutase
Methylmalonic acidemia
LC-MS/MS
S-Trap digestion
MaxQuant
Perseus

ABSTRACT

Methylmalonic acidemia is a rare inborn error of metabolism with severe clinical complications and poor outcome. The present data article is related to a proteomic investigation conducted on a HEK 293 cell line which has been genetically modified using CRISPR-CAS9 system to knockout the methylmalonyl-CoA mutase enzyme (MUT-KO). Thus, the generated cell model for methylmalonic acidemia was used for a proteomic comparison with respect to HEK 293 wild type cells performing a label-free quantification (LFQ) experiment. A comparison between FASP and S-Trap digestion methods was performed on protein extracts before to proceed with the proteomic analysis of the samples. Four biological replicates were employed for LC-MS/MS analysis and each was run in technical triplicates. MaxQuant and Perseus platforms were used to perform the LFQ of the proteomes and carry out statistical analysis, respectively. Globally, 4341 proteins were identified, and 243 as differentially regulated,

* Corresponding author.
** Corresponding author at: Department of Molecular Medicine and Medical Biotechnology, School of Medicine, University of Naples Federico II, 80131 Naples, Italy.
*E-mail addresses:* chiara.guerrera@inserm.fr (I.C. Guerrera), margherita.ruoppolo@unina.it (M. Ruoppolo).
[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.dib.2020.106453

of which 150 down-regulated and 93 up-regulated in the MUT-KO condition. MS proteomics data have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD017977. The information provided in this dataset shed new light on the cellular mechanisms altered in this rare metabolic disorder, highlighting quantitative unbalances in proteins acting in cell structure and architecture organization and response to the stress. This article can be used as a new source of protein actors to be validated and a starting point for the identification of clinically relevant therapeutic targets.



## Specifications Table

| | |
|---|---|
| Subject | Biochemistry |
| Specific subject area | Proteomics, inborn errors of metabolism |
| Type of data | Figures<br>Table |
| How data were acquired | LC-MS/MS analysis on a nanoRSLC-Q Exactive PLUS (RSLC Ultimate 3000) (Thermo Scientific) |
| Data format | Analyzed<br>Filtered |
| Parameters for data collection | Total proteins were extracted from two different HEK 293 cell lines. One was a wild type (WT) cell line. The second one was a cell line developed as cellular model for methylmalonic acidemia after knockout of methylmalonyl-CoA mutase enzyme (MUT-KO). |
| Description of data collection | The knockout of methylmalonyl-CoA mutase was tested to ensure it was stable over the time. MUT-KO and WT cells were cultured for the same time and in the same conditions before to be collected for proteome extraction. 4 biological replicates per condition were cultured independently. |
| Data source location | CEINGE – Biotecnologie Avanzate s.c.ar.l., Naples, Italy<br>Proteomics Platform Necker, Paris, France |
| Data accessibility | Raw MS proteomics data have been deposited to the ProteomeXchange Consortium (available at http://www.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD017977 |
| Related research article | Costanzo M, Caterino M, Cevenini A, Jung V, Chhuon C, Lipecka J, Fedele R, Guerrera IC, Ruoppolo M. Proteomics Reveals that Methylmalonyl-CoA Mutase Modulates Cell Architecture and Increases Susceptibility to Stress. Int J Mol Sci. 2020 Jul 15;21(14):E4998. https://doi.org/10.3390/ijms21144998 |

## Value of the Data

- This proteomic dataset has been generated in a cellular model of methylmalonic acidemia. We provide a list of differentially regulated proteins potentially acting in the processes underlying the cellular damage in methylmalonic acidemia.
- All the scientists that work on methylmalonic acidemias using different approaches may implement their research with our proteomics data. In addition, also researchers involved in the study of other organic acidemias could benefit from these data to find common and specific alterations within different but metabolically related diseases.
- Our proteomic research has provided findings regarding the damage mechanisms underlying methylmalonic acidemia, and these data could be reused with the purpose of validating new targets for the disease or develop further experimentations aimed at discovering new therapeutic strategies for clinical applications.

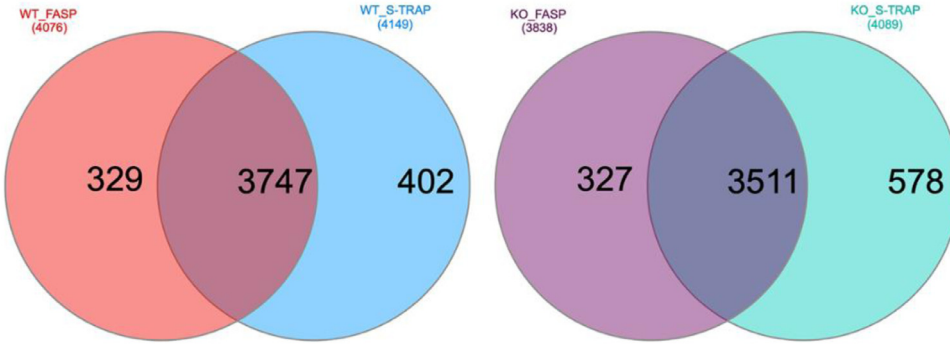


**Fig. 1.** Comparison of protein identification using both FASP and S-Trap digestion columns. The number of identified proteins was compared between the same WT and MUT-KO samples.

## 1. Data Description

On the basis that isolated methylmalonic acidemia is caused by deficiency of methylmalonyl-CoA mutase (MUT) enzyme, we generated a MUT knockout (MUT-KO) cell model to study *in vitro* alterations of the disease.

In order to profile quantitative changes in the proteome of the cell model, a comparative proteomic experiment was performed with respect to wild type (WT) cells. To this aim, the cellular proteomes were extracted, digested and analyzed by liquid chromatography – tandem mass spectrometry (LC-MS/MS).

First of all, the choice of the digestion strategy was taken looking at the number of proteins identified digesting the same samples with FASP (Filter-Aided Sample Preparation) and S-Trap supports (Fig. 1), in favour of S-Trap ones. Label-free quantification (LFQ) was performed as quantitative strategy employing MaxQuant software. Normal distribution and Pearson's correlation graphs (Fig. 2) were used as quality control of the acquired proteomic dataset. The raw proteomic dataset, with details of the differentially regulated proteins quantified by LC-MS/MS, including the UniProt IDs, gene names, protein names, and the statistically significant log2 Difference values, is reported in Supplementary Table 1. Moreover, profile plots were used to show the trends of abundance of the differentially (down- and up-) regulated proteins in the MUT-KO condition (Fig. 3).

These data have been used to interpret a previous study on a MUT-knockdown cell line [1] and to explore the cellular damage in methylmalonic acidemia [2].

## 2. Experimental Design, Materials and Methods

### *2.1. Choice of the trypsin-based digestion strategy*

Before proceeding with the proteomic analysis of all the samples, two samples of WT and MUT-KO cell lines were digested using both FASP and S-Trap columns and analyzed by LC-MS/MS in order to choose the strategy that retrieved the highest number of proteins identified [3]. The results from this comparison were analyzed using the Thermo Proteome Discoverer (version 1.4.1.14) platform (Thermo Scientific, Bremen, Germany), set as reported [4]. Taxonomy was selected for *Homo sapiens*. The identified proteins in WT and MUT-KO from both digestion methods were filtered retaining only those identified by a number equal or higher than 1 unique peptide. Then, the resulting proteins were counted and analyzed using Eulero-Venn diagram [5].

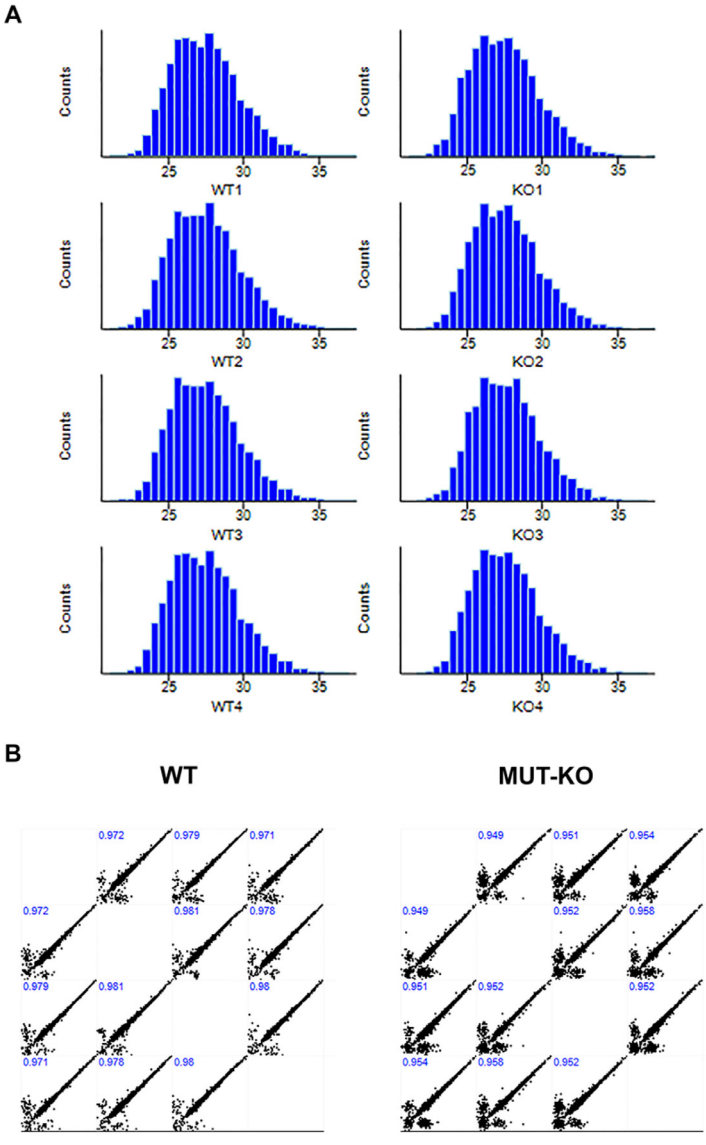


**Fig. 2.** The quality of the proteomic dataset acquired by LC-MS/MS analysis was checked in each replicate analysing (A) the normal distributions of LFQ intensities and (B) the correlation coefficients using Pearson's correlation.

### *2.2. Proteomic analysis*

Cellular protein lysates of 4 biological replicates from each experimental condition were digested on S-Trap columns and analyzed in technical triplicates by nanoLC-MS/MS, using a Q-Exactive Plus Orbitrap mass spectrometer coupled with a nanoLC (RSLC Ultimate 3000) system. Details of the tryptic digestion and the LC-MS/MS measurements are reported in [2]. The raw MS files have been deposited to the ProteomeXchange Consortium (available at http://www.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD017977.

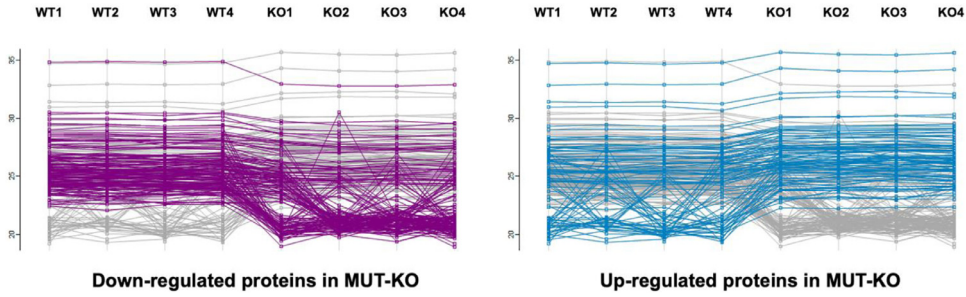


**Fig. 3.** Profile plots showing the trend of abundance of the differentially down- and up-regulated proteins in the MUT-KO condition.

### *2.3. Data processing*

MaxQuant software (version 1.5.8.3) was employed to perform the LFQ analysis of the proteomes, while the Perseus platform (version 1.6.0.7) to carry out data visualization and statistical analysis [2]. Within Perseus, log2-transformed LFQ intensities were reported as histograms to ensure that data were normally distributed. Additionally, multi scatter plots were obtained using Pearson's correlation as measure of the correlation between the replicates of the same condition. Finally, Perseus was further employed to obtain profile plots for data visualization. The trend of the differentially down- and up-regulated proteins was highlighted with two different colors.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

## Acknowledgments

This research was funded by PRIN (Progetti di Ricerca di Rilevante Interesse Nazionale—grant number Prot. 2017SNRXH3)

## Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2020.106453.

## References

[1] M. Costanzo, A. Cevenini, E. Marchese, E. Imperlini, M. Raia, L. Del Vecchio, M. Caterino, M. Ruoppolo, Label-free quantitative proteomics in a methylmalonyl-CoA mutase-silenced neuroblastoma cell line, Int. J. Mol. Sci. 19 (2018) 3580 https://doi.org/10.3390/ijms19113580.
[2] M. Costanzo, M. Caterino, A. Cevenini, V. Jung, C. Chhuon, J. Lipecka, R. Fedele, I.C. Guerrera, M. Ruoppolo, Proteomics reveals that methylmalonyl-CoA mutase modulates cell architecture and increases susceptibility to stress, Int. J. Mol. Sci. 21 (2020) 4998 https://doi.org/10.3390/ijms21144998.
[3] K.R. Ludwig, M.M. Schroll, A.B. Hummon, Comparison of in-solution, FASP, and S-trap based digestion methods for bottom-up proteomic studies, J. Proteome Res. (2018) https://doi.org/10.1021/acs.jproteome.8b00235.
[4] V. De Pasquale, M. Costanzo, R. Siciliano, M. Mazzeo, V. Pistorio, L. Bianchi, E. Marchese, M. Ruoppolo, L. Pavone, M. Caterino, Proteomic analysis of mucopolysaccharidosis IIIB mouse brain, Biomolecules 10 (2020) 355 https://doi.org/10.3390/biom10030355.
[5] H. Heberle, V.G. Meirelles, F.R. da Silva, G.P. Telles, R. Minghim, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams, BMC Bioinform. (2015) https://doi.org/10.1186/s12859-015-0611-3.