Data in Brief 41 (2022) 107843

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Proteome data of neuroblastoma cells overexpressing Neuroglobin

Michele Costanzo[a,b,*], Marianna Caterino[a,b], Illari Salvatori[c], Valeria Manganelli[d], Alberto Ferri[c], Roberta Misasi[d], Margherita Ruoppolo[a,b]

[a] Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples 80131, Italy
[b] CEINGE-Biotecnologie Avanzate s.c.ar.l., Naples 80145, Italy
[c] Fondazione Santa Lucia IRCCS, c/o CERC, Rome 00143, Italy
[d] Department of Experimental Medicine, "Sapienza" University of Rome, Rome 00161, Italy

ARTICLE INFO

Article history:
Received 28 December 2021
Revised 11 January 2022
Accepted 14 January 2022
Available online 20 January 2022

Keywords:
Proteomics
Neuroglobin
Label-free quantification
LC-MS/MS
S-Trap
MaxQuant
Perseus
MetaboAnalyst

ABSTRACT

In this article, we present data on the proteome of human neuroblastoma cells stably overexpressing Neuroglobin (NGB). The neuroprotective role of NGB is clearly established, nevertheless the related mechanistic processes, which are dependent on NGB overexpression, are not known. To address this question, we performed shotgun label-free quantification (LFQ) proteomics using an SH-SY5Y cell model of neuroblastoma that overexpresses an NGB-FLAG construct, and wild type cells transfected with an empty vector as control (CTRL). The proteomes from six biological samples per condition were digested using the S-Trap sample preparation followed by LC-MS/MS analysis with a LTQ-Orbitrap XL mass spectrometer. The quantitative analysis was performed using the LFQ algorithm of MaxQuant, leading to 1654 correctly quantified proteins over 2580 identified proteins. Finally, the statistic comparison of the two analyzed groups within Perseus platform identified 178 differential proteins (107 up- and 71 down-regulated). In addition, multivariate statistical analysis was carried out using MetaboAnalyst 5.0

* Corresponding author at: Department of Molecular Medicine and Medical Biotechnology, University of Naples Federico II, Naples 80131, Italy.
E-mail address: michele.costanzo@unina.it (M. Costanzo).

https://doi.org/10.1016/j.dib.2022.107843

software. MS proteomics data are available via ProteomeXchange with the dataset identifier PXD029012.

Specifications Table

Subject	Omics: Proteomics Biochemistry
Specific subject area	Dataset of the differential proteome of Neuroglobin-overexpressing neuroblastoma SH-SY5Y cells
Type of data	Figures Table
How the data were acquired	The data were acquired via liquid chromatography-tandem mass spectrometry (LC-MS/MS): EASY-nLC + LTQ-Orbitrap XL
Data format	Analyzed Filtered Raw
Description of data collection	The overexpression of Neuroglobin in the form of a FLAG-tagged protein (NGB-FLAG) was repeatedly tested over the time to confirm its stability. NGB-FLAG and CTRL cells (six replicates per condition) were analyzed by LC-MS/MS using MaxQuant, Perseus, and MetaboAnalyst platforms for statistical and comparative analysis. Raw LC-MS/MS files were made publicly available within the ProteomeXchange Consortium.
Data source location	CEINGE–Biotecnologie Avanzate s.c.ar.l., Naples, Italy
Data accessibility	The raw mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029012 (available at https://www.ebi.ac.uk/pride/archive/projects/PXD029012)
Related research article	V. Manganelli, I. Salvatori, M. Costanzo, A. Capozzi, D. Caissutti, M. Caterino, C. Valle, A. Ferri, M. Sorice, M. Ruoppolo, T. Garofalo, R. Misasi, Overexpression of Neuroglobin Promotes Energy Metabolism and Autophagy Induction in Human Neuroblastoma SH-SY5Y Cells, Cells. 10 (2021) 3394. https://doi.org/10.3390/cells10123394

Value of the Data

- This work provides the first documentation of the proteome of neuroblastoma cells when Neuroglobin (NGB) expression is up-regulated.
- The proteome dataset highlights the quantitative differences of significant up- or down-regulated proteins as direct effect of NGB overexpression.
- The scientist interested in NGB-based research or in neurodegenerative disorders may find these data a valuable resource for their studies.

1. Data Description

In this study, we used a shotgun label-free proteomic approach to investigate the differential protein expression profiles of neuroblastoma SH-SY5Y cells overexpressing the NGB-FLAG protein [1]. In parallel, wild type cells transfected with an empty vector were employed as control cells (CTRL). The proteomes of these cells were digested using S-Trap columns and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a LTQ-Orbitrap XL mass spectrometer. The raw LC-MS/MS runs were elaborated with MaxQuant according to the LFQ (Labe-Free Quantification) approach, while data analysis was carried out adopting Perseus and MetaboAnalyst platforms. The LFQ analysis revealed a total of 2580 identified proteins, of which

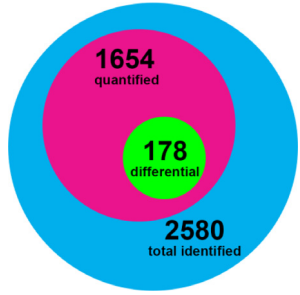

Fig. 1. Eulero-Venn diagram illustrating with different colours the qualitative results of the proteomic analysis in the NGB-FLAG-overexpressing SH-SY5Y cells. Blue section includes the number of total identified proteins, the pink area contains the number of quantified proteins, while the green one the significant differentially abundant proteins obtained after statistical analysis of NGB-FLAG versus CTRL samples.

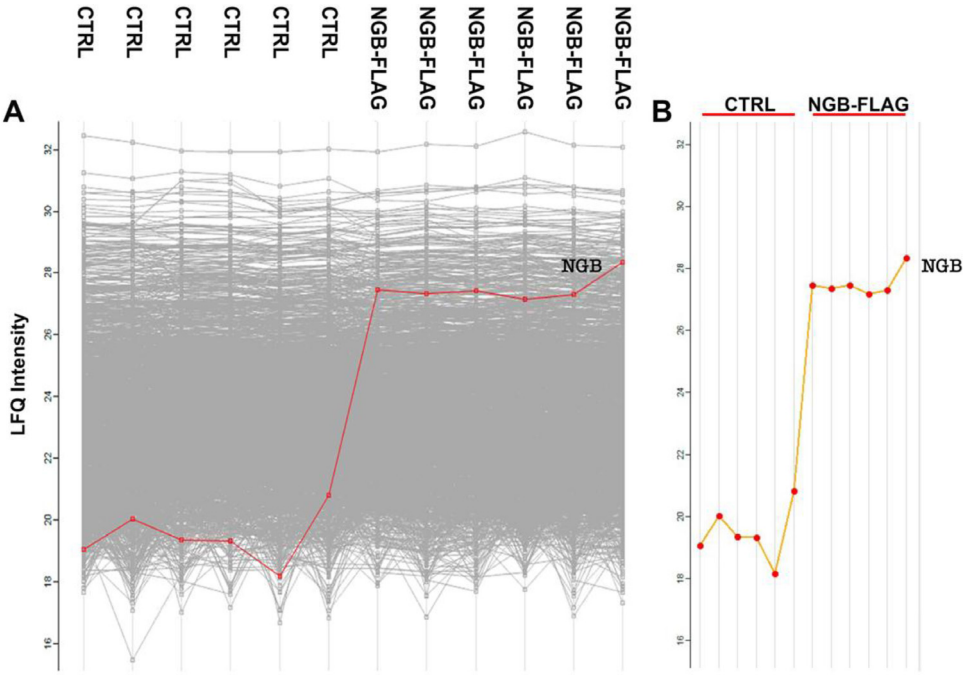

Fig. 2. Profile plots showing the trend of abundance of the proteins identified in the LFQ experiment, with focus on NGB (A) in the total proteome and (B) taken alone.

1654 correctly quantified. Of these, 178 proteins were selected as differentially regulated (107 up and 71 down) between NGB-FLAG and CTRL samples (Fig. 1). Supplementary Table S1 contains the list of the 178 differential proteins characterizing the proteome of NGB-overexpressing SH-SY5Y cells. This list includes for each protein: UniProt ID, Gene name, and Difference value. Supplementary Table S2 contains all the details relative to protein identification and quantification as generated from MaxQuant and Perseus elaborations. Then, the proteomic dataset was explored to check the trend of NGB levels within the two compared cell populations, being NGB protein the experimental internal control of our proteomic analysis. The profile plots showed in Fig. 2 highlight the trend of abundance (LFQ intensity) of NGB among the replicates. Finally, both Perseus and MetaboAnalyst software were used to generate heatmaps in order to visualize the

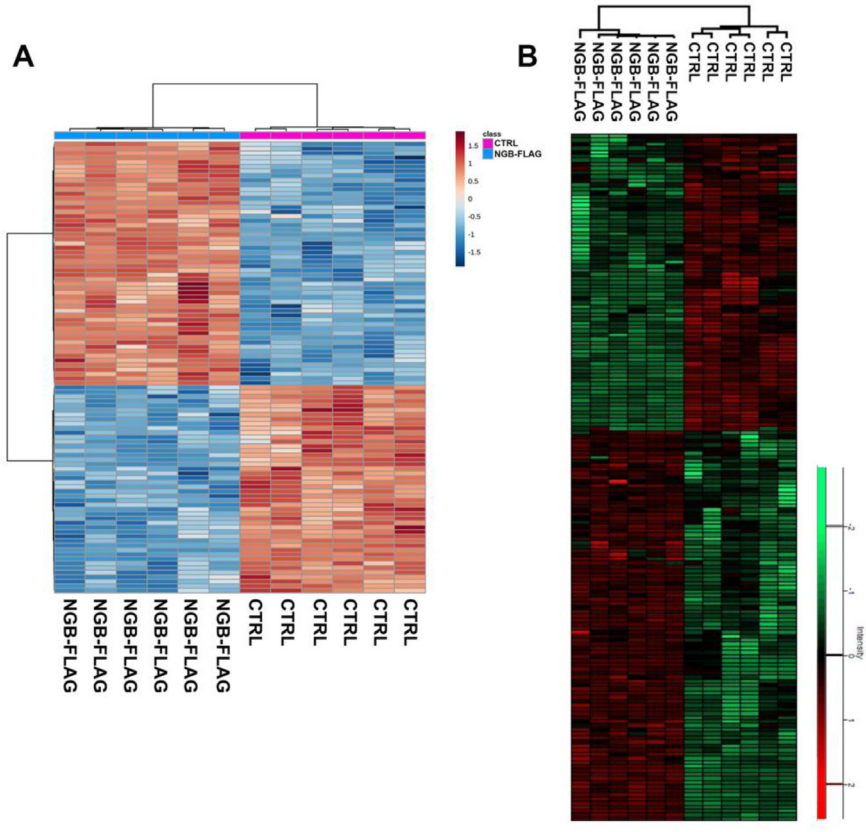

Fig. 3. Heatmaps of LFQ protein intensities to show the separation of the two analyzed groups (NGB-FLAG versus CTRL) using (A) MetaboAnalyst 5.0 and (B) Perseus.

distribution of protein abundances within the replicates. As shown in Fig. 3, both the heatmaps were built to have double confirmation of a clear separation of the two analyzed groups.

2. Experimental Design, Materials and Methods

2.1. Cell samples preparation and LC-MS/MS analysis

The proteomes of six biological replicates from both the experimental conditions (NGB-FLAG and CTRL) were extracted and digested with trypsin using S-Trap digestion protocol [2], and analyzed by LC-MS/MS, using a LTQ-Orbitrap XL mass spectrometer coupled with a EASY-nLC system (Thermo Scientific, Bremen, Germany) [3]. In more detail, the cells were lysed in RIPA buffer (Sigma-Aldrich, St. Louis, MO, USA) and mechanical homogenization of the samples was obtained with the TissueLyser II homogenizer (Qiagen, Duesseldorf, Germany). After, the protein lysates were incubated with 1% Benzonase (E8263–5KU, Sigma-Aldrich, St. Louis, MO, USA) in 2 mM $MgCl_2$ at 37 °C for 30 min to degrade nucleic acids, and then centrifuged at 18,000 rpm for 30 min at 4 °C to clarify the samples. Proteome hydrolysis was performed onto S-Trap micro spin filters (Protifi, Huntington, NY, USA), using trypsin (Promega, Madison, WI, USA) at 47 °C for four hours, in a ratio 1:25 per 50 µg of protein extracts. Eluted peptides were vacuum-dried, resuspended in 0.2% formic acid, and the volume corresponding to 750 ng of sample was

injected onto the LC system. Samples were concentrated and desalted onto a C18 trapping column (2 cm × 100 µm ID, 5 µm) and then fractionated onto a C18 capillary reverse-phase column (20 cm × 75 µm ID, 5 µm) working at the flow rate of 250 nL/min, using chromatography solvent A (0.1% formic acid in water) and solvent B (80% ACN, 0.08% formic acid). The gradient used was: 5% to 40% B (180 min), 40% to 80% (5 min), 80% (5 min), 5% (20 min). The LTQ-Orbitrap XL mass spectrometer equipped with an ESI (Electrospray Ionization) ion source was used for MS/MS analysis setting a Data-Dependent Acquisition (DDA) method. MS analysis was performed on a 400–1800 m/z mass range for the precursor ions with a resolution set to 30,000. For the MS/MS analysis, the five most abundant ions in each MS scan were selected and fragmented using a Collision-Induced Dissociation (CID) fragmentation, applying a dynamic exclusion window of 40 s. One blank after each sample was run for preventing sample carryover. The LC-MS/MS raw files were deposited to ProteomeXchange via the PRIDE partner repository with the dataset identifier PXD029012 [4,5].

2.2. Proteomic data analysis

The LC-MS/MS raw files were elaborated with MaxQuant (version 1.6.17.0) for the processes of protein identification and quantification, according to the LFQ algorithm [6,7]. The runs were analyzed with the Andromeda search engine against the freely available reference proteome of Homo sapiens (Organism ID:9606; Proteome ID: UP000005640; total proteins: 79,038) downloaded from the UniProtKB database (January 2021). The precursor and the fragment mass tolerance allowed was set to 4.5 and 20 ppm, respectively. The minimum peptide length was set to seven amino acids and trypsin was selected as proteolytic enzyme, allowing up to two missed cleavage sites. Carbamidomethylation (Cys) was set as fixed modification, while oxidation (Met) and N-term acetylation were the variable modifications. The false discovery rate (FDR) at both the protein and peptide level was set to 1%. In this analysis, the match between runs option was selected. According to the MaxLFQ algorithm, proteins were quantified on the basis of the extracted ion currents (XIC) of precursor ion peptides. Then, the results of this analysis were first imported onto Perseus (version 1.6.14.0) and, then, onto MetaboAnalyst 5.0 for univariate and multivariate statistical data analysis, and data visualization [8–10]. Within Perseus, the initial proteomic dataset was filtered removing proteins only identified by site, potential contaminants, and reverse hits. Further filtering of the dataset was carried out considering for each protein a number of valid LFQ values equal to or higher than the 50% in all the replicates [11]. LFQ intensities were $\log_2$-transformed, and missing values imputed using random numbers from an ideal gaussian distribution (width = 0.5; down shift = 1.8). Finally, the comparison between the NGB-FLAG and CTRL conditions was carried out selecting significant proteins with values of S0 = 0.5 and Student's t-test with FDR = 0.01. The differential abundance (Difference) was calculated as difference of the average $\log_2$ protein LFQ intensities of NGB-FLAG and CTRL samples. A profile plot analysis was performed to individuate the trend of the identified proteins throughout the replicates, with a particular focus on NGB, whose trend was highlighted with a red line [12]. The heatmap produced using Perseus was generated after Z-score normalization of the protein dataset, and minimum and maximum intensity values were represented by default green and red colours, respectively. On the other hand, for the analysis performed with MetaboAnalyst the dataset underwent no further normalization or transformation, but data were Pareto-scaled. The heatmap generated was represented with default blue and red colours, indicating the minimum and the maximum intensity range, respectively [13].

Ethics Statement

Not applicable.

Funding

This work was supported by PRIN (Progetti di Ricerca di Rilevante Interesse Nazionale-grant number Prot. 2017SNRXH3).

CRediT Author Statement

Michele Costanzo: Data curation, Formal analysis, Methodology, Visualization, Writing – original draft; **Marianna Caterino:** Data curation, Formal analysis, Methodology; **Illari Salvatori:** Data curation, Formal analysis; **Valeria Manganelli:** Data curation, Formal analysis; **Alberto Ferri:** Conceptualization, Supervision; **Roberta Misasi:** Conceptualization, Funding acquisition, Supervision; **Margherita Ruoppolo:** Conceptualization, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2022.107843.

References

[1] V. Manganelli, I. Salvatori, M. Costanzo, A. Capozzi, D. Caissutti, M. Caterino, C. Valle, A. Ferri, M. Sorice, M. Ruoppolo, T. Garofalo, R. Misasi, Overexpression of Neuroglobin promotes energy metabolism and autophagy induction in human neuroblastoma SH-SY5Y Cells, Cells 10 (2021) 3394, doi:10.3390/cells10123394.
[2] 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), doi:10.1021/acs.jproteome.8b00235.
[3] 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, doi:10.3390/ijms19113580.
[4] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yılmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data, Nucleic Acids Res 47 (2019) D442–D450, doi:10.1093/nar/gky1106.
[5] A.F. Jarnuczak, J.A. Vizcaíno, Using the PRIDE Database and ProteomeXchange for Submitting and Accessing Public Proteomics Datasets, Curr. Protoc. Bioinforma. 59 (2017), doi:10.1002/cpbi.30.
[6] 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, doi:10.3390/ijms21144998.
[7] M. Costanzo, M. Fiocchetti, P. Ascenzi, M. Marino, M. Caterino, M. Ruoppolo, Proteomic and bioinformatic investigation of altered pathways in Neuroglobin-deficient breast cancer cells, Molecules 26 (2021) 2397, doi:10.3390/molecules26082397.
[8] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The perseus computational platform for comprehensive analysis of (prote)omics data, Nat. Methods 13 (2016) 731–740, doi:10.1038/nmeth.3901.
[9] Z. Pang, J. Chong, G. Zhou, D.A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.É. Jacques, S. Li, J. Xia, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic Acids Res. (2021), doi:10.1093/nar/gkab382.
[10] M. Caterino, M. Ruoppolo, M. Costanzo, L. Albano, D. Crisci, G. Sotgiu, L. Saderi, A. Montella, F. Franconi, I. Campesi, Sex affects human premature neonates' blood metabolome according to gestational age, parenteral nutrition, and caffeine treatment, Metabolites 11 (2021) 158, doi:10.3390/metabo11030158.
[11] M. Gonzalez Melo, A. Fontana, D. Viertl, G. Allenbach, J.O. Prior, S. Rotman, R.G. Feichtinger, J.A. Mayr, M. Costanzo, M. Caterino, M. Ruoppolo, O. Braissant, F. Barbey, D. Ballhausen, A knock-in rat model unravels acute and chronic renal toxicity in glutaric aciduria type I, Mol. Genet. Metab. (2021), doi:10.1016/j.ymgme.2021.10.003.

[12] M. Costanzo, M. Caterino, A. Cevenini, V. Jung, C. Chhuon, J. Lipecka, R. Fedele, I.C. Guerrera, M. Ruoppolo, Dataset of a comparative proteomics experiment in a methylmalonyl-CoA mutase knockout HEK 293 cell model, Data Br. 33 (2020) 106453, doi:10.1016/j.dib.2020.106453.
[13] M. Caterino, M. Costanzo, R. Fedele, A. Cevenini, M. Gelzo, A. Di Minno, I. Andolfo, M. Capasso, R. Russo, A. Annunziata, C. Calabrese, G. Fiorentino, M. D'Abbraccio, C. Dell'Isola, F.M. Fusco, R. Parrella, G. Fabbrocini, I. Gentile, G. Castaldo, M. Ruoppolo, The serum metabolome of moderate and severe COVID-19 patients reflects possible liver alterations involving carbon and nitrogen metabolism, Int. J. Mol. Sci. 22 (2021) 9548, doi:10.3390/ijms22179548.