Data in Brief 45 (2022) 108634

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Dataset for proteomic analysis of arylamine *N*-acetyltransferase 1 knockout MDA-MB-231 breast cancer cells

Kyung U. Hong[a], Jonathan Q. Gardner[a], Mark A. Doll[a], Marcus W. Stepp[a], Daniel W. Wilkey[b], Frederick W. Benz[a], Jian Cai[b], Michael L. Merchant[b], David W. Hein[a,*]

[a] *Department of Pharmacology & Toxicology, University of Louisville School of Medicine, Louisville, KY, USA*
[b] *Department of Medicine, University of Louisville School of Medicine, Louisville, KY, USA*

ARTICLE INFO

Article history:
Received 10 August 2022
Revised 16 September 2022
Accepted 19 September 2022
Available online 24 September 2022

Dataset link: Proteomic Analysis of Arylamine N-Acetyltransferase 1 Knockout Breast Cancer Cells (Original data)

Keywords:
Tumor immunity
Major histocompatibility complex I
Antigen presentation
ATP synthase
Complex V
Mitochondria
Cell cycle
Cell death

ABSTRACT

Arylamine *N*-acetyltransferase 1 (*NAT1*) is frequently upregulated in breast cancer. An unbiased analysis of proteomes of parental MDA-MB-231 breast cancer cells and two separate *NAT1* knockout (KO) cell lines were performed. Among 4,890 proteins identified, 737 and 651 proteins were found significantly ($p < 0.01$) upregulated and downregulated, respectively, in *NAT1* KO cells, compared to the parental cells. Each set of proteins was analyzed to identify Gene Ontology biological processes, molecular functions, and cellular components that were enriched in the set. Among the proteins upregulated in *NAT1* KO cells, processes associated with MHC major histocompatibility complex I-mediated antigen presentation were significantly enriched. Multiple processes involved in mitochondrial functions were collectively downregulated in *NAT1* KO cells, including multiple subunits of mitochondrial ATP synthase (Complex V of the electron transport chain). This was accompanied by a reduction in cell cycle-associated proteins and an increase in pro-apoptotic pathways in *NAT1* KO cells. The current dataset

DOI of original article: 10.1016/j.toxrep.2022.07.010
* Corresponding author.
E-mail address: david.hein@louisville.edu (D.W. Hein).
Social media: @Peptidome (M.L. Merchant)

https://doi.org/10.1016/j.dib.2022.108634

contains additional representations of the biological processes and components that are differentially enriched in *NAT1* KO MDA-MB-231 cells and will serve as a basis for generating novel hypotheses regarding the role of NAT1 in breast cancer. Data are available via ProteomeXchange with identifier PXD035953.

Specifications Table

Subject	Cancer research
Specific subject area	Role of arylamine *N*-acetyltransferase 1 (*NAT1*) in breast cancer growth
Type of data	Figure Chart Graph
How the data were acquired	An unbiased proteomic study was conducted on parental and two *NAT1* KO MDA-MB-231 breast cancer cell lines. The raw data was acquired using liquid chromatography-mass spectrometry (LC-MS). The proteins that were significantly ($p < 0.01$) upregulated and downregulated in *NAT1* KO cells were analyzed to identify biological processes, molecular functions, and cellular components that were significantly enriched in each set using GOnet [1], PANTHER [2], and Reactome [3].
Data format	Analyzed Filtered
Description of data collection	Global proteomes of parental MDA-MB-231 cells and two *NAT1* KO cell lines [4,5] were analyzed. Proteomic analysis was conducted using a TMT-labeling approach and LC/MS to identify proteins that are differentially regulated in *NAT1* KO cells. Intensity based normalization of reporter ions was done using the median calculation type, unique peptides blocking level, individual spectrum references, and normalization between samples.
Data source location	• Institution: University of Louisville, School of Medicine • City/Town/Region: Louisville, Kentucky • Country: U.S.A.
Data accessibility	Repository name: ProteomeXchange Consortium via the PRIDE [4] partner repository with the dataset identifier PXD035953 and 10.6019/PXD035953.
Related research article	*K. U. Hong, J. Q. Gardner, M. A. Doll, M. W. Stepp, D. W. Wilkey, F. W. Benz, J. Cai, M. L. Merchant, D. W. Hein,* Proteomic Analysis of Arylamine N-Acetyltransferase 1 Knockout Breast Cancer Cells: Implications in immune evasion and mitochondrial biogenesis, *Toxicology Reports* 9: 1566-1573, 2022. 10.1016/j.toxrep.2022.07.010 [5]

Value of the Data

- *NAT1* is frequently upregulated in breast cancer and has been shown to contribute to cancer growth in both *in vitro* and *in vivo*. However, the mechanism by which NAT1 contributes to breast cancer growth remains unknown. In attempt to generate novel hypotheses, the global proteomes of a breast cancer cell line, MDA-MB-231 (triple-negative) and its *NAT1* knockout (KO) counterparts were analyzed and compared.
- The resulting lists of proteins that are differentially abundant in *NAT1* KO cells (compared to the parental cells) are useful in understanding the molecular and cellular processes that NAT1 plays in breast cancer cells.
- These data can 1) directly benefit researchers studying the novel link between NAT1 and breast cancer, and 2) others in the breast cancer field, for they suggest biological pathways that may be targeted for cancer therapeutics.

- The current dataset describes the biological processes, molecular functions, and cellular components that are differentially regulated in *NAT1* KO cells (e.g., antigen presentation and mitochondrial complex V). These data can be used to generate new hypotheses regarding the novel roles of NAT1 in breast cancer or to support and complement future experimental findings in this field.

1. Data Description

The current dataset including 18 raw data files has been deposited in the ProteomeXchange Consortium via the PRIDE [4] partner repository with the dataset identifier PXD035953 and 10.6019/PXD035953. It represents analyses of the data generated from proteomic analysis of parental MDA-MB-231 breast cancer cells and two *NAT1* KO cell lines. It is separated into two groups of figures: (1) analyses of proteins *upregulated* in *NAT1* KO cells (Figs. 1–4) and (2) analyses of proteins *downregulated* in *NAT1* KO cells (Figs. 5–8).

Gene Ontology (GO) biological processes, molecular functions, and cellular components that are significantly enriched (q [FDR] < 0.01) among proteins *upregulated* or *downregulated* in *NAT1* KO cells were identified using GOnet [1] and PANTHER Overrepresentation Test [2]. Fig. 1A shows false discovery rate (FDR)-adjusted p values (i.e., q values) for selected GO biological processes (GO terms and their IDs) that are significantly enriched among the *upregulated* proteins. The GO biological process terms were grouped (and separated using dotted lines and colors) according to a common or larger biological process/category they belong to (e.g., 'Antigen Presentation' or 'Proteolysis'). Fig. 1B and 1C show fold enrichments of selected, top GO molecular function and GO cellular component terms enriched among the *upregulated* proteins, respectively. Figs. 2–4 show flow charts that allow visualization of the inter-relationships between the GO terms enriched among the *upregulated* proteins. The flow charts were generated using GOrilla [5]. Figs. 2, 3, and 4 represent GO biological processes, molecular functions, and cellular components, enriched among the *upregulated* proteins, respectively. White, yellow and orange colors in Figs. 2–4 represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

Fig. 5A shows FDR-adjusted p values for selected GO biological processes, significantly enriched among the *downregulated* proteins. The GO biological process terms were grouped according to a common or larger biological process/category they belong to (e.g., 'Mitochondria' or 'Cell Cycle'). Fig. 5B and 5C show fold enrichments of selected, top GO molecular function and GO cellular component terms enriched among the *downregulated* proteins, respectively. Figs. 6–8 show flow charts that that allow visualization of the inter-relationships between the GO terms enriched among the *downregulated* proteins, based on analyses by GOrilla [5]. Figs. 6, 7, and 8 represent the GO biological processes, molecular functions, and cellular components enriched among the *downregulated* proteins, respectively. White, yellow and orange colors in Figs. 6–8 represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

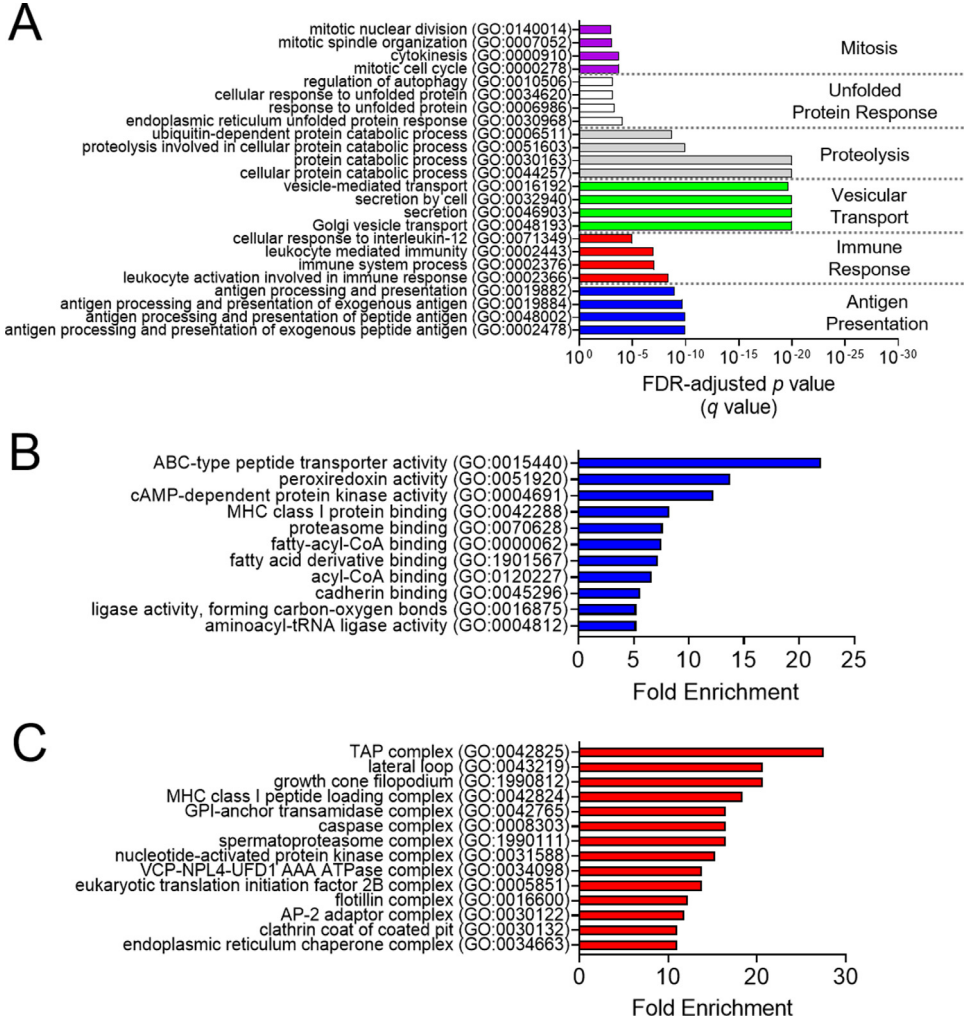

Fig. 1. Gene Ontology (GO) biological processes, molecular functions, and cellular components significantly enriched among proteins *upregulated* in *NAT1* KO MDA-MB-231 cells. Panel A shows false discovery rate (FDR)-adjusted *p* values (i.e., *q* values) for selected GO biological processes (GO terms and their IDs) that are significantly enriched among the *upregulated* proteins. The GO biological process terms were grouped (and separated using dotted lines and colors) according to a common or larger biological process/category they belong to (e.g., 'Antigen Presentation' or "Proteolysis"). Panels B and C show fold enrichments of selected, top GO molecular function and GO cellular component terms enriched among the *upregulated* proteins, respectively.

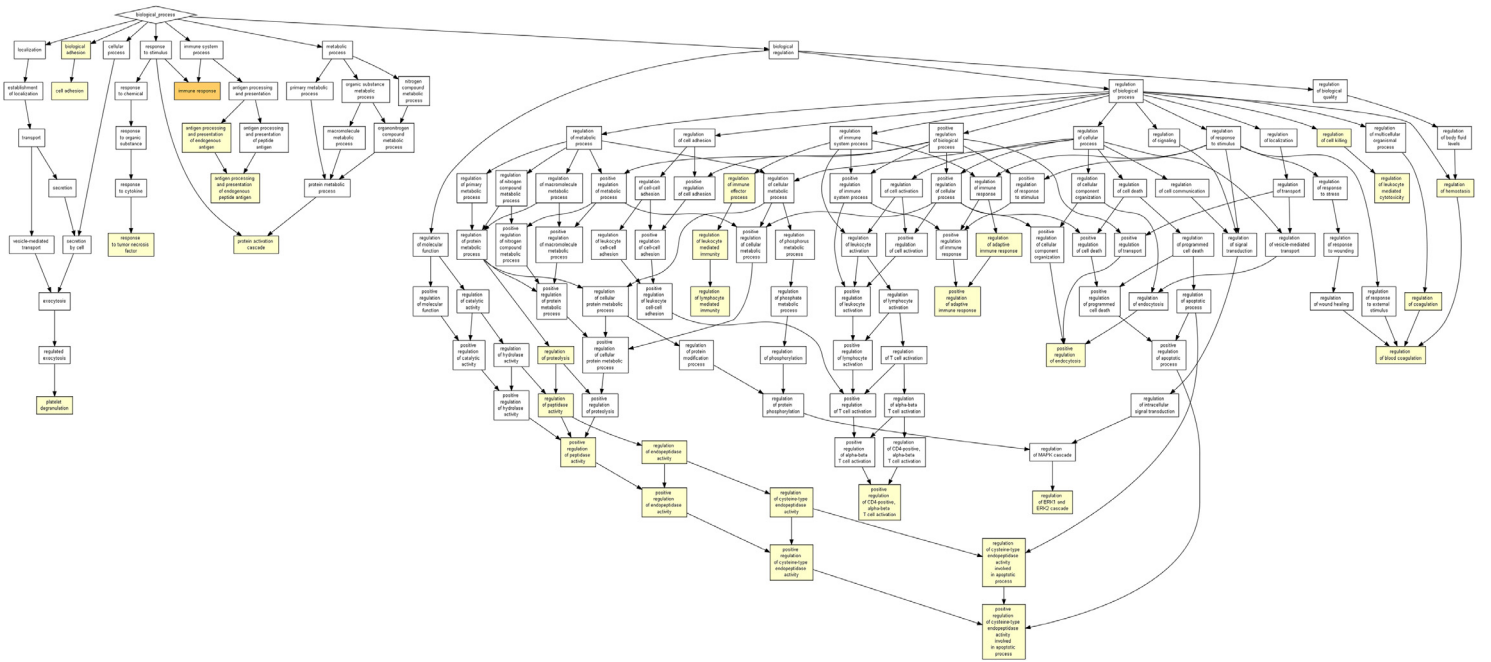

Fig. 2. Relationships and connections between the GO biological processes enriched among the proteins *upregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

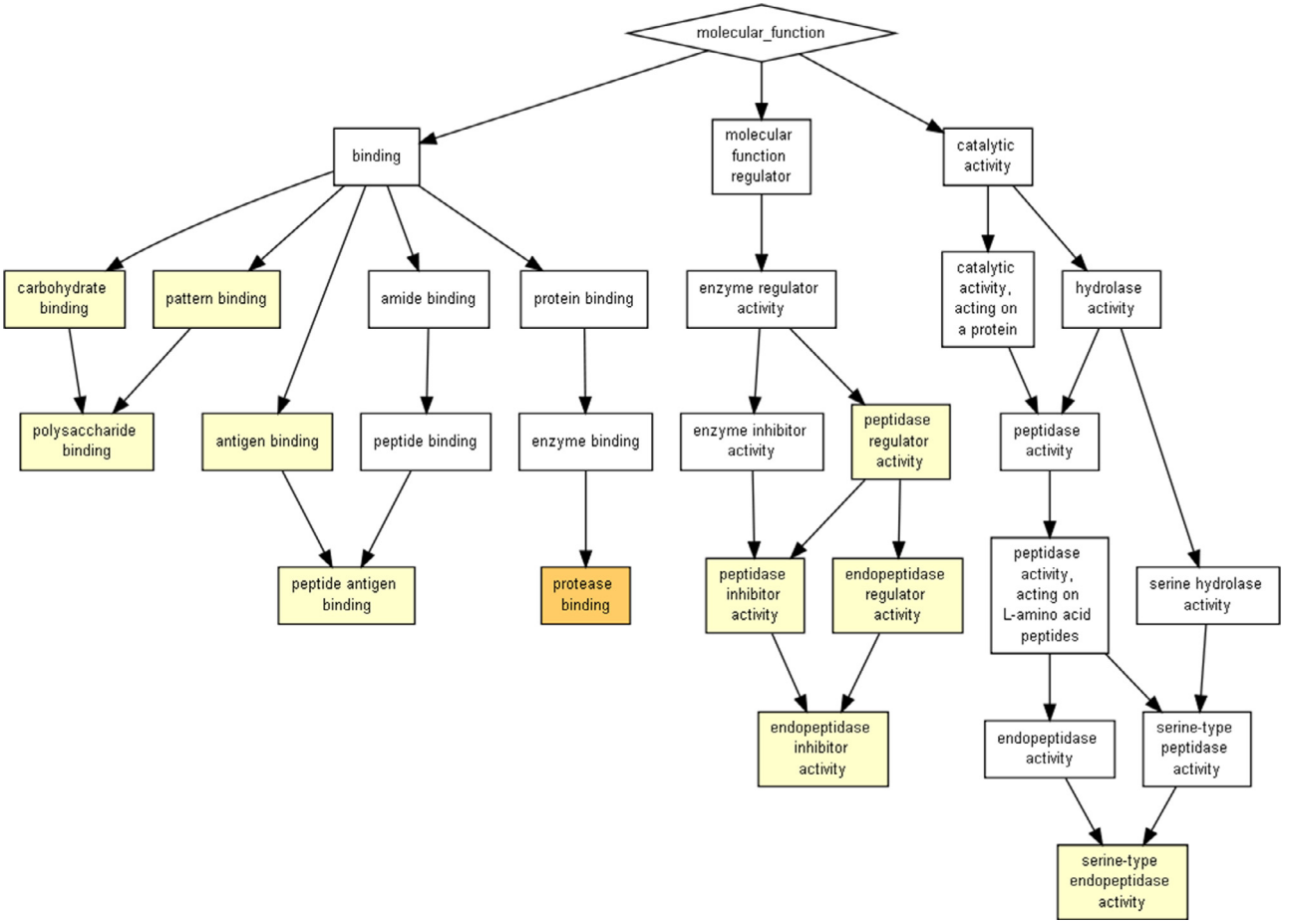

Fig. 3. Relationships and connections between the GO molecular functions enriched among the proteins *upregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

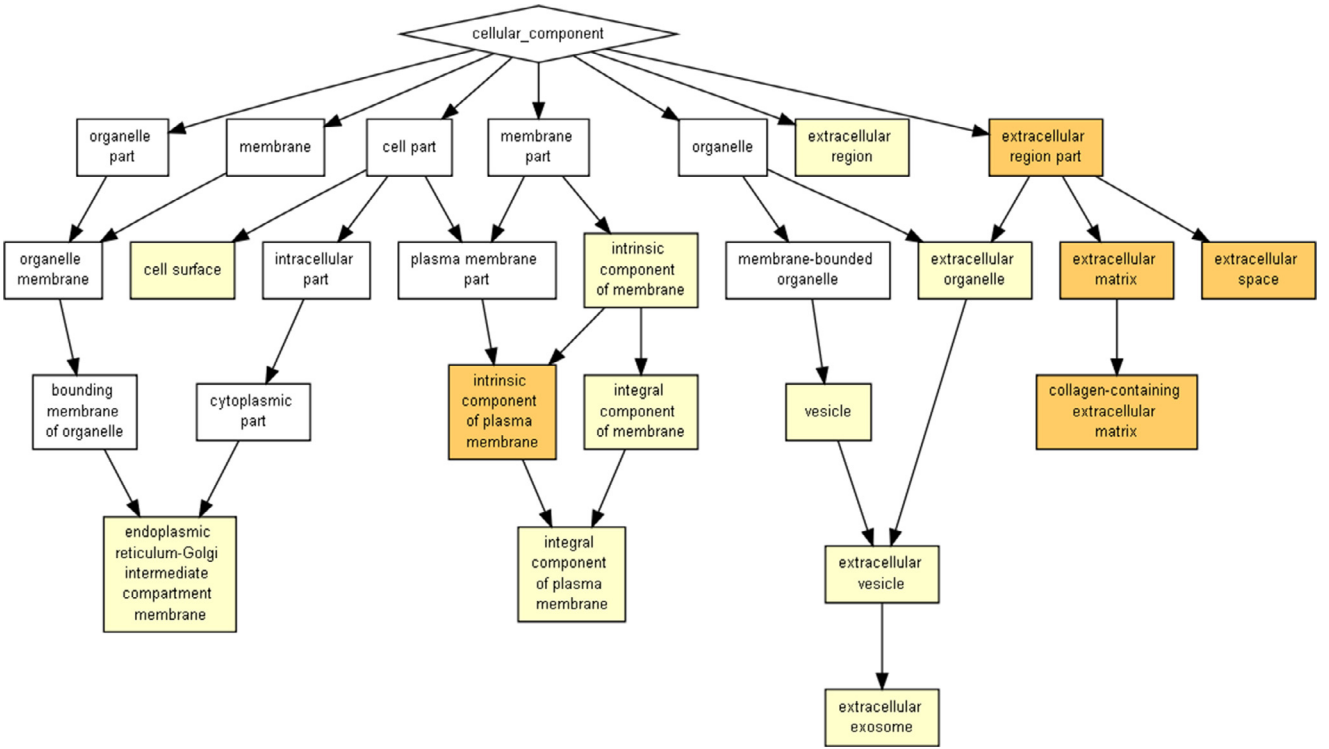

Fig. 4. Relationships and connections between the GO cellular components enriched among the proteins *upregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

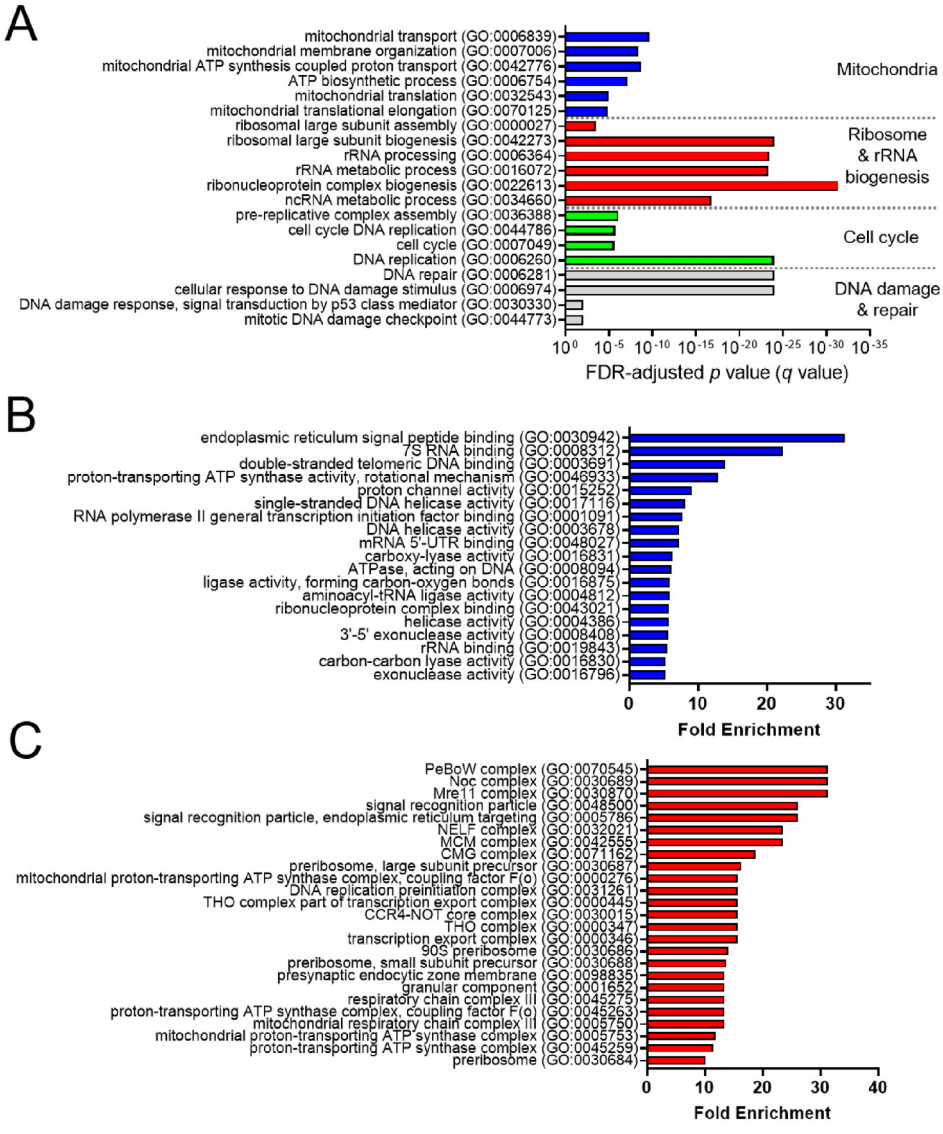

Fig. 5. Gene Ontology (GO) biological processes, molecular functions, and cellular components significantly enriched among proteins *downregulated* in *NAT1* KO MDA-MB-231 cells. Panel A shows false discovery rate (FDR)-adjusted *p* values (i.e., *q* values) for selected GO biological processes (GO terms and their IDs) that are significantly enriched among the *downregulated* proteins. The GO biological process terms were grouped (and separated using dotted lines and colors) according to a common or larger biological process/category they belong to (e.g., 'Mitochondria' or 'Cell Cycle'). Panels B and C show fold enrichments of selected, top GO molecular function and GO cellular component terms enriched among the *downregulated* proteins, respectively.

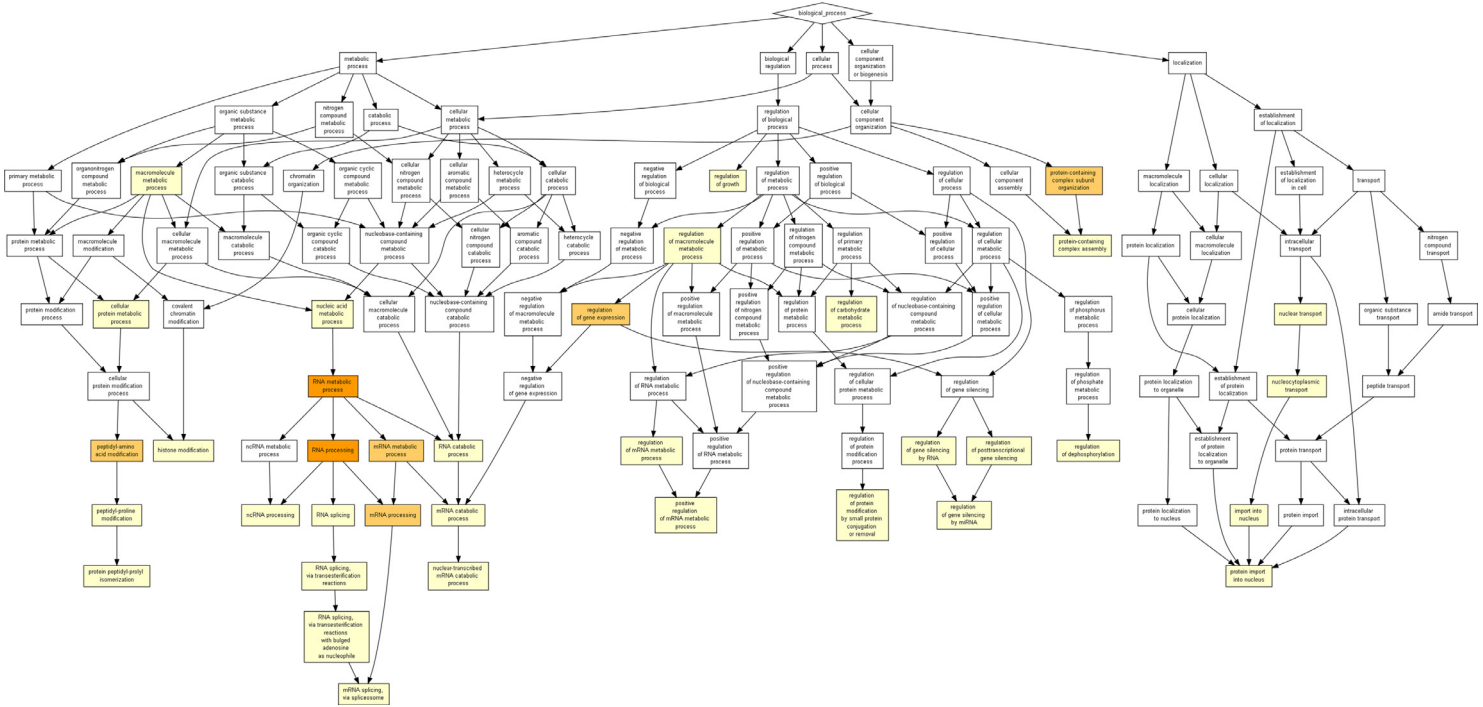

Fig. 6. Relationships and connections between the GO biological processes enriched among the proteins *downregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

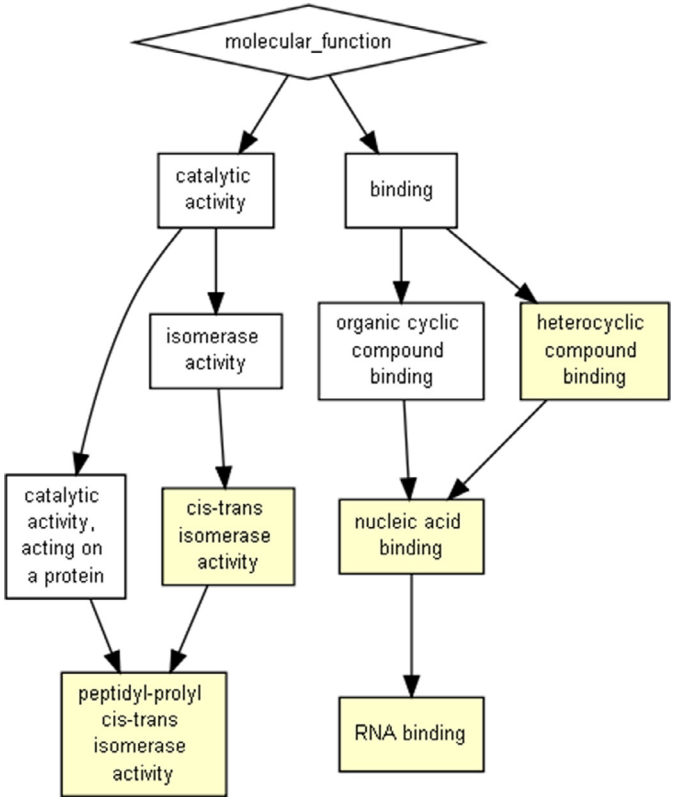

Fig. 7. Relationships and connections between the GO molecular functions enriched among the proteins *downregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

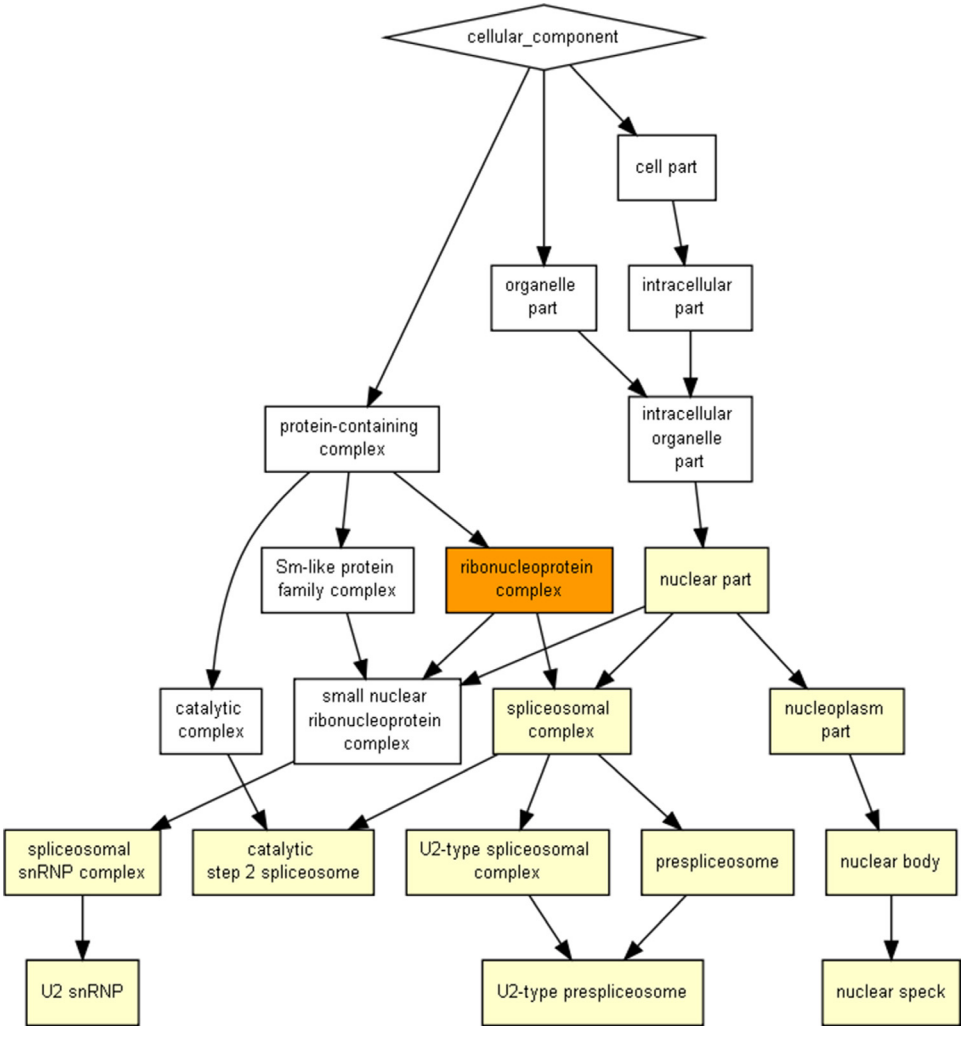

Fig. 8. Relationships and connections between the GO cellular components enriched among the proteins *downregulated* in *NAT1* KO MDA-MB-231 cells. The chart was generated using GOrilla [5]. White, yellow and orange colors in represent p values of $<10^{-3}$, 10^{-3} to 10^{-5}, and 10^{-5} to 10^{-7}, respectively.

2. Experimental Design, Materials and Methods

2.1. Cell Culture

Two different *NAT1* knockout (KO) MDA-MB-231 cell lines were generated using CRISPR/Cas9 technology and two unique guide RNAs. Generation and characterization of *NAT1* KO MDA-MB-231 cell lines have been previously described [6]. Parental and *NAT1* KO MDA-MB-231 cells were cultured in DMEM media (Cytiva) containing glucose (4.5 g/L), fetal bovine serum (10%), L-glutamine (4 mM), pyruvate (1 mM) and pen/strep (1%). Cells were cultured at 37 °C with 5% CO_2 in a humidified incubator.

3. Proteomic Analysis

Proteomic studies were conducted using a TMT-labeling approach as previously described [5] to compare the effects of CRISPR/Cas9 KO of NAT1 activity on wild-type and two *NAT1* KO MDA-MB-231 cell lines ($n = 3$ technical replicates per condition). Cells were lysed in 2% SDS dissolved in 100 mM Tris HCl, pH 8.5 and supplemented with 10 mM sodium butyrate and 1X HALT™ (Thermo Fisher) protease inhibitors. Protein concentrations estimated using a detergent compatible assay (BioRad, Cat. No. 500-0116) against a bovine serum albumin standard curve. Protein lysates (200 μg) were reduced with tris(2-carboxyethyl)phosphine (TCEP) (Thermo Fisher), denatured with 8 M urea and alkylated with iodoacetamide followed by centrifugation through a high molecular weight cutoff centrifugal filter (MilliporeSigma; 10k MWCO) [7]. After digestion with recombinant LysC and then sequencing grade Trypsin (Promega, Madison, WI), the digested proteins were collected and quantified by Nanodrop (Thermo Fisher) A205 nm prior to concentration adjustment to 0.5 μg/μL using 100 mM triethylammonium bicarbonate. Protein digested samples (50 μg) were labeled with tandem mass tag (TMT) using TMT10plex™ Isobaric Label Reagent Set (Thermo Fisher); samples were then concentrated and desalted with Oasis HLB Extraction cartridges (Waters Corporation) using a modified protocol for extraction of the digested peptides [8,9]. Samples were then subjected to high pH reversed phase separation with fraction concatenation on a Beckman System Gold LC system supplemented with 126 solvent module and 166 detector in tandem with a BioRad Model 2110 Fraction Collector. A total of 18 concatenated fractions were collected and analyzed by liquid chromatography-mass spectrometry (LC-MS) to detect and quantify TMT-labeled peptides. Briefly, every high pH reversed phase fraction was dissolved in 50 µL solution of 2% v/v acetonitrile/0.1% v/v formic acid and 2 µL of each fraction was analyzed on EASY-nLC 1000 UHPLC system (Thermo Fisher) and an Orbitrap Elite – ETD mass spectrometer (Thermo Fisher). Peptides were loaded onto an Acclaim PepMap 100 75 µm x 2 cm, nanoViper (C18, 3 µm, 100 Å) trap prior to separation at 250 nl/min on an Acclaim PepMap RSLC 50 µm x 15 cm, nanoViper (C18, 2 µm, 100 Å) separating column (Thermo Fisher) with a 2 to 40% acetonitrile/0.1% formic acid gradient using an EASY n-LC UHPLC system (Thermo Fisher). Eluted peptides were introduced into the Orbitrap ELITE mass spectrometer using a Nanospray Flex source (Thermo Fisher) with an ion transfer capillary temperature of 225 °C and 1.6 kV spray voltage. An Nth Order Double Play was created in Xcalibur v2.2. Scan event one of the method obtained an FTMS MS1 scan (normal mass range; 60,000 resolution, full scan type, positive polarity, centroid data type) for the range 300–2000 m/z. Scan event two obtained FTMS MS2 scans (HCD activation, 1.0 m/z isolation width, normalized collision energy of 40.0, default charge state of 2, 0.1 msec activation time, normal mass range, 60,000 resolution, centroid data type) on up to 10 peaks that had a minimum signal threshold of 5,000 counts from scan event one. The lock mass option was enabled (0% lock mass abundance) using the 371.101236 m/z polysiloxane peak as an internal calibrant. Proteome Discoverer v1.4.1.14 was used to direct the Mascot v2.5.1 (Matrix Science Ltd, London, UK) and SequestHT (Thermo Fisher) searches of the raw data using the 5/12/2017 version of the UniprotKB Homo sapiens reference proteome reviewed canonical and isoform sequences. The search criteria included: enzyme specified was trypsin (maximum two missed cleavages; inhibition by P) with Carbamidomethyl(C) and TMT10plex (K, N-term) set as a static modifications and Oxidation(M) as dynamic. Fragment tolerance was 0.05 Da (monoisotopic) and parent tolerance was 50 ppm (monoisotopic). A Target Decoy PSM Validator node was included in the Proteome Discoverer workflow. The result files from Proteome Discoverer were loaded into Scaffold Q+S v4.4.5. Scaffold was used to calculate the false discovery rate using the Scaffold Local FDR and Protein Prophet algorithms. Peptides were accepted if the identification had probability greater than 99.9% and parent mass error within 2 ppm. Proteins were accepted if they had a probability greater than 99.9% and at least one peptide. Proteins were grouped into clusters to satisfy the parsimony principle. TMT purity correction factors were obtained from Thermo. Intensity based normalization of reporter ions was done using the median calculation type, unique peptides blocking level, individual spectrum references, and normalization between samples. The refer-

ence value was required for use in the Scaffold Q module, and the minimum dynamic range of reporter ions was set to 1%. The results were annotated with human gene ontology information from the Gene Ontology Annotations Database (ftp.ebi.ac.uk). This yielded 4890 proteins with a false discovery rate of 0.389%.

4. Data Analysis

Levels of protein species in *NAT1* KOs were transformed to $\log_2$ fold change relative to those in parental cells. MetaCore™ (https://clarivate.com/products/metacore/) software were used to analyze protein fold changes between cell lines and predict probable cellular pathways affected by NAT1 deletion. In addition to MetaCore™, we analyzed our data through PANTHER Classification System [2], GOnet [1], and Reactome (ver 65) [3,10]. These online algorithms provide gene function, ontology, pathways and statistical analysis tools.

For analyses of data by GOnet [1], the proteins upregulated in *NAT1* KO cells and those downregulated were analyzed separately via online algorithm available at https://tools.dice-database.org/GOnet/. Each list containing gene symbols of differentially regulated proteins was uploaded as a CSV file. Under 'GO namespace', 'biological_process,' 'molecular_function,' or 'cellular_component' was selected, and for analysis type, 'GO term enrichment' was selected. For 'q-value threshold' and 'Background,' default options were chosen, which were '* ($\leq$ 0.05)' and 'all annotated genes', respectively. The data was in a CSV format which was exported into a MS Excel file.

For analyses of data by the PANTHER Classification System [2], the proteins upregulated in *NAT1* KO cells and those downregulated were analyzed separately via online algorithm available at http://www.pantherdb.org/ (version released on 10/24/2021). Each list containing gene symbols ('ID list') of differentially regulated proteins was uploaded as a CSV file. Homo sapiens was selected under 'Select organism.' For type of analysis, 'Statistical enrichment test' was chosen, and subsequently 'PANTHER Protein Class' . The tabulated data output was then exported into a MS Excel file.

5. Data Presentation

Bar graphs showing *q* values or fold enrichments were created using GraphPad Prism (ver 9). Flow charts showing relationships between the significantly enriched GO terms were generated using GOrilla [5].

Ethics Statements

The manuscript adheres to ethics in publishing standards.

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Data Availability

Proteomic Analysis of Arylamine N-Acetyltransferase 1 Knockout Breast Cancer Cells (Original data) (ProteomeXchange via the PRIDE database).

CRediT Author Statement

Kyung U. Hong: Writing – original draft, Formal analysis, Validation, Visualization; **Jonathan Q. Gardner:** Investigation, Data curation, Formal analysis, Validation, Visualization, Project administration, Writing – review & editing; **Mark A. Doll:** Investigation, Data curation, Formal analysis, Validation, Visualization, Project administration, Writing – review & editing; **Marcus W. Stepp:** Investigation, Data curation, Formal analysis, Validation, Visualization, Project administration, Writing – review & editing; **Daniel W. Wilkey:** Investigation, Data curation, Formal analysis; **Frederick W. Benz:** Investigation, Data curation, Formal analysis; **Jian Cai:** Investigation, Data curation, Formal analysis; **Michael L. Merchant:** Conceptualization, Resources, Supervision, Project administration, Writing – review & editing; **David W. Hein:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Acknowledgments

This work was supported by the National Institutes of Health [NIEHS T32- ES011564, NCI R25- CA134283, NIEHS P30-ES030283 and NIGMS P20-GM113226 to D.W.H., and NIAAA P50-AA024337 and NIAAA R01-AA028436 to M.L.M.].

References

[1] M. Pomaznoy, B. Ha, B. Peters, GOnet: a tool for interactive gene ontology analysis, BMC Bioinform. 19 (2018) 470, doi:10.1186/s12859-018-2533-3.
[2] H. Mi, A. Muruganujan, J.T. Casagrande, P.D. Thomas, Large-scale gene function analysis with PANTHER classification system, Nat. Protoc. 8 (2013) 1551–1566, doi:10.1038/nprot.2013.092.
[3] A. Fabregat, K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein, H. Hermjakob, Reactome pathway analysis: a high-performance in-memory approach, BMC Bioinform. 18 (2017) 142, doi:10.1186/s12859-017-1559-2.
[4] Y. Perez-Riverol, J. Bai, Bandla C, S. Hewapathirana, D. García-Seisdedos, S. Kamatchinathan, D. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J.A. Vizcaíno, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, Nucl. Acids Res. 50 (D1) (2022) D543–D552 (PubMed ID: 34723319).
[5] K.U. Hong, J.Q. Gardner, M.A. Doll, M.W. Stepp, D.W. Wilkey, F.W. Benz, J. Cai, M.L. Merchant, D.W. Hein, Proteomic analysis of arylamine *N*-acetyltransferase 1 knockout breast cancer cells: implications in immune evasion and mitochondrial biogenesis, Toxicol. Rep. 9 (2022) 1566–1573, doi:10.1016/j.toxrep.2022.07.010.
[6] M.W. Stepp, R.A. Salazar-González, K.U. Hong, M.A. Doll, D.W. Hein, *N*-acetyltransferase 1 knockout elevates acetyl coenzyme a levels and reduces anchorage-independent growth in human breast cancer cell lines, J. Oncol. 2019 (2019) 3860426, doi:10.1155/2019/3860426.
[7] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nat. Methods 6 (2009) 359–362, doi:10.1038/nmeth.1322.
[8] H. Keshishian, M.W. Burgess, M.A. Gillette, P. Mertins, K.R. Clauser, D.R. Mani, E.W. Kuhn, L.A. Farrell, R.E. Gerszten, S.A. Carr, Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury, Mol. Cell Proteom. 14 (2015) 2375–2393, doi:10.1074/mcp.M114.046813.
[9] G.S. McDowell, A. Gaun, H. Steen, iFASP: combining isobaric mass tagging with filter-aided sample preparation, J. Proteome Res. 12 (2013) 3809–3812, doi:10.1021/pr400032m.
[10] J. Griss, G. Viteri, K. Sidiropoulos, V. Nguyen, A. Fabregat, H. Hermjakob, ReactomeGSA - efficient multi-omics comparative pathway analysis, Mol. Cell Proteom. 19 (2020) 2115–2125, doi:10.1074/mcp.TIR120.002155.