



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data Article

Dataset of target mass spectromic proteome profiling for human chromosome 18



Ekaterina V. Ilgisonis*, Arthur T. Kopylov, Victor G. Zgoda

Orekhovich Institute of Biomedical Chemistry, Moscow, Russia

ARTICLE INFO

Article history:

Received 8 June 2016

Received in revised form

30 June 2016

Accepted 19 July 2016

Available online 26 July 2016

ABSTRACT

Proteome profiling is a type of quantitative analysis that reveals level of protein expression in the sample. Proteome profiling by using selected reaction monitoring is an approach for the Chromosome-centric Human Proteome Project (C-HPP). Here we describe dataset generated in the course of the pilot phase of Russian part of C-HPP, which was focused on human Chr 18 proteins. Proteome profiling was performed using stable isotope-labeled standards (SRM/SIS) for plasma, liver tissue and HepG2 cells. Dataset includes both positive and negative results of protein detection.

These data were partly discussed in recent publications, “Chromosome 18 Transcriptome Profiling and Targeted Proteome Mapping in Depleted Plasma, Liver Tissue and HepG2 Cells” [1] and “Chromosome 18 transcriptoproteome of liver tissue and HepG2 Cells and targeted proteome mapping in depleted plasma: Update 2013” [2], supporting the accompanying publication “State of the Chromosome 18-centric HPP in 2016: Transcriptome and Proteome Profiling of Liver Tissue and HepG2 Cells” [3], and are deposited at the ProteomeXchange via the PASSEL repository with the dataset identifier PASSEL: PASS00697 for liver and HepG2 cell line.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail address: ilgisonis.ev@gmail.com (E.V. Ilgisonis).

Specifications Table

Subject area	Biology
More specific sub- ject area	Targeted mass-spectrometric proteome profiling of liver and HepG2 cell line
Type of data	Figure, table, raw files (.d), skyline files (.sky)
How data was acquired	Proteome profiling was performed using stable isotope-labeled standards (SRM/SIS) for liver tissue and HepG2 cells
Data format	Raw
Experimental factors	The trypsin digestion was used.
Experimental features	Digested samples were separated using the HPLC Agilent 1290 system including pump and autosampler. Internal Standard were produced using Overture (Protein Technologies, USA) or Hamilton Microlab STAR devices. The quantitative SRM analysis was performed using Agilent 6495 Triple Quadrupole (Agilent, USA) equipped with Jet Stream ionization source.
Data source location	Institute of Biomedical Chemistry, Moscow, Russia
Data accessibility	Data is available within this article and at the ProteomeXchange via PASSEL (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD004407).

Value of the data

- This data characterizes the diversity of chromosome 18 protein species in liver tissue and HepG2 cell line using SRM.
- This data could be of interest to laboratories studying protein reference levels and cross-tissue biological variability of proteome.
- This data could be useful for protein, peptide and transition selection for SRM-assay development.
- Dataset may be used as a test for automated SRM-data processing algorithms.

1. Data

This dataset describes conditions of liver tissue and HepG2 cell line proteome profiling. Targeted protein list included 268 proteins of chromosome 18. Data were automatically processed to quantify proteins in the biosample. Dataset includes raw data, transition list, skyline files and sample preparation instructions, available in PASSEL, 2 figures and [Supplementary table](#) with protein copy numbers in liver tissue and HepG2 cell line.

2. Experimental design, materials and methods

2.1. Sample preparation

The trypsin digestion of liver tissue and cell lysates was performed as described in Ponomarenko et al. [2].

2.2. Peptide synthesis

The peptides were produced using the SOLiD-phase peptide synthesis on the Overture (Protein Technologies, USA) or Hamilton Microlab STAR devices according to the method published in Hood

et al. [4]. The isotope-labeled leucine (Fmoc-Leu-OH-13C6,15N) was used for isotope-labeled peptide synthesis instead of the unlabeled leucine (Fmoc-Leu-OH) [5].

2.3. Transition list

List of peptides for 268 chromosome 18 proteins was generated manually using data about occurrence of proteotypic peptides from proteomic repositories GPMdb, ProteinAtlas and PRIDE and MaRiMba-criteria (protocol was described earlier in Supplementary note 2, Zgoda et al. [1]). For each protein one “best-flyer” peptide was chosen. For each peptide 3 the most intensive transitions [6] were chosen using previous research results.

All 268 peptides were distributed over 3 SRM-assays (A–C) in equal parts according to their calculated retention time to avoid interference.

LC-SRM Analysis was held as described earlier in Supplementary note 2, Zgoda et al. [1]. Each SRM experiment was repeated in 3 technical runs. Each transition peak was characterized with the following variables: retention time, peak height, SIS/endogenous peak area ratio. No manual inspection for to find transitions that were similar to those in the target peptides or to reveal detected peptides was held.

2.4. How to use data

Dataset is represented by several file types (Fig. 1). For re-using of the dataset and extraction relevant information from it one can install freely-available and open source Windows client application Skyline [7] (for each biosample there is one skyline file, including transitions and technical runs info). It provides an opportunity to open raw data, visualize (Fig. 2) and analyze SRM data. Besides it is possible to use proprietary software (Agilent MassHunter Workstation Software) for data visualization.

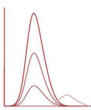



File type	HepG2 cell line	Liver tissue
 Raw data from the instrument (.d format)	1 sample, 3 technical runs for each SRM-assay (A, B, C)	1 sample, 3 technical runs for each SRM-assay (A, B, C)
 Skyline files (sky)	1 for sample, includes 268 peptides, 268 SIS, 3 technical runs for each SRM-assay (A, B, C)	1 for sample, includes 268 peptides, 268 SIS, 3 technical runs for each SRM-assay (A, B, C)
 Transition list (merged assays A,B,C)	3 SRM assays (A, B, C) for 268 endogenous peptides and 268 SIS	3 SRM assays (A, B, C) for 268 endogenous peptides and 268 SIS
 Sample preparation protocol	✓	✓

Fig. 1. Chromosome 18 proteome profiling dataset scheme.

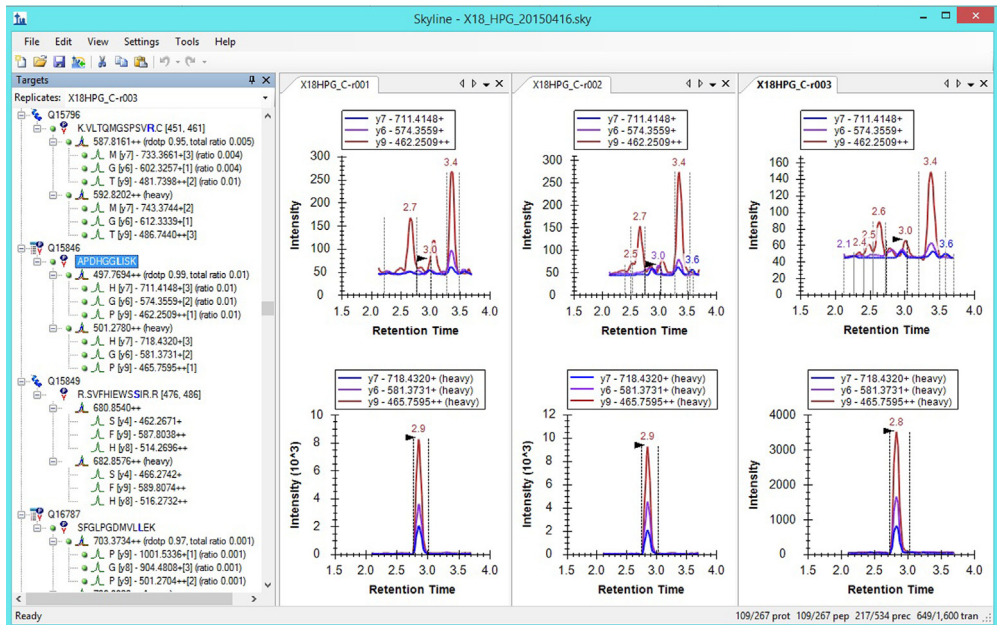


Fig. 2. Screenshot of raw data visualization for HepG2 cell line, SRM-assay C using Skyline.

All raw file are named using the following template:

- 1) Chromosome number: *X18*;
- 2) Sample type : *HLV* (liver tissue) or *HPG* (HepG2 cell line);
- 3) SRM-assay id: *A*, *B* or *C*; and
- 4) Technical run number: *r001*, *r002*, *r003*.

For example, raw file name *X18HPG_C-r002.d* (Fig. 2). Skyline files also are named using these keys.

2.5. Quantification

Calibration curves were obtained for each of the desired peptides using the mixtures of purified synthetic non-labeled peptides in the concentration range of 100–100 fmole/ μ l and its isotopically labeled standards (SIS) were added at the concentration of 2 fmole/ μ l. All calibration curves were linear in the range of 100–0.1 fmole/ μ l and showed the coefficient of linear regression equal to 0.95.

Prior to the sample processing, the performance of the LC-SRM platforms used was validated by obtaining the calibration curves of the corresponding set of SIS and synthetic non-labeled peptides. Moreover, after five LC-SRM runs we verified the relevance of calibration by analyzing one of the calibration peptide solution at 10 fmole/ μ l.

The detection limit was defined as the lowest concentration determined on the linear part of calibration curve. It varies for different peptides in the range from 100 amole/ μ l to 200 amole/ μ l.

Labeled (SIS)/target peptide peak area ratios were used to calculate the concentration of the targeted peptide in a sample. Peak area ratios were obtained using Skyline software.

$C_{\text{pept}} = C_{\text{lab}} \cdot S_{\text{pept}} / S_{\text{lab}}$ where C_{pept} – target peptide concentration, C_{lab} – labeled peptide (SIS) concentration (see *Quantification*), S_{pept} – area of target peptide peak, and S_{lab} – area of labeled peptide peak.

All calculated target peptide copy numbers for liver tissue and HepG2 are listed in the [Supplementary Table 1](#).

Conflicts of interest

None

Acknowledgments

This work was supported by Russian Science Foundation (#15-15-30041).

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.07.034>.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.07.034>.

References

- [1] V.G. Zgoda, A.T. Kopylov, O.V. Tikhonova, et al., Chromosome 18 transcriptome profiling and targeted proteome mapping in depleted plasma, liver tissue and HepG2 cells, *J. Proteome Res.* 12 (1) (2013) 123–134.
- [2] E.A. Ponomarenko, A.T. Kopylov, A.V. Lisitsa, et al., Chromosome 18 transcriptoproteome of liver tissue and HepG2 Cells and targeted proteome mapping in depleted plasma: update 2013, *J. Proteome Res.* 13 (1) (2014) 183–190.
- [3] E.V. Poverennaya, A.T. Kopylov, E.A. Ponomarenko, et al., State of the chromosome 18-centric HPP in 2016: transcriptome and proteome profiling of liver tissue and HepG2 cells, *J. Proteom Res.* (2016) (submitted for publication).
- [4] C.A. Hood, G. Fuentes, H. Patel, K. Page, M. Menakuru, J.H. Park, Fast conventional Fmoc solid-phase peptide synthesis with HCTU, *J. Pept. Sci.* 14 (1) (2008) 97–101.
- [5] D. Fekkes, State-of-the-art of high-performance liquid chromatographic analysis of amino acids in physiological samples, *J. Chromatogr. B. Biomed. Appl.* 682 (1) (1996) 3–22.
- [6] C. Ludwig, M. Claassen, A. Schmidt, R. Aebersold, Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry, *Mol. Cell. Proteom.* [Internet] 11 (3) (2012) (M111.013987).
- [7] B. MacLean, D.M. Tomazela, N. Shulman, et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics* [Internet] 26 (7) (2010) 966–968.