Data Article

Protein dataset of immortalized keratinocyte HaCaT cells and normal human keratinocytes

A.L. Rusanov, D.D. Romashin, V.G. Zgoda, T.V. Butkova*, N.G. Luzgina

V. N. Orekhovich Institute of Biomedical Chemistry, Russian Federation

A R T I C L E   I N F O

Article history:
Received 23 October 2020
Revised 5 February 2021
Accepted 10 February 2021
Available online 13 February 2021

Keywords:
Normal human keratinocytes
Keratinocytes HaCaT
Proteomic analysis
Experimental model

A B S T R A C T

Learning of the molecular mechanisms of the pathological processes development in the normal human keratinocytes (NHK) are difficult. Immortalized keratinocytes HaCaT are often used as an analogue of NHK since they have a number of advantages over the latter - they do not require the presence of growth and differentiation factors in the medium, have unlimited potential for proliferation, demonstrate stable phenotype regardless of the number of passages [1]. Taking into account the properties and characteristics of the HaCaT line, these cells can be considered as a promising experimental model for research of various physiological processes occurring in human keratinocytes. However, to understand the limitations of such an experimental model, a detailed comparative characterization of HaCaT and NHK is required, which can be obtained by carrying out its proteomic analysis.

In this article we present datasets obtained through the high-throughput shotgun proteomics analysis of normal human keratinocytes and immortalized HaCaT keratinocytes. As a protocol for proteomic profiling of cells, we used the approach of obtaining LC-MS / MS measurements followed by their processing with Progenesis LC-MS software (Nonlinear Dynamics Ltd.). The mzML files were deposited to the Mendeley Data.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications Table

| Subject | Biology  |
|---------|---------|
| Specific subject area | Biochemistry, omics analysis, Biotechnology |
| Type of data | Table |
| How data was acquired | Liquid chromatography-tandem mass spectrometric analysis was carried out using Q Exactive high-resolution mass spectrometer (Thermo Fisher Scientific, USA) coupled with an Ultimate 3000 Nano-flow HPLC system (Thermo Fisher Scientific, USA) |
| Data format | mzML, filtered |
| Parameters for data collection | LC/MS-MS spectra were collected for five biological replicates of formal human keratinocytes and four biological replicates of immortalized HaCaT cells. Three technical replicates were made for each biological replicate |
| Description of data collection | - Cell cultivation.  
- Protein extraction.  
- LC-MS/MS analysis.  
- Data processing. |
| Data source location | V. N. Orekhovich Institute of Biomedical Chemistry, Moscow, Russia |
| Data accessibility | Data are available via Mendeley Data (http://dx.doi.org/10.17632/bpck4nwdv9.1). |

Value of the Data

• Dataset represents proteomes of samples from normal human keratinocytes as well as HaCaT cells which can be compared to reveal differences between them.  
• These data may be of value to the scientists involved in the development of skin models in vitro.  
• These data may be of interest to the researchers interested in the processes of differentiation of keratinocytes.  
• Protein profiles are available in the form of “*.mzML” and “*.txt” data that can be further processed by researchers using their own bioinformatics algorithms and analyzed together with their own data.

1. Data Description

The dataset contains “*.mzML” and “*.txt” files obtained through the high-throughput shotgun proteomics analysis of normal human keratinocytes and immortalized HaCaT keratinocytes. Data are available via Mendeley Data. Information about samples is presented in Table 1. Dataset covers 9 biological samples (see Table 2).

2. Experimental Design, Materials and Methods

2.1. Reagents

Acetonitrile, taurocholic acid sodium salt (TCA) and sodium chloride were from Merck (Germany). Formic acid was from ACROS Organics (USA). Modified trypsin was from Promega

| Table 1 |
|---------|
| Data of cell samples. |
| Parameter | NHK | HaCaT |
| Number of samples | 5 | 4 |
| Number of technical repeats per sample | 3 | 3 |
Table 2
Sample description.

| Sample ID                  | Files "".txt", "". mzML"" | Size of"". mzML", MB | Type of set |
|----------------------------|--------------------------|---------------------|-------------|
| HaCaT_Sample1_TR1. mzML    | mzML                     | 261                 | HaCaT       |
| HaCaT_Sample1_TR2. mzML    | mzML                     | 255                 | HaCaT       |
| HaCaT_Sample1_TR3. mzML    | mzML                     | 262                 | HaCaT       |
| HaCaT_Sample2_TR1. mzML    | mzML                     | 259                 | HaCaT       |
| HaCaT_Sample2_TR2. mzML    | mzML                     | 316                 | HaCaT       |
| HaCaT_Sample2_TR3. mzML    | mzML                     | 265                 | HaCaT       |
| HaCaT_Sample3_TR1. mzML    | mzML                     | 264                 | HaCaT       |
| HaCaT_Sample3_TR2. mzML    | mzML                     | 265                 | HaCaT       |
| HaCaT_Sample3_TR3. mzML    | mzML                     | 264                 | HaCaT       |
| HaCaT_Sample4_TR1. mzML    | mzML                     | 268                 | HaCaT       |
| HaCaT_Sample4_TR2. mzML    | mzML                     | 308                 | HaCaT       |
| HaCaT_Sample4_TR3. mzML    | mzML                     | 265                 | HaCaT       |
| proteinGroups1_4.txt       | SEARCH                   | 2                   | HaCaT       |
| NHK_Sample1_TR1. mzML      | mzML                     | 246                 | NHK         |
| NHK_Sample1_TR2. mzML      | mzML                     | 254                 | NHK         |
| NHK_Sample1_TR3. mzML      | mzML                     | 249                 | NHK         |
| NHK_Sample2_TR1. mzML      | mzML                     | 261                 | NHK         |
| NHK_Sample2_TR2. mzML      | mzML                     | 263                 | NHK         |
| NHK_Sample2_TR3. mzML      | mzML                     | 261                 | NHK         |
| NHK_Sample3_TR1. mzML      | mzML                     | 243                 | NHK         |
| NHK_Sample3_TR2. mzML      | mzML                     | 247                 | NHK         |
| NHK_Sample3_TR3. mzML      | mzML                     | 252                 | NHK         |
| NHK_Sample4_TR1. mzML      | mzML                     | 245                 | NHK         |
| NHK_Sample4_TR2. mzML      | mzML                     | 247                 | NHK         |
| NHK_Sample4_TR3. mzML      | mzML                     | 249                 | NHK         |
| NHK_Sample5_TR1. mzML      | mzML                     | 276                 | NHK         |
| NHK_Sample5_TR2. mzML      | mzML                     | 295                 | NHK         |
| NHK_Sample5_TR3. mzML      | mzML                     | 276                 | NHK         |
| proteinGroups5_9.txt       | SEARCH                   | 2                   | NHK         |

(USA). Tris-(2-carboxyethyl)-phosphine (TCEP), methanol, trifluoroacetic acid (TFA) were from Fluka (Germany). DMEM:F:12 medium (1:1), EpiLife™ medium, GlutaMAX, fetal bovine serum, PBS, bovine pituitary extract, human epidermal growth factor and penicillin/streptomycin were obtained from Gibco, USA.

2.2. Cell cultivation

NHK were cultured in EpiLife™ medium supplemented with bovine pituitary extract (0,2% v/v), human epidermal growth factor (0,2 ng/mL), penicillin/streptomycin (100 UI/mL and 100 μg/mL) and 1% GlutaMAX. HaCaT cells were cultured in DMEM:F:12 medium supplemented with 1% GlutaMAX, penicillin/streptomycin (100 UI/mL and 100 μg/mL) and 10% v/v fetal bovine serum. Before the experiment, HaCaT cells were harvested by trypsinization and plated in the same culture medium as used for NHK. Cells were grown in 25 cm² tissue treated flasks (Corning, USA). Culture medium was replaced by fresh every other day in both cell lines. After reaching confluency of 75% cells were washed three times with PBS and then harvested mechanically with cell scrapers (Corning, USA).

2.3. Protein extraction

After harvesting the cells pellet was lysed with 100 μL lysis buffer (4% SDS in PBS, pH 7.4). Cell lysates were incubated on an orbital shaker at room temperature for 20 min with subsequent 5 min incubation at 95 °C. Then, the samples were cooled down at room temperature and
sonicated using Bandelin 2070 (Bandelin, Germany) as below: three 20 s cycles at 90% power. Protein precipitation was performed with methanol-chloroform method [2]. Briefly, 400 μL of methanol were added to 100 μL of each sample, the reaction mixtures were shortly vortexed, then 100 μL of chloroform were added. 300 μL of ddH2O were added to the mixtures, which after samples were mixed by vortexing and centrifuged at 14,000 g for 2 min. The upper aqueous layer was discarded and 400 μL of methanol were added. Samples were vigorously shaken and centrifuged for 5 min at 14,000 g. Methanol was carefully pipetted off and the peptide pellets were speed-dried for 5 min using Eppendorf Concentrator 5301 (Eppendorf, Germany) at 45 °C.

2.4. Sample preparation for Ms analysis

Protein concentration was determined with BCA assay (Pierce™ BCA Protein Assay Kit) following the manufacturer’s instructions. Aliquots of each sample containing 50 μg of protein extracts were dissolved in 20 μL denaturation buffer (5 M urea, 1% TCA, 15% acetonitrile, 50 mM phosphate buffer pH 6.3, 300 mM sodium chloride). Protein reduction was performed by adding 5 μL of 25 mM TCEP in 0.1 M ammonium bicarbonate followed by 45 min incubation at room temperature, then 5 μL of 300 mM IAA in 0.1 M ammonium bicarbonate was added. The reaction mixtures were incubated in the dark for 30 min. Remaining IAA was quenched by adding 5 μL of 300 mM DTT in 0.1 M ammonium bicarbonate. Subsequently, the reaction mixtures were diluted up to 200 μL with 0.1 M ammonium bicarbonate. In-solution digestion was performed by adding trypsin in enzyme:protein ratio 1:50. Samples were digested overnight while shaking at 37 °C. Then samples were centrifuged at 10 °C and 12,000 × g for 10 min, after which the supernatant was collected and cleaned up with C18 ZipTip according to manufacturer’s instructions [3]. In brief, Zip-Tip column was washed with 0.1% trifluoroacetic acid (TFA) in acetonitrile and equilibrated twice with 0.1% TFA in ddH2O, then the samples were passed through the Zip-Tips repeatedly by pipetting. The columns were washed three times with 0.1% TFA and 5% methanol in ddH2O. The peptides were eluted by 70% acetonitrile with 0.1% formic acid.

2.5. LC-MS/MS analyses

Prior to separation the peptides were concentrated on an Acclaim μ-Precolumn (0.5 mm × 3 mm, particle size 5 μm, inner diameter 75 μm; Thermo Scientific, USA) by direct loading of 1 μg of peptide dissolved in a volume of 1–4 μL of 0.1% formic acid. The procedure was performed in the isocratic mode of Mobile Phase C (2% acetonitrile, 0.1% formic acid) at flow rate of 10 μL/min for 4 min.

Then the peptides were separated with high-performance liquid chromatography (HPLC, Ultimate 3000 Nano LC System, Thermo Scientific, Rockwell, IL, USA) using a 15-cm long C18 column (Acclaim® PepMapTM RSLC with inner diameter of 75 μm, Thermo Fisher Scientific, USA) followed by elution in a gradient mode. The gradient was formed by the Mobile Phase A (0.1% formic acid) and buffer B (80% acetonitrile, 0.1% formic acid) at flow rate of a 0.3 μL/min. The column was equilibrated to Mobile Phase A for 12 min, then buffer B concentration was linearly increased from 5 to 35% for over 95 min. Then, the concentration of buffer B was linearly increased to reach concentration of 99% for 6 min. The columns were flushed for 10 min with 99% buffer B and repeatedly equilibrated with buffer A for 7 min.

Mass spectrometric analysis was performed at least in three technical repeats with a Q Exactive HF-X mass spectrometer (Q Exactive HF-X Hybrid Quadrupole-OrbitrapTM Mass spectrometer, Thermo Fisher Scientific, USA) in accordance with the following parameters: the capillary temperature was 240 °C and ionizing voltage was 2.1 kV. Mass spectra were acquired at the 300–1500 m/z range at resolution of 120,000 (MS). Tandem mass spectra of fragments were acquired in the range from 140 m/z to 2000 m/z (the precise m/z value was determined by the charge state of the precursor) at a resolution of 15,000 (MS/MS).
The maximum integration time was 50 ms and 110 ms for precursor and fragment ions, respectively. AGC target for precursor and fragment ions was set to $1 \times 10^6$ and $2 \times 10^5$, respectively. An isolation intensity threshold of 50,000 counts was determined for precursor's selection and up to top 20 precursors were chosen for fragmentation with high-energy collisional dissociation (HCD) at 29 NCE. Precursors with a charge state of +1 and more than +5 were rejected and all measured precursors were dynamically excluded from triggering of a subsequent MS/MS for 20 s.

The MS/MS spectra in a RAW format were processed in SearchGUI v.3.3.20 [4]. The RAW files were converted to mzML with ProteoWizard v.3 [5]. Obtained data were deposited to the Mendeley Data. Mass spectrometric measurements were performed using the equipment of “Human Proteome” Core Facility of the Institute of Biomedical Chemistry (Russia, Moscow).

### 2.6. Protein identification

Peak lists obtained from MS/MS spectra were identified using SearchGUI software v.3.3.20 [4]. Proteins were identified against a concatenated target/decoy version of the Homo sapiens Complement of the UniProtKB [6,7]. Decoy proteins sequences were obtained by reversing original target sequences using SearchGUI v.3.3.20 [4]. The identification parameters used for the database search were: enzyme specificity - Trypsin, maximum cleavages allowed – 2, MS1 and MS2 tolerances – 5.0 ppm and 0.01 Da, respectively, fixed modification - carbamidomethylation (Cys), variable modifications - N-terminal proteins acetylation and methionine oxidation (Met), false discovery rate estimated using the decoy hit distribution for Peptide Spectrum Matches (PSMs), peptides and proteins identification – 1.0%.

### Ethics Statement

The study was carried out in accordance with the World Medical Association Declaration of Helsinki. NHK were isolated from the human tissue material discarded after surgery, which was obtained with signed informed consent of the patients and treated anonymously.

### CRediT Author Statement

**Rusanov Alexander:** Conceptualization, Methodology, Writing – Original draft preparation; **Romashin Daniil:** Resources; **Zgoda Victor:** Investigation, Data curation; **Butkova Tatiana:** Data curation, Writing – Original draft preparation; **Luzgina Natalia:** Supervision, Writing - Review & Editing.

### 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.

### Acknowledgements

This analysis was funded by the Program for Basic Research of State Academies of Sciences for 2013–2020.
References

[1] I. Colombo, E. Sangiovanni, R. Maggio, C. Mattozzi, S. Zava, Y. Corbett, M. Fumagalli, C. Carlino, P.A. Corsetto, D. Scaccabarozzi, S. Calvieri, A. Gismondi, D. Taramelli, M. Dell’Agli, HaCaT cells as a reliable in vitro differentiation model to dissect the inflammatory/repair response of human keratinocytes, Mediators Inflamm 2017 (2017) 7435621.

[2] D. Wessel, U.I. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, Anal. Biochem. 138 (1984) 141–143.

[3] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (8) (2007) 1896–1906.

[4] H. Barsnes, M. Vaudel, A highly adaptable common interface for proteomics search and de novo engines, Proteome Res 17 (7) (2018) 2552–2555.

[5] R. Adusumilli, P. Mallick, Data conversion with ProteoWizard msConvert, Methods Mol Biol 1550 (2017) 339–368.

[6] J.R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties, Bioinformatics 33 (18) (2017) 2938–2940.

[7] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, Mol. Cell. Proteom. 4 (9) (2005) 1265–1272.