JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ORIGINAL ARTICLE | DOI: 10.5584/jiomics.v4i2.172

Proteomic analysis of low quantities of cellular material in the range obtainable from scarce patient samples

Katja Parapatics[1,2], Marie L. Huber[1], Dorit Lehmann[3], Christian Knoll[1], Giulio Superti-Furga[1], Keiryn L. Bennett[1], Elena L. Rudashevskaya*[1]

[1]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. kparapatics@cemm.oeaw.ac.at (K.P.), mhuber@cemm.oeaw.ac.at (M.L.H.), gsuperti@cemm.oeaw.ac.at (G.S.F.), kbennett@cemm.oeaw.ac.at (K.L.B.); [2]Department of Ophthalmology, Medical University of Vienna, Vienna, Austria. [3]Department of Vascular Biology and Thrombosis Research, Medical University of Vienna, Vienna, Austria.

Received: 16 June 2014 **Accepted:** 21 December 2014 **Available Online**: 28 January 2015

Abstract

The application of proteomics to patient material is increasingly widespread, however, a major shortcoming still are the number of cells or protein material that can be obtained. This study explores the lower limit of cell numbers that can be successfully analysed by liquid chromatography mass spectrometry to determine the protein expression profile that is specific to, and indicative of, the investigated cell type. The aim was to analyse an equivalent quantity of cellular material that can be obtained from, *e.g.*, a fine-needle aspiration biopsy (FNAB). Fifteen thousand and 30,000 cells from adherent (HEK293) and suspension (U937) cell lines were lysed under two different conditions: a 'native' and a denaturing buffer. To extend the study to clinical material, human whole PBMCs were also lysed under identical conditions. Proteins from 5,000 and 10,000 cells were analysed by both 1D and 2D-LC-MSMS on an LTQ Orbitrap XL mass spectrometer. In total, 3,219; 1,693 and 659 unique proteins were identified from HEK293, U937 and total PBMCs, respectively. Additionally, an iTRAQ 4-plex experiment was performed to determine the relative quantity of the proteins in the three cell types. In this study, we show that it is feasible to obtain a deep, yet cell-specific protein profile from a very low number of cultured and primary cells. This advancement will enable proteomic-profiling of cellular material from fine needle aspiration biopsies that ultimately can assist cytopathologists in the diagnosis of disease.

Keywords: mini-proteome; iTRAQ; FNAB; PBMC; HEK293; U937.

Abbreviations

1D-LC-MSMS: one-dimensional liquid chromatography tandem mass spectrometry; **1D-SDS-PAGE**: one-dimensional sodium dodecyl-sulphate polyacrylamide gel electrophoresis; **2D-LC-MSMS**: two-dimensional liquid chromatography tandem mass spectrometry; **CID**: collision-induced dissociation; **DMEM:** Dulbecco's modified eagle medium; **FASP**: filter-aided sample preparation; **FCS**: fetal calf serum; **FNAB**: fine-needle aspiration biopsy; **GO**: gene ontology; **HCD**: higher-energy collision-induced dissociation; **iTRAQ**: isobaric tag for relative and absolute quantitation; **LCM:** laser-capture microdissection; **LCMS**: liquid chromatography mass spectrometry; **LTQ**: linear trap quadrupole; **PBMC**: peripheral blood mononucleocytes; **PBS**: phosphate-buffered saline; **RPMI**: Roswell Park Memorial Institute medium; **SD**: standard deviation; **SDS**: sodium dodecylsulfate; **TEAB**: triethylammonium bicarbonate.

1. Introduction

Clinical samples from patients are extremely valuable and the amount of material and cells available is very low, often difficult to obtain and thus extremely precious. A fine-needle aspiration biopsy (FNAB) is a relatively painless and straightforward minor surgical diagnostic procedure used to extract cellular material from percutaneous masses and/or drain fluid -filled cysts. A very fine gauge needle is inserted and the cells and/or fluid are aspirated (suctioned) from the mass into the

***Corresponding author:** Institute of Pharmacology, Center for Physiology and Pharmacology, Medical University of Vienna, Waehringerstrasse 13A. 1090 Vienna, Austria. Tel: +43 1 40160 31366. E-Mail address: elena.rudashevskaya@meduniwien.ac.at

needle, stained, examined by a cytopathologist and a clinical diagnosis made. The number of cells obtained via FNAB is exceptionally low and proteomic analysis of the material in the aspirated sample is extremely challenging. Thus, only a limited number of studies exist on proteomic profiling of FNAB [1-5]. In a study by Rapkiewicz et al. [6], fine-needle aspiration samples from breast tumors were analyzed by quantitative protein microarray technology. These researchers calculated that the number of cells in frozen aspirate samples was in the range of 1,000–50,000 cells.

The analysis of 'core' proteomes by liquid chromatography mass spectrometry (LCMS) has been performed on large quantities (30-50 μg) of protein from cell lysates of immortalised, cultured cells [7-9]. Protein numbers ranged from approximately 1,800 [7] to between 2,000-4,000 [8] and up to 10,500 [9] non-redundant proteins from a single cell line. The historical increase in protein numbers reflects the combination of improved MS instrumentation technology and sample preparation methodology. With limited protein quantities, e.g., from clinical samples, older generation mass spectrometers were unable to delve deeply into the proteome of a low number of patient cells. The advent of the hybrid linear trap quadrupole (LTQ) and quadrupole only Orbitrap series of mass spectrometers has meant that it is now possible to analyse low numbers of cells from primary sources. More importantly, such analyses do not just result in the identification of the usual abundant, 'uninteresting' housekeeping proteins but actually lead to the identification of cell-specific proteins. Some key examples from the literature include: 3,800 proteins identified from an equivalent of 5,000 FACS-sorted colon stem cells [10]; single islets of Langerhans and ~10,000 laser-capture microdissected (LCM) mouse kidney glomeruli containing 2,000-4,000 and ~2,400 proteins, respectively [11]; 900-1,900 proteins from 250-10,000 FACS-sorted short-term culture melanoma cells [12, 13]; ~1,000 proteins from 3,000 LCM-derived breast carcinoma tumor cells [14]; and 3,600-4,400 proteins from 20,000 microdissected formalin-fixed, paraffin-embedded colon carcinoma cells [15].

The current study was designed to ascertain the depth that could be obtained from a low number of cells using our sample preparation approach and MS instrumentation. In particular, we were interested in establishing appropriate experimental conditions for a 'native' lysis buffer with the goal of analysing cellular proteomes from small quantities of material. The data generated would be used to assess the feasibility of a planned projection into routine proteomic analyses of FNAB. The 'native' lysis buffer, that was ultimately chosen, closely resembled buffers commonly used for investigating native protein complexes and protein interactions. Considering the medical impact and relevance for translational research, there is an increasing interest in investigating alterations in protein complex formation or pathway composition and also protein-drug interactions directly from clinical material [16]. In all cases, such experiments require cellular lysis conditions that maintain and preserve native protein conformations and interactions, i.e., non-denaturing. In parallel, denaturing conditions that utilise buffers containing urea can supplement the data by 'releasing' proteins from cellular compartments that are not completely accessible with a 'native', non-denaturing buffer alone. Furthermore, urea buffers are included in commercially-available kits for enrichment of post-translationally-modified (PTM) peptides (e.g., phosphotyrosine-enrichment). It has also been reported that urea-based buffers have a certain advantage over SDS-containing buffers in the analysis of cancer tumors [17].

The well-known and extensively-studied adherent HEK-293 cell line was chosen as a model, the methodology extended to a human macrophage suspension cell line (U937) and finally to readily-obtainable primary human total peripheral blood mononucleocytes (PBMCs) to mimic a clinical setting. Two lysis conditions were chosen and evaluated on the three different cell types: (i) a 'native' buffer consisting of 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1mM PMSF, pH 8.0; and (ii) a denaturing buffer comprised of 9 M urea, 20 mM HEPES, 1mM PMSF, pH 8.0. Fifteen thousand and 30,000 cells were lysed and the protein equivalent of 5,000 or 10,000 cells analyzed by one-dimensional liquid chromatography tandem mass spectrometry (1D-LC-MSMS). The peptide digests from 15,000 and 30,000 lysed cells were also separated by reversed-phase chromatography, 20 fractions collected and analyzed by LC-MSMS. Finally, to add relative quantitative information to the proteins identified from the different cell types, iTRAQ labeling coupled to 2D-LC-MSMS was used to compare the three different cell types lysed under two different conditions. Functional annotation analysis was utilised to evaluate the obtained protein lists to determine the specificity of the individual cellular proteomes.

2.Material and Methods

2.1 *Reagents*

Iodoacetamide, dithiothreitol (DTT), HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 1 M triethylammonium bicarbonate (TEAB), formic acid, urea, NaCl, EDTA (ethylenediaminetetraacetic acid), thiourea, DMEM, protease inhibitor cocktail, PMSF (Phenylmethanesulfonyl fluoride), ≥99.0% (SIGMA-Aldrich, St. Louis, MO); trypsin (Promega Corp., Madison, WI); RPMI, PBS, penicillin/streptomycin (PAA, Pasching, Austria), iTRAQ (ABI, Framingham, MA); FCS (Gibco, Grand Island, NY), NP-40 Alternative (CALBIOCHEM, San Diego, CA), Krypton™ Protein Stain (Thermo Scientific, Austria, Vienna).

2.2 *Collection of HEK293 and U937 cells*

HEK293 and U937 cells were grown as adherent (in DMEM medium) and suspension cultures (in RPMI medium) respectively. Each media was complemented with 10%

foetal calf serum (FCS) and 1% penicillin/streptomycin. Cells were harvested in posphate-buffered saline (PBS) solution and counted with a CASY® cell counting system (Roche Diagnostics, Rotkreuz, Switzerland). Multiple aliquots of 15,000 and 30,000 cells were collected in 1.5 mL eppendorf tubes and centrifuged at 1,800×g for 5 min at 4°C. The supernatant was removed and discarded; and the cell pellets stored at -80°C until required.

2.3 *Collection of peripheral blood mononucleocytes (PBMCs)*

Venous blood from a healthy donor was collected into BD Vacutainer tubes containing EDTA and diluted 1:1 in PBS. PBMCs were separated using LSM 1077 lymphocyte separation media (PAA Austria) by density gradient centrifugation at 2,100 r.p.m. for 30 min (Sorvall RT6000B) and washed with PBS by centrifugation for 10 min at 1,200 r.p.m. (Sorvall RT6000B). Contaminating erythrocytes were removed by incubation in a lysis buffer (containing 0.15 M NH_4Cl, 0.01 M $NaHCO_3$ and 0.1 mM EDTA, pH 7.2) for 10 min at 4°C and washing with PBS by 20 min centrifugation at 800 r.p.m. (Sorvall RT6000B), which also eliminated remaining platelets. PBMCs (4×10^7 cells) were counted with a Neubauer counting chamber (Karl Roth GmbH, Karlsruhe, Germany) and aliquots of 15,000 and 30,000 cells collected. Cells were centrifuged, the supernatant removed, and pellets frozen at -80°C and stored until required.

2.4 *Assessing lysis conditions for large quantities of cells*

HEK293 cells (8.7×10^6) were collected and lysed in 1 mL under the following six different conditions: (i) 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1 mM PMSF, pH 8.0 (no sonication). Cells were incubated for 30 min at 4°C; and then centrifuged at 20,000 × g for 15 min at 4°C. (ii) 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 1 mM PMSF, pH 8.0 (+ sonication). Cells were sonicated at 70% output with 3 bursts of 30 s; and cooled on ice for 30 s between each burst. The resultant lysates were centrifuged at 20,000 × g for 15 min at 4°C. (iii) 20 mM HEPES, 9 M urea, pH 8.0 (no sonication). Cells were incubated for 30 min at room temperature, with vortexing every 10 min, and centrifuged at 20,000 × g for 15 min at RT. (iv) 20 mM HEPES, 9 M urea, pH 8.0 (+ sonication). Cells were incubated for 30 min at room temperature, with vortexing every 10 min, and with sonication at 70% output with 3 bursts of 30 s; and centrifuged at 20,000 × g for 15 min at RT. (v) 20 mM HEPES, 7 M urea, 2 M thiourea, pH 8.0 (+ sonication). Cells were incubated for 30 min at RT, with vortexing every 10 min, and with sonication at 70% output with 3 bursts of 30 s; and centrifuged at 20,000 × g for 15 min at RT.

2.5 *Protein concentration measurement*

The total protein content of HEK293 cell lysates generated from 8.7×10^6 cells was determined by the Bradford assay (using bovine serum albumin as a protein standard).

2.6 *'Native' buffer lysis and in situ tryptic digestion*

Fifteen thousand and 30,000 HEK293, U937 and human total PBMC cells were individually lysed in 100 µL 'native' buffer: 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1 mM PMSF, pH 8.0 and incubated for 30 min at 4° C. Cell lysates were centrifuged at 20,000 × g for 15 min at 4° C and the protein extracts (supernatants) were collected. Laemmli buffer was added, and the samples were reduced and alkylated with dithiothreitol and iodoacetamide, respectively. 1D-SDS-PAGE of the lysates was performed on Novex Bis-Tris 4-12% gels at 200 V for 15 min only. Gels were stained with colloidal Coomassie blue. Stained gel regions containing the proteins were excised from the gel and digested in situ with trypsin at 37°C overnight. Tryptic digests were concentrated and purified by solid phase extraction (SPE) (UltraMicroSpin columns 3-30 µg capacity, Nest Group Inc., Southboro, MA, USA), the volumes reduced to approximately 2 µL in a vacuum centrifuge and reconstituted in 26 µL 5% formic acid for 1D-LC-MSMS triplicate analysis.

2.7 *Denaturing lysis and solution tryptic digestion*

Fifteen thousand and 30,000 HEK293, U937 and human total PBMC cells were individually lysed in 100 µL denaturing buffer: 20 mM HEPES, 9 M urea, pH 8.0 for 30 min at room temperature. Samples were vortexed every 10 min. Cell lysates were centrifuged at 20,000 × g for 15 min at RT and the protein extracts (supernatants) were collected. Samples were diluted with 100 mM TEAB to a final concentration of 1.4 M urea, reduced with dithiothreitol, alkylated with iodoacetamide and digested with 0.5 µg modified porcine trypsin at 37°C overnight. Tryptic digests were concentrated and purified by solid phase extraction (SPE) (UltraMicroSpin columns 3-30 µg capacity, Nest Group Inc., Southboro, MA, USA), the volumes reduced to approximately 2 µL in a vacuum centrifuge and reconstituted in 26 µL 5% formic acid for 1D-LC-MSMS triplicate analysis.

2.8 *iTRAQ derivatisation*

The tryptic digests from 30,000 cells lysed in either the 'native' or denaturing buffer were derivatised with the 4-plex iTRAQ reagent (ABI, Framingham, MA) [18] and labelled according to the instructions provided by the manufacturer. Two iTRAQ 4-plex experiments were prepared. One experiment was performed on cells lysed in the 'native' buffer. iTRAQ labels 114, 115, corresponded to 30,000 HEK293 cells; 30,000 U937 cells; and 116 and 117 corresponded each to 30,000 PBMC cells, respectively. In the other experiment, the same labeling was performed but on the tryptic digests from the denaturing buffer.

2.9 *Reversed-phase reversed-phase (RPRP) separation for 2D-LC-MSMS [19]*

Tryptic digests from the 'native' and denaturing individual cell lysates and also for the two iTRAQ experiments were concentrated and purified by SPE (UltraMicroSpin columns 3-30 µg capacity, Nest Group Inc., Southboro, MA, USA) and reconstituted in 23 µL 100 mM TEAB, pH 10, prior to injection onto a Phenomenex column (150 × 2.0 mm Gemini-NX 3 µm C18 110Å, Phenomenex, Torrance, CA, USA) on an Agilent 1200 series HPLC (Agilent Biotechnologies, Palo Alto, CA). Twenty and 40 fractions were collected for expression and quantitative proteomics, respectively. All fractions were acidified with 5 µL 5% formic acid and the volumes reduced to approximately 2 µL in a vacuum centrifuge. Samples were reconstituted to 26 µL with 5% formic acid and analysed as technical duplicates by LC-MSMS. Details of the methodology are as previously described [20].

2.10 *Liquid chromatography mass spectrometry*

Mass spectrometry was performed on a hybrid linear trap quadrupole (LTQ) Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA) using the Xcalibur version 2.0.7 (expression proteomics). iTRAQ experiments were analysed on a hybrid LTQ Orbitrap Velos (ThermoFisher Scientific, Waltham, MA) using the Xcalibur version 2.1.0.1140 (relative iTRAQ quantitation). The Orbitrap mass spectrometers were coupled to an Agilent 1200 HPLC nanoflow system (dual pump system with one precolumn and one analytical column) (Agilent Biotechnologies, Palo Alto, CA) via a nanoelectrospray ion source using liquid junction (Proxeon, Odense, Denmark). Solvents for LCMS separation of the digested samples were as follows: solvent A consisted of 0.4% formic acid in water and solvent B consisted of 0.4% formic acid in 70% methanol and 20% isopropanol. From a thermostatted microautosampler, 8 µL of the tryptic peptide mixture were automatically loaded onto a trap column (Zorbax 300SB-C18 5 µm, 5×0.3 mm, Agilent Biotechnologies, Palo Alto, CA) with a binary pump at a flow rate of 45 µL/min. 0.1% TFA was used for loading and washing the pre-column. After washing, the peptides were eluted by back-flushing onto a 16 cm fused silica analytical column with an inner diameter of 50 µm packed with C18 reversed phase material (ReproSil-Pur 120 C18-AQ, 3 µm, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The peptides were eluted from the analytical column with a 27 minute gradient ranging from 3 to 30% solvent B, followed by a 25 minute gradient from 30 to 70% solvent B and, finally, a 7 minute gradient from 70 to 100% solvent B at a constant flow rate of 100 nL/min [20]. The analyses were performed in a data-dependent acquisition mode using a top 6 collision-induced dissociation (CID) method for peptide identification alone (LTQ Orbitrap XL); or a top 10 high-energy collision-induced dissociation (HCD) method for peptide identification plus relative quantitation of iTRAQ reporter ions (LTQ Orbitrap Velos). Dynamic exclusion for selected ions was 60s. A single lock mass at m/z 445.120024 was employed on the LTQ OrbitrapVelos [21], but no lock mass was used on the LTQ Orbitrap XL. Maximal ion accumulation time allowed on the LTQ Orbitrap in CID mode was 150 ms for MS^n in the LTQ and 1,000 ms in the C-trap. Automatic gain control was used to prevent overfilling of the ion traps and were set to 5,000 (CID) in MSn mode for the LTQ, 10^6 ions for a full FTMS scan and 10^5 ions for HCD. Maximum ion time for HCD was set to 1,000 ms for acquiring 1 microscan at a resolution of 7,500. Intact peptides were detected in the Orbitrap at 100,000 resolution for CID fragmentation and 30,000 for HCD fragmentation experiments. The threshold for switching from MS to MSMS was 2,000 counts.

2.11 *Data analysis*

The acquired raw MS data files were converted into Mascot generic format (mgf) files with msconvert (ProteoWizard Library v2.1.2708). The resultant peak lists were searched against the human SwissProt database version v2010.09_20100812 (35,149 sequences, including isoforms as obtained from varsplic.pl) with the search engines Mascot (v2.3.02, MatrixScience, London, UK, www.matrixscience.com) and Phenyx (v2.6, GeneBio, Geneva, Switzerland) [22]. Submission to the search engines was via a Perl script that performs an initial search with relatively broad mass tolerances (Mascot only) on both the precursor and fragment ions (±10 ppm and ±0.6 Da, respectively). High-confidence peptide identifications are used to recalibrate all precursor and fragment ion masses prior to a second search with narrower mass tolerances (±4 ppm and ±0.3 Da for CID and ±4 ppm and ±0.025 Da for HCD, respectively). One missed tryptic cleavage site was allowed. Carbamidomethyl cysteine and N-terminal and lysine residue iTRAQ labelling were set as fixed modifications, and oxidised methionine was set as a variable modification. For the two-dimensional LC-MSMS samples, 10 individual analyses were merged into a single .mgf file prior submission to the search engines.

To validate the proteins, Mascot and Phenyx output files were processed by internally developed parsers. Proteins with ≥2 unique peptides above a score T1, or with a single peptide above a score T2, were selected as unambiguous identifications. Additional peptides for these validated proteins with score >T3 were also accepted. For Mascot and Phenyx, T1, T2 and T3 were equal to 12, 45, 10 and 5.5, 9.5, 3.5, respectively (p-value $<10^{-3}$). Following the selection criteria, proteins were grouped on the basis of shared peptides, and only the group reporters are considered in the final output of identified proteins. Spectral conflicts between Mascot and Phenyx peptide identifications were discarded. The whole procedure was repeated against a reversed database to assess the protein group false discovery rate (FDR). Peptide and protein group identifications were<0.1 and <1% FDR, respectively.

Comparisons between analytical methods involved comparisons between the corresponding sets of identified proteins. This was achieved by an internally-developed program that simultaneously computes the protein groups in all samples and extracts statistical data such as the number of distinct peptides, number of spectra, and sequence coverage.

2.12 *iTRAQ Quantitation*

The quantitation module of Proteome Discoverer 1.4, version 1.4.0.288 (Thermo Fisher Scientific, Waltham, MA) was used to assess the ratios for the individually-tagged tryptic digests of 30,000 HEK293, U937 and PBMCs. The intensity of the iTRAQ 4-plex reporter ions were integrated using the default settings for centroid peak detection at the highest confidence and a mass tolerance of 20 ppm. Correction for isotopic impurities was not performed. In addition, spectra with reporter ion intensities below 100 counts; and spectra with co-isolation of contaminating peptides exceeding 40% of the selected precursor ion were excluded from the protein ratio calculations. The median ratios for all peptides was calculated for each pair of cell lines (114/115, 114/116 114/117, 115/114, 115/116, 115/117, 116/114, 116/115, 117/114, 117/115), but shared peptides were excluded from quantitation. Protein ratios for the two combined technical replicates were calculated using the arithmetic mean of the protein ratios (median ratio of all used peptide ratios) for each replicate.

2.13 *Functional annotation analysis*

Functional annotation analysis of the identified proteins was performed using the DAVID Bioinformatics Resources 6.7 [23, 24]. Enrichment of GENETIC_ASSOCIATION_DB_DISEASE categories were performed with the thresholds: Count = 2, EASE = 0.05. Enrichment of KEGG_PATHWAY and UP_TISSUE categories were performed with the thresholds: Count = 2, EASE = 0.01. REVIGO was used to summarize the GO analysis [25].

3. Results and Discussion

3.1 *Selection of the optimal cell lysis conditions.*

Standard laboratory buffers for lysing cells usually include solubilising agents (e.g., NP-40 or other detergents) and/or denaturing agents (e.g., urea, thiourea or SDS) to facilitate protein extraction. Additionally, sonication is often advised to aid in clarification of the cell lysate. To determine the optimal conditions for the present study, a series of standard laboratory lysis conditions were evaluated on cultured, adherent HEK293 cells. The chosen conditions were: (i) 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, pH 8.0; (ii) 50 mM HEPES, 150 mM NaCl, 5 mM EDTA, pH 8.0 (without NP-40) but with sonication; (iii) and (iv) 9 M urea; 20 mM HEPES, pH 8.0 ± sonication; and (v) 7 M urea, 2 M thiourea, 20 mM HEPES, pH 8.0 + sonication. In 1 mL of each of the six buffers described above, 8.7×10^6 cells were lysed. Following lysis, aliquots corresponding to 1,000 cells were removed to simulate low quantities of cellular material. Samples lysed in buffers containing NP-40 (i.e., 'native' buffers) were analysed by 1D-gel-LC-MSMS and samples lysed in urea (i.e., denaturing buffers) were analysed as technical duplicates by gel-free 1D-LC-MSMS.

The total number of proteins identified by 1D-LC-MSMS from the aliquots of cell lysate corresponding to 1,000 cells was evaluated (Supplementary Figure S1 A). The two buffers that gave the highest number of identified proteins were: (i) HEPES with NP-40; and (iii) urea buffer. For the proteins identified in these samples, the Gene Ontology classification for cellular component (GOTERM_CC) was performed (P-value <0.01, Supplementary Figure S1 B). As anticipated, the protein extracts generated with the buffer containing NP-40 were enriched for the GO terms: membrane-bound organelles, mitochondria, endoplasmic reticulum, and organelle lumen. Conversely, analysis of the extracts from the urea lysis buffer complemented the 'native' buffer data set by providing additional ribosomal and nuclear proteins. Thus, these two lysis conditions: (i) a 'native' buffer comprised of 50 mM HEPES, 150 mM NaCl, 5 mM EDTA with 0.5% NP-40); and (iii) a denaturing buffer containing 9 M urea (both without sonication) generated the highest number of identified proteins that together cover the broadest range of cellular components. Subsequently, both buffers were chosen for further evaluation.

3.2 *Determining the protein content from low cell numbers*

All the standard laboratory methods for determining protein concentration failed to provide accurate and reproducible values for the low numbers of cells analysed in this study. Nevertheless, we were curious to attempt to estimate the quantity of protein from a low number of cells. Thus, the quantity of protein in 10,000 HEK293, U937 and human primary PBMCs was determined as follows. Thirty thousand cells of each type were lysed in 'native' buffer; one third of each sample resolved by 1D-SDS-PAGE (simulation of 10,000 lysed cells), stained with fluorescent Krypton Protein Stain and visualised with ChemiDoc™ MP Imaging System (BIO-RAD) (Figure 1A). In addition, one third of the extract from 30,000 HEK293 cells was loaded onto a gel together with serial dilutions of the cell lysate from 8.7×10^6 HEK293. Both samples were lysed under identical conditions. For the larger sample, an accurate determination of the protein content was achieved using the Bradford assay. A dilution series was used to calibrate the protein content, and thus, the protein quantity in the lysate of 10,000 cells was estimated by fluorescence intensity of the gel lanes analysed with Image Lab 5.0 software (BIO-RAD) (Figure 1B). For HEK293, 10,000 cells equated to 780 ng total protein. Perhaps as a consequence of the smaller size (ranging from 7-20 μm in diameter) and mixed cell population of human primary

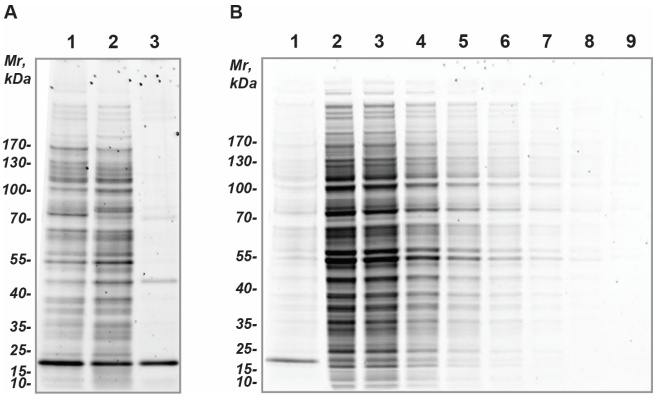

Figure 1. Calibration and determination of the protein content in 10,000 lysed cells. Samples were separated on a 4-12% Bis-Tris gel and proteins visualized by Krypton™ Protein Stain. (A) Protein extract from 30,000 lysed cells with one third of the extract loaded. Lane 1: 10,000 HEK293 cells; lane 2: 10,000 U937 cells; and lane 3: 10,000 primary human PBMCs. (B) Comparison of 10,000 HEK-293 cells from 30,000 cells extract with serial dilution of the protein extract from 8.7×10^6 lysed HEK293 cells. Protein content in the larger sample was determined by the Bradford assay. Lane 1: ⅓ protein extract from 30,000 cells (10,000 cells); Lane 2: 10 µg; Lane 3: 5 µg; Lane 4: 2.5 µg; Lane 5: 1.25 µg; Lane 6: 0.625 µg; Lane 7: 0.313 µg; Lane 8: 0.156 µg; Lane 9:0.078 µg.

PBMCs (consisting of neutrophils, eosinophils, basophils, lymphocytes, monocytes); the quantity of protein for 10,000 PBMCs is approximately 4 times less than for both HEK293 and U937 cells, as estimated by comparison of fluorescense intensity of the lanes on the gel (Figure 1A). All subsequent experiments were based on the number of cells rather than on the quantity of protein in the lysate.

3.3 *Determining the proteome of low cell quantities*

To emulate the situation of low sample availability, e.g., human FNAB and/or core needle biopsies, aliquots of 3,000, 15,000 and 30,000 HEK293 cells were lysed under (i) 'native' and (iii) denaturing conditions. Each sample was analysed as technical triplicates by 1D-LC-MSMS (each replicate thus corresponds to the analysis of 1,000, 5,000 and 10,000 cells). Interestingly, the result obtained from 5,000 cells lysed in the urea denaturing buffer (191 proteins) (Supplementary Figure S2) was comparable to a 1,000 cell aliquot from a higher number of cells lysed in the same buffer (148 proteins) (Supplementary Figure S1 A). The result from the 1,000 cells lysed under these conditions was unsatisfactory due to the extremely low number of identified proteins (Supplementary Figure S2). This is a critical point when reviewing data from the literature and designing proteomic experiments involving low cell numbers. It is important to distinguish between the types of approaches used in terms of the quantity of cellular material analysed. The total lysate of a low quantity of cells results in less protein identifications, than the same quantity of cells taken as an aliquot from a significantly larger number of lysed cells. This observation is simply explained by the fact that larger relative losses from small starting quantities are incurred during the sample preparation. Based on these findings, further experiments were performed on aliquots of 15,000 and 30,000 cells only.

Fifteen thousand and 30,000 HEK293 cells, U937 cells and primary human PBMCs were lysed in the two chosen buffers: (i) 'native' and (iii) denaturing. One third of each sample corresponding to 5,000 and 10,000 cells, respectively, was analysed by 1D-gel-based ('native' lysis) and 1D-gel-free (denaturing lysis) LC-MSMS. A schematic of the experimental design is shown in Figure 2. In parallel, a second set of samples were analysed by 2D-LC-MSMS. The proteins identified in all conditions from HEK293, U937 and PBMC cells are presented in Supplementary Tables S1, S2 and S3. In almost all instances, fractionation of the samples prior to LC -MSMS analysis gave a higher number of identified proteins compared to the 1D approach (Figure 3A). Only the PBMCs lysed in the urea denaturing buffer gave a poorer result (Figure 3A, left panel). Obviously, with a very low quantity of starting material (estimated below 100 ng protein from Fig.1) the additional fractionation step only resulted in fur-

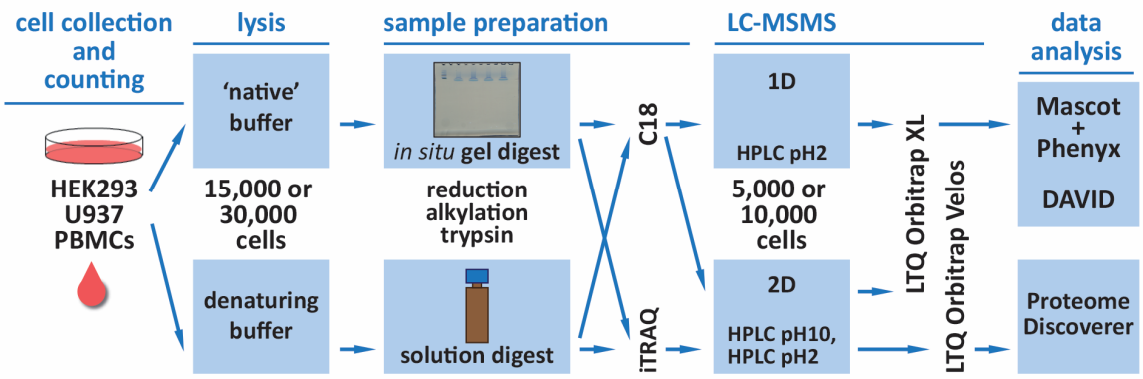

Figure 2. Schematic of the experimental design.

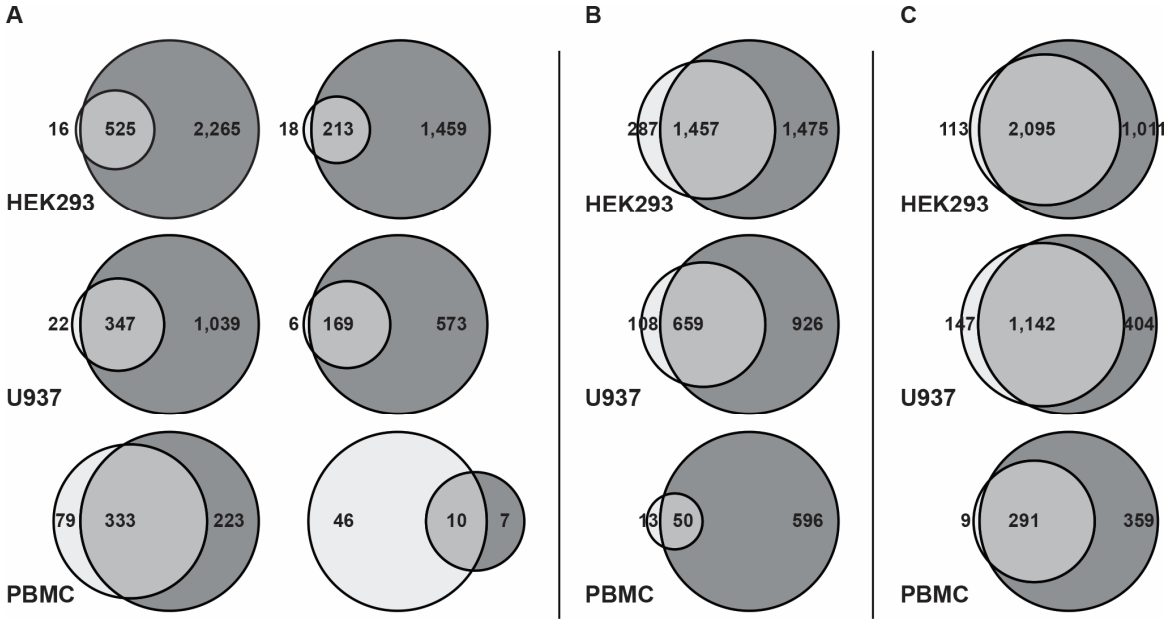

Figure 3. Comparison of the number of unique proteins identified from lysates of the three different cell lines: HEK293, U937, and PBMCs, according to additional sample fractionation before LC-MSMS analysis, the two lyses conditions and the amount of starting cellular material. (A) Comparison of 1D- and 2D-LC-MSMS of 10,000 cells (proteins extracted from 30,000 lysed cells). Left panel: protein extractions obtained with the 'native' NP-40 buffer. Right panel: protein extractions obtained with the denaturing urea buffer. Light grey circle, 1D-LC-MSMS; dark grey circle, 2D-LC-MSMS. (B) Comparison of the denaturing urea buffer (light grey circle) versus the 'native' buffer (dark grey circle). For each cell line, the number of unique protein identifications was generated from the analysis of 5,000 and 10,000 cells (protein extract obtained from 15,000 and 30,000 lysed cells, respectively) combined with the analyses by 1D- and 2D-LC-MSMS. (C) Comparison of 5,000 (light grey circle) and 10,000 cells (dark grey circle) (protein extract obtained from 15,000 and 30,000 lysed cells, respectively). For each cell line, the number of unique protein identifications was generated from the two lyses conditions ('native' and denaturing) combined with the analyses by 1D- and 2D-LC-MSMS.

ther dilution of the sample and/or sample loss leading to the observed decrease in the number of identified proteins compared to the 1D-LC-MSMS analysis.

As is often the case for proteomic experiments, using two conditions leads to an additive effect. The results of the different lysis buffers with respect to the number of proteins identified by 1D- and 2D- analyses of 5,000 and 10,000 cells are shown in Fig. 3B. As can be seen from this figure, the 'native' extraction conditions were more successful than the urea denaturing conditions; as evidenced by the higher number of proteins identified from the cells. Nevertheless, the results of the two lysis conditions are complementary with respect to the number of unique proteins identified.

When the quantity of material used for each experiment was compared, naturally 10,000 cells (initially 30,000 cells lysed) resulted in a higher number of identified proteins in all three cell types (Figure 3C). Approximately two thirds of the total number of proteins identified in HEK293 cells was common to the 5,000 and 10,000 cells; and three quarters of the proteins from the U937 cells were also shared. This appeared to occur even when applying a 2D approach that utilised additional sample fractionation. In contrast, only approximately one third of all the identified proteins from the PBMCs were evident for the results of different cell quantities. It is possible to surmise that for this cell type, 5,000 cells challenges the lower limits of successfully extracted protein material and sample fractionation and does not lead to an improvement in the number of proteins identified.

In total, 3,219; 1,693 and 659 unique proteins were identified from HEK293, U937 and total PBMCs, respectively. From 5,000 cells, 2,208; 1,289; and 300 proteins were identified and from 10,000 cells, 3,107; 1,546; and 650 proteins were identified. The most ideal condition for sample processing and subsequent analysis was to lyse 30,000 cells with the 'native' buffer followed by 2D-LC-MSMS. From this choice of experimental design alone, 2,790; 1,387; and 556 proteins were identified for the HEK293, U937 and PBMCs, respectively.

3.4 *'Mini-proteomes' complement known cell line proteomes and provide additional proteins*

Some of the earlier studies on HEK293 and U937cultured cell lines were performed utilising comparative studies of the 2D-gel approach combined with LC-MSMS [26-28] or MALDI-TOF analysis [26-28]. The earliest profile of HEK-293 proteins analysed by 1D-SDS-PAGE combined with LC-MSMS reported 1,111 and 1,063 proteins from the cytoplasmic and nuclear fractions, respectively [7]. A more recent publication identified 8,543 proteins from HEK293 cells. To date, this is the largest reported proteome of the eleven cell lines that were compared [9]. Another study that aimed at

describing the 'core' or 'central' proteome of different cultured cell lines, reported 4,154 and 2,073 protein groups for HEK293 and U937 cells, respectively [8]. The overall, central 'core' proteome of the five cell lines analysed in this work [8] consisted of 1,124 proteins. To the best of our knowledge, this is the only publication where HEK293 and U937 cell lines were proteomically-compared. As the cell lysis conditions were similar to those used in the current study, our HEK293 and U937 proteomes were evaluated against this work [8]. The 'mini-proteome' of both cell lines did not cover the 'core' proteome in entirety (Fig. 4 A, C). At the same time, however, the 'mini-proteome' generated from 5,000 and 10,000 HEK293 cells lysed in the 'native' buffer and analysed by 2D-LC-MSMS revealed that 54% and 63%, respectively, of the proteins identified were not part of the 'core' proteome. Under the same conditions, 42% and 45% of the proteins identified from 5,000 and 10,000 U937 cells, respectively, also did not overlap with 'core' proteome. This observation suggested that these non-'core' proteins represent a specific component of the proteome profile for each of the two cell lines. Interestingly, even from the proteome of 5,000 cells many proteins were identified that were not apparent from the proteomes generated from a larger amount (6 µg) of material (Fig. 4 B, D). Not unexpectedly, the total number of proteins identified in the individual HEK293 and U937 'mini-proteomes' decreased proportionally when compared to the larger proteomes. Namely, a decrease of 1.7 and 1.8 times for 10,000 HEK293 and U937 cells, respectively; and a decrease of 2.3 and 2.0 times for 5,000 cells (Fig 4 E). Comparison of each 'mini-proteome' with the 'core' proteome revealed that the U937 cells have a lower total number of protein identifications overlapping with the 'core' proteome than the HEK293 cells. For the HEK293 cells, however, the portion of the 'core' proteome proteins increased from 24% in the large proteome [8] to 46% in the 5,000 cell 'mini-proteome'. For the U937 cells, 46% of the large proteome overlaps with the 'core' proteome. This overlap increased to 58% when a lower amount of material was analysed (Fig 4 F). Thus, overall the smaller U937 proteome retained a larger portion of the 'core' proteome when the amount of analysed material was decreased.

It is important to note that although decreasing the amount of analysed material leads to an expected decrease in the total number of proteins identified, the obtained 'mini-proteome' still included many proteins from the 'core' proteome but also still remained representative of the cells under investigation. The relative abundance of the proteins expressed as emPAI values (Supplementary Table S4) showed that the most abundant proteins in the samples are from the 'core' proteome and many of the non-'core' proteins have a low relative abundance. At the same time, the relative abundance values have a large degree of variation for the 'core' proteins and also for the cell-specific proteins. As such, no direct connection between the abundance of a protein and the specificity of that particular protein for certain cell type can be made. As many of the cell-specific proteins have an abundance similar to the 'core' proteins, the former thus have an equal opportunity to be identified from low amounts of cellular material and therefore contribute to the cell-specific proteome.

There are a few publications describing the number of proteins identified in proteomic studies of PBMCs by 1D-LC-MSMS. These include: 1,432 [29]; and 514 proteins from 50 µg protein [30]. In our work, a total of 652 proteins were identified. Nevertheless, it is important to note here that this number of protein identifications was generated from a low amount of material. From just one of the analysis conditions (i.e., 10,000 cells, 'native' buffer, 2D-LC-MSMS) 556 proteins were identified. Interestingly, 17% of the PBMC proteome (115 protein identifications) from this current study did not correspond with the results from the two previously mentioned publications [29, 30] (Fig. 4 G).

As PBMCs are primary cells and by extrapolation, primary cells of any origin are of relevance and interest in clinical studies, the total proteome of these cells will be described and characterized in more detail. In particular, we will focus on the newly-identified proteins.

The mentioned above, 115 proteins identified from PBMCs were analysed using DAVID and the results revealed enrichment of the term IMMUNE as GENETIC_ASSOCIATION_DB_DISEASE_CLASS (Count = 13, p-value 0.01); and hsa05322:Systemic lupus erythematosus as KEGG_PATHWAY (Count = 5, p-value 0.05) (Supplementary Table S5). These categories indicate a possible relevance of the proteins involved in immune responses for clinical studies. Gene Ontology analysis also revealed features that are characteristic of the function of blood cells. The most highly-enriched GO biological process categories in the DAVID analyses with subsequent enrichment in REVIGO were: immune system process, immune response, cellular component assembly, response to stimulus, and multi-organism process (Fig 4H) (Supplementary Table S5). Response to biotic stimulus, killing cells of other organisms, defense response to bacterium were also revealed.

3.5 Identifying salient features of specific cell proteomes

Summarised in Fig. 5A are the combined, complementary results for the 30,000 cells lysed under both the 'native' and denaturing conditions and analyzed by 2D-LC-MSMS. The largest and smallest proteome coverage was obtained for the HEK293 and PBMC cells, respectively. The overlap between all three cell types contained the majority of the PBMC proteome, with around one third of the U937 proteome and almost one seventh of the HEK293 proteome. A similar outcome was obtained for the three different proteomes when the results from all the experimental conditions applied to each cell line were combined (Supplementary Figure S3). At the same time, both HEK293 and U937 have a specific and separate overlap with the proteome of the PBMCs. This was confirmed by functional annotation of the obtained proteomes for the 'pathways' (Fig. 5B; Supplementary Table S6)

and 'disease' categories (Fig. 5C; Supplementary Table S7).

For the HEK293 cell proteome, among the specific categories of diseases that were enriched are bladder cancer, head and neck cancer, and leukemia. This is in accordance with the nature of the HEK cells being originally endothelial or epithelial cells immortalized from embryonic kidney tissue

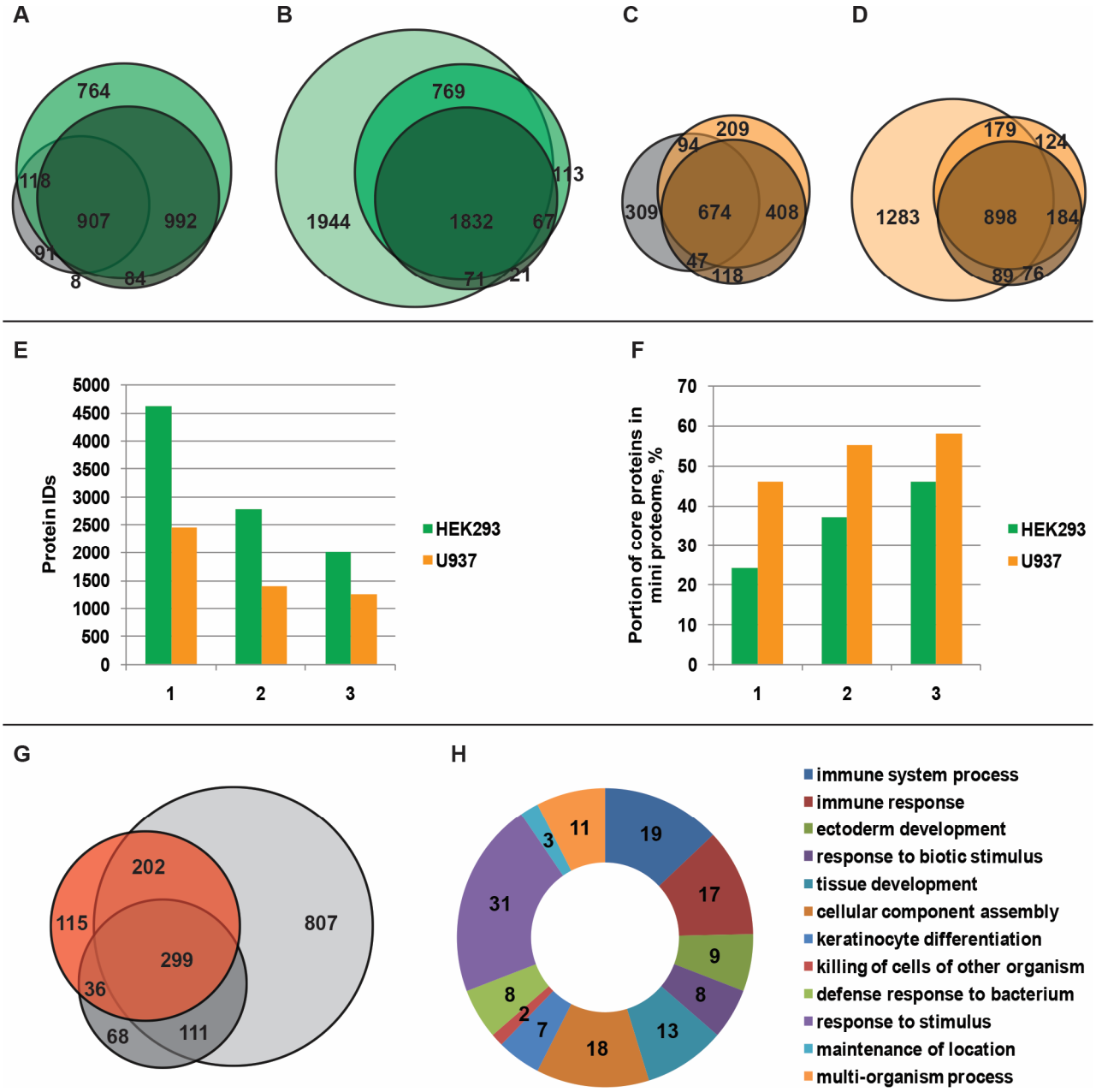

Figure 4. Comparison of the 'mini-proteomes' of HEK293, U937 and PBMC cells with the proteomes generated from larger amounts of protein. (A) (B) (C) (D) Comparison of the number of protein identifications from this current study with Burkard et al.[8] 'Core' proteome from 5 cell lines[8] (grey); large proteome of HEK293 cells (light green) and U937 cells (light orange) (50 µg protein lysate, 6 µg analysed;[8]) 'mini-proteomes' of 10,000 HEK293 (medium green) and 10,000 U937 cells (medium orange) (30,000 cells lysed in the 'native' buffer and material from 10,000 cells analysed by 2D-LC-MSMS); 'mini-proteomes' of 5,000 HEK293 (dark green) and 5,000 U937 cells (dark orange) (15,000 cells lysed in the 'native' buffer and material from 5,000 cells analysed by 2D-LC-MSMS). (A) (C) HEK293 proteomes and the 'core' proteome from;[8] (B) (D) U937 proteomes and the 'core' proteome from.[8] (E) Comparison of the number of proteins identified from (1) large proteome; (2) 'mini-proteome' from 10,000 cells; and (3) 'mini-proteome' from 5,000 HEK293 (green) and U937 cells (orange). (F) Comparison of the portion of the 'core' proteome observed in HEK293 (green) and U937 (orange) proteome obtained from the same amount of analysed material as indicated in (E). (G) Number of proteins identified in the proteome of PBMCs from three independent studies. Current study (red) representing the total number of protein identified from all combined experimental conditions; Haudek et al. (light grey);[29] Maccarrone et al. (dark grey).[30] (H) Enriched biological processes for the part of the PBMC 'mini-proteome' identified in our work that does not overlap with published data. GO analysis from DAVID and REVIGO.

[31]. Enrichment of neural tube defects and ALS/amyotrophic lateral sclerosis, both neurodegenerative diseases [32], is in accordance with the opinion that HEK cells originate from the transformation of neuronal cells of kidney tissue [33]. In the ALS/amyotrophic lateral sclerosis disease category, SOD1 (SODC_HUMAN) was found and this is known to be associated with ALS [34].

U937 cells originate from blood cells [35], so it is not surprising that these cells share common features with PBMCs. Both cell types were enriched for the disease category 'catalase activity'. Superoxide dismutase, catalase and glutathione peroxidase (SODM_HUMAN, CATA_HUAMN and GPX1_HUMAN) are important in the protection against oxidative stress. Several HLA (human leukocyte antigene) histocompatibility antigens were found in the PBMCs and U937 cells (1A68, 1A03, 1B53, 1B51, 1B58, 1B73, 1C12_HUMAN) that led to the enrichment of three other disease categories: hypothyroidism, spondyloarthropathies, and vitiligo. Although the PBMC proteome is the smallest of the three, many more disease categories were enriched from the data set (Fig. 5B, Supplementary Table S7).

In the pathway analysis, the lower number of pathway categories enriched for U937 and PBMCs in comparison with HEK293 cells may be a reflection of the size of the identified proteome (Fig 5C). Nevertheless, U937 cells were specifically-enriched for lysosome and Fc gamma R-mediated phagocytosis pathways (Supplementary Table S6). This is in accordance with the origin of the cell line and coincides with the monocytic (in particular macrophage) cell characteristics [36]. Interestingly, the tight junction pathway was enriched in HEK293 which may reflect the adherent growth pattern of these cells in culture compared to the U937 cell line.

The main energy releasing metabolic pathways (e.g., glycolysis, pentose phosphate pathway, pyruvate metabolism and oxidative phosphorylation) were enriched in PBMCs. Excluding glutathione metabolism, however, nucleotide and amino acid metabolism pathways were not enriched. This coincides with the important role of glutathione metabolism in reductive processes to counteract oxidative stress. The decrease in many metabolic processes and also DNA and translation-related processes is complementary to the status of PBMCs as highly-differentiated cells. Despite the lower number of proteins identified from PBMCs, there are several pathways specifically enriched in this cell type. These included cell motility, cell communication and immune system pathways. From the PBMC 'min-proteome' a range of proteins were also identified that are specific for certain cell types [37] (Supplementary Figure S5). These included monocytes, T cells, platelets and neutrophils.

It is important to highlight that from such a low number of cells combined with the described cell lyses conditions and mass spectrometry approaches; it was possible to identify specific characteristics inherent to each cell type. This was particularly evident for the human PBMCs, which are primary, differentiated cells, in contrast to the two immortalised cultured cells.

3.6 *Quantitative differences between specific cell proteomes*

For the quantitative analysis of the cell proteomes, 30,000 cells were lysed with denaturing or 'native' buffers as described above and shown in Fig. 2. After tryptic digestion, peptides from HEK293 and U937 were labelled with the 114 and 115 iTRAQ reagents, respectively. In addition, two independent PBMC preparations were labeled with the 116 and 117 iTRAQ reagents (Fig. 6A). Samples were prepared and injected in amounts equivalent to those in the experiment without labeling. At the same time, the quantitative MSMS analysis of the labeled samples was performed by implementation of HCD ion fragmentation in contrast to the CID fragmentation for the non-quantitative approach.

Taking into account the results from the two lysis conditions, the total number of proteins identified in these experiments was 1,647 (Supplementary Table S8). In contrast, in the unlabelled experiments analysed by collision-induced dissociation (CID), a total of 3,374 proteins were identified

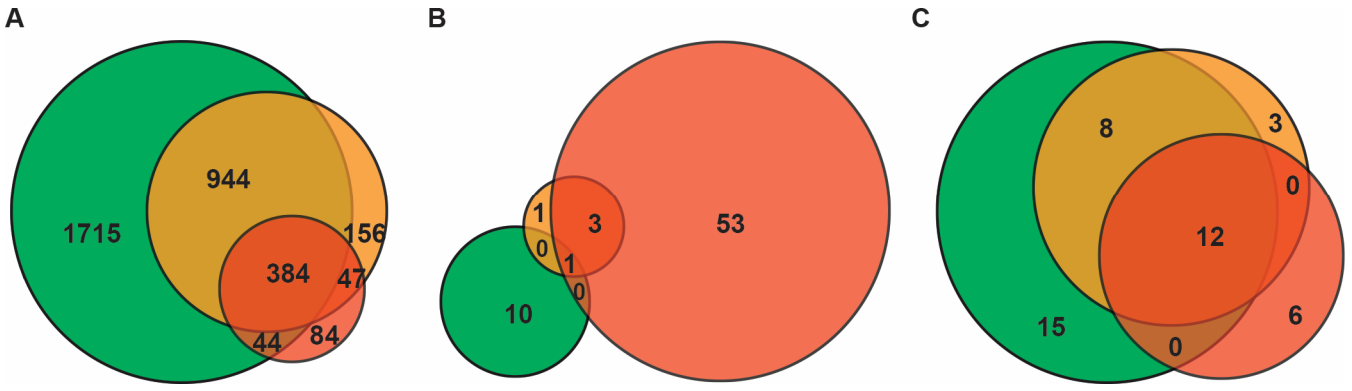

Figure 5. Comparison of the three cell types: HEK293 (green); U937 (orange); and human primary PBMCs (red). Data were generated on the proteins identified from 10,000 cells lysed in the two buffers and analysed by 2D-LCMSMS. (A) Number of unique proteins identified in the cell lines. (B) Number of disease categories. (C) Number of pathways enriched via functional annotation analysis.

from 10,000 cells lysed under the two conditions and analysed by 2D-LC-MSMS. The observed decrease in the total number of identified proteins between the two approaches can be explained by: (i) additional losses incurred during the extra sample preparation steps; and (ii) the slower duty cycle of high-energy collision-induced dissociation (HCD) compared to the duty cycle of CID. It has already been shown that the number of identified proteins in plasma decreased around 1.5 times in TMT-labeled samples compared to the unlabelled counterpart [38]. Care should obviously be taken when a quantitative, chemical labelling approach is performed on small amounts of material.

Shown in Fig. 6B and 6C are the distributions of the iTRAQ relative ratio quantitation between the three cell lines

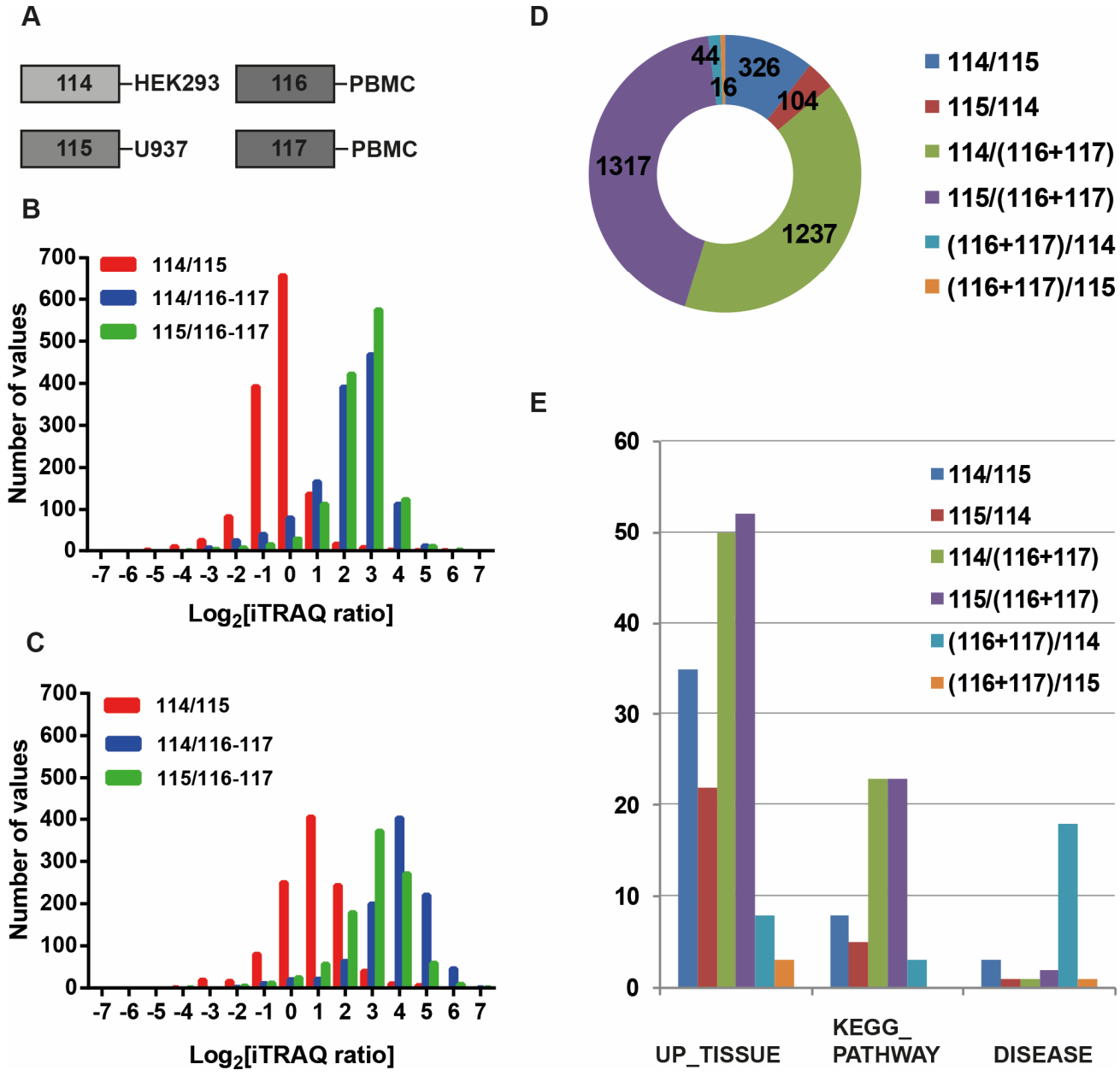

Figure 6. Quantitative analysis of the 'mini-proteomes' of the three cell lines: HEK293, U937 and PBMCs. (A) Scheme of the iTRAQ labelling. (B) Distribution of iTRAQ ratios in the cells lysed with the NP-40 'native' buffer. (C) Distribution of iTRAQ ratios in the cells lysed with the urea denaturing buffer. (D) Number of proteins identified that were more abundant in each cell line in a pair-wise analysis. Mean of the ratios and SD were calculated using GraphPad Prism 6. Proteins are discussed as more abundant if the ratios are 2-fold different and ratios are calculated in more than one replica. (E) Functional annotation analysis of the proteins from (B). Comparison of UP_TISSUE, KEGG_PATHWAY and GENETIC_ASSOCIATION_DB_DISEASE enrichment using DAVID Bioinformatics Resources 6.7. For UP_TISSUE and KEGG_PATHWAY, Count = 2, EASE = 0.01; for disease categories, Count = 2, EASE = 0.05; for the KEGG_PATHWAY of the PBMCs, Count = 2, EASE = 0.05.

for the 'native' buffer and the denaturing urea buffer, respectively. Data is expressed as a $\log_2$[iTRAQ ratio]-fold difference in protein abundance between the cell types. For both lyses conditions, the differences between the cultured and blood-derived cells are much greater than between the two cultured cell lines. From the 114/115 ratio (HEK293/U937), there are proteins more abundant in HEK293 cells and vice versa. Although the majority of the proteins obviously have a higher abundance in HEK293 and U937 cells compared to the PBMCs, there are still proteins that are more abundant in the PBMCs than either the HEK293 or the U937 cells. Some of these proteins exhibit up to a 3-fold difference in expression levels. This observation was particularly evident for the 'native' buffer (Fig. 6B).

A pair-wise comparison between the three cell proteomes was performed to determine the total number of identified proteins that are more abundant in each cell type (Fig. 6D). Again, it can be seen that the greatest difference between the proteomes is observed between the cultured cell lines and the primary PBMC cells. Nevertheless, 44 and 16 proteins from the PBMCs had a higher abundance compared to the HEK293 and U937 cells, respectively. The smaller difference between PBMCs and U937 might potentially reflect the closer origin between U937 cells and PBMCs. Numerous proteins showed differing degrees of abundance between the two types of cultured cells (Supplementary table S9).

Figure 6E further shows the observed differences between the cell lines and also that it is feasible to distinguish the specificity of each cell type. There are particular pathways enriched for each cell line. Due to the points noted earlier in this section that led to a decrease number of proteins identified, the total number of enriched categories is also lower than from the unlabelled experiments. Additionally, for each cell type, only the proteins with a significantly-higher expression (Supplementary table S9) were used for the enrichment analysis (Supplementary table S10). The quantitative data confirmed an enrichment of metabolic pathways (Supplementary table S11) in the cultured cells compared to the PBMCs. With a lower confidence (EASY>0.1), only a few categories were enriched in the blood-derived cells. These were ECM receptor interaction, regulation of actin cytoskeleton, and hematopoietic cell lineage. Although the number of proteins that were shown to be different in the PBMCs are relatively low, these are disease-related; and overall, there are more disease categories for the PBMCs in comparison to the other cells assessed in this study (Supplementary table S12). Among the proteins with an increased ratio in PBMCs are trombospondin-1 (P07996), gelsolin (P06396), several antigens CD41 (P08514), CD61 (P05106), CD11b (P11215), fibrinogens (P06271, P06279), arachidonate 5-lipoxygenase-activating protein (P20292). Thus, the approach confirms the applicability of this type of study in analysing and determining salient features of proteomes from samples of low cell number.

4. Concluding Remarks

Comparative proteomic analysis of three different cell types from low quantities of cellular material revealed that the data obtained contained information that was specific to, and representative of, the individual proteomes. The proteomes obtained from a limited amount of cellular material ('mini-proteomes') were smaller with respect to the total number of proteins identified than those obtained under similar conditions from larger quantities of cellular material. Although these 'mini-proteomes' contained a portion of the 'core' proteome shared between many different cell lines, the data also revealed many cell type specific proteins.

The specificity of the smallest possible proteome from a cell type and the minimal (critical) amount of material required to enable identification is highly-dependent on the source of the cells. A portion of specific cell line proteins was reduced in the smaller proteome, e.g., U937 in comparison with HEK293. From the same number of cells as the cultured cell lines, the highly-differentiated PBMCs provided less protein material which was reflected in the lower total number of identified proteins. Despite this, however, the PBMCs proteome contained a larger percentage of specific proteins than either the HEK293 or the U937 cells. Thus, the limitation caused by the lower amount of material was compensated in these differentiated cells by a higher relative level of cell-specific proteins. This finding provides enormous potential and relevance to clinical studies from low amounts of cellular material, e.g., the proteomic analysis of scarce patient samples such as a fine-needle aspiration biopsy (FNAB). Although out of the main focus of this study, our data from a small amount of cells supplements the characterisation of the PBMC proteome previously analysed by 1D-LC-MSMS. We identified and characterised an additional 115 proteins for this cell type.

Quantitation by chemical labelling and mass spectrometry is somewhat compromised when using small amounts of cells and protein material. At the same time, even for PBMCs it was possible to detect specific proteins when comparing the cell types to each other. It can be envisaged that with next generation approaches that combine improvements in protein solubilisation, sample preparation techniques and MS instrumentation (e.g., Q-Exactive) there will be an even deeper proteome of between 5,000-10,000 cells. Potentially, the analysis of material from even fewer cells will be facilitated and even deeper proteomic analyses, e.g., improved quantitative or posttranslational modifications will be feasible.

5. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/172/0

Supplementary-material.pdf file contains the table of content of all supplementary files, Supplementary Figures S1-S4, and Supple-

mentary Tables S6,S7 and S10-S12. Supplementary Tables S1-S5, S8 and S9 can be found in the separate Excel files with corresponding names.

Supplementary Figure S1 - Evaluation of buffers and lysis conditions for analysis of proteomes from small amounts of cellular material.
Supplementary Figure S2 - Number of unique protein identifications from different amounts of HEK293 cells lysed with the urea denaturing buffer.
Supplementary Figure S3 - Comparison of number of unique proteins identified in three cell types.
Supplementary Figure S4 - Proteins in PBMCs that are specific for certain blood cell types.
Supplementary Table S1 - List of proteins identified from HEK293 cells.
Supplementary Table S2 - List of proteins identified from U937 cells.
Supplementary Table S3 - List of proteins identified from PBMCs.
Supplementary Table S4 - emPAI values extracted from Mascot for the 30,000 cells analysed by 2D-LC-MSMS.
Supplementary Table S5 - List of proteins identified in our work from PBMCs (summary of all experimental approaches). These do not overlap with data previously published by Haudek et al., 2008 and Maccarrone et al., 2013; and Gene Ontology analysis of these proteins. Excel file.
Supplementary Table S6 - Comparison of functional annotation results for the three different cell types: HEK293, U937 and PBMCs. Enrichment of KEGG_PATHWAY categories.
Supplementary Table S7 - Comparison of functional annotation results for 3 different cell types: HEK293, U937 and PBMCs. Enrichment of GENETIC_ASSOCIATION_DB_DISEASE categories.
Supplementary Table S8 - Results from the iTRAQ quantitative experiments for the three cell types: HEK293, U937 and PBMCs. The table was generated from Proteome Discoverer. iTRAQ channels: 114, HEK293; 115, U937; 116 and 117, PBMCs. A and B, cells were lysed in denaturing buffer; C and D, cells were lysed in 'native' buffer. Excel file.
Supplementary Table S9 - Lists of proteins with a significantly higher level in each cell type. Sheet 'summary' contains calculation of fold differences in the level of all proteins identified in the iTRAQ experiment for the three cell types. Other sheets in the file contain lists of protein with significantly higher protein content in each pair of cell types.
Supplementary Table S10 - List of proteins with a significantly higher level in each cell type in comparison with the others in a pairwise comparison.
Supplementary Table S11 - Comparison of functional annotation results for quantitative analysis of the three different cell types: HEK293, U937 and PBMCs. Enrichment of KEGG_PATHWAY categories.
Supplementary Table S12 - Comparison of functional annotation results for quantitative analysis of HEK293, U937 and PBMCs. Enrichment of GENETIC_ASSOCIATION_ DB_DISEASE categories.

Acknowledgements

The authors would like to thank Elisabeth Salzer from the Boztug laboratory at CeMM for blood collection, and Melanie Planyavsky and André C. Müller from the Bennett laboratory for technical support with sections of the experimental work and assistance with the analysis of the iTRAQ data, respectively. The authors are grateful to members of the Colinge Bioinformatic group at CeMM: Florian P. Breitwieser and Jacques Colinge for assistance with the analysis of the iTRAQ data, Alexey Stukalov for advice on the DAVID analyses, and Peter Májek for technical assistance in data processing. In addition, the authors would like to thank Ulrike Resch from the Department of Vascular Biology and Thrombosis Research at the Medical University of Vienna and Edit Szodorai from Lubec laboratory at the Medical University of Vienna for consultation on the preparation of human PBMCs and assistance with the ChemiDoc™ MP Imaging System (BIO-RAD) and Image Lab software, respectively.

Funding Sources

Research in our laboratory is supported by the Austrian Academy of Sciences, the Austrian Federal Ministry for Science and Research (Gen-Au projects, APP-III and BIN-III), and the Austrian Science Fund FWF. E.L.R. is supported by the Gen-Au APP-III program (No. 820965).

References

[1] L. Giusti, P. Iacconi, F. Ciregia, G. Giannaccini, F. Basolo, G. Donatini, P. Miccoli, A. Lucacchini, Journal of Endocrinological Investigation, 30 (2014) 865-869. doi: bf03349229

[2] L. Giusti, P. Iacconi, F. Ciregia, G. Giannaccini, G.L. Donatini, F. Basolo, P. Miccoli, A. Pinchera, A. Lucacchini, J Proteome Res, 7 (2008) 4079-4088. doi: 10.1021/pr8000404

[3] L. Giusti, P. Iacconi, A. Lucacchini, Proteomics Clin Appl, 5 (2011) 24-29. doi: 10.1002/prca.201000091

[4] P. Iacconi, L. Giusti, Y. Da Valle, F. Ciregia, G. Giannaccini, L. Torregrossa, A. Proietti, G. Donatini, S. Mazzeo, F. Basolo, A. Lucacchini, J Med Case Rep, 6 (2012) 103. doi: 10.1186/1752-1947-6-103

[5] F. Ciregia, L. Giusti, A. Molinaro, F. Niccolai, P. Agretti, T. Rago, G. Di Coscio, P. Vitti, F. Basolo, P. Iacconi, M. Tonacchera, A. Lucacchini, PLoS One, 8 (2013) e72911. doi: 10.1371/journal.pone.0072911

[6] A. Rapkiewicz, V. Espina, J.A. Zujewski, P.F. Lebowitz, A. Filie, J. Wulfkuhle, K. Camphausen, E.F. Petricoin, 3rd, L.A. Liotta, A. Abati, Cancer, 111 (2007) 173-184. doi: 10.1002/cncr.22686

[7] M. Schirle, M.A. Heurtier, B. Kuster, Mol Cell Proteomics, 2 (2003) 1297-1305. doi: 10.1074/mcp.M300087-MCP200

[8] T.R. Burkard, M. Planyavsky, I. Kaupe, F.P. Breitwieser, T. Burckstummer, K.L. Bennett, G. Superti-Furga, J. Colinge, BMC Syst Biol, 5 (2011) 17. doi: 10.1186/1752-0509-5-17

[9] T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, Mol Cell Proteomics, 11 (2012) M111 014050. doi: 10.1074/

mcp.M111.014050

[10] S. Di Palma, D. Stange, M. van de Wetering, H. Clevers, A.J. Heck, S. Mohammed, J Proteome Res, 10 (2011) 3814-3819. doi: 10.1021/pr200367p

[11] L.F. Waanders, K. Chwalek, M. Monetti, C. Kumar, E. Lammert, M. Mann, Proc Natl Acad Sci U S A, 106 (2009) 18902-18907. doi: 10.1073/pnas.0908351106

[12] M. Maurer, A.C. Muller, C. Wagner, M.L. Huber, E.L. Rudashevskaya, S.N. Wagner, K.L. Bennett, J Proteome Res, 12 (2013) 1040-1048. doi: 10.1021/pr301009u

[13] M. Maurer, A.C. Muller, K. Parapatics, W.F. Pickl, C. Wagner, E.L. Rudashevskaya, F.P. Breitwieser, J. Colinge, K. Garg, J. Griss, K.L. Bennett, S.N. Wagner, J Proteome Res, 13 (2014) 2830-2845. doi: 10.1021/pr401270y

[14] A. Umar, T.M. Luider, J.A. Foekens, L. Pasa-Tolic, Proteomics, 7 (2007) 323-329. doi: 10.1002/pmic.200600293

[15] J.R. Wisniewski, P. Ostasiewicz, M. Mann, J Proteome Res, 10 (2011) 3040-3049. doi: 10.1021/pr200019m

[16] E.B. Haura, R. Sacco, J. Li, A.C. Muller, F. Grebien, G. Superti -Furga, K.L. Bennett, J Integr OMICS, 2 (2012) 55-68. doi: 10.5584/jiomics.v2i1.81

[17] L.C. Ngoka, Proteome Sci, 6 (2008) 30. doi: 10.1186/1477-5956-6-30

[18] P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin, Mol Cell Proteomics, 3 (2004) 1154-1169. doi: 10.1074/mcp.M400129-MCP200

[19] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Journal of Separation Science, 28 (2005) 1694-1703. doi: 10.1002/jssc.200500116

[20] K.L. Bennett, M. Funk, M. Tschernutter, F.P. Breitwieser, M. Planyavsky, C. Ubaida Mohien, A. Muller, Z. Trajanoski, J. Colinge, G. Superti-Furga, U. Schmidt-Erfurth, J Proteomics, 74 (2011) 151-166. doi: 10.1016/j.jprot.2010.10.002

[21] J.V. Olsen, L.M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Mol Cell Proteomics, 4 (2005) 2010-2021. doi: 10.1074/mcp.T500030-MCP200

[22] J. Colinge, A. Masselot, M. Giron, T. Dessingy, J. Magnin, Proteomics, 3 (2003) 1454-1463. doi: 10.1002/pmic.200300485

[23] W. Huang da, B.T. Sherman, R.A. Lempicki, Nat Protoc, 4 (2009) 44-57. doi: 10.1038/nprot.2008.211

[24] D.W. Huang, B.T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M.W. Baseler, H.C. Lane, R.A. Lempicki, Nucleic Acids Res, 35 (2007) W169-175. doi: 10.1093/nar/gkm415

[25] F. Supek, M. Bosnjak, N. Skunca, T. Smuc, PLoS One, 6 (2011) e21800. doi: 10.1371/journal.pone.0021800

[26] M. Qasim, H. Rahman, M. Oellerich, A.R. Asif, Proteome Sci, 9 (2011) 57. doi: 10.1186/1477-5956-9-57

[27] S. Bartel, J. Doellinger, K. Darsow, D. Bourquain, R. Buchholz, A. Nitsche, H.A. Lange, Virol J, 8 (2011) 380. doi: 10.1186/1743-422X-8-380

[28] L. Minafra, G. Di Cara, N.N. Albanese, P. Cancemi, Leuk Res, 35 (2011) 226-236. doi: 10.1016/j.leukres.2010.07.040

[29] V.J. Haudek, N.C. Gundacker, A. Slany, H. Wimmer, E. Bayer, K. Pable, C. Gerner, J Proteome Res, 7 (2008) 5138-5147. doi: 10.1021/pr800438f

[30] G. Maccarrone, C. Rewerts, M. Lebar, C.W. Turck, D. Martins-de-Souza, Proteomics, 13 (2013) 893-897. doi: 10.1002/pmic.201200377

[31] F.L. Graham, J. Smiley, W.C. Russell, R. Nairn, Journal of General Virology, 36 (1977) 59-72. doi: 10.1099/0022-1317-36-1-59

[32] V. Meininger, Arch Ital Biol, 149 (2011) 29-37. doi: 10.4449/aib.v149i1.1265

[33] G. Shaw, S. Morse, M. Ararat, F.L. Graham, FASEB J, 16 (2002) 869-871. doi: 10.1096/fj.01-0995fje

[34] L.I. Bruijn, Science, 281 (1998) 1851-1854. doi: 10.1126/science.281.5384.1851

[35] C. Sundström, K. Nilsson, International Journal of Cancer, 17 (1976) 565-577. doi: 10.1002/ijc.2910170504

[36] J.S. Passmore, P.T. Lukey, S.R. Ress, Immunology, 102 (2001) 146-156. doi: 10.1046/j.1365-2567.2001.01164.x

[37] V.J. Haudek, A. Slany, N.C. Gundacker, H. Wimmer, J. Drach, C. Gerner, J Proteome Res, 8 (2009) 3834-3843. doi: 10.1021/pr801085g

[38] L. Dayon, M. Kussmann, EuPA Open Proteomics, 1 (2013) 8-16. doi: 10.1016/j.euprot.2013.08.001