



Article

# Blood Plasma Proteome: A Meta-Analysis of the Results of Protein Quantification in Human Blood by Targeted Mass Spectrometry

Anna A. Kliuchnikova <sup>1,2</sup>, Svetlana E. Novikova <sup>1</sup>, Ekaterina V. Ilgisonis <sup>1</sup>, Olga I. Kiseleva <sup>1</sup>, Ekaterina V. Poverennaya <sup>1</sup>, Victor G. Zgoda <sup>1</sup>, Sergei A. Moshkovskii <sup>2,3</sup>, Vladimir V. Poroikov <sup>1</sup>, Andrey V. Lisitsa <sup>1</sup>, Alexander I. Archakov <sup>1</sup> and Elena A. Ponomarenko <sup>1,\*</sup>

<sup>1</sup> Institute of Biomedical Chemistry, 119121 Moscow, Russia

<sup>2</sup> Federal Research and Clinical Center of Physical-Chemical Medicine, 119435 Moscow, Russia

<sup>3</sup> Department of Biochemistry, Medico-Biological Faculty, Pirogov Russian National Research Medical University, 117997 Moscow, Russia

\* Correspondence: 2463731@gmail.com

**Abstract:** A meta-analysis of the results of targeted quantitative screening of human blood plasma was performed to generate a reference standard kit that can be used for health analytics. The panel included 53 of the 296 proteins that form a “stable” part of the proteome of a healthy individual; these proteins were found in at least 70% of samples and were characterized by an interindividual coefficient of variation <40%. The concentration range of the selected proteins was  $10^{-10}$ – $10^{-3}$  M and enrichment analysis revealed their association with rare familial diseases. The concentration of ceruloplasmin was reduced by approximately three orders of magnitude in patients with neurological disorders compared to healthy volunteers, and those of gelsolin isoform 1 and complement factor H were abruptly reduced in patients with lung adenocarcinoma. Absolute quantitative data of the individual proteome of a healthy and diseased individual can be used as the basis for personalized medicine and health monitoring. Storage over time allows us to identify individual biomarkers in the molecular landscape and prevent pathological conditions.

**Keywords:** proteome; targeted mass spectrometry analysis; human proteome project; knowledge databases



**Citation:** Kliuchnikova, A.A.; Novikova, S.E.; Ilgisonis, E.V.; Kiseleva, O.I.; Poverennaya, E.V.; Zgoda, V.G.; Moshkovskii, S.A.; Poroikov, V.V.; Lisitsa, A.V.; Archakov, A.I.; et al. Blood Plasma Proteome: A Meta-Analysis of the Results of Protein Quantification in Human Blood by Targeted Mass Spectrometry. *Int. J. Mol. Sci.* **2023**, *24*, 769. <https://doi.org/10.3390/ijms24010769>

Academic Editors: Alexey Kononikhin, Vladimir Frankevich and Eugene Nikolaev

Received: 14 November 2022

Revised: 14 December 2022

Accepted: 27 December 2022

Published: 1 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Blood biochemistry tests have long been used to assess the human health status. The complete set of plasma proteins (proteome) plays a special role in diagnosing socially significant diseases, including cardiovascular diseases, diabetes mellitus, and different types of cancer. Researchers have long been interested in answering questions of what proteins are present in blood plasma, what their concentrations are, and how these plasma proteomes differ between a healthy and sick individual [1]. The use of monoclonal antibodies is currently the gold standard for detecting and quantifying proteins in blood plasma. This approach has several drawbacks, including its limited specificity owing to the cross-reactivity of antibodies and low potential for multiplex assays. Qualitative and quantitative assessments of protein markers in blood plasma will significantly improve diagnostic efficiency. Currently, SOMAscan and Olink are prevalent affinity-based techniques that have excellent potential for multiplexity and are often preferred over mass spectrometry [2,3].

Shotgun and targeted mass spectrometry approaches, whose development has been rapidly advancing, are excellent alternatives to antibody-based techniques for studying the plasma proteome. High-resolution shotgun mass spectrometry allows the simultaneous study of several hundreds of proteins with high specificity. Approximately 3000 proteins

have been detected using this technology under the HUPO Human Plasma Proteome Project (HPPP) owing to the joint efforts of 35 laboratories; 900 of these proteins have been identified with high confidence. More than 1500 proteins have been detected in blood plasma at concentrations ranging from  $10^{-12}$  to  $10^{-3}$  M [4]. This observation provides grounds for considering blood plasma as a complex biological matrix. In 2005, the research team led by Aebersold demonstrated the important role of meta-analysis, that is, summarizing the results of several experiments to obtain a complete proteomic map. Different proteomic techniques, sample preparation options, and algorithms for analyzing the resulting data lead to the identification of different sets of proteins in the same sample [5]. In 2011, as part of creating the Peptide Atlas resource, this research team compiled a list of 1929 master proteins identified in blood plasma with high confidence using LC-MS/MS [6,7]. According to the Human Protein Atlas, 4072 proteins have been identified over a broad concentration range in the plasma of healthy people using mass spectrometry without stable isotope labeling [8].

Targeted methods, including selected/multiple reaction monitoring (SRM/MRM), enable the detection of proteins in the biomaterial and quantification of their absolute concentrations, which is critical for medical applications. Using synthetic isotope-labeled peptide standards as reference samples improves the identification accuracy and measured protein concentration [9]. Furthermore, SRM analysis can be used to detect proteins present at low ( $10^{-14}$ – $10^{-12}$  M) and ultralow ( $10^{-18}$ – $10^{-15}$  M) concentrations via the irreversible binding of minor plasma proteins to biogranules [10] or fractionation [11–13].

Fewer studies on blood plasma proteins have utilized the SRM method compared to shotgun analysis. Hüttenhain et al. [14] used the SRM method to investigate candidate biomarkers for malignant neoplasms. They identified 182 plasma proteins, presumably associated with tumor diseases, with spectral counts as their primary characteristic. Domanski et al. [15] employed SRM to design a panel consisting of 67 potential marker proteins for diagnosing cardiovascular diseases. The normal concentrations of the proteins estimated using various methods, including mass spectrometry, were obtained from the literature.

As the SRM method enables structural analysis of the amino acid sequence of peptides, it is increasingly used to detect point amino acid substitutions in proteins. Thus, this method can be used to quantify proteoforms [16] or validate mRNA editing events by ADAR enzymes at the protein level [17].

SRM technology has entered the global market as multiplex test kits for research. Such kits are manufactured by Biognosys AG (Schlieren, Switzerland) and Cambridge Isotope Laboratories (Tewksbury, MA, USA) (PeptiQuant kit) [18]. The evolution of these test kits involves increasing the number of detectable analytes and improving analysis sensitivity, thus making them an efficient tool for protein marker detection. This method can quantify up to 500 proteins per sample; however, the reference concentrations in these kits are not provided to users.

The next logical step is to adapt targeted mass spectrometry for clinical applications. Designing and improving the available SRM test kits can enable the personalized prediction of disease risks, diagnosis, and assessment of treatment efficacy based on rational drug prescription and individual treatment selection to generate human digital molecular images [19]. A combination of labeling and targeted MS in the SRM mode is currently widely used as an alternative to immunoassays for accurate protein quantification and confirmation of the findings of earlier studies where biomarkers had a controversial clinical significance [16].

Identifying the reference ranges for analyte proteins is an important challenge for introducing SRM test kits in clinical practice. For example, such information is not available for the Biognosys AG kits. In the SRM studies performed by Hüttenhain et al. and Domanski et al., the normal concentrations of the proteins under study were obtained from the literature; the reference values were obtained using different methods, including those unrelated to mass spectrometry [14,15]. Notably, the differences between the methods employed for collecting serum or plasma from patients (e.g., using test tubes coated with

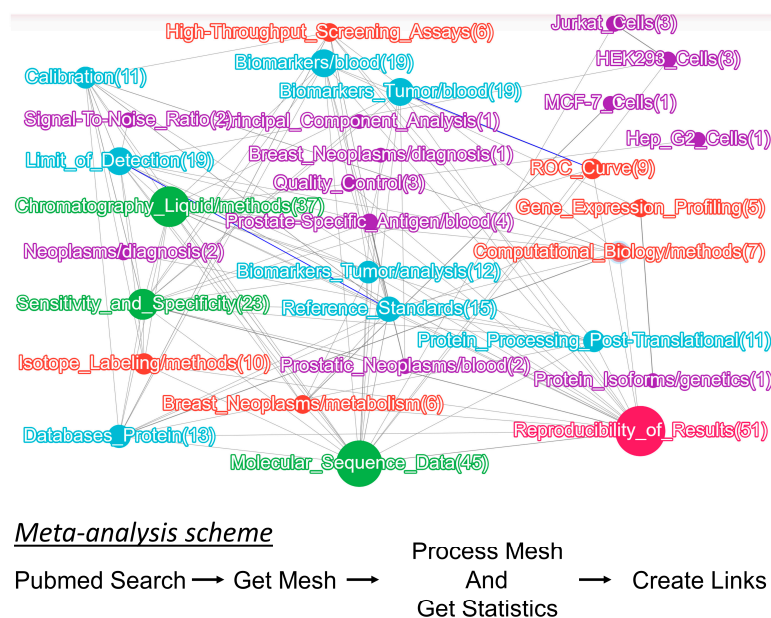
different anticoagulants) may lead to statistically significant variations in protein levels, thereby impeding biomarker research [20].

In the present study, a set of reference concentrations of proteins that could be used for SRM analysis was generated according to a meta-analysis of the SRM data for plasma collected from healthy volunteers. Furthermore, the fundamental principles for selecting potential biomarker candidates and designing panels for human health monitoring were formulated.

## 2. Results and Discussion

### 2.1. Meta-Analysis of the SRM Data for Plasma from Healthy Volunteers

Publications for the meta-analysis of the quantitative proteomics data were selected using two strategies. The ScanBious platform [21], which allows classification of research publications from the PubMed database according to keywords and the selection of semantically similar publications based on the integrated semantic similarity analysis system, was employed for the first strategy. Figure 1 shows the search results obtained with the query “Selected (or Multiple) Reaction Monitoring Proteomics Human”.



**Figure 1.** Results of the search for publications reporting the selected/multiple reaction monitoring (SRM/MRM) data for humans over the past ten years (2012–2022) using the ScanBious system. The size of the nodes reflects the occurrence of the keywords in the abstracts. The identified key clusters involve search terms related to tumor biomarkers, their sensitivity, specificity, and reproducibility. Of note, SRM analysis was frequently used to study cancer cell lines, such as Jurkat, MCF-7, and HepG2.

As the internal search algorithm data across the PubMed database included enrichment with synonyms, the retrieved information included data from both SRM and MRM analyses. The list included 197 entries on a specified topic published over the past ten years. The literature primarily comprised studies using the SRM method to measure the concentration of single proteins. Kumar et al. [22] showed that the plasma concentration of apolipoprotein F in healthy individuals was 445.1 ng/mL, with a coefficient of variation (CV) <12%. In many studies, this method was used as a supplementary technique for validating active pharmaceutical substances in human plasma [23–25] and assessing pharmacokinetics [26]. After 2012, the MRM analysis began to be used as a method that allows targeted quantification of plasma proteins, including proteoforms encoded by a single gene (e.g., quantification of the Apoe3 and Apoe4 isoforms differing by one or two amino acid residues [16]). Therefore, these methods are currently used more frequently for detailed investigations of individual proteins and their proteoforms than for quantitative proteome

screening, which is promising for medical applications. Nonetheless, this screening in the gene-centric format was proposed as the experimental method of choice for implementing the International Human Proteome Project, which is of fundamental significance for the evolution of medicine [27].

The second strategy for selecting research publications involved the biocuration of PubMed literature sources. Based on these findings, researchers have been selecting the SRM/MRM method for proteomic analysis less frequently (Figure 2a). One plausible reason is that other methods, such as tandem mass tag (TMT) labeling, which enables multiplex investigation of the samples and determination of the relative and absolute peptide concentrations, are now commonly used [28,29]. Moreover, affine protein enrichment with aptamers on the SOMAscan platform has been implemented, which enables the simultaneous detection of several hundred to thousand proteins [30]. The quality of shotgun proteomics data quantification can also be improved using the data-independent acquisition (DIA) method instead of data-dependent acquisition (DDA) [31].

The comparison of SRM and aptamer-based methods is of particular interest, as aptamer-based methods have been promoted in recent years as a powerful alternative to the “gold standard” antibody-based approach. Table 1 presents the important features of the SOMAscan platform and SRM assays.

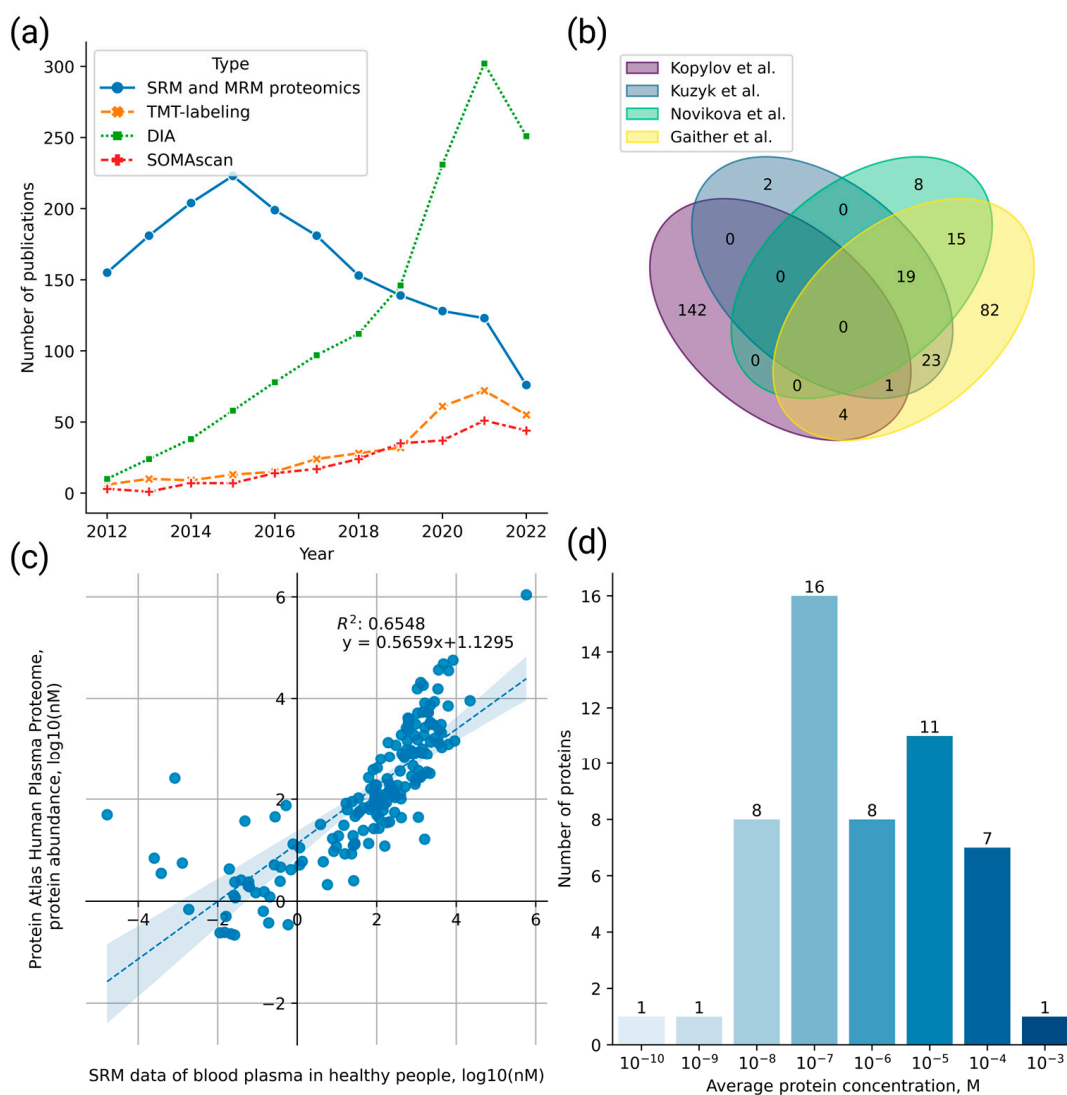
**Table 1.** Comparison of the mass spectrometric SRM (selected reaction monitoring) approach and aptamer-based SOMAscan platform for protein measurement in human serum/plasma.

Method Feature	SOMAscan	SRM
Volume of crude serum/plasma, $\mu$ L	15–50 [2,32]	2.5–5 [33]
Sample preparation	Not needed	Tryptic digestion
Special reagent	DNA-based aptamers	SIS peptides
Multiplexity of protein analytes	813–7288 [2,32,34]	111–329 [33,35]
Quantification type	Relative [36]	Absolute
Direct/indirect protein quantification	DNA serves as an intermediary	Peptide amino acid sequence serves as an intermediary
Sensitivity	$10^{-14}$ – $10^{-3}$ [32]	$10^{-16}$ – $10^{-3}$ [34,35]
Dynamic range in serum/plasma, orders of magnitude	8–10 [34]	4–8 [34,35]
High throughput array multiplexing	Yes	No
Structure analysis, analysis of SAP, isoforms, etc.	No	Yes [17]

A comprehensive comparison revealed that the SRM method has advantages, such as high specificity, sensitivity, low reagent development cost, absolute quantitation capability, and amino acid sequence structural analysis capability. However, rapid analysis of as many individual proteomes as possible is crucial for clinical applications. In this regard, SOMAscan platforms that employ high-throughput microarrays are still more desirable than the relatively slow MRM method. However, problems have recently been observed with aptamer-based cross-platform correlation [36]. Further developments in mass spectrometry, such as ultrafast analysis [37], hold promise for the retention of mass spectrometry on the “waiting list” as an alternative to antibody-based methods.

Finally, despite its high selectivity and sensitivity, the evaluation of SRM data is only particularly automated compared to shotgun proteomics methods. Careful visual assessment of the spectra by qualified personnel is required [38].

As a descending trend (see Figure 2a) exists for the application of the SRM method, most datasets obtained using this method with the current level of analytical sensitivity have already been published and are available in databases [13].



**Figure 2.** Meta-analysis of the SRM data for plasma from healthy volunteers. **(a)** Comparison of the number of published studies employing different proteomic methods in the PubMed database over the past ten years. A descending trend for selected and multiple reaction monitoring over the past years is highlighted in blue. **(b)** Venn diagram based on a comparison of the datasets obtained using the SRM method by Kopylov et al., Novikova et al., Kuzyk et al., and Gaither et al. [18,33,35,39]. **(c)** Correlation between protein concentrations (Log<sub>10</sub> (nM)) measured in the plasma of healthy volunteers using the SRM method in four studies [18,33,35,39] and concentrations of the same proteins reported in the Human Protein Atlas database. **(d)** Range of protein concentrations detected in more than 70% of samples and characterized by the interindividual coefficient of variation <40% obtained via a meta-analysis of studies [18,33,35,39].

Therefore, in the present study, a meta-analysis of publications was performed to generate a summary table of the plasma protein concentrations in healthy volunteers based on the SRM method. The already-known but insufficiently annotated peptide concentrations that vary in the plasma of healthy volunteers can be helpful in further research and searches for biomarkers of various diseases.

Following a review of the literature sources and criteria for dataset inclusion in the meta-analysis, we selected publications reporting the absolute concentrations of at least ten proteins measured in the plasma of healthy volunteers. The results of the meta-analysis of the data obtained by quantitative SRM screening of blood plasma from healthy volunteers are reported in a tabular format (Table S1a), specifying protein identifiers in the UniProt

database, the names of the protein and its encoding gene, the average concentration, and the CV characterizing interindividual variability.

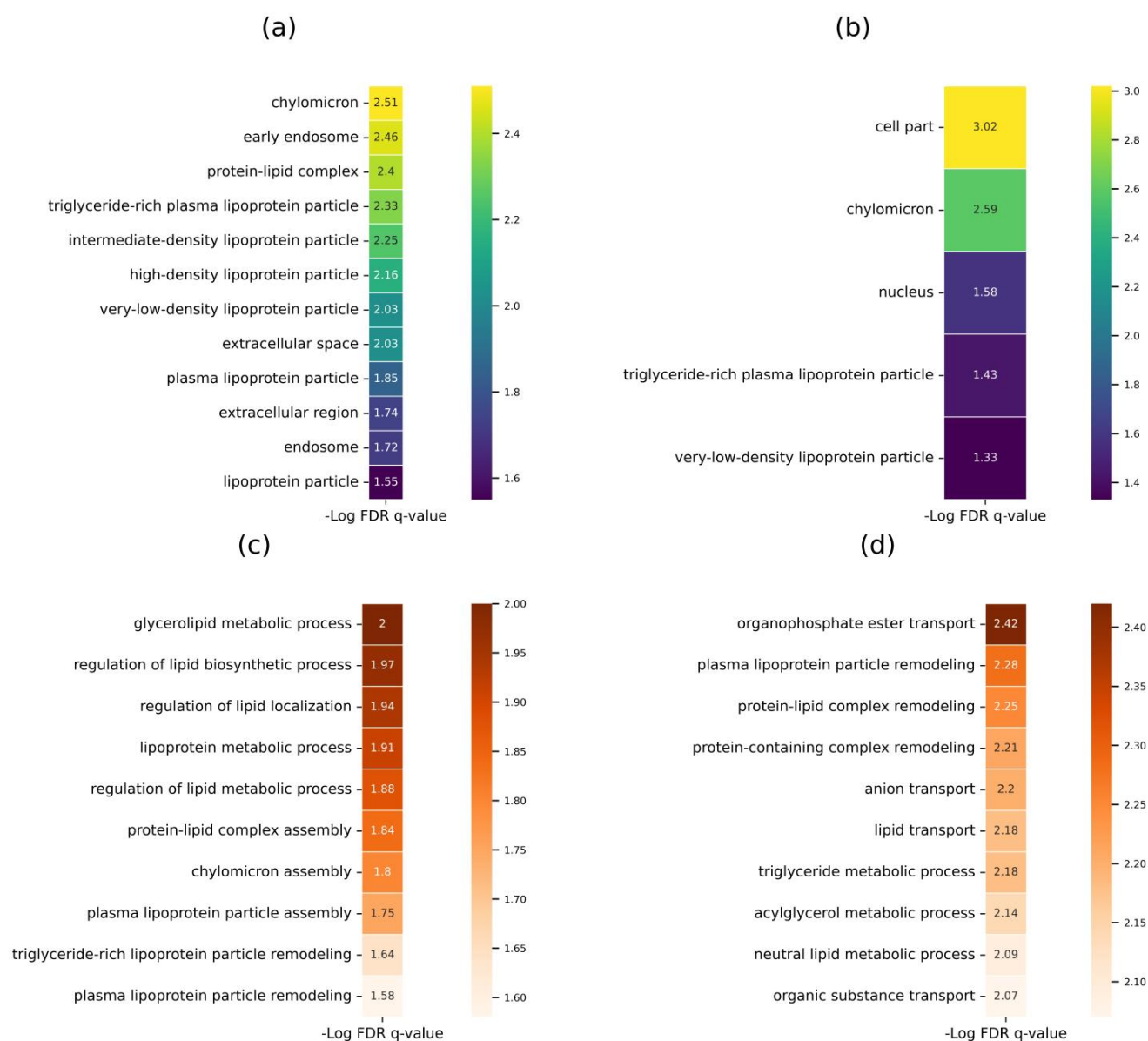
Among all the analyzed publications, only four datasets obtained using the SRM method with synthetic isotope-labeled peptide standards met the specified criteria. Protein concentrations in these studies were obtained by measuring a single synthetic peptide. The first dataset reported by Kopylov et al. [35] contained data on the concentrations of 147 peptides (measured in 54 samples), whose genes belong to chromosomes 18 and 13, and the Y chromosome and mitochondrial DNA. A second publication by Novikova et al. [33] presented data on the concentrations of 42 peptides (measured in 31 samples) belonging to FDA-approved proteins. The third dataset was obtained from Kuzyk et al. [39]. The measured concentrations of 45 proteins in 20 plasma samples were previously reported. The fourth publication by Gaither et al. [18] reported quantitative data on the maximum and minimum concentrations of 144 proteins measured in at least five of the 20 analyzed commercial human plasma samples and one pool. EDTA was used as an anticoagulant to obtain plasma in all studies. Therefore, the comprehensive list contained concentrations of 296 proteins, 20 of which were shared by three datasets, and 42 by two datasets (Figure 2b). By comparing the absolute concentrations of the shared proteins, we identified the maximum and minimum values and calculated their fold changes. For 55 proteins, the range of concentration variations was within one order of magnitude. However, the concentrations of proteins, such as antithrombin III, alpha-2-macroglobulin, complement component C3, transthyretin, transferrin, and apolipoprotein B100, differed by a factor of 10 or more. The comparison results are presented in Table S1b.

The final list containing 296 proteins was compared with the data for 4072 proteins from the Human Protein Atlas project obtained via mass spectrometry based on open-source Peptide Atlas project data. Two datasets shared the 172 values shown in Figure 2c, and the values shown in Table S1c were shared by the two datasets. As shown in Figure 2c, the correlation between the values decreased significantly only at low concentrations ( $\leq 10^{-12}$  M for the Human Protein Atlas project data). The medium and high concentrations (from  $10^{-11}$  to  $10^{-5}$  M) measured according to the number of spectral identifications were comparable to the results obtained using the SRM method. The Human Protein Atlas project data can be used to design the diagnostic panels.

To create the overall list reported herein (Table S1a), we selected the values with the highest CV showing interindividual variability when concentrations were available from at least two sources. Data with high CV values were obtained for numerous samples and may provide a more realistic assessment of the value when selecting potential biomarkers. To identify the proteins with the most stable concentrations, CVs were used to show interindividual differences. We measured the percentage of samples containing a particular protein whenever possible. Proteins found in at least 70% of the analyzed samples with an interindividual CV <40% were considered as a stable characteristic of the proteome of a healthy human.

In the resulting list, only 53 proteins met the formulated criteria and could be considered a stable characteristic of the proteome of a healthy human (Table S1a). Figure 2d illustrates the range of the protein concentrations. The highest concentration was in the range of  $10^{-8}$  to  $10^{-4}$  M. The remaining 243 proteins measured using the SRM method in the four analyzed datasets were examined as a variable group.

Biological data annotation was used to analyze the enrichment of the most stable proteins (core proteins) and variable proteins according to the categories in the Gene Ontology (GO) database (Figure 3).



**Figure 3.** Gene Ontology enrichment analysis of the lists of respective proteins obtained using the SRM method. **(a)** Core proteins ( $N = 53$ ) belonging to the GO category "Cell Component", **(b)** Variable proteins ( $N = 243$ ) belonging to the GO category "Cell Component", **(c)** Core proteins ( $N = 53$ ) belonging to the GO category "Biological Processes", **(d)** Variable proteins ( $N = 243$ ) belonging to the GO category "Biological Processes". Core proteins (the most stable ones) are the proteins found in more than 70% of the samples with  $CV < 40\%$ . Variable proteins are those found in less than 70% of the samples or have a  $CV > 40\%$ .

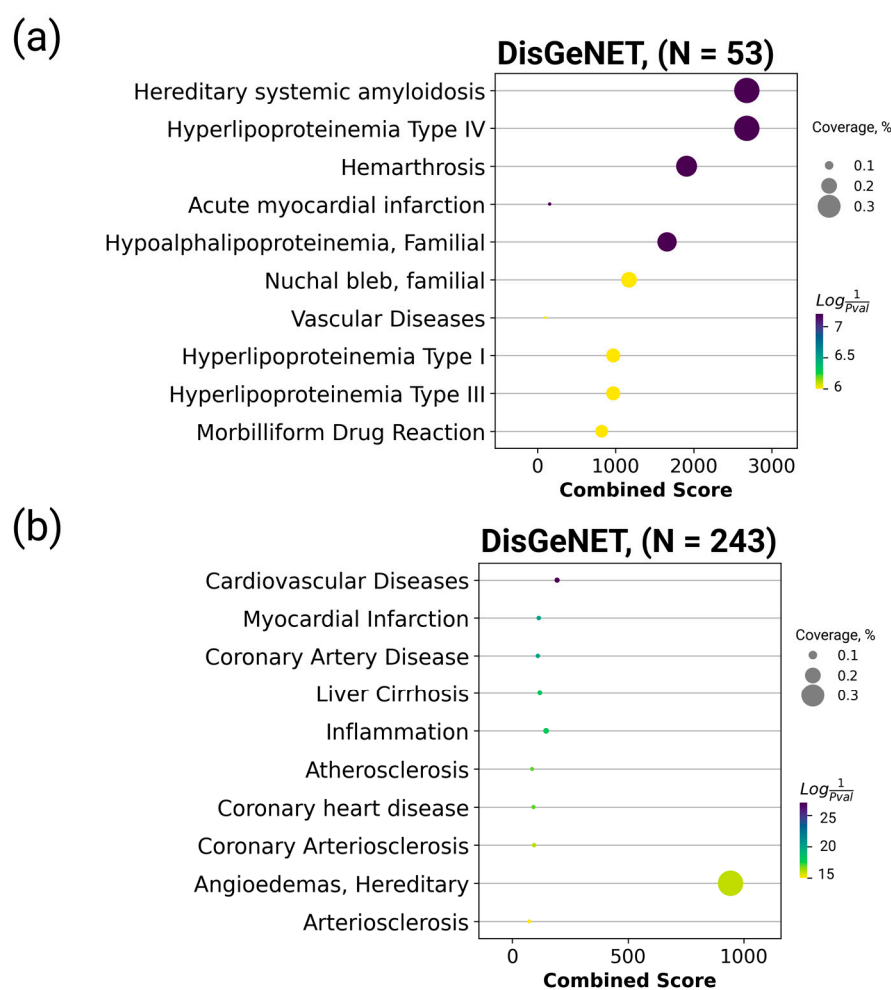
The analysis demonstrated that in terms of their cell localization, the most stable proteins (core proteins) are components of chylomicron with high confidence (FDR  $q$ -value =  $3.1 \times 10^{-3}$ ) (Figure 3A). According to the GO category "Biological Processes," stable proteins were involved in lipid metabolism regulation (Figure 3C), and their primary function was binding receptors of lipoprotein particles (FDR  $q$ -value =  $2.57 \times 10^{-2}$ ).

For the group of variable proteins consisting of 243 proteins, along with cellular localization, proteins were identified to be primarily located in chylomicrons (Figure 3B), and to participate in organophosphate ester transport, lipoprotein particle remodeling (Figure 3D), and in the activity of phosphatidylcholine-sterol O-acyltransferase activator and anion binding (FDR  $q$ -value =  $1.18 \times 10^{-2}$  and  $2.07 \times 10^{-2}$ , respectively).

Data on the enrichment of a stable set of proteins with the lipoprotein fraction can be used to design panels for monitoring lipid metabolism in healthy humans and individuals with a high risk of atherosclerosis (e.g., genetic predisposition, obesity, hypertension, and smoking history).

## 2.2. Panel for Monitoring Human Health Status and Potential Biomarker Selection

Here, we sought to demonstrate how data on the healthy human plasma proteome can be evaluated in the context of diseases. The list of the most stable proteins was annotated using the DisGeNET platform (Figure 4), the largest database of gene–disease associations [40]. The 30 proteins found to be associated with different diseases are listed in Table S1a,d.



**Figure 4.** Enrichment analysis of stable proteins found to be associated with diseases according to the DisGeNET database obtained for (a) the list of the most stable proteins and (b) the variable dataset. The combined score is the decimal logarithm of the  $p$ -value determined using the Fisher’s test and multiplied by the  $z$ -score [41]. “Coverage, %” is the percentage of target protein associated with this disease.

Figure 4 indicates that enrichment was more significant for the set of stable proteins than for variable proteins. Thus, an association was revealed between some stable proteins, hereditary systemic amyloidosis, and type IV hyperlipoproteinemia, which are rare familial disorders. The concentration variation for the set of stable proteins revealed in this study may be rare and may require further study. Proteins with variable concentrations were primarily associated with cardiovascular and inflammatory diseases (Table S1e).



The set of proteins with the most stable concentrations can be used to identify associations with various familial disorders. The composition of potential protein panels designed for MRM studies can be easily modified (e.g., by adding or replacing the protein analytes). Owing to such plasticity, researchers can employ the MRM technology to perform test repositioning (similar to drug repositioning) using well-studied and diagnostically relevant protein analytes in new combinations, which enables the assessment of the biological status for which they have never been used before.

The results of the meta-analysis of protein concentrations in the blood of healthy individuals were compared to the measured concentrations of the same proteins in human plasma for two diagnosed diseases. We selected data obtained using the SRM method for plasma samples collected from patients with cancer and neurological diseases. The first dataset comprised the protein concentrations in plasma samples collected from patients with neurological diseases [42]. The second dataset for comparative analysis was obtained from a study that focused on biomarkers of lung adenocarcinoma [43]. The results are presented in Table 2.

**Table 2.** Comparison of protein concentrations in healthy individuals and patients with pathologies obtained using the multiple reaction monitoring method. The cells where protein concentrations differed by more than two orders of magnitude are shown in blue. The most stable proteins in the plasma of healthy volunteers retrieved via meta-analysis are displayed in bold [42,43].

Protein Name	Gene Name	Log 10 (Average Concentration, fM, $M \times 10^{-15}$ )		
		Healthy Human Plasma	Neurological Diseases ( $n = 19$ , Kiseleva et al., Clin Trans Med, 2015) *	Lung Adenocarcinoma ( $n = 102$ , Wu et al., Proteomics Clin Appl, 2020)
<b>Alpha-1-antitrypsin</b>	<i>A1AT</i>	<b>10.0</b>	<b>9.5</b>	
<b>Alpha-2-macroglobulin</b>	<i>A2MG</i>	<b>10.2</b>	<b>8.5</b>	
<b>Apolipoprotein A-I</b>	<i>APOA1</i>	<b>10.6</b>	<b>9.1</b>	
<i>Ceruloplasmin</i>	<i>CERU</i>	9.1	6.5	
<b>Complement C3</b>	<i>CO3</i>	<b>9.2</b>	<b>8.4</b>	
<i>Cystatin-C</i>	<i>CST3</i>	7.7	7	
<b>Fibrinogen alpha chain</b>	<i>FIBA</i>	<b>10.3</b>	<b>8.9</b>	
<i>Haptoglobin</i>	<i>HPT</i>	9.9	9	
<b>Hemopexin</b>	<i>HEMO</i>	<b>9.4</b>	<b>8.3</b>	
<i>Insulin-like growth factor-binding protein 3</i>	<i>IGFBP3</i>	7.4	7	
<b>Plasma protease C1 inhibitor</b>	<i>IC1</i>	<b>8.6</b>	<b>7.5</b>	
<i>Platelet basic protein</i>	<i>CXCL7</i>	8.0	6.8	
<b>Serotransferrin</b>	<i>TRFE</i>	<b>10.3</b>	<b>9.3</b>	
<i>Serum albumin *</i>	<i>ALB</i>	11.8	10	
<b>Transthyretin</b>	<i>TTR</i>	<b>9.9</b>	<b>8.5</b>	
<b>von Willebrand factor</b>	<i>VWF</i>	<b>7.8</b>	<b>7</b>	
<i>Complement factor H</i>	<i>CFH</i>	9.0		5.8
<b>Desmoglein-2</b>	<i>DSG2</i>	<b>7.0</b>		<b>6.4</b>
<i>Gelsolin, isoform 1</i>	<i>GSN</i>	9.2		6.8
<i>Lambda-crystallin homolog</i>	<i>CRYL1</i>	6.6		6.2
<i>Lumican</i>	<i>LUM</i>	8.5		6.7
<i>Mucin-16</i>	<i>MUC16</i>	7.7		5.8

\* The average concentrations calculated using the data of three protocols were used for the dataset in the publication by Kiseleva et al. Albumin concentration was determined using the values obtained for the protocol without depletion.

As shown in Table 2, proteins with the most stable concentrations in healthy volunteers differed from those in patients with a pathology by no more than two orders of magnitude. The maximum difference (1.7-fold) was demonstrated for alpha-2-macroglobulin, a component of the innate immune system; variation in its concentration was detected

earlier in individuals with neuronal damage [44,45]. A comparison of the remaining values in Table 2 revealed that the concentration of ceruloplasmin (*CERU*) was lower by almost three orders of magnitude in patients with neurological diseases than in healthy volunteers. This protein is known to regulate iron efflux, whereas ceruloplasmin deficiency may cause iron overload and can be associated with neurodegenerative disorders [46,47]. Notably, the concentrations of gelsolin isoform 1 (*GSN*) and complement factor H (*CFH* gene) were significantly reduced in patients with lung adenocarcinoma. Gelsolin is an actin-binding protein that affects cell mobility and maintains cytoskeletal integrity. The expression of gelsolin is variable in some cancer cell lines and malignancies, such as non-small-cell lung and breast cancer [48]. Complement factor H, a component of the complement system, plays a crucial role in tumor immune surveillance [49]. Recent studies have demonstrated that *CFH* can play a pivotal role in the resistance to complement-mediated lysis in different cancer cells, including lung cancer cells [50]. Detection of such drastic variations in concentrations can be an incentive for further research.

For proteins not shown with color in Table 2, concentrations differed by 1–1.5 orders of magnitude. For most of them (displayed in bold), concentration values were also most stable in the plasma of healthy volunteers. Therefore, the data listed in Tables 2 and S1a can be used to select potential single proteins or entire protein arrays for designing diagnostic panels. If the differences in concentrations between the samples of healthy individuals and patients with pathology are minimal, another protein should be selected for further research.

### 3. Materials and Methods

Publications for the analysis were selected using the ScanBious platform [21], which allows publications to be retrieved from the PubMed database upon inquiry and visualizes related keywords as a semantic network. Semantically similar studies can be classified based on an integrated semantic similarity analysis system.

The data from the Human Protein Atlas project [8] obtained using mass spectrometry, available as open-source data of the Peptide Atlas project, were used to compare the absolute protein concentrations. GO enrichment analysis was performed using the GOrilla tool [51].

An analysis to determine the enrichment of the most stable and variable proteins with molecular components associated with the diseases was performed using the Enrichr module of the Gseapy library (v. 0.13.0) across categories of the DisGeNET database (<https://www.disgenet.org/>, accessed on 10 November 2022); the *p*-value cutoff was <0.05 [51]. The ten most robust categories for each protein group (stable and variable proteins) were visualized.

### 4. Conclusions

Absolute quantitative proteomic data are essential for future biomarker discovery and the creation of personal digital molecular images for healthy and sick people. Information on the quantitative proteome of the blood plasma of healthy people will generate a standard for comparison with the proteome of a diseased person, significantly expanding the possibilities of diagnostics via biological markers.

Currently, proteomic methods have limitations in sensitivity, that is, the MS-detector is triggered in the presence of at least one million molecules in 1  $\mu$ L of sample ( $10^{-12}$  M). Perhaps in the future, analytical proteomics may allow us to detect and measure one molecule in 1  $\mu$ L ( $10^{-18}$  M), which will significantly increase the number of analyzed proteins in the framework of health monitoring.

Estimating the ranges of interindividual and individual variability will enable the transfer of medical approaches to personalized diagnostics, risk assessment of diseases, and therapy. In addition to the early diagnosis of diseases and evaluation of treatment efficacy, such data can serve as a basis for identifying new drug targets. Although the trend of targeted proteomics methods is declining, the use of such methods has resulted

in the accumulation of an impressive quantitative data array for various human tissues and organs and an understanding of the difference between proteomic landscapes of healthy and diseased persons. These data can be used for cross-validation of the results obtained with new, more sensitive methods and can become the basis for the formation of hypotheses for subsequent studies. In the present study, we selected a dataset of human plasma quantitative measurements to provide researchers with reference values for protein concentrations. Until enough experiments are accumulated using new methods, SRM data will serve as the optimal basis for estimating variability and reliability. We believe that the efforts of the global community will be directed toward the development of cheaper methods with better analytical sensitivity and reproducibility.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24010769/s1>.

**Author Contributions:** A.A.K., E.V.I., O.I.K. and E.A.P. conceived the paper. A.A.K., S.E.N., E.V.I., O.I.K. and E.A.P. wrote the draft of the manuscript. A.A.K. and S.E.N. prepared the figures. E.V.P., V.G.Z., S.A.M. and V.V.P. reviewed and edited the manuscript. A.V.L. and A.I.A. were responsible for funding acquisition. All authors revised and edited the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The study was performed employing “Avogadro” large-scale research facilities and was financially supported by the Ministry of Education and Science of the Russian Federation, Agreement No. 075-15-2021-933, unique project ID: RF00121X0004.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable. No new data were created or analyzed in this study.

**Acknowledgments:** We would like to thank “Human Proteome” Core Facility of the Institute of Biomedical Chemistry (Moscow, Russia) which provided computational resources for this project.

**Conflicts of Interest:** The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Skates, S.J.; Gillette, M.A.; LaBaer, J.; Carr, S.A.; Anderson, L.; Liebler, D.C.; Ransohoff, D.; Rifai, N.; Kondratovich, M.; Težak, Ž.; et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J. Proteome Res.* **2013**, *12*, 5383–5394. [[CrossRef](#)] [[PubMed](#)]
2. Candia, J.; Cheung, F.; Kotliarov, Y.; Fantoni, G.; Sellers, B.; Griesman, T.; Huang, J.; Stuccio, S.; Zingone, A.; Ryan, B.M.; et al. Assessment of variability in the somascan assay. *Sci. Rep.* **2017**, *7*, 14248. [[CrossRef](#)]
3. Cui, M.; Cheng, C.; Zhang, L. High-throughput proteomics: A methodological mini-review. *Lab. Investig.* **2022**. [[CrossRef](#)] [[PubMed](#)]
4. Anderson, N.L.; Anderson, N.G.; Pearson, T.W.; Borchers, C.H.; Paulovich, A.G.; Patterson, S.D.; Gillette, M.; Aebersold, R.; Carr, S.A. A human proteome detection and quantitation project. *Mol. Cell Proteom.* **2009**, *8*, 883–886. [[CrossRef](#)]
5. Deutsch, E.W.; Eng, J.K.; Zhang, H.; King, N.L.; Nesvizhskii, A.I.; Lin, B.; Lee, H.; Yi, E.C.; Ossola, R.; Aebersold, R. Human Plasma PeptideAtlas. *Proteomics* **2005**, *5*, 3497–3500. [[CrossRef](#)] [[PubMed](#)]
6. Farrah, T.; Deutsch, E.W.; Omenn, G.S.; Campbell, D.S.; Sun, Z.; Bletz, J.A.; Mallick, P.; Katz, J.E.; Malmström, J.; Ossola, R.; et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell Proteom.* **2011**, *10*, M110.006353. [[CrossRef](#)] [[PubMed](#)]
7. Archakov, A.; Lisitsa, A.; Ponomarenko, E.; Zgoda, V. Recent advances in proteomic profiling of human blood: Clinical scope. *Expert Rev. Proteom.* **2015**, *12*, 111–113. [[CrossRef](#)]
8. The Human Protein Atlas. Available online: <https://www.proteinatlas.org/humanproteome/blood+protein/proteins+detected+in+ms> (accessed on 17 October 2022).
9. Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **2010**, *7*, 43–46. [[CrossRef](#)]
10. Kopylov, A.T.; Zgoda, V.G.; Lisitsa, A.V.; Archakov, A.I. Combined use of irreversible binding and MRM technology for low- and ultralow copy-number protein detection and quantitation. *Proteomics* **2013**, *13*, 727–742. [[CrossRef](#)]

11. Vavilov, N.E.; Zgoda, V.G.; Tikhonova, O.V.; Farafonova, T.E.; Shushkova, N.A.; Novikova, S.E.; Yarygin, K.N.; Radko, S.P.; Ilgisonis, E.V.; Ponomarenko, E.A.; et al. Proteomic analysis of chr 18 proteins using 2D fractionation. *J. Proteome Res.* **2020**, *19*, 4901–4906. [[CrossRef](#)]
12. Vavilov, N.E.; Ilgisonis, E.V.; Lisitsa, A.V.; Ponomarenko, E.A.; Farafonova, T.E.; Tikhonova, O.V.; Zgoda, V.G.; Archakov, A.I. Deep proteomic dataset of human liver samples obtained by two-dimensional sample fractionation coupled with tandem mass spectrometry. *Data Brief* **2022**, *42*, 108055. [[CrossRef](#)] [[PubMed](#)]
13. Vavilov, N.; Ilgisonis, E.; Lisitsa, A.; Ponomarenko, E.; Farafonova, T.; Tikhonova, O.; Zgoda, V.; Archakov, A. Number of detected proteins as the function of the sensitivity of proteomic technology in human liver cells. *Curr. Protein Pept. Sci.* **2022**. [[CrossRef](#)]
14. Hüttenhain, R.; Soste, M.; Selevsek, N.; Röst, H.; Sethi, A.; Carapito, C.; Farrah, T.; Deutsch, E.W.; Kusebauch, U.; Moritz, R.L.; et al. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* **2012**, *4*, 142ra94. [[CrossRef](#)] [[PubMed](#)]
15. Domanski, D.; Percy, A.J.; Yang, J.; Chambers, A.G.; Hill, J.S.; Freue, G.V.C.; Borchers, C.H. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* **2012**, *12*, 1222–1243. [[CrossRef](#)] [[PubMed](#)]
16. Simon, R.; Girod, M.; Fonbonne, C.; Salvador, A.; Clément, Y.; Lantéri, P.; Amouyel, P.; Lambert, J.C.; Lemoine, J. Total ApoE and ApoE4 isoform assays in an Alzheimer's disease case-control study by targeted mass spectrometry (n = 669): A pilot assay for methionine-containing proteotypic peptides. *Mol. Cell Proteom.* **2012**, *11*, 1389–1403. [[CrossRef](#)] [[PubMed](#)]
17. Kliuchnikova, A.A.; Goncharov, A.O.; Levitsky, L.I.; Pyatnitskiy, M.A.; Novikova, S.E.; Kuznetsova, K.G.; Ivanov, M.V.; Ilina, I.Y.; Farafonova, T.E.; Zgoda, V.G.; et al. Proteome-Wide Analysis of ADAR-Mediated Messenger RNA Editing during Fruit Fly Ontogeny. *J. Proteome Res.* **2020**, *19*, 4046–4060. [[CrossRef](#)] [[PubMed](#)]
18. Gaither, C.; Popp, R.; Mohammed, Y.; Borchers, C.H. Determination of the concentration range for 267 proteins from 21 lots of commercial human plasma using highly multiplexed multiple reaction monitoring mass spectrometry. *Analyst* **2020**, *145*, 3634–3644. [[CrossRef](#)] [[PubMed](#)]
19. Balashova, E.E.; Lokhov, P.G.; Ponomarenko, E.A.; Markin, S.S.; Lisitsa, A.V.; Archakov, A.I. Metabolomic diagnostics and human digital image. *Per. Med.* **2019**, *16*, 133–144. [[CrossRef](#)]
20. Geyer, P.E.; Voytik, E.; Treit, P.V.; Doll, S.; Kleinhempel, A.; Niu, L.; Müller, J.B.; Buchholtz, M.-L.; Bader, J.M.; Teupser, D.; et al. Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **2019**, *11*, e10427. [[CrossRef](#)]
21. Tarbeeva, S.; Lyamtseva, E.; Lisitsa, A.; Kozlova, A.; Ponomarenko, E.; Ilgisonis, E. Scanbious: Survey for obesity genes using pubmed abstracts and disgenet. *J. Pers. Med.* **2021**, *11*, 246. [[CrossRef](#)]
22. Kumar, A.; Gangadharan, B.; Zitzmann, N. Multiple reaction monitoring and multiple reaction monitoring cubed based assays for the quantitation of apolipoprotein F. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2016**, *1033–1034*, 278–286. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, D.; Teng, Y.; Chen, K.; Liu, S.; Wei, C.; Wang, B.; Yuan, G.; Zhang, R.; Liu, X.; Guo, R. Determination of salbutamol in human plasma and urine using liquid chromatography coupled to tandem mass spectrometry and its pharmacokinetic study. *Biomed. Chromatogr.* **2012**, *26*, 1176–1182. [[CrossRef](#)]
24. Guo, B.; Zhang, J.; Yu, J.; Wu, X.; Shi, Y.; Tsai, C.-Y. A liquid chromatography-tandem mass spectrometry assay for the determination of nemonoxacin (TG-873870), a novel nonfluorinated quinolone, in human plasma and urine and its application to a single-dose pharmacokinetic study in healthy Chinese volunteers. *Biomed. Chromatogr.* **2012**, *26*, 1333–1340. [[CrossRef](#)] [[PubMed](#)]
25. Patel, D.S.; Sharma, N.; Patel, M.C.; Patel, B.N.; Shrivastav, P.S.; Sanyal, M. Application of a rapid and sensitive liquid chromatography-tandem mass spectrometry method for determination of bumetanide in human plasma for a bioequivalence study. *J. Pharm. Biomed. Anal.* **2012**, *66*, 365–370. [[CrossRef](#)] [[PubMed](#)]
26. Singhal, P.; Yadav, M.; Winter, S.; Guttikar, S.; Patel, D.; Mills, M.; Shrivastav, P.S. Enantiomeric separation of verapamil and its active metabolite, norverapamil, and simultaneous quantification in human plasma by LC-ESI-MS-MS. *J. Chromatogr. Sci.* **2012**, *50*, 839–848. [[CrossRef](#)]
27. Archakov, A.I.; Aseev, A.L.; Bykov, V.A.; Grigoriev, A.I.; Govorun, V.M.; Ilgisonis, E.V.; Ivanov, Y.D.; Ivanov, V.T.; Kiseleva, O.I.; Kopylov, A.T.; et al. Challenges of the Human Proteome Project: 10-Year Experience of the Russian Consortium. *J. Proteome Res.* **2019**, *18*, 4206–4214. [[CrossRef](#)] [[PubMed](#)]
28. Hutchinson-Bunch, C.; Sanford, J.A.; Hansen, J.R.; Gritsenko, M.A.; Rodland, K.D.; Piehowski, P.D.; Qian, W.-J.; Adkins, J.N. Assessment of TMT Labeling Efficiency in Large-Scale Quantitative Proteomics: The Critical Effect of Sample pH. *ACS Omega* **2021**, *6*, 12660–12666. [[CrossRef](#)] [[PubMed](#)]
29. Stopfer, L.E.; Mesfin, J.M.; Joughin, B.A.; Lauffenburger, D.A.; White, F.M. Multiplexed relative and absolute quantitative immunopeptidomics reveals MHC I repertoire alterations induced by CDK4/6 inhibition. *Nat. Commun.* **2020**, *11*, 2760. [[CrossRef](#)]
30. Joshi, A.; Mayr, M. In Aptamers They Trust: The Caveats of the SOMAscan Biomarker Discovery Platform from SomaLogic. *Circulation* **2018**, *138*, 2482–2485. [[CrossRef](#)]
31. Brown, L.M. Quantitative shotgun proteomics with data-independent acquisition and traveling wave ion mobility spectrometry: A versatile tool in the life sciences. *Adv. Exp. Med. Biol.* **2014**, *806*, 79–91. [[CrossRef](#)]

32. Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E.N.; Carter, J.; Dalby, A.B.; Eaton, B.E.; Fitzwater, T.; et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **2010**, *5*, e15004. [[CrossRef](#)] [[PubMed](#)]
33. Novikova, S.E.; Farafonova, T.E.; Tikhonova, O.V.; Shushkova, N.A.; Pyatnitsky, M.A.; Zgoda, V.G.; Ponomarenko, E.A.; Lisitsa, A.V.; Grigoryev, A.I.; Tutelyan, V.A.; et al. Mass-spectrometric MRM analysis of FDA-verified proteins in the blood plasma of healthy volunteers. *Biomed. Khim.* **2020**, *66*, 294–316. [[CrossRef](#)] [[PubMed](#)]
34. Candia, J.; Daya, G.N.; Tanaka, T.; Ferrucci, L.; Walker, K.A. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci Rep.* **2022**, *12*, 17147. [[CrossRef](#)]
35. Kopylov, A.T.; Ponomarenko, E.A.; Ilgisonis, E.V.; Pyatnitskiy, M.A.; Lisitsa, A.V.; Poverennaya, E.V.; Kiseleva, O.I.; Farafonova, T.E.; Tikhonova, O.V.; Zavalova, M.G.; et al. 200+ protein concentrations in healthy human blood plasma: Targeted quantitative SRM SIS screening of chromosomes 18, 13, Y, and the mitochondrial chromosome encoded proteome. *J. Proteome Res.* **2019**, *18*, 120–129. [[CrossRef](#)] [[PubMed](#)]
36. Raffield, L.M.; Dang, H.; Pratte, K.A.; Jacobson, S.; Gillenwater, L.A.; Ampleford, E.; Barjaktarevic, I.; Basta, P.; Clish, C.B.; Comellas, A.P.; et al. NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics*. **2020**, *20*, e1900278. [[CrossRef](#)] [[PubMed](#)]
37. Ivanov, M.V.; Bubis, J.A.; Gorshkov, V.; Tarasova, I.A.; Levitsky, L.I.; Lobas, A.A.; Solovyeva, E.M.; Pridatchenko, M.L.; Kjeldsen, F.; Gorshkov, M.V. DirectMS1: MS/MS-Free Identification of 1000 Proteins of Cellular Proteomes in 5 Minutes. *Anal Chem.* **2020**, *92*, 4326–4333. [[CrossRef](#)] [[PubMed](#)]
38. Filimonov, A.; Kopylov, A.; Lisitsa, A.; Archakov, A. Manual method of visually identifying candidate signals for a targeted peptide. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2018**, *1083*, 258–270. [[CrossRef](#)] [[PubMed](#)]
39. Kuzyk, M.A.; Smith, D.; Yang, J.; Cross, T.J.; Jackson, A.M.; Hardie, D.B.; Anderson, N.L.; Borchers, C.H. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell Proteomics* **2009**, *8*, 1860–1877. [[CrossRef](#)]
40. Piñero, J.; Saüch, J.; Sanz, F.; Furlong, L.I. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2960–2967. [[CrossRef](#)]
41. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma’ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [[CrossRef](#)]
42. Kiseleva, O.I.; Romashova, Y.A.; Moskaleva, N.E.; Petushkova, N.A.; Teryaeva, N.B.; Belyaev, A.Y.; Lisitsa, A.V. Plasma preparation to measure FDA-approved protein markers by selected reaction monitoring. *Clin. Transl. Med.* **2015**, *4*, 32. [[CrossRef](#)] [[PubMed](#)]
43. Wu, H.-Y.; Pan, Y.-Y.; Kopylov, A.T.; Zgoda, V.; Ma, M.-C.; Wang, C.-H.; Su, W.-C.; Lai, W.-W.; Cheng, P.-N.; Liao, P.-C. Assessment of Serological Early Biomarker Candidates for Lung Adenocarcinoma by using Multiple Reaction Monitoring-Mass Spectrometry. *Proteomics Clin. Appl.* **2020**, *14*, e1900095. [[CrossRef](#)] [[PubMed](#)]
44. Arandjelovic, S.; Dragojlovic, N.; Li, X.; Myers, R.R.; Campana, W.M.; Gonias, S.L. A derivative of the plasma protease inhibitor alpha(2)-macroglobulin regulates the response to peripheral nerve injury. *J. Neurochem.* **2007**, *103*, 694–705. [[CrossRef](#)] [[PubMed](#)]
45. Varma, V.R.; Varma, S.; An, Y.; Hohman, T.J.; Seddighi, S.; Casanova, R.; Beri, A.; Dammer, E.B.; Seyfried, N.T.; Pletnikova, O.; et al. Alpha-2 macroglobulin in Alzheimer’s disease: A marker of neuronal injury through the RCAN1 pathway. *Mol. Psychiatry* **2017**, *22*, 13–23. [[CrossRef](#)] [[PubMed](#)]
46. Texel, S.J.; Xu, X.; Harris, Z.L. Ceruloplasmin in neurodegenerative diseases. *Biochem. Soc. Trans.* **2008**, *36*, 1277–1281. [[CrossRef](#)] [[PubMed](#)]
47. Vassiliev, V.; Harris, Z.L.; Zatta, P. Ceruloplasmin in neurodegenerative diseases. *Brain Res. Brain Res. Rev.* **2005**, *49*, 633–640. [[CrossRef](#)]
48. Yang, J.; Tan, D.; Asch, H.L.; Swede, H.; Bepler, G.; Geradts, J.; Moysich, K.B. Prognostic significance of gelsolin expression level and variability in non-small cell lung cancer. *Lung Cancer* **2004**, *46*, 29–42. [[CrossRef](#)]
49. Ajona, D.; Castaño, Z.; Garayoa, M.; Zudaire, E.; Pajares, M.J.; Martinez, A.; Cuttitta, F.; Montuenga, L.M.; Pio, R. Expression of complement factor H by lung cancer cells: Effects on the activation of the alternative pathway of complement. *Cancer Res.* **2004**, *64*, 6310–6318. [[CrossRef](#)]
50. Cui, T.; Chen, Y.; Knösel, T.; Yang, L.; Zöllner, K.; Galler, K.; Berndt, A.; Mihlan, M.; Zipfel, P.F.; Petersen, I. Human complement factor H is a novel diagnostic marker for lung adenocarcinoma. *Int. J. Oncol.* **2011**, *39*, 161–168. [[CrossRef](#)]
51. Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **2009**, *10*, 48. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.