






## Article

# Absolute Quantification of Pan-Cancer Plasma Proteomes Reveals Unique Signature in Multiple Myeloma

David Kotol <sup>1,2</sup>, Jakob Woessmann <sup>1,2</sup>, Andreas Hober <sup>1,2</sup>, María Bueno Álvarez <sup>1,2</sup>, Khue Hua Tran Minh <sup>1,2</sup>, Fredrik Pontén <sup>3</sup>, Linn Fagerberg <sup>1,2</sup>, Mathias Uhlén <sup>1,2</sup> and Fredrik Edfors <sup>1,2,\*</sup>

<sup>1</sup> Science For Life Laboratory, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden; david.kotol@scilifelab.se (D.K.); jakob.woessmann@scilifelab.se (J.W.); andreas.hober@scilifelab.se (A.H.); maria.bueno@scilifelab.se (M.B.Á.); khue.hua@scilifelab.se (K.H.T.M.); linn.fagerberg@scilifelab.se (L.F.); mathias.uhlen@scilifelab.se (M.U.)

<sup>2</sup> Department of Protein Science, Division of Systems Biology, School of Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

<sup>3</sup> Rudbeck Laboratory, Uppsala University, 752 36 Uppsala, Sweden; fredrik.ponten@igp.uu.se

\* Correspondence: fredrik.edfors@scilifelab.se

**Simple Summary:** A precise mass spectrometry-based method was utilized to study proteins in the blood samples of over a thousand cancer patients. By accurately identifying and measuring protein levels using mass spectrometry, we focused on multiple myeloma and found potential markers for diagnosing the disease. These markers, including the complement C1 complex, JCHAIN, and CD5L, were combined in a prediction model with high accuracy for identifying multiple myeloma patients. Our findings could significantly impact cancer research by improving diagnostic tools.

**Abstract:** Mass spectrometry based on data-independent acquisition (DIA) has developed into a powerful quantitative tool with a variety of implications, including precision medicine. Combined with stable isotope recombinant protein standards, this strategy provides confident protein identification and precise quantification on an absolute scale. Here, we describe a comprehensive targeted proteomics approach to profile a pan-cancer cohort consisting of 1800 blood plasma samples representing 15 different cancer types. We successfully performed an absolute quantification of 253 proteins in multiplex. The assay had low intra-assay variability with a coefficient of variation below 20% (CV = 17.2%) for a total of 1013 peptides quantified across almost two thousand injections. This study identified a potential biomarker panel of seven protein targets for the diagnosis of multiple myeloma patients using differential expression analysis and machine learning. The combination of markers, including the complement C1 complex, JCHAIN, and CD5L, resulted in a prediction model with an AUC of 0.96 for the identification of multiple myeloma patients across various cancer patients. All these proteins are known to interact with immunoglobulins.

**Keywords:** DIA; multiple myeloma; precision medicine; targeted proteomics



**Citation:** Kotol, D.; Woessmann, J.; Hober, A.; Álvarez, M.B.; Tran Minh, K.H.; Pontén, F.; Fagerberg, L.; Uhlén, M.; Edfors, F. Absolute Quantification of Pan-Cancer Plasma Proteomes Reveals Unique Signature in Multiple Myeloma. *Cancers* **2023**, *15*, 4764. <https://doi.org/10.3390/cancers15194764>

Academic Editor: Taketo Yamada

Received: 17 June 2023

Revised: 31 August 2023

Accepted: 20 September 2023

Published: 28 September 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

Cancer accounts for more than ten million deaths worldwide and is considered the second most common cause of mortality today. It consists of more than 200 different subgroups, making it a very heterogeneous disease. However, disease progression typically follows the same trajectory in diverse cancers. These general features of the disease can be described as events in which cells undergo dysregulated and autonomous growth, eventually disrespecting tissue boundaries, leading to metastasis [1]. Molecular events resulting from alterations in the genome of cancer cells are critical components for identifying cancer type-specific biomarkers and have been studied and mapped for several decades [2–4]. Recent advances in high-throughput sequencing technologies have enabled large-scale efforts to accurately map genomic alterations across many different types and subtypes of

cancer [5–7]. This has paved the way to study altered gene expression and immunogenic reactions, which provides new opportunities for developing biomarker panels to cover the need for early and more precise cancer diagnostics [8].

Blood plasma and the analysis of its constituents is the most general diagnostic procedure in modern medicine. It is easily accessible and provides essential information about the healthy physiological state and disease, such as cancer-induced alterations in the human body [9,10]. Alterations in gene expression offer opportunities for the early detection of circulating cancer biomarkers, which can increase the chances of long-term survival and provide a more precise diagnosis. The levels of circulating cancer biomarkers such as the prostate-specific antigen (PSA) are well studied as indicators of tissue-specific growth or damage but lack clinical specificity for cancer diagnosis [11]. This presents opportunities for novel multiplex technologies providing a way to modernize tomorrow's clinical laboratory tests by focusing on biomarker panels with high cancer specificity. Recently, transcriptome analysis performed on cell-free RNA collected from patients with various cancers has revealed that panels of biomarkers can be used to effectively subclassify cancer types [12]. In contrast to sequencing approaches, mass spectrometry-based proteomics performed on liquid biopsies is one of the most powerful strategies for quantifying proteins in multiplex. It has rapidly developed alongside widely used sequencing technologies [13]. Liquid chromatography coupled with mass spectrometry (LC-MS/MS)-based targeted proteomics has emerged as an attractive alternative to antibody-based immunoassays due to its accurate and precise measurements of protein concentrations in complex sample matrices [14,15]. Its quantitative performance can be further enhanced by adding stable isotope standard protein epitope signature tags (SIS-PrESTs) [16] which enable absolute protein measurements needed for clinical diagnostics [17].

In this study, we performed a targeted proteomics analysis of 1800 human plasma samples utilizing 276 SIS-PrESTs towards 253 proteins providing a comprehensive plasma proteome map of 15 different cancer types. We present a view of the molecular phenotypes that distinguish patients with different cancers based on their signature plasma protein levels. The strategy for precise protein quantification was deployed utilizing SIS-PrESTs in combination with data-independent acquisition (DIA). Medium- to high-abundant blood plasma proteins were absolutely quantified, including 40 Food and Drug Administration (FDA)-approved markers. Using differential expression analysis and machine learning, the complement C1 complex, JCHAIN, and CD5L were identified as potential biomarkers for diagnosing multiple myeloma (MM).

## 2. Results

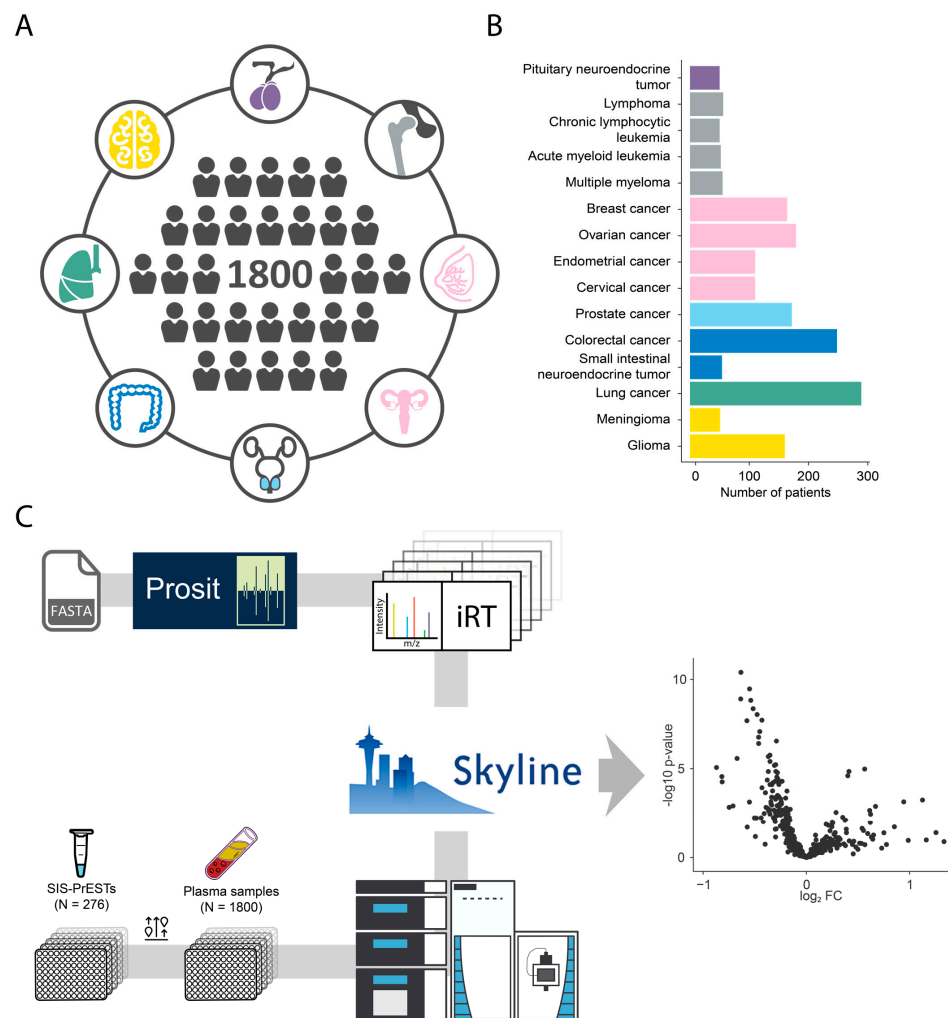
### 2.1. Cohort and Analytical Strategy

Blood plasma samples were collected from 1800 patients diagnosed with one out of eight cancers (Figure 1A). All samples were provided by the Uppsala–Umeå Comprehensive Cancer Consortium (UCAN) biobank. The broader cancer classification can be further stratified into fifteen cancer types, ranging from the most prevalent lung cancer ( $n = 289$ ) followed by colorectal cancer ( $n = 248$ ) down to the least prevalent pituitary neuroendocrine tumor and chronic lymphocytic leukemia ( $n = 50$ ) (Figure 1B). All liquid biopsies were subjected to the same targeted proteomics bottom-up analytical workflow previously described [18] and were spiked with a mixture of 276 SIS-PrESTs (Supplementary Table S1) representing the same number of proteins (Figure 1C).

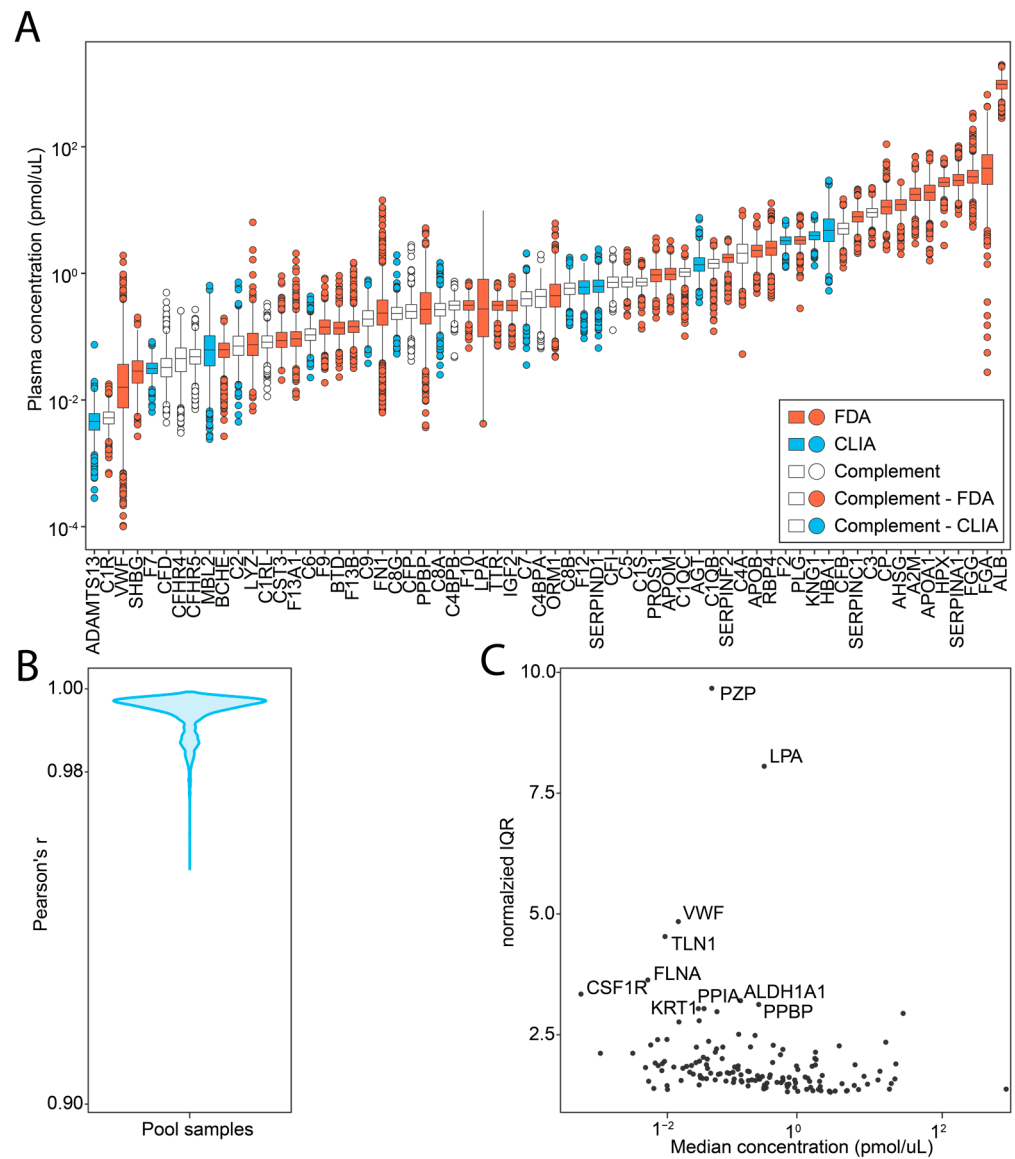
### 2.2. Investigated Targets and Analytical Performance

An assay covering 1013 peptides from 253 proteins was established using spiked SIS-PrESTs in a plasma background. In total, 146 proteins were absolutely quantified spanning more than six orders of magnitude in concentration range (Figure S1) with 395 SIS-PrEST peptides. Quantified proteins included 40 FDA- and 16 Clinical Laboratory Improvement Amendments (CLIA)-approved clinical targets. Moreover, 23 protein members of the complement cascade were part of the analysis (Figure 2A). Within the quantified proteins,

128 (51%) were actively secreted proteins in blood according to the annotation of the human secretome [19]. To assess the performance and reproducibility of the targeted assay, we determined the intra-assay variation using pool samples randomly distributed onto all plates. The median intra-assay coefficient of variation (CV) was between 7.16 and 20% CVs (median CV = 11.3%) per plate and 17.2% CV across all 31 plates. The pool samples across all plates were highly correlated with a median Pearson's  $r$  of 0.99 (Figure 2B). The overall biological variation between all subjects was low, with a median normalized IQR = 1.7 (Figure 2C). This signifies that even with the large variety of cancer types, most of the absolute protein variation was less than two-fold when compared between different cancers. However, two proteins with interindividual variation were observed. Those were pregnancy zone protein (PZP) and apolipoprotein(a) (LPA). The large difference in PZP was observed due to its 10-fold higher concentration in the female population compared to males. The LPA variability was caused by its quantification peptide overlapping with a repeated kringle domain whose count is dependent on the individual's genotype, as reported in [20]. Overall, it was possible to report absolute concentration measurements of plasma proteins with low bias with respect to their levels.



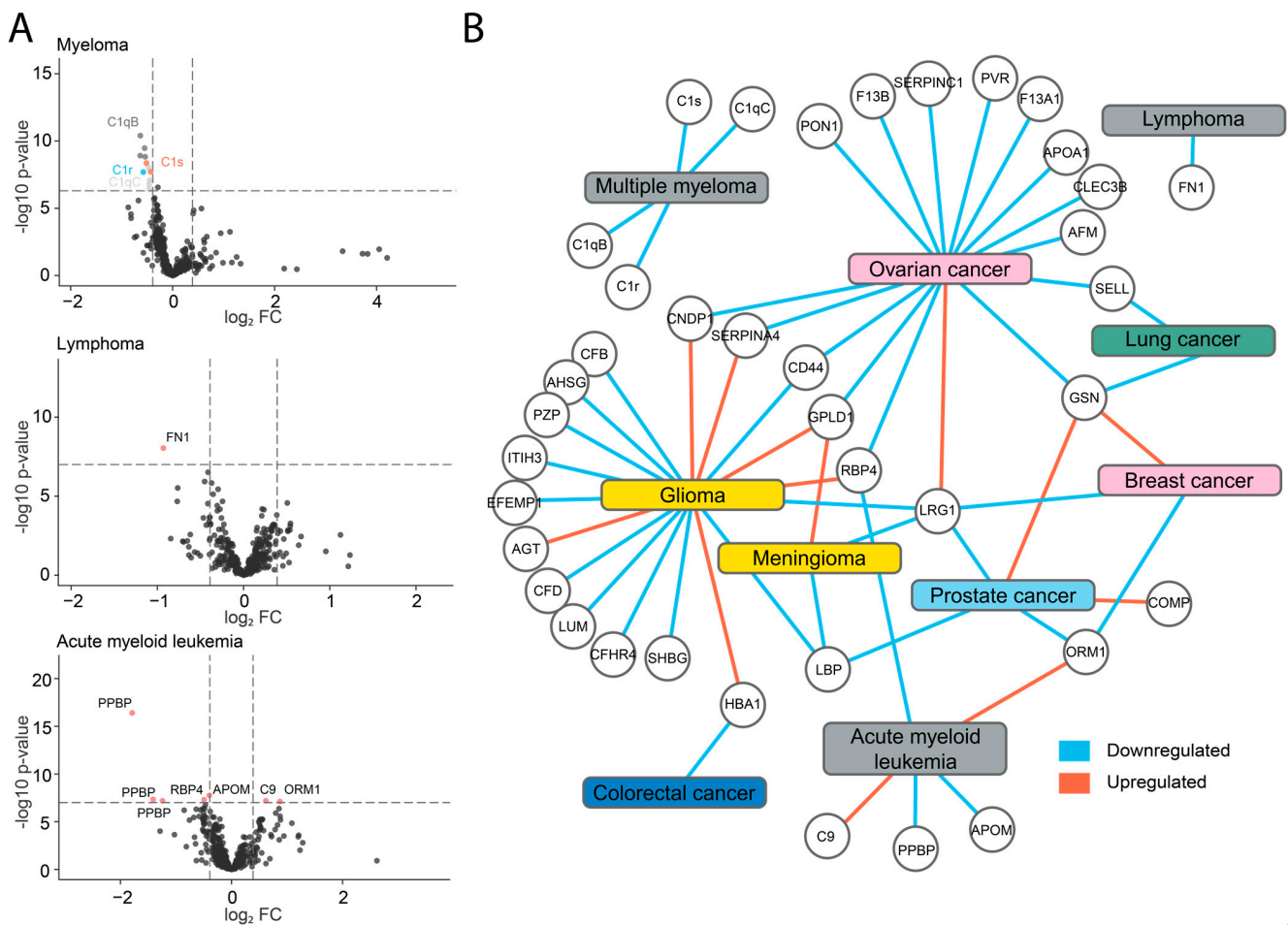
**Figure 1.** Overview of the pan-cancer cohort and study design. (A) Cohort of investigated cancer patients ( $n = 1800$ ) distributed across eight different organs (each color highlights tissue origin of the cancer). (B) Distribution of patient samples ( $n = 1800$ ) across fifteen different cancer subtypes. (C) Overview of analytical workflow, which includes library generation (top row) and flow of patient samples with spike in SIS-PrESTs (bottom row) prior to chromatography extraction and data analysis (middle).



**Figure 2.** Proteins quantified in the targeted proteomics workflow. (A) The dynamic range and concentration (y-axis) of all FDA (red) and CLIA (blue) markers, supplemented with complement system proteins (white) (x-axis). (B) Density distribution of the cross-correlation (Pearson's *r*) between all pooled technical replicates in the dataset. (C) Inter-individual variation in the human plasma proteome visualized as the normalized interquartile range (IQR, y-axis) plotted versus the median protein concentration (pmol/ $\mu$ L, x-axis).

### 2.3. Identification of Potential Biomarkers

Differential expression analysis was performed comparing each cancer type against the rest on the peptide level (Figure 3A). In cases of male- and female-specific cancers, only the patients of the relevant sexes were compared. We could observe that in cases where proteins were quantified with multiple peptides, all of them were significantly downregulated or upregulated. Here, cases such as the platelet basic protein (PPBP) in acute myeloid leukemia, which was previously identified as a dysregulated hub gene, can be highlighted [21]. Other examples are C1qB and C1qC, for which four peptides were identified as significantly downregulated in MM. All the targets identified as differentially expressed have been summarized in a protein network (Figure 3B). In further investigation, we focused on the unique protein pattern of MM patients, which formed an isolated island of four proteins, all part of the complement C1 complex.



**Figure 3.** Identification of differentially expressed proteins in the pan-cancer cohort. (A) Comparisons of plasma protein levels between myeloma, lymphoma, and acute myeloid leukemia versus all other cancers group (Student's *t*-test), visualized in volcano plots with labels on the topmost significant proteins. (B) Network visualization of all protein targets identified as differentially expressed in one cancer compared to all the other cancers ( $p$ -value < 0.0005, Bonferroni adj.). Blue and red connections signify up- or downregulation, respectively. Each cancer is colored by tissue of origin.

#### 2.4. Downregulation of Components of the C1 Complex in Multiple Myeloma

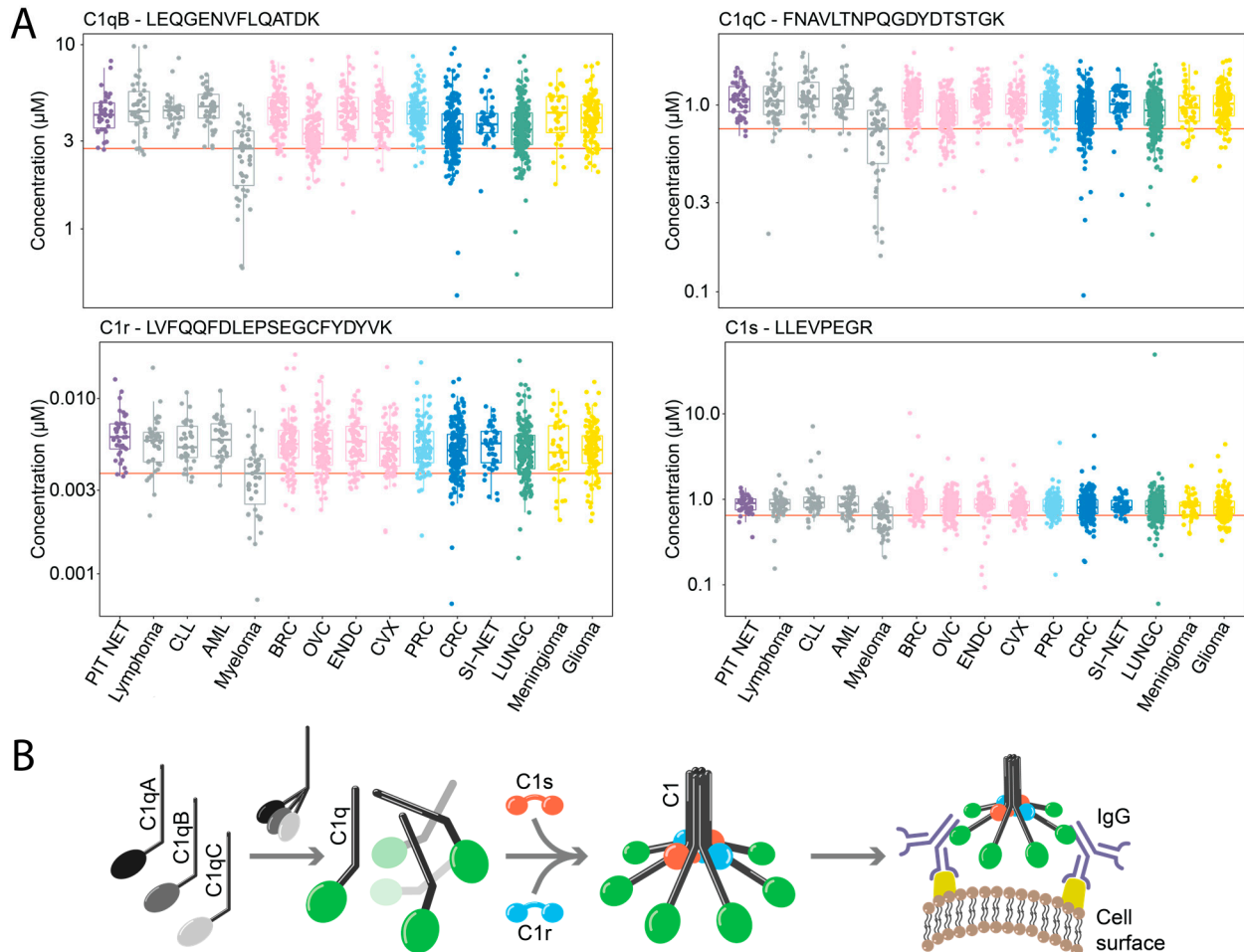
Multiple myeloma, characterized by its heterogeneous nature as a hematologic malignancy, displays dysregulation within the plasma proteome due to the accumulation of plasma cells within the bone marrow, ultimately displacing healthy blood cells. In the context of this study, a profound alteration in the plasma proteome of multiple myeloma patients was observed in the label free MS data acquired alongside the absolute quantification MS data. Notably, this alteration stemmed from a widespread reduction in peptides linked to IGHM and IGHA1 as seen in the volcano plot (Figure S2A). The label free data also showed a couple of non-IgG related proteins that undergo dysregulation in multiple myeloma, among them being albumin and APOA1, aligning with established literature [22].

The diagnosis of multiple myeloma relies on the detection of a monoclonal spike (M spike), often originating from lambda or kappa light variable chains detected by electrophoresis. Interestingly, the kappa variable chain KV37 exhibited the most substantial surge, accompanied by a fold change of 82.6. However, it is essential to note that this elevation of a singular light chain was not universally present among all patients, and did therefore not reach above the statistical threshold.

Differential expression analysis revealed a significant decrease in plasma levels of four components of the complement C1 complex in MM patients (Figure 4A), namely the



proteins C1qB, C1qC, C1r and C1s. Interestingly, this observation was specific to MM and was not detected in any other cancer, including the three immune cell malignancies: lymphoma, acute myeloid leukemia (AML), and chronic lymphocytic leukemia (CLL). We could observe this effect in our label free data as well (Figure S2B).



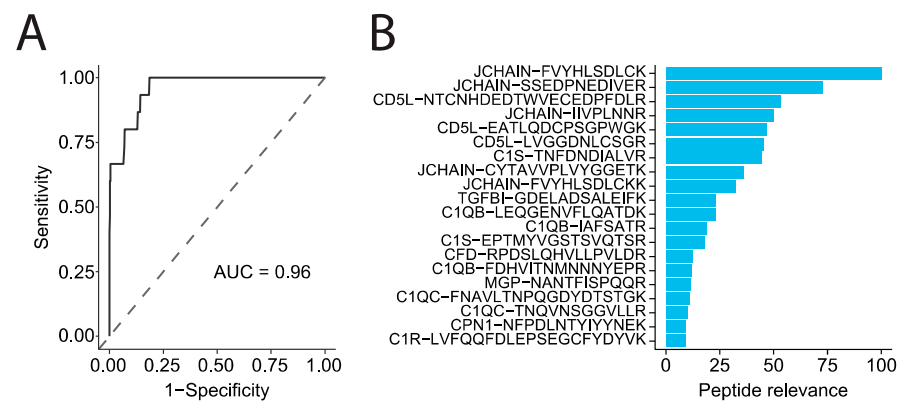
**Figure 4.** Quantification of the complement system and its circulating components across 15 different cancers. (A) Boxplots that visualize the protein levels (y-axis, log<sub>2</sub> concentration) of C1qB, C1qC, C1r and C1s across all patients (n = 1800), grouped by cancer type (y-axis) and colored by tissue origin. (B) General overview of the C1 complex with genes involved in its architecture.

The C1 complex, as part of the innate immune system, initializes the classical complement pathway activation. It consists of five components, C1q built up from C1qA (not quantified), C1qB and C1qC [23]. Further components are peptidase C1r and serine protease C1s. Complement activation occurs after the binding of the globular domain of C1q to target molecules, including IgM and IgG (Figure 4B). The binding of C1q initializes the activation of C1r, which in turn leads to C1s activation. The activated C1s initializes the following proteolytic complement cascade, leading to cell lysis, the activation of phagocytes, and the induction of inflammation [23]. The complement system and its activation or suppression have been related to pro- as well as anti-tumoral effects in a wide variety of cancers [24].

### 2.5. Decreased JCHAIN and CD5L Plasma Levels Distinguish Multiple Myeloma

To further identify the unique patterns within the plasma proteome of MM patients we trained a model based on a random forest algorithm to predict disease outcome. We could distinguish MM patients from all other cancer diagnoses with high confidence based on their plasma protein signature (AUC = 0.96) (Figure 5A). Here, the model identified the

downregulation of JCHAIN and CD5L plasma as the most powerful proteins to separate MM from 14 other cancer types (Figure 5B). Further proteins that defined the plasma protein signature of MM in our study were the previously described decreased levels of complement proteins C1q, C1r, and C1s as well as upregulated TGFBI, CFD, and MGP and downregulated CBPN (Figure S2). Eight out of the nine of these proteins are linked to the regulation of the complement system and interaction with immunoglobulins. As a joining chain, JCHAIN connects the Fc regions of IgM and IgA and is necessary for the transport of these polymeric immunoglobulins across epithelial cells [25]. JCHAIN-negative IgM has been reported to induce a stronger complement activation than JCHAIN-positive IgM [25–27].



**Figure 5.** Data-driven analysis reveals a panel of biomarkers suitable for MM prediction. (A) ROC curve visualizing the performance and classification specificity of a model generated by a random forest algorithm, which was used to distinguish MM from all other fourteen cancers, with an AUC of 0.96. (B) Feature selection based on peptides, visualizing their relevance score (y-axis) and impact on the model used to separate MM from the fourteen other cancers (color intensity represents the impact level of each analyte, ranging from 0 to 100).

As quantitative information was not available on either IgM or IgA in the analyzed patients, it was not possible to specify whether decreased JCHAIN levels were accompanied by decreased IgM levels. However, we also found the circulating protein CD5L, also called apoptosis inhibitor of macrophage, to be downregulated. CD5L has been reported to be an integral part of IgM and binds to the Fc region of IgM and utilizes immunoglobulin as a carrier which prevents its renal excretion [28–30]. We found CD5L and JCHAIN to be highly dependent (Figure S4) as recently shown by Oskam and coworkers [31]. As for TGFBI, CFD, and MGP, MM patients displayed the highest median plasma concentration in comparison to the other cancer patients. TGFBI has been reported to be both tumor-suppressive as well as tumor-promoting in multiple cancers depending on the cancer progression [32].

### 3. Discussion

Cancer is the second-highest cause of mortality in the world. Improved methods that can detect changes in cancer-associated proteins are needed. Our study presents a large pan-cancer initiative in which proteins were absolutely quantified in human plasma using SIS-PrESTs technology. The analytical strategy based on SIS-PrESTs was capable of quantifying proteins with high precision, as they are long polypeptides added as the first step of sample preparation. Therefore, the biological variance across protein targets could be accurately measured. A disadvantage of using internal standards for quantification lies in the fact that they have to be spiked prior to the sample preparation and DIA analysis. Therefore, the availability of targets for absolute quantification can be a limiting factor. However, the DIA strategy allows to acquire all detectable proteins in samples, which can be used to explore the label-free part of the dataset to identify proteins without the precision of absolute quantification. Furthermore, the analytical sensitivity of today's non-depleted targeted proteomics measurements is limiting. Today's data acquisition of blood-based

tests is comprehensive and has a promising quantitative performance, but the assay is restricted by the dynamic range of plasma, limiting its full potential. Notably, more than 56 FDA- or CLIA-approved biomarkers could be measured in multiplex using this strategy, and by not depleting the plasma, the quantitative integrity of the samples can be assured. This shows that targeted proteomics is an attractive alternative to more sensitive methods based on affinity reagents.

Within this study, we identified proteins that are implicated in the regulation of complement activation and interaction with immunoglobulins and which we suggest as a plasma biomarker panel for MM. These target proteins include JCHAIN, CD5L and four proteins of the C1 complex. Notably, the protein most predictive for MM in the random forest model was JCHAIN, which links two monomer units of either IgM or IgA together. In the case of IgM, the JCHAIN dimerizes and acts as a nucleating unit for the IgM pentamer. The work of Wang et al. [33] has shown that the CD5L loss turns safe Th17 cells into pathogenic cells, causing autoimmunity. By altering lipids, CD5L affects Ror $\gamma$ t, the master regulator, shifting the immune balance. As CD5L is a major switch of Th17 cell functional states *in vivo*, this may indicate that Th17 cell functions are dysregulated in Myeloma Patients.

The role of the complement system and its components as potential biomarkers for cancer has been debated in the literature. Here, the role of C1 in cancer has been a double-edged sword. In clear-cell renal-cell carcinoma, the tumor-induced formation of the complement C1 complex in its microenvironment has been described [24]. In contrast, the protein C1q has previously been related to pro-apoptotic and anti-tumor activities in prostate, breast or ovarian cancers [34–36]. Furthermore, decreased serum levels of C1q have been described in patients with MM and have been suggested as a potential biomarker for the tumor burden [37]. The systemic decrease of C1q levels in plasma highlighted in our study supports previous findings. Furthermore, we not only observed a decrease of C1q but also of C1r and C1s, suggesting a downregulation of all the proteins forming the C1 complex. Interestingly, this decrease does not extend to other complement proteins, which highlight the proteins related to the C1 complex as possible biomarkers in MM.

The level of TGFBI has been suggested as a biomarker for tumor progression [32]. Another upregulated protein, CFD, is part of the alternative complement pathway. CFD is a serine protease, which cleaves factor B to form C3-convertase [38]. Whereas decreased CFD levels have been reported in obesity [39], there are scant reports of its direct involvement in cancer. CFD has been suggested as a biomarker for cutaneous squamous-cell carcinoma [40]. Matrix Gla protein (MGP) has been connected to the inhibition of calcification and there is evidence of its relation to the progression of different cancers [41]. Finally, Carboxypeptidase N catalytic chain (CPN1) is part of the Carboxypeptidase N complex, which has been shown to lead to the inactivation of C3a, C4a, and C5a [42–44]. CPN has been suggested as a prognostic biomarker in breast cancer and it has been reported that MM patients sensitive to bortezomib treatment have lower CPN levels [45,46]. Therefore, we suggest that these four proteins might not be unique identifiers for the classification of MM patients. However, the overall protein levels of these targets provide an interplay that is unique for MM and requires further investigation. Yet, it must be noted that the majority of these proteins interact with the complement cascade and immunoglobulins.

In conclusion, we describe a targeted proteomics approach capable of measuring hundreds of proteins with their concentrations reported on an absolute scale. This multiplex approach provides a complementary strategy to standardized clinical assays and provides the absolute concentrations of a large number of plasma proteins. Here, we show that this approach can be used with liquid biopsies to identify protein targets which are unique for the detection of multiple myeloma patients.



## 4. Materials and Methods

### 4.1. Ethical Statement

The research adheres to all pertinent ethical guidelines. This pan-cancer study received approval from the Swedish Ethical Review Authority (EPM dnr 2019-00222) and aligned with donor consents in U-CAN (28631533, EPN Uppsala 2010-198 with amendments), with all participants providing written informed consent. The study protocol is in accordance with the ethical principles outlined in the 1975 Declaration of Helsinki.

### 4.2. Cohort

A sample cohort consisting of blood plasma from 1800 cancer patients was provided by the biobank of the Uppsala–Umeå Comprehensive Cancer Consortium (UCAN). The samples were collected following the same protocol. Briefly, blood was collected by venipuncture in 6 mL EDTA tubes (Vacuette Cat. no.456243, Greiner-bio One; Kremsmünster, Austria) and centrifuged at 3000 rcf at room temperature (RT) immediately after sample collection. Plasma was transferred to 0.5 mL tubes and frozen and stored at  $-80^{\circ}\text{C}$ . The plasma samples were fully randomized into thirty-three 96-well plates and deidentified. Plasma from 3 males and 2 females was pooled and added to each plate in triplicate. The cohort consisted of patients diagnosed with one out of fifteen cancers: pituitary neuroendocrine tumors (PIT NET,  $n = 50$ ), lymphoma ( $n = 56$ ), chronic lymphocytic leukemia (CLL,  $n = 50$ ), acute myeloid leukemia (AML,  $n = 52$ ), multiple myeloma (MM,  $n = 55$ ), breast cancer (BRC,  $n = 164$ ), ovarian cancer (OVC,  $n = 179$ ), endometrial cancer (ENDC,  $n = 110$ ), cervical cancer (CVX,  $n = 110$ ), prostate cancer (PRC,  $n = 172$ ), colorectal cancer (CRC,  $n = 248$ ), small intestinal neuroendocrine tumor (SI-NET,  $n = 54$ ), lung cancer (LUNG,  $n = 289$ ), meningioma ( $n = 51$ ), and glioma ( $n = 160$ ).

### 4.3. Sample Preparation

A set of 276 absolutely quantified SIS-PrESTs of 276 proteins was pooled at close-to-endogenous levels in healthy blood plasma (Table S1), creating an artificial heavy labeled plasma. The SIS-PrEST pool was aliquoted in 96-well plates and vacuum dried for 16 h at  $35^{\circ}\text{C}$  and stored at  $-20^{\circ}\text{C}$ . Patient plasma samples were thawed on ice for 1 h and 2  $\mu\text{L}$  diluted 20 $\times$  with 1 $\times$  phosphate-buffered saline (PBS, Sigma Aldrich, St. Louis, MI, USA), RapiGest (Waters, Milford, MA, USA), dithiothreitol (DTT, Sigma Aldrich, St. Louis, MI, USA), and diluted plasma corresponding to 1  $\mu\text{L}$  of raw plasma was added to the vacuum-dried SIS-PrESTs to final concentrations of 0.1% RapiGest and 10 mM DTT. The samples were reduced at  $37^{\circ}\text{C}$  for 1 h and alkylated with 50 mM chloroacetamide (CAA) for 30 min in the dark. Digestion was performed overnight using SOLu-Trypsin (Sigma-Aldrich, St. Louis, MI, USA) in an enzyme:substrate ratio of 1:50 and quenched with trifluoroacetic acid (TFA) to a final concentration of 0.5% ( $v/v$ ). Half of each sample was desalted using in-house packed C18 StageTips, according to Rappsilber et al. [47]. In brief, the matrix from 3 layers of Empore C18 disks (Supelco, Sigma Aldrich, St. Louis, MI, USA) was activated with 100% acetonitrile (ACN) and equilibrated with 0.1% TFA. The digest was loaded into the StageTip and washed twice with 80  $\mu\text{L}$  of 0.1% TFA and eluted twice with 30  $\mu\text{L}$  80% acetonitrile and 0.1% formic acid (FA). The StageTips were centrifuged for 2 min at 1000 rcf after each addition. Eluted peptides were vacuum dried at  $45^{\circ}\text{C}$  for 30 min. Prior to analysis, samples were dissolved in Solvent A (3% ACN, 0.1% FA). The samples were processed in batches of two to four plates per digestion.

### 4.4. Mass Spectrometry Analysis

Peptides were quantified in an online system of Ultimate 3000 (Thermo Fisher Scientific, Santa Clara, CA, USA) LC connected to QExactive HF (Thermo Fisher Scientific, Santa Clara, CA, USA) MS. A sample corresponding to 2  $\mu\text{g}$  of raw plasma was loaded onto a trap column (PN 164535, Thermo Fisher Scientific, Santa Clara, CA, USA) and washed for 3 min at 7  $\mu\text{L}/\text{min}$  with 100% Solvent A. The peptides were separated on an analytical column (PN ES902, Thermo Fisher Scientific, Santa Clara, CA, USA) using a

40 min linear gradient of 1–32% Solvent B (95% ACN, 0.1% FA) at 0.7  $\mu\text{L}/\text{min}$ . The columns were washed with 3 two-minute seesaw gradients of 1–99% Solvent B and equilibrated for 9 min. MS analysis was performed using a DIA method with cycles consisting of a full MS scan (30,000 resolution, AGC =  $3 \times 10^6$ , 300–1200  $m/z$ , IT = 105 ms) followed by 30 DIA scans (30,000 resolution, AGC =  $1 \times 10^6$ , NCE = 26, 10  $m/z$  isolation window, IT = 55 ms).

#### 4.5. Absolute Quantification

A list of proteotypic peptides was generated by *in silico* digestion of the fasta file including the amino acid sequences of all 276 spiked-in SIS-PrESTs using EncyclopeDIA (ver. 1.2.2) [48] and whole human proteome as background (Homo Sapiens, UniProt ID: #UP000009606, 20,370 entries, accessed on 26 October 2020). One missed cleavage was allowed, and other parameters were adjusted according to the MS method. A spectral library was generated for all peptides using a ProSight machine learning algorithm [49]. The background proteome and the first 6 analyzed raw files were imported into Skyline (ver. 20.2.0.286) [50] and the peaks were manually inspected. Peptides in which both light and heavy signals were not detected were deleted together with interfering transitions. Peptide retention times were predicted by an indexed retention time library which included the 12 most intensive APOA1 peptides. A so prepared Skyline file was used to import all of the resulting raw files using 3 min windows of predicted peptide retention time with mass accuracy set to 5 ppm. Data for both light and heavy signals were exported for further analysis.

Exported results were imported into RStudio (ver. 1.4.1717). First, the data were filtered to contain only quantified peptides (rdotp > 0.7, dotp > 0.5, 1000 > ratio to standards > 0.01). Samples that failed APOA1 iRT regression or had fewer than 120 proteins quantified were excluded from the analysis. Additionally, peptides with a quantification rate of less than 50% across all the samples were excluded. Non-paired transitions were filtered out and the heavy to light ratio was calculated from the summed AUCs of transitions present in both light and heavy channels. Further, the data were median-normalized using the pool samples.

#### 4.6. Label-Free Data Extraction

Label-free data was extracted using EncyclopeDIA. First, mzML files were generated from raw files using msConvert within ProteoWizard [51] followed by EncyclopeDIA [52] search against a spectral library generated from list of blood plasma proteins with ProSight integrated into ProteomicsDB [49]. A whole human proteome (Homo Sapiens UniProt ID: # UP000005640, reviewed, 20,371 entries, accessed 11 August 2021) was used as a background proteome.

#### 4.7. Disease Prediction

A random forest prediction model was built aiming to classify multiple myeloma patients based on peptide levels in plasma. First, peptide levels for missing values were imputed using the impute.knn function from the impute R package (ver. 1.64.0). The model was built using the train function in the caret R package (ver. 6.0.90) using 70% of the cohort and 5-fold cross validation. The model was tested on the remaining 30% and specificity, sensitivity, and AUC scores were summarized in a receiver operating characteristic (ROC) curve.

## 5. Conclusions

In conclusion, our study pioneers a targeted proteomics approach with a precise quantification of hundreds of plasma proteins, offering a complementary strategy to standard clinical assays. We identified potential biomarkers, including JCHAIN, C1 complex proteins, and others, for multiple myeloma detection. This work underscores the promise of targeted proteomics in cancer diagnostics and biomarker discovery.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15194764/s1>. Figure S1: All 146 quantified proteins showing dynamic range over 6 orders of magnitude. Figure S2: Label free quantification data showed protein dysregulation for multiple myeloma compared to all other cancers. A. Confirmed downregulated hallmark proteins for multiple myeloma linked to mainly immunoglobulins. B. Complement complex C1 proteins were observed downregulated in label free data, though, less significantly than in absolute quantification data. Figure S3: Peptides identified as the most important by the random forest models for the classification of multiple myeloma. Figure S4: Correlation plot between all quantified peptides from CD5L and JCHAIN. Table S1: Stable isotope standards with corresponding spike in levels.

**Author Contributions:** F.E. and M.U. conceived of and designed the analysis. D.K., J.W., A.H., M.B.Á., K.H.T.M., F.P. and L.F. collected and contributed data to this study. D.K. and M.B.Á. performed the data analysis. F.E., D.K., J.W. and M.B.Á. drafted the manuscript. D.K., J.W., M.B.Á., A.H., M.U. and F.E. revised the manuscript. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The main funding was provided by the Erling Persson Foundation (M.U.) and the Knut and Alice Wallenberg Foundation (M.U. 2019.0341).

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki and approved by the Swedish Ethical Review Authority (EPM dnr 2019-00222) for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in this study.

**Data Availability Statement:** The MS raw data as well as Skyline files and libraries are available on Panorama public "<https://panoramaweb.org/QKcywA.url> (accessed on 19 September 2023) and ProteomeXchange, ID PXD037946.

**Acknowledgments:** We acknowledge the entire staff of the Human Protein Atlas program and the Science for Life Laboratory (SciLifeLab) for their valuable contributions.

**Conflicts of Interest:** M.U. is a co-founder of Atlas Antibodies AB. D.K., A.H., F.E., and M.U. are co-founders of ProteomeEdge AB.

## References

1. Ramaswamy, S.; Ross, K.N.; Lander, E.S.; Golub, T.R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **2003**, *33*, 49–54. [[CrossRef](#)] [[PubMed](#)]
2. Golub, T.R.; Slonim, D.K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J.P.; Coller, H.; Loh, M.L.; Downing, J.R.; Caligiuri, M.A.; et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **1999**, *286*, 531–537. [[CrossRef](#)] [[PubMed](#)]
3. Cline, M.S.; Craft, B.; Swatloski, T.; Goldman, M.; Ma, S.; Haussler, D.; Zhu, J. Exploring TCGA Pan-Cancer Data at the UCSC Cancer Genomics Browser. *Sci. Rep.* **2013**, *3*, 2652. [[CrossRef](#)]
4. International Cancer Genome Consortium; Hudson, T.J.; Anderson, W.; Artez, A.; Barker, A.D.; Bell, C.; Bernabé, R.R.; Bhan, M.K.; Calvo, F.; Eerola, I.; et al. International network of cancer genome projects. *Nature* **2010**, *464*, 993–998. [[PubMed](#)]
5. Meyerson, M.; Gabriel, S.; Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **2010**, *11*, 685–696. [[CrossRef](#)] [[PubMed](#)]
6. Frampton, G.M.; Fichtenholtz, A.; Otto, G.A.; Wang, K.; Downing, S.R.; He, J.; Schnall-Levin, M.; White, J.; Sanford, E.M.; An, P.; et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **2010**, *31*, 1023–1031. [[CrossRef](#)]
7. Hoshino, A.; Kim, H.S.; Bojmar, L.; Gyan, K.E.; Cioffi, M.; Hernandez, J.; Zambirinis, C.P.; Rodrigues, G.; Molina, H.; Heissel, S.; et al. Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **2020**, *182*, 1044–1061.e18. [[CrossRef](#)]
8. Ludwig, J.A.; Weinstein, J.N. Biomarkers in Cancer Staging, Prognosis and Treatment Selection. *Nat. Rev. Cancer* **2005**, *5*, 845–856. [[CrossRef](#)]
9. Haber, D.A.; Velculescu, V.E. Blood-based analyses of cancer: Circulating tumor cells and circulating tumor DNA. *Cancer Discov.* **2014**, *4*, 650–661. [[CrossRef](#)]
10. Anderson, N.L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R.S.; Conrads, T.P.; Veenstra, T.D.; Adkins, J.N.; Pounds, J.G.; Fagan, R.; et al. The human plasma proteome. *Mol. Cell Proteom.* **2004**, *3*, 311–326. [[CrossRef](#)]
11. Ankerst, D.P. di Societa Italiana di ITOU2006 Sensitivity and Specificity of Prostate-Specific Antigen for Prostate Cancer Detection with High Rates of Biopsy Verification—Abstract—Europe PMC. Available online: <https://europepmc.org/article/med/17269614> (accessed on 10 March 2022).

12. Larson, M.H.; Pan, W.; Kim, H.J.; Mauntz, R.E.; Stuart, S.M.; Pimentel, M.; Zhou, Y.; Knudsgaard, P.; Demas, V.; Aravanis, A.M.; et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat. Commun.* **2021**, *12*, 2357. [[CrossRef](#)] [[PubMed](#)]
13. Percy, A.J.; Byrns, S.; Pennington, S.R.; Holmes, D.T.; Anderson, N.L.; Agreste, T.M.; Duffy, M.A. Clinical translation of MS-based, quantitative plasma proteomics: Status, challenges, requirements, and potential. *Expert Rev. Proteom.* **2016**, *13*, 673–684. [[CrossRef](#)] [[PubMed](#)]
14. Liotta, L.A.; Ferrari, M.; Petricoin, E. Clinical proteomics: Written in blood. *Nature* **2003**, *425*, 905. [[CrossRef](#)] [[PubMed](#)]
15. Marx, V. Targeted proteomics. *Nat. Methods* **2012**, *10*, 19–22. [[CrossRef](#)] [[PubMed](#)]
16. Zeiler, M.; Straube, W.L.; Lundberg, E.; Uhlén, M.; Mann, M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol. Cell Proteom.* **2012**, *11*, O111.009613. [[CrossRef](#)]
17. Nakayasu, E.S.; Gritsenko, M.; Piehowski, P.D.; Gao, Y.; Orton, D.J.; Schepmoes, A.A.; Fillmore, T.L.; Frohnert, B.I.; Rewers, M.; Krischer, J.P.; et al. Tutorial: Best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **2021**, *16*, 3737–3760. [[CrossRef](#)]
18. Kotol, D.; Hober, A.; Strandberg, L.; Svensson, A.-S.; Uhlén, M.; Edfors, F. Targeted proteomics analysis of plasma proteins using recombinant protein standards for addition only workflows. *BioTechniques* **2021**, *71*, 473–483. [[CrossRef](#)]
19. Uhlén, M.; Karlsson, M.J.; Hober, A.; Svensson, A.-S.; Scheffel, J.; Kotol, D.; Zhong, W.; Tebani, A.; Strandberg, L.; Edfors, F.; et al. The human secretome. *Sci. Signal.* **2019**, *12*, eaaz0274. [[CrossRef](#)]
20. Lassman, M.E.; McLaughlin, T.M.; Zhou, H.; Pan, Y.; Marcovina, S.M.; Laterza, O.; Roddy, T.P. Simultaneous quantitation and size characterization of apolipoprotein(a) by ultra-performance liquid chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 1101–1106. [[CrossRef](#)]
21. Cai, D.; Liang, J.; Cai, X.-D.; Yang, Y.; Liu, G.; Zhou, F.; He, D. Identification of six hub genes and analysis of their correlation with drug sensitivity in acute myeloid leukemia through bioinformatics. *Transl. Cancer Res.* **2021**, *10*, 126–140. [[CrossRef](#)]
22. Liang, L.; Li, J.; Fu, H.; Liu, X.; Liu, P. Identification of High Serum Apolipoprotein A1 as a Favorable Prognostic Indicator in Patients with Multiple Myeloma. *J. Cancer* **2019**, *10*, 4852–4859. [[CrossRef](#)] [[PubMed](#)]
23. Merle, N.S.; Church, S.E.; Fremeaux-Bacchi, V.; Roumenina, L.T. Complement System Part I—Molecular Mechanisms of Activation and Regulation. *Front. Immunol.* **2015**, *6*, 262. [[CrossRef](#)] [[PubMed](#)]
24. Roumenina, L.T.; Daugan, M.V.; Petitprez, F.; Sautès-Fridman, C.; Fridman, W.H. Context-dependent roles of complement in cancer. *Nat. Rev. Cancer* **2019**, *19*, 698–715. [[CrossRef](#)] [[PubMed](#)]
25. Johansen, F.E.; Braathen, R.; Brandtzaeg, P. Role of J chain in secretory immunoglobulin formation. *Scand J. Immunol.* **2000**, *52*, 240–248. [[CrossRef](#)] [[PubMed](#)]
26. Davis, A.C.; Roux, K.H.; Shulman, M.J. On the structure of polymeric IgM. *Eur. J. Immunol.* **1988**, *18*, 1001–1008. [[CrossRef](#)]
27. Wiersma, E.J.; Collins, C.; Fazel, S.; Shulman, M.J. Structural and functional analysis of J chain-deficient IgM. *J. Immunol.* **1998**, *160*, 5979–5989. [[CrossRef](#)]
28. Hiramoto, E.; Tsutsumi, A.; Suzuki, R.; Matsuoka, S.; Arai, S.; Kikkawa, M.; Miyazaki, T. The IgM pentamer is an asymmetric pentagon with an open groove that binds the AIM protein. *Sci. Adv.* **2018**, *4*, eaau1199. [[CrossRef](#)]
29. Arai, S.; Miyazaki, T. Impacts of the apoptosis inhibitor of macrophage (AIM) on obesity-associated inflammatory diseases. *Semin Immunopathol.* **2013**, *36*, 3–12. [[CrossRef](#)]
30. Miyazaki, T.; Yamazaki, T.; Sugisawa, R.; Gershwin, M.E.; Arai, S. AIM associated with the IgM pentamer: Attackers on stand-by at aircraft carrier. *Cell Mol. Immunol.* **2018**, *15*, 563–574. [[CrossRef](#)]
31. Oskam, N.; den Boer, M.A.; Lukassen, M.V.; Ooijevaar-de Heer, P.; Veth, T.S.; van Mierlo, G.; Lai, S.-H.; Derksen, N.I.L.; Yin, V.C.; Streutker, M.; et al. CD5L is a canonical component of circulatory IgM. *bioRxiv* **2023**. [[CrossRef](#)]
32. Corona, A.; Blobel, G.C. The role of the extracellular matrix protein TGFBI in cancer. *Cell. Signal.* **2021**, *84*, 110028. [[CrossRef](#)] [[PubMed](#)]
33. Wang, C.; Yosef, N.; Gaublumme, J.; Wu, C.; Lee, Y.; Clish, C.B.; Kaminski, J.; Xiao, S.; Zu Horste, G.M.; Pawlak, M.; et al. CD5L/AIM Regulates Lipid Biosynthesis and Restrains Th17 Cell Pathogenicity. *Cell* **2015**, *163*, 1413–1427. [[CrossRef](#)] [[PubMed](#)]
34. Hong, Q.; Sze, C.-I.; Lin, S.-R.; Lee, M.-H.; He, R.-Y.; Schultz, L.; Chang, J.-Y.; Chen, S.-J.; Boackle, R.J.; Hsu, L.-J.; et al. Complement C1q Activates Tumor Suppressor WWOX to Induce Apoptosis in Prostate Cancer Cells. *PLoS ONE* **2019**, *4*, e5755. [[CrossRef](#)] [[PubMed](#)]
35. Bandini, S.; Macagno, M.; Hysi, A.; Lanzardo, S.; Conti, L.; Bello, A.; Riccardo, F.; Ruiu, R.; Merighi, I.F.; Forni, G.; et al. The non-inflammatory role of C1q during Her2/neu-driven mammary carcinogenesis. *Oncol Immunology* **2016**, *5*, e1253653. [[CrossRef](#)] [[PubMed](#)]
36. Kaur, A.; Sultan, S.H.A.; Murugaiah, V.; Pathan, A.A.; Alhamlan, F.S.; Karteris, E.; Kishore, U. Human C1q Induces Apoptosis in an Ovarian Cancer Cell Line via Tumor Necrosis Factor Pathway. *Front. Immunol.* **2016**, *7*, 599. [[CrossRef](#)] [[PubMed](#)]
37. Yang, R.; Huang, J.; Ma, H.; Li, S.; Gao, X.; Liu, Y.; Shen, J.; Liao, A. Is complement C1q a potential marker for tumor burden and immunodeficiency in multiple myeloma? *Leuk. Lymphoma* **2019**, *60*, 1812–1818. [[CrossRef](#)]
38. Barratt, J.; Weitz, I. Complement Factor D as a Strategic Target for Regulating the Alternative Complement Pathway. *Front. Immunol.* **2021**, *12*, 712572. [[CrossRef](#)]

39. Flier, J.S.; Cook, K.S.; Usher, P.; Spiegelman, B.M. Severely Impaired Adipsin Expression in Genetic and Acquired Obesity. *Science* **1987**, *237*, 405–408. [[CrossRef](#)]
40. Nezhad, P.R.; Riihilä, P.; Knuutila, J.S.; Viiklepp, K.; Peltonen, S.; Kallajoki, M.; Meri, S.; Nissinen, L.; Kähäri, V.-M. Complement Factor D Is a Novel Biomarker and Putative Therapeutic Target in Cutaneous Squamous Cell Carcinoma. *Cancers* **2022**, *14*, 305. [[CrossRef](#)]
41. Gheorghe, S.R.; Crăciun, A.M. Matrix Gla protein in tumoral pathology. *Clujul. Med.* **2016**, *89*, 319–321. [[CrossRef](#)]
42. Bokisch, V.A.; Müller-Eberhard, H.J. Anaphylatoxin inactivator of human plasma: Its isolation and characterization as a carboxypeptidase. *J. Clin. Investig.* **1970**, *49*, 2427–2436. [[CrossRef](#)] [[PubMed](#)]
43. Matthews, K.W.; Mueller-Ortiz, S.L.; Wetsel, R.A. Carboxypeptidase N: A pleiotropic regulator of inflammation. *Mol. Immunol.* **2004**, *40*, 785–793. [[CrossRef](#)] [[PubMed](#)]
44. Skidgel, R.A.; Erdös, E.G. Structure and function of human plasma carboxypeptidase N, the anaphylatoxin inactivator. *Int. Immunopharmacol.* **2007**, *7*, 1888–1899. [[CrossRef](#)] [[PubMed](#)]
45. Cumová, J.; Jedličková, L.; Potěšil, D.; Sedo, O.; Stejskal, K.; Potáčová, A.; Zdráhal, Z.; Hájek, R. Comparative plasma proteomic analysis of patients with multiple myeloma treated with bortezomib-based regimens. *Klin. Onkol.* **2012**, *25*, 17–25.
46. Cui, R.; Wang, C.; Li, T.; Hua, J.; Zhao, T.; Ren, L.; Wang, Y.; Li, Y. Carboxypeptidase N1 is anticipated to be a synergy metrics for chemotherapy effectiveness and prognostic significance in invasive breast cancer. *Cancer Cell Int.* **2021**, *21*, 571. [[CrossRef](#)]
47. Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906. [[CrossRef](#)]
48. Searle, B.C.; Swearingen, K.E.; Barnes, C.A.; Schmidt, T.; Gessulat, S.; Kuster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1–10. [[CrossRef](#)]
49. Gessulat, S.; Schmidt, T.; Zolg, D.P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. Prosit: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509–518. [[CrossRef](#)]
50. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [[CrossRef](#)]
51. Chambers, M.C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D.L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30*, 918–920. [[CrossRef](#)]
52. Searle, B.C.; Pino, L.K.; Egertson, J.D.; Ting, Y.S.; Lawrence, R.T.; MacLean, B.X.; Villén, J.; MacCoss, M.J. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **2018**, *9*, 5128. [[CrossRef](#)] [[PubMed](#)]

**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.