



Article

Plasma Proteomics Enable Differentiation of Lung Adenocarcinoma from Chronic Obstructive Pulmonary Disease (COPD)

Thilo Bracht ^{1,2,*} , Daniel Kleefisch ^{2,3} , Karin Schork ^{2,3} , Kathrin E. Witzke ^{2,3} , Weiqiang Chen ², Malte Bayer ^{1,2}, Jan Hovanec ⁴, Georg Johnen ⁴ , Svetlana Meier ⁴, Yon-Dschun Ko ⁵, Thomas Behrens ⁴, Thomas Brüning ⁴, Jana Fassunke ⁶ , Reinhard Buettner ⁶ , Julian Uszkoreit ^{2,3} , Michael Adamzik ¹, Martin Eisenacher ^{2,3,†} and Barbara Sitek ^{1,2,*,†}

- ¹ Clinic for Anesthesiology, Intensive Care and Pain Therapy, University Medical Center Knappschaftskrankenhaus Bochum, 44892 Bochum, Germany
- ² Medizinisches Proteom-Center, Ruhr-University Bochum, 44801 Bochum, Germany
- ³ Center for Protein Diagnostics (PRODI), Medical Proteome Analysis, Ruhr-University Bochum, 44801 Bochum, Germany
- ⁴ Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), 44789 Bochum, Germany
- ⁵ Department of Internal Medicine, Johanniter-Kliniken Bonn GmbH, Johanniter Krankenhaus, 53113 Bonn, Germany
- ⁶ Institute of Pathology, Medical Faculty and Center for Molecular Medicine (CMMC), University of Cologne, 50924 Cologne, Germany
- * Correspondence: thilo.bracht@rub.de (T.B.); barbara.sitek@rub.de (B.S.); Tel.: +49-234-32-29985 (T.B.); +49-234-32-24362 (B.S.)
- † These authors contributed equally to this work.



Citation: Bracht, T.; Kleefisch, D.; Schork, K.; Witzke, K.E.; Chen, W.; Bayer, M.; Hovanec, J.; Johnen, G.; Meier, S.; Ko, Y.-D.; et al. Plasma Proteomics Enable Differentiation of Lung Adenocarcinoma from Chronic Obstructive Pulmonary Disease (COPD). *Int. J. Mol. Sci.* **2022**, *23*, 11242. <https://doi.org/10.3390/ijms231911242>

Academic Editors: I-Lin Tsai and Michael X. Chen

Received: 27 July 2022

Accepted: 20 September 2022

Published: 24 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Abstract: Chronic obstructive pulmonary disease (COPD) is a major risk factor for the development of lung adenocarcinoma (AC). AC often develops on underlying COPD; thus, the differentiation of both entities by biomarker is challenging. Although survival of AC patients strongly depends on early diagnosis, a biomarker panel for AC detection and differentiation from COPD is still missing. Plasma samples from 176 patients with AC with or without underlying COPD, COPD patients, and hospital controls were analyzed using mass-spectrometry-based proteomics. We performed univariate statistics and additionally evaluated machine learning algorithms regarding the differentiation of AC vs. COPD and AC with COPD vs. COPD. Univariate statistics revealed significantly regulated proteins that were significantly regulated between the patient groups. Furthermore, random forest classification yielded the best performance for differentiation of AC vs. COPD (area under the curve (AUC) 0.935) and AC with COPD vs. COPD (AUC 0.916). The most influential proteins were identified by permutation feature importance and compared to those identified by univariate testing. We demonstrate the great potential of machine learning for differentiation of highly similar disease entities and present a panel of biomarker candidates that should be considered for the development of a future biomarker panel.

Keywords: lung cancer; plasma proteomics; machine learning; artificial intelligence; random forest; Ig kappa light chain; SERPINA3; SAA1

1. Introduction

Lung cancer is the leading cause of death in all cancer types and makes up about 18% of cancer-related deaths worldwide while contributing to about 11% of all diagnosed cancer cases [1,2]. The main cause for lung cancer is smoking [3], which is associated with more than 90% of all lung cancer cases [2]. Smoking is also a major risk factor in the development of a plethora of other diseases related to the respiratory tract, especially

lung cancer and COPD. Chronic obstructive pulmonary disease (COPD) develops in about 50% of smokers over time [4]. Smokers with COPD have a two- to five-times higher risk of developing lung cancer [5,6] and COPD is found in about 50–90% of lung-cancer patients [7]. Five-year survival rates for lung cancer are low with about 20%, reflecting the large number of diagnoses at late stages, when 57% of the patients already show metastatic progress. For patients presenting with metastatic disease, the 5-year survival rate is only 5% in comparison to 57% for localized stages [8]. Due to the fatality of lung cancer and the high risk of COPD patients for developing AC, biomarkers for detection of AC and differentiation from COPD are urgently needed, preferably in body fluids that can be obtained by minimal-invasive methods. Currently, diagnosis is commonly performed via invasive biopsy, mostly after observation of visible changes in CT scans or bronchoscopy.

Mass-spectrometry (MS)-based proteomics is a powerful tool for high-throughput discovery approaches in the area of biomarker research. A wide range of materials can be used for analysis including serum or plasma, which, in addition to tissue, are major sources for biomarker discovery. Minimally- or noninvasive biomarkers would improve cancer assessment as diagnostic or prognostic markers, as well as a tool for monitoring [9]. Numerous proteomics studies analyzing tissue, bronchoalveolar lavage fluid, and blood have been performed for identifying biomarkers for AC, COPD, and the differentiation of both [10]. Several candidate biomarkers have been proposed, including AGR2 [11,12], SAA [11], HER2 [11,13,14], APOE [15], or SCGB3A2 [16] for AC; YKL-40 [17], MFAP4 [18], GRP78, soluble CD163, IL1AP, and MSPT9 for COPD [11]; CRP, VEGF [13,14,19], IL-8, and MMP9 [14] for the differentiation of AC and COPD. However, none of those proposed biomarker candidates have yet taken the step into clinical practice; therefore, reliable biomarkers for these applications are still highly demanded.

With the rise of a multitude of bioinformatics methods, machine learning approaches for the separation of disease groups became applicable to a larger life science community to aid in the identification of protein candidates [20]. Logistic regression and linear discriminant analysis (LDA) are classical modeling statistics. Logistic regression and LDA are easy to interpret with regard to the influence of the predictor variables, but can only perform a linear separation of the data. The support vector machines (SVMs) [21] can use transformation in a higher-dimensional space by the use of a polynomial kernel function to learn more complex (separation) patterns than logistic regression or LDA. Random forest is based on tree-building algorithms [22] and is robust against overfitting. All the stated approaches can be used for large numbers of predictor variables in limited sample sizes and with heterogeneous data, making them ideal tools for clinical proteomics data analysis.

In this study, we used mass spectrometry analysis of plasma samples for the identification of biomarker candidates using univariate statistics and machine-learning-based classification algorithms. We analyzed samples of AC patients with or without COPD, COPD patients, and hospital controls (HCs) with the aim of identifying proteins differentiating AC from COPD (Figure 1). We developed classification models to (A) separate all AC patients from COPD patients (AC vs. COPD), and (B) specifically differentiate AC patients with a COPD background from COPD patients (AC with COPD vs. COPD). We optimized the feature selection and tested several machine learning algorithms for classification. After choosing the best model, we calculated the permutation feature importance to determine the most influential proteins. To further validate the results of the model, we repeated the analysis in 50 random train-test-splits of the dataset for model (A) and compared the precision on the test sets as well as the most often used features to the model with the full dataset. This allowed us to compare the results of univariate tests and machine learning. We demonstrated that plasma proteomics is a powerful technique to distinguish patient groups that are challenging to discriminate in clinical routine. We identified a set of proteins using complementary approaches, which might serve as a starting point for the development of a clinical biomarker panel for diagnosis and differentiation of AC and COPD.

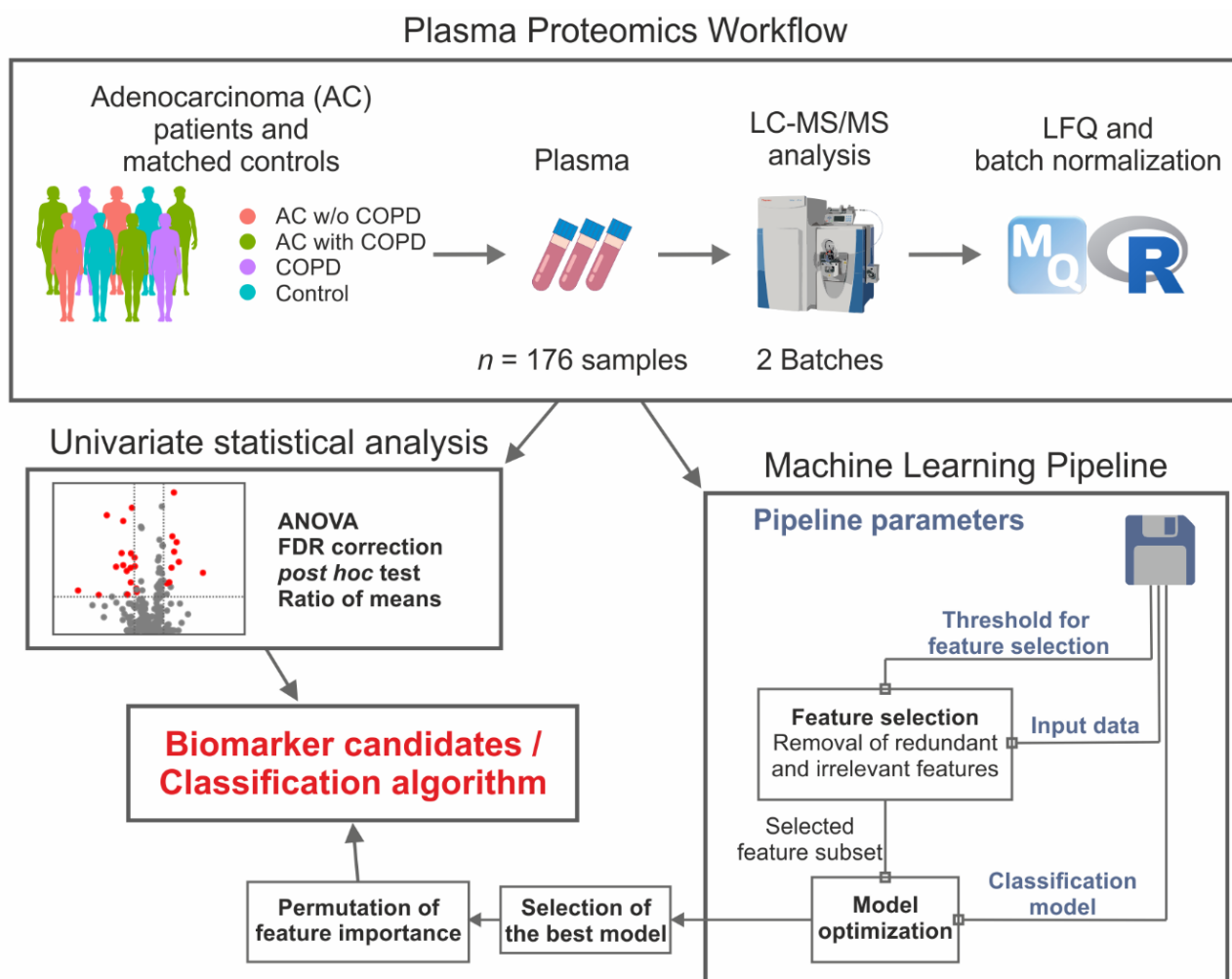


Figure 1. Schematic representation of the project workflow. Plasma samples from $n = 176$ patients were analyzed in two batches using LC-MS/MS. Proteins were quantified label-free and the resulting intensities were normalized to account for batch effects. Normalized intensities were analyzed by univariate statistics and machine learning approaches. Five different machine learning algorithms were compared using the same modeling pipeline.

2. Results

2.1. Successful Normalization of Label-Free Proteomics Data

Due to the reception of clinical samples at different time points, the 176 clinical plasma samples were analyzed in two separate batches, with two different instruments and different LC gradients. Principal component analysis (PCA) of nonnormalized intensities showed two major clusters, which, however, did not represent the two analyzed batches (Figure 2A). The exact technical parameters that were responsible for the clustering of the data were unfortunately not identified. The applied normalization method keeps connections to biological effects while reducing systematic and technical errors. After normalization, samples of the two batches were much more congruent, which was highlighted by the reduced influence of PC1 (67.2% before normalization, 19.5% after normalization). Normalization was further evaluated using boxplots (Supplementary Figure S1) and MA plots (Supplementary Figure S2). Boxplots showed a decrease in inter-sample variability and MA plots showed good agreement between the samples that were doubly measured in both batches. The normalized protein intensities can be found in Supplementary Table S2.

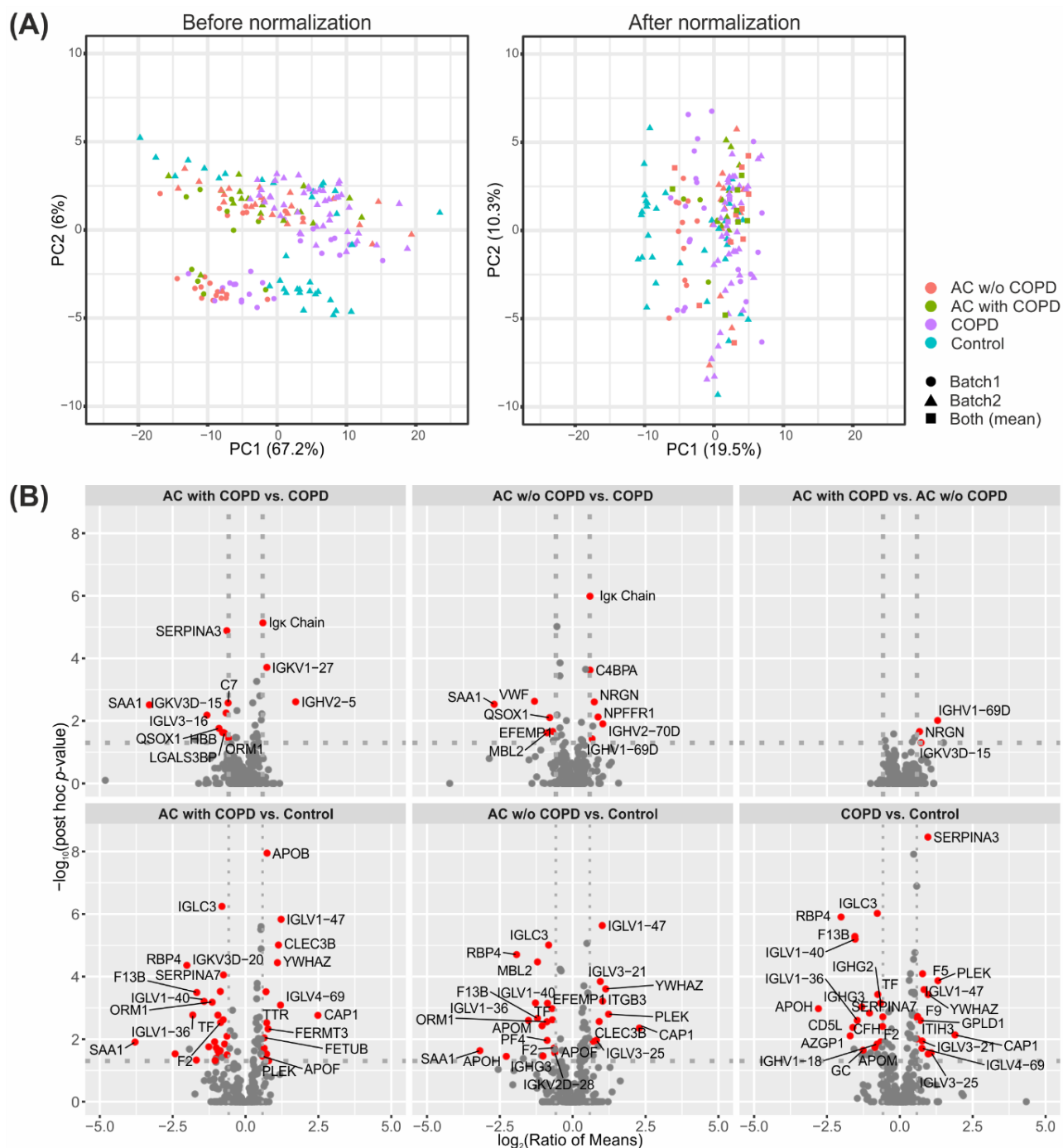


Figure 2. Normalization and statistical analysis of proteomics data. **(A)** Principal component analysis (PCA) plots of label-free LC–MS/MS data before and after batch normalization. Each data point corresponds to a sample measured in either batch 1, batch 2, or both batches (colors representing patient groups). **(B)** Volcano plots representing the results of statistical analysis using Welch-ANOVA. Significant proteins highlighted in red (ANOVA p_{FDR} -value ≤ 0.05 (corrected according to Benjamini–Hochberg); post hoc p_{FDR} -value ≤ 0.05 (corrected according to Bonferroni–Holm); absolute ratio of means ≥ 1.5) and labeled with gene names (except Igk Chain).

2.2. Univariate Statistics Reveal Proteins Discriminating AC and COPD

The combination of both analyzed batches led to the quantification of 397 protein groups in plasma. Of these, 83 protein groups were exclusively measured in batch 1, whereas 55 protein groups were exclusive for batch 2. At first, we analyzed the univariate

protein differences between patient groups by means of ANOVA and the post hoc test. The statistical analysis showed the overall greatest differences for the comparisons between the diseased patient groups and hospital controls (Figure 2B, lower panel). This illustrates the need for appropriate control groups in clinical proteomics as the use of an unspecific control group—in this case, hospital controls—leads to many false findings. A maximum of 39 significantly differentially abundant proteins was observed for AC with COPD vs. Control (Table 1). In the following, the comparisons between the AC and COPD patient groups were emphasized, considering their clinical relevance. Here, both comparisons of AC with or w/o COPD vs. COPD resulted in 11 significantly differentially abundant proteins of which, however, only three proteins were regulated in both comparisons (Sulfhydryl oxidase 1 (QSOX1), Serum amyloid A-1 protein (SAA1), and Ig kappa light chain (no gene name; Supplementary Tables S3 and S4)). The comparison between AC with COPD and AC w/o COPD resulted in three significantly altered proteins. We performed hierarchical cluster analysis based on 42 proteins that were found with a significant p_{FDR} value between both AC groups and COPD by ANOVA. In addition to remarkable sample heterogeneity, the corresponding heatmap showed two major protein clusters separating COPD and AC (Figure 3).

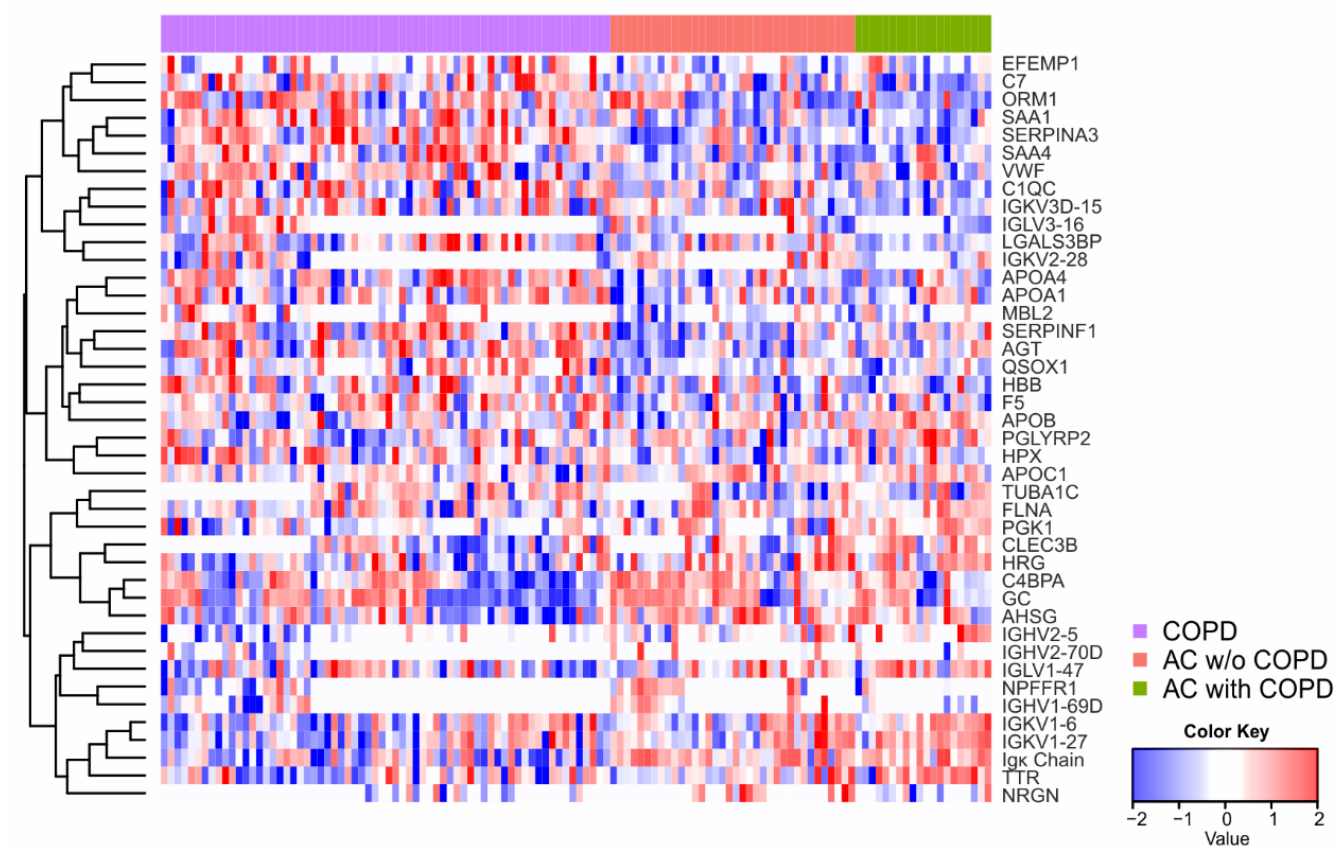


Figure 3. Hierarchical cluster analysis and two-group comparison of proteomics data. Heatmap illustrating hierarchical cluster analysis (distance based on Pearson’s correlation, complete linkage) considering 42 proteins, which passed a p_{FDR} -value threshold ≤ 0.05 calculated using Welch-ANOVA for comparisons between either AC with or without COPD vs. COPD.

Table 1. Significantly differentially abundant protein groups.

Comparison (Condition A vs. Condition B)	Protein Groups Considered for Statistical Testing ¹	Significantly Differentially Abundant Protein Groups ²	Higher Abundance in Condition A	Higher Abundance in Condition B
AC with COPD vs. COPD	325	11	3	8
AC w/o COPD vs. COPD	349	11	6	5
AC w/o COPD vs. AC with COPD	324	3	3	0
AC with COPD vs. Control	271	39	14	25
AC w/o COPD vs. Control	278	26	9	17
COPD vs. Control	283	31	14	18

¹ Protein groups were filtered for a minimum of five valid quantifications per patient group. ² Significance filter criteria: ANOVA p_{FDR} and post hoc p_{FDR} -values ≤ 0.05 ; absolute RoM ≥ 1.5 .

2.3. Machine Learning Yields Highly Predictive Classification Models

Five classification algorithms were assessed for the correct classification of (A) AC vs. COPD and (B) AC with COPD vs. COPD. For feature selection, increasing p -value thresholds were tested, resulting in increasing numbers of proteins considered for model optimization. Accordingly, the numbers of considered proteins were lowest for a p -value threshold of 0.05 with 24 proteins (A) and 12 proteins (B). Without any cut-off, 194 proteins were considered for both (A) and (B). For the differentiation of (A) AC vs. COPD, random forest classification outperformed the other models, achieving the maximum AUC when a p -value threshold of 0.02 was applied (Figure 3, Supplementary Table S5). The differences between individual p -value thresholds were small, however, and all models resulted in AUCs over 0.9. The highest AUC was 0.935 with a sensitivity of 0.848 and a specificity of 0.879 (Figure 4B, Table 2). The other classification models showed AUCs between 0.8 and 0.9 with SVM with the polynomial kernel, which is superior to SVM with the linear kernel and LDA. Logistic regression was outperformed by all models and showed a constant decrease in AUCs with increasing p -value threshold.

For the differentiation of (B) AC with COPD vs. COPD, random forest and both SMV algorithms performed very similarly. Notably, SVM with the linear kernel performed better than random forest for some tested p -value thresholds (Figure 4A, Supplementary Table S5). The maximum AUC, however, resulted from random forest classification without feature selection by a statistical test (AUC = 0.916, sensitivity = 0.57, specificity = 0.965; Figure 4B, Table 3). Notably, LDA performed significantly poorer than the other models and logistic regression showed a striking performance decrease with increasing p -value thresholds. The subsequent feature importance analysis of the best-performing models revealed the greatest influence for Alpha-1-antichymotrypsin (SERPINA3) in both comparisons (Figure 4C). For (A), Ig kappa light chain was the second-most informative feature followed by Pigment epithelium-derived factor (SERPINF1), Apolipoprotein A-IV (APOA4), and the uncharacterized protein C16orf46. For (B), the feature importance was generally lower with SERPINA3 followed by C16orf46, Ig kappa variable 1–27 (IGKV1–27), Haptoglobin-related protein (HPR), and Transthyretin (TTR). Additional results from train-test-set validation for analysis (A) supported these results with mean AUC values of 0.823 over 50 repetitions on the test set and 0.901 in the cross-validation of the trainsets (Table 3, Figure 5). SERPINA3 and Ig κ Chain were selected in 49 and 50 repetitions, respectively; SERPINF1 and C16orf46 were used in over 70% of the repetitions supporting the significance of these features. On the contrary, APOA4 was only used in three repetitions, but AHSG, which was not selected by the model for the full dataset, was selected in 29 (58%) repetitions.

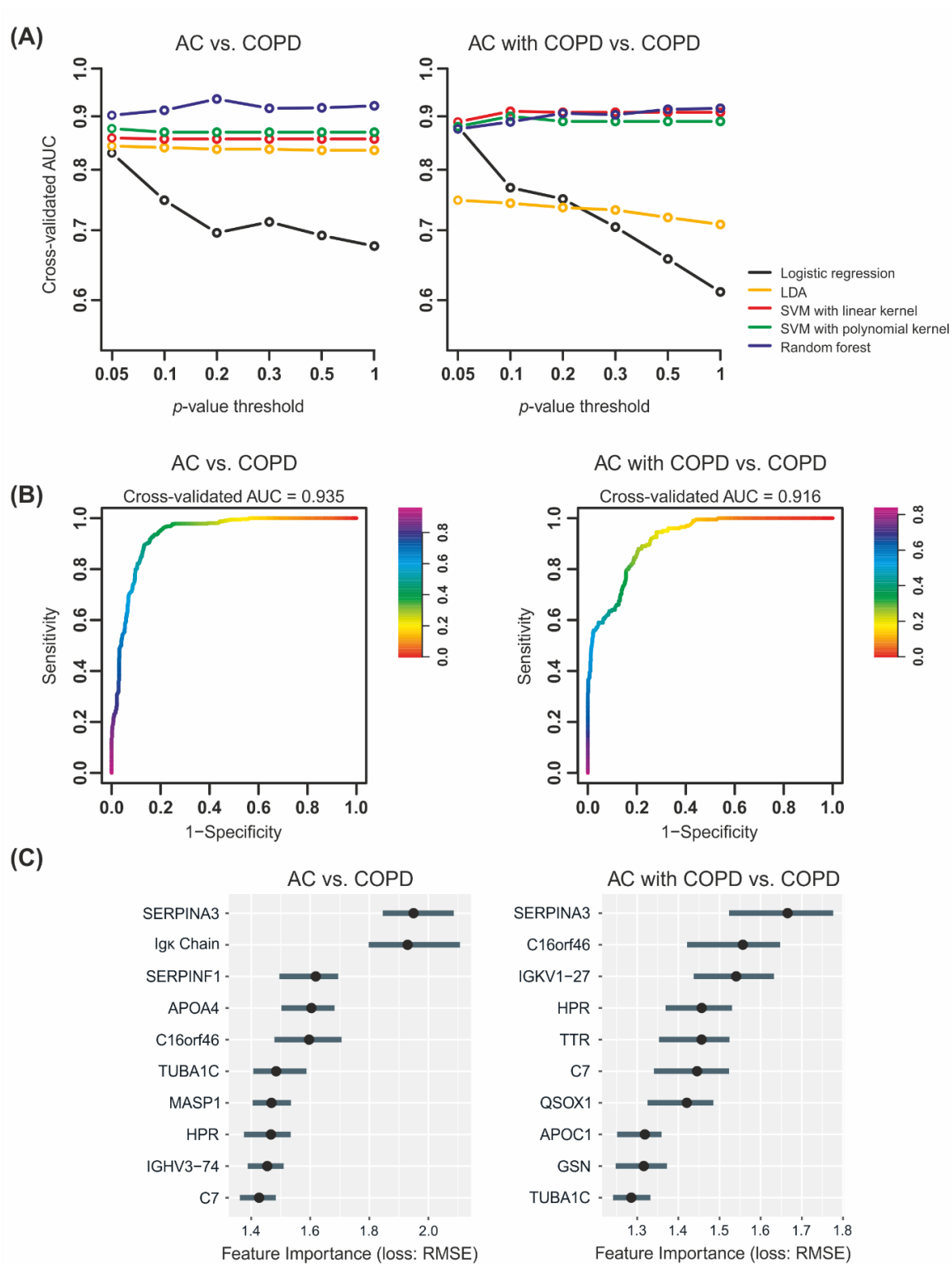


Figure 4. Results of machine learning approaches. (A) Five machine learning approaches were compared for classification of AC vs. COPD (left panel) and AC with COPD vs. COPD (right panel). Different *p*-value thresholds were assessed for feature selection and plotted against the respective ten-times-repeated 10-fold-cross-validated AUCs. (B) Receiver operating characteristic (ROC) curves for the best-performing random forest classifiers (AC vs. COPD: *p*-value threshold = 0.2; AC with COPD vs. COPD: no *p*-value threshold). (C) Feature importance plots illustrating the relative influence of individual proteins on multivariate classification models. Proteins represented by gene names (except Igκ Chain).

Table 2. Characteristics of random forest classifiers with the highest AUCs *.

(1) AC vs. COPD ¹		(2) AC with COPD vs. COPD ²	
AUC	0.935	AUC	0.916
PRAUC	0.928	PRAUC	0.882
Accuracy	0.865	Accuracy	0.873
Sensitivity ³	0.848	Sensitivity ⁵	0.570
Specificity ⁴	0.879	Specificity ⁴	0.965

* Ten-times-repeated 10-fold cross-validated. ¹ For feature selection, a *p*-value threshold of 0.2 was used. ² No *p*-value threshold was applied for feature selection. ³ Sensitivity corresponds to true classification of AC. ⁴ Specificity corresponds to true classification of COPD. ⁵ Sensitivity corresponds to true classification of AC with COPD.

Table 3. Metrics for train-test-split validation for the comparison AC vs. COPD.

		Minimum ¹	Mean ¹	Maximum ¹
AUC	Train set ²	0.85	0.901	0.965
	Test set ³	0.667	0.823	0.936
PRAUC	Train set	0.763	0.864	0.968
	Test set	0.554	0.766	0.931
Accuracy	Train set	0.76	0.831	0.91
	Test set	0.65	0.753	0.85
Sensitivity	Train set	0.726	0.815	0.905
	Test set	0.5	0.763	1
Specificity	Train set	0.759	0.844	0.941
	Test set	0.455	0.745	1

¹ Model metrics represent 50 repetitions of random train-test-splits. ² Model built on train set with ten-times-repeated 10-fold cross-validation. ³ Model validated on test set representing 1/3 of the whole dataset.

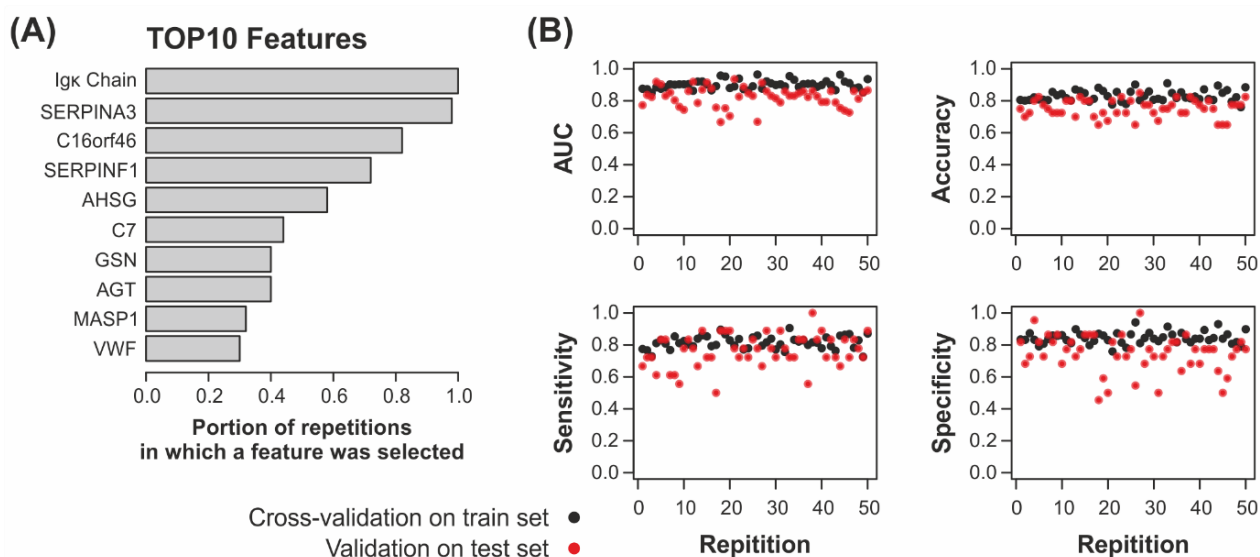


Figure 5. Results of intra-set validation for the comparison AC vs. COPD. The dataset was randomly split into train and test sets for 50 repetitions. The random forest model was developed with a ten-times-repeated 10-fold-cross-validation on the train set and validated on the test set. (A) Top 10 list of the most frequently selected features (i.e., proteins, represented by gene names (except Igκ Chain)). (B) Characteristics of the random forest classifier for cross-validation on the train sets (black) and validation on the test sets (red), respectively, were plotted against the number of repetitions.

2.4. Univariate Statistics and Machine Learning Reveal Candidates for a Biomarker Panel

The analysis of plasma proteomics data using univariate statistics revealed proteins that were significantly regulated between the AC subgroups with and without COPD, and COPD. The overlap between these proteins was small, illustrating the differences between the AC subgroups and the necessity of their individual consideration. Taken together, univariate analyses led to a comprehensive picture of protein regulations between the individual patient (sub-)groups and a list of robustly regulated proteins. The feature importance analysis of the best performing classification model resulted in complementary lists of candidates. Here, the most informative feature for both comparisons was Alpha-1-antichymotrypsin (SERPINA3), which was also significantly regulated in AC with COPD vs. COPD. Igκ Chain, which was regulated in both comparisons of AC subgroups vs. COPD, was also found to be highly informative using machine learning. These two features were also selected in 49 and 50 repetitions of the intra-set validation for AC vs. COPD. Further highly influential candidates were C7 and QSOX1, which were also identified by univariate testing. For univariate testing, a RoM threshold was applied to select candidates with distinct changes in intensities. For machine learning, no such filter was applied and, without considering the RoM threshold, additional proteins overlapped between both lists (i.e., SERPINF1, TTR, TUB1C, APOA4, and APOC1; Supplementary Table S3). Interestingly, C16orf46 and HPR, which were included in both random forest models, were not found to be significantly regulated in any univariate comparison. C16orf46 was selected in 41 of 50 repetitions during intra-set validation for AC vs. COPD.

3. Discussion

Identification of diagnostic biomarkers, especially in easily accessible body fluids such as blood, is the ultimate aim of clinical biomarker research. While the idealized concept of a single biomarker promises to be easily applicable, biomarker panels might realistically achieve higher precision, also taking into account the heterogeneity of the disease. On the pre-clinical level, mass spectrometry offers multiplex analysis of several hundred (in the case of plasma) to thousands (in the case of tissue) of proteins per sample [23,24]. State-of-the-art instrumentation and analytical strategies allow for sufficient sample throughput and accurate quantification, rendering LC-MS/MS a powerful technique for clinical biomarker discovery [25–27]. Discovery approaches are often based on tissue analysis and the subsequent transfer of candidates to other assays and easily accessible samples such as plasma. Naturally, many candidates are lost during this attempt and never successfully measured in blood. The analysis of plasma samples circumvents this step, allowing for direct analysis of the sample that is used in routine diagnostics. In addition to this advantage, the enormous dynamic range of the plasma proteome still makes it a challenging kind of sample [28]. Consequently, plasma proteomics studies mostly cover typical plasma proteins, which, nevertheless, might contain valuable information for diagnostics, especially when combined in a multi-biomarker panel [29,30].

We performed a mass-spectrometry-based analysis of plasma and a combined data analysis using univariate statistics and machine learning. We addressed the problem of minimal-invasive differentiation of AC and COPD with the aim of identifying novel biomarker candidates to be implemented in a biomarker panel. To this end, samples from AC patients with and without COPD, COPD patients, and hospital controls (HC) without the target disease were analyzed. COPD and lung cancer are closely linked diseases with a common etiology and similarities in their molecular pathogenesis [31], with COPD being an independent risk factor of lung cancer [7,32]. The effect of inflammatory signaling, for example, has already been shown in COPD [33], but in lung adenocarcinoma, there are also always sites of inflammation [34]. As a consequence, the comparison between AC and HC might result in regulated proteins that might be associated with, e.g., inflammation, but not be specific for AC. In our study, the respective univariate comparisons between diseased groups and HC lead to most significantly differentially abundant proteins, supporting the assumption that many of these are not specific for AC. This highlights the need for the

direct comparison of clinically relevant AC and COPD patients to identify proteins that are specific for the respective differentiation. Hence, we focused on the discrimination of AC vs. COPD and the more challenging differentiation of AC with COPD vs. COPD using machine learning.

We constructed the pipeline for model optimization with respect to avoiding overfitting of the trained models. The selected classification algorithms were capable of handling low numbers of observations, and artificial neural networks were not considered, due to their known tendency toward overfitting when used with small datasets. In addition, we selected proteins for model development based on their univariate relevance on the target variable and removed redundant proteins to reduce the overall number of features. Finally, the algorithms were optimized for only ten features and the ten-times-repeated 10-fold cross-validation approach was applied to prevent overoptimistic evaluation of the model's precision. The comparison of machine learning approaches led to the best results for random forest classification; however, for the comparison of AC with COPD vs. COPD, support vector machines performed comparably well. Both best-performing classifiers yielded cross-validated AUCs over 0.9. To further evaluate our results, we repeated the development of the random forest model for (A) AC vs. COPD on 50 randomly selected train-test-set splits with one third of the dataset used for testing. We compared the model characteristics on the test sets to the results on the complete dataset to ensure that the models are not overoptimistic. As expected, the mean AUC of 0.823 in intra-set validation was slightly lower than on the complete dataset but still showed a very good precision. In addition, we compared the most frequently selected features to the results from the complete dataset. Here, four of the five most important features were selected in over 70% of the train-test-splits. For (B) AC with COPD vs. COPD, we did not perform the same analysis, because of the limited group size of AC with COPD ($n = 21$) and the unbalanced nature of the dataset. Although our machine learning approach addressed the problem of overfitting consequently, we must assume that the classifiers are still over-optimistic and should be evaluated with an independent patient cohort before considering them in clinical application. However, the excellent differentiation of clinically very similar patient groups clearly illustrates the great potential of plasma proteomics for biomarker discovery and diagnostics. In addition, the feature importance suggests several proteins to be considered for biomarker panel development.

The proteins that were found to be significantly regulated using univariate statistics or included in the multivariate algorithms were classical plasma proteins. Thus, a pathomechanistic interpretation of the observed differences in abundance is difficult and mostly speculative. Obviously, this was the case for Ig kappa light chain, which was more abundant in AC with or without COPD vs. COPD and found to be a highly informative feature in the classifier for AC vs. COPD. Bottom-up mass spectrometry approaches do not allow routine sequencing of antibody binding sites and the specificity of the regulated Ig chains remains unknown. Free light chains, however, which are synthesized in excess during antibody production, were also shown to be highly bioactive molecules modulating the immune response, e.g., in COPD [35,36]. Although the exact nature of the Ig kappa light chain cannot be clarified based on our data, our findings suggest a robust difference between AC and COPD patients, suggesting its consideration for a biomarker panel. Alpha-1-antichymotrypsin (SERPINA3) belongs to a superfamily of serine proteinase inhibitors. Although its exact biological function is not known, it has been described in association with diverse pathologies, such as acute kidney injury, cardiovascular disease, and (lung) cancer [37–40]. It was described to promote metastasis and the epithelial–mesenchymal transition in breast cancer [41] and was found to be upregulated in lung cancer in comparison to control patients [42]. In our study, the highest SERPINA3 abundance was found in COPD patients (Supplementary Figure S3), again demonstrating the need for the direct comparison of clinically relevant patient groups. Although SERPINA3 was less abundant compared to COPD, it might be included in a biomarker panel. Its value for differentiation of AC and COPD has been underlined by both univariate statistics and random forest

classification. In addition, several other protein candidates should be taken into consideration for the development of a biomarker panel. Apolipoprotein A-IV (APOA4) and Transthyretin (TTR), which were included in random forest classifiers, were both described to be dysregulated in adenocarcinoma before [43]. While APOA4 was less abundant in comparison to nontumorous tissue, TTR was upregulated, which is in concordance with our observations (Supplementary Figure S3). SAA1, which was one of the significantly regulated proteins overlapping between both comparisons of AC with and without COPD vs. COPD, was reported as a biomarker for lung cancer before [44–46]. Contrary to most previous reports, SAA1 was significantly downregulated in AC compared to COPD and hospital controls. Notably, most studies reported in the literature used healthy donors as controls, which is not fully comparable with our control group of subjects that were hospitalized under suspicion of lung cancer. Thus, plasma levels of SAA1 should be further studied in clinically relevant patient groups to investigate its true value as a biomarker for lung adenocarcinoma. Notably, intra-set validation supported the value of Ig kappa light chain, SERPINA3, and others, but also highlighted the relevance of the uncharacterized protein C16orf46, which might be further studied in the context of AC and COPD.

4. Materials and Methods

4.1. Patient Plasma Samples and Clinical Data

Samples and data were collected at the Johanniter-Clinics Bonn and the Malteser Krankenhaus Seliger Gerhard Bonn/Rhein-Sieg. The local ethics committees at the Ruhr-University Bochum, Ärztekammer Nordrhein, and Cologne University approved the study (approval numbers 17-5970, 2017133, 17-162). Written, informed consent was obtained from each patient, and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Peripheral blood was collected in 9 mL S-Monovette EDTA gel tubes (Sarstedt, Nümbrecht, Germany). Within 30 min after blood collection, samples were centrifuged at $2000\times g$ for ten minutes at room temperature. Plasma was separated subsequently and frozen immediately at $-80\text{ }^{\circ}\text{C}$ until analyses. Samples were measured separately in two batches ($n = 63$ and $n = 133$), which were subsequently combined for analysis. In total, plasma samples from 43 AC patients without COPD, 21 AC patients with COPD, 77 COPD patients, and 35 hospital controls (HCs) were analyzed ($n = 176$, Table 4, Supplementary Table S1). Control subjects were patients with suspicious malignant disease, which was not confirmed subsequently.

Table 4. Composition of the analyzed patient cohorts.

Group	Description	Mean Age (Years)	Sex	Smoking Behavior	
AC* (n = 64)	AC w/o COPD (n = 43)	AC-patients without diagnosed COPD	67.17 ± 9.43, min. 41, max. 85	25 female, 18 male	20 smokers, 10 ex-smokers, 13 never-smokers
	AC with COPD (n = 21)	AC-patients with diagnosed COPD	64.48 ± 8.89, min. 52, max. 84	12 female, 9 male	11 smokers, 8 ex-smokers, 2 never-smokers
COPD § (n = 77)	COPD-patients without AC	68.61 ± 10.43, min. 38, max. 87	36 female, 41 male	36 smokers, 33 ex-smokers, 6 never-smokers, 2 NA	
HC (n = 35)	Hospital controls	65.34 ± 12.40 min. 41, max. 82	16 female, 19 male	14 smokers, 13 ex-smokers, 8 never-smokers	

* Lung adenocarcinoma. § Chronic obstructive pulmonary disease.

4.2. Sample Preparation for LC–MS/MS Analysis

Briefly, 1 μL of plasma per sample was mixed with 24 μL of buffer (1% sodium deoxycholate, 10 mM tris (2-carboxyethyl) phosphine, 40 mM chloroacetamide, and 100 mM Tris; pH 8.5) and incubated at 95 °C for 10 min. After cooling down to room temperature, 10 μL of 50 mM ammonium bicarbonate and 10 μL of paramagnetic beads were added according to the SP3 protocol [47]. For protein binding, 140 μL of acetonitrile was added and samples were incubated at room temperature for 10 min. Afterward, the beads were washed twice with 200 μL of 70% ethanol and once with 200 μL of acetonitrile using a magnetic rack. After a short airdrying process, 70 μL of 50 mM ammonium bicarbonate containing trypsin (1:50, *w/w*, SERVA Electrophoresis, Heidelberg, Germany) was added to digest the protein overnight at 37 °C. The supernatant was transferred to a glass vial and dried in a vacuum centrifuge. An amount of 100 μL of 0.1% TFA was used to dissolve the peptides and the resulting peptide concentration was about 1 $\mu\text{g}/\mu\text{L}$.

4.3. LC–MS/MS Analysis

The LC–MS/MS analysis was carried out using an Ultimate 3000 RSLCnano liquid chromatography system coupled online to a Q Exactive HF (batch 1) or Q Exactive (batch 2) mass spectrometer (all Thermo Fisher Scientific). Next, 100 ng of plasma peptides per sample were injected for analysis. The peptides were pre-concentrated for 7 min on a trap column (Acclaim[®] PepMap 100, 75 $\mu\text{m} \times 2 \text{ cm}$, C18, 5 μm , 100 Å) using 30 $\mu\text{L}/\text{min}$ of 0.1% TFA as the loading solvent. Subsequent separation on an analytical column (Acclaim[®] PepMap RSLC, 75 $\mu\text{m} \times 50 \text{ cm}$, nano Viper, C18, 5 μm , 100 Å) was carried out using a gradient from 5 to 40% solvent B in solvent A over 98 min (batch 1) or 38 min (batch 2) (solvent A: 0.1% formic acid; solvent B: 0.1% formic acid, 84% acetonitrile). A flow rate of 400 nL/min was used with a column oven temperature of 60 °C. Data-dependent acquisition mode was used. Full scans were acquired in the Orbitrap analyzer (mass range: 350–1400 *m/z*, resolution: 60,000). The Fourier Transform Mass Spectrometry full-scan Automatic Gain Control target was set to 3×10^6 with a maximum injection time of 80 ms. The number of micro-scans was set to 1. The 10 most abundant ions of a spectrum acquired at the MS1 level were fragmented using HCD (higher-energy collisional dissociation) with a normalized collision energy of 30% and an isolation width of 2 *m/z*. Fragment mass spectra were acquired in the orbitrap with a maximum injection time of 100 ms. Samples were measured in a random sequence. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035120 and 10.6019/PXD035120".

4.4. Protein Identification and Quantification

Protein identification and label-free quantification were performed using MaxQuant (version 2.0.1.0; MPI of Biochemistry, Martinsried, Germany) separately for both analyzed batches. Unless specifically mentioned, the MaxQuant default parameters were used. Spectra were searched against the Uniprot/Swissprot database version 2021_03 restricted to *Homo sapiens* (20,362 entries) and Biognosys iRT peptides (1 entry). Variable modifications of methionine (oxidation) and protein N-termini (acetylation) were considered as well as fixed modification of cysteine (carbamidomethyl). Modified peptides were considered for quantification and the match between runs function was enabled. The false discovery rate (FDR) was set to 0.01 for peptide-spectrum matches (PSMs) and proteins. Protein groups were used for further processing and reverse hits, proteins were identified by site, and identified iRT peptides were omitted. Potential contaminants were manually reviewed to avoid exclusion of typical plasma proteins.

4.5. Batch Normalization

LFQ intensities were normalized to reduce technical variation and the batch effects between the two batches to allow analyzing them together. First, each batch was normalized separately using the LOESS normalization method [48] using the R package limma

(version 3.44.3) [49] on the log-transformed protein intensities. For each protein, a linear regression model was calculated by using the batch number as a categorical independent variable. The obtained coefficients were an estimation of the underlying batch effect, which were subsequently subtracted from the protein intensities. This normalization procedure reduces the batch differences for better comparability of the samples. As the LC gradient was adapted for the second batch due to the higher sample amount and a different instrument, 20 samples of the first batch were also measured in the second batch and taken as a quality control for normalization. The quality of the normalization was evaluated using boxplots, PCA (principal component analysis), and MA plots (Figure 2, Supplementary Figures S1 and S2).

4.6. Statistical Analysis

The statistical analysis was conducted using R version 4.0.3 (R Core Team 2020, Vienna, Austria). For this, the intensities of the doubly measured samples were averaged after the batch normalization. Protein groups quantified in a minimum of five patients in the compared patient groups were considered for testing. In order to identify significantly different protein groups between the experimental groups, normalized intensities were analyzed by application of a Welch-ANOVA (R package car version 3.0-10). This was followed by single Welch tests for each protein as a post hoc method in all possible pairwise group comparisons. Ratios of means (RoMs) between groups were determined on the delogarithmized intensities. The FDR was controlled by adjusting ANOVA p -values using the method of Benjamini and Hochberg [50]. Post hoc p -values were adjusted by the method of Bonferroni–Holm [51] for each protein separately. Proteins were considered significant with p_{FDR} -values ≤ 0.05 (from ANOVA and post hoc test) and an absolute RoM ≥ 1.5 . For the comparison of AC vs. COPD, a separate Welch test was used and corrected according to Benjamini–Hochberg.

4.7. Machine Learning

The normalized protein intensities were used to develop classification models. For the two research questions, separation of (1) AC vs. COPD and (2) AC with COPD vs. COPD, the same pipeline for model optimization, consisting of feature selection and model optimization, was used (Figure 1). Proteins with more than 30% missing values or a variance near zero across all samples were omitted. In addition, redundant proteins were identified according to Spearman's rank correlation coefficient >0.7 and removed from analysis. For feature selection, proteins were selected according to their influence on the respective target variable. Therefore, Kruskal–Wallis and Levene tests were calculated. p -value thresholds of 0.05, 0.1, 0.2, 0.3, and 0.5 were applied to optimize this feature selection. In addition, the performance without feature selection by statistical tests was assessed. For model optimization, logistic regression, linear discriminant analysis (LDA), and support vector machines (SVMs) with linear as well as polynomial kernel and random forest were applied and compared. We optimized the random forest and SVM models for the best subset of ten proteins from the selected features to reduce overfitting. For LDA and logistic regression, the best subset was determined by the recursive feature selection procedure. The models were trained with the ten-times-repeated 10-fold cross-validation approach using the area under the precision recall curve (PRAUC) as the optimization criterion, which was chosen to cope with the unbalanced datasets. The models with the highest AUC (area under the ROC curve) and PRAUC were analyzed according to permutation feature importance (<https://doi.org/10.48550/arXiv.2006.04628> (accessed on 9 February 2022)). Permutation feature importance calculates the increase in the root-mean-square error (RSME) of the prediction model when a single feature value is randomly shuffled. The increase in RSME is indicative of how much the model depends on the feature. In addition, we repeated the development of the final model on 50 randomly selected train-test-splits with 2/3 of the dataset selected for model development and 1/3 as the test set for validation.

We compared the model characteristics on the test sets to the results on the complete dataset and evaluated the number of repetitions in which a feature was selected by the final model.

5. Conclusions

A realistic protein biomarker panel in clinical use needs to be limited to a manageable number of proteins, which, at least in the nearer future, will be measured by multiplex ELISA or comparable techniques. These requirements limit the variety of biomarker candidates for assay development. Therefore, the considered proteins need to address a specific clinical problem, which was analyzed by use of the appropriate clinical groups. Here, we describe a panel of proteins that were differentially abundant between patients with AC and COPD, while taking into consideration that many AC patients also suffer from underlying COPD. Addressing this problem represents the first and urgently necessary step in discovery of reliable markers that distinguish between early AC, COPD, and their overlaps. The combination of univariate statistics and machine learning adds complementary information and, thus, provides insights that would not be available with either approach alone: Classification algorithms strive for the highest diagnostic precision while permutation feature importance or univariate statistics allow identification of the most influential proteins. Those proteins that were identified in this study might serve as starting points for development of a biomarker panel, which addresses the heterogeneity of AC and COPD, ideally in an early detection setting.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms231911242/s1>.

Author Contributions: Conceptualization, T.B. (Thomas Behrens), T.B. (Thomas Bruning), and B.S.; methodology, S.M., W.C. and J.H.; software, D.K., K.S., J.U. and M.E.; formal analysis, M.A.; investigation, R.B. and J.F.; data curation, D.K., K.S., J.U. and M.E.; writing—original draft preparation, T.B. (Thilo Bracht) and K.E.W.; writing—review and editing, G.J., T.B. (Thomas Behrens), M.E., M.B. and B.S.; visualization, T.B. (Thilo Bracht) and K.S.; supervision, M.E. and B.S.; project administration, T.B. (Thomas Behrens) and Y.-D.K.; funding acquisition, T.B. (Thomas Behrens), T.B. (Thomas Bruning) and Y.-D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the German Social Accident Insurance (DGUV; project FP339A) and de. NBI, a project of the Federal Ministry of Education and Research (BMBF) (FKZ 031 A 534 A).

Institutional Review Board Statement: The local ethics committees at the Ruhr-University Bochum, Ärztekammer Nordrhein, and Cologne University approved the study (approval numbers 17-5970, 2017133, 17-162).

Informed Consent Statement: Written, informed consent was obtained from each patient, and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035120 and 10.6019/PXD035120.

Acknowledgments: The authors would like to thank Kristin Fuchs and Birgit Zülch for their excellent technical assistance. We acknowledge support by the Open Access Publication Funds of the Ruhr-Universität Bochum.

Conflicts of Interest: R.B. has received honoraries for lectures and advisory boards from AbbVie, Amgen, AstraZeneca, Bayer, BMS, Boehringer-Ingelheim, Illumina, Janssen, Lilly, Merck-Serono, MSD, Novartis, Qiagen, Pfizer, and Roche. R.B. is a scientific Director and Co-Founder of Targos MP Inc./Kassel Germany, Gnothis Inc./Stockholm Sweden.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Adcock, I.M.; Caramori, G.; Barnes, P.J. Chronic obstructive pulmonary disease and lung cancer: New molecular insights. *Respir. Int. Rev. Thorac. Dis.* **2011**, *81*, 265–284. [[CrossRef](#)] [[PubMed](#)]
3. Parris, B.A.; O'Farrell, H.E.; Fong, K.M.; Yang, I.A. Chronic obstructive pulmonary disease (copd) and lung cancer: Common pathways for pathogenesis. *J. Thorac. Dis.* **2019**, *11*, S2155–S2172. [[CrossRef](#)] [[PubMed](#)]
4. Lundback, B.; Lindberg, A.; Lindstrom, M.; Ronmark, E.; Jonsson, A.C.; Jonsson, E.; Larsson, L.G.; Andersson, S.; Sandstrom, T.; Larsson, K. Not 15 but 50% of smokers develop copd?—Report from the obstructive lung disease in northern sweden studies. *Respir. Med.* **2003**, *97*, 115–122. [[CrossRef](#)]
5. Burney, P.G.; Patel, J.; Newson, R.; Minelli, C.; Naghavi, M. Global and regional trends in copd mortality, 1990–2010. *Eur. Respir. J.* **2015**, *45*, 1239–1247. [[CrossRef](#)]
6. Young, R.P.; Duan, F.; Chiles, C.; Hopkins, R.J.; Gamble, G.D.; Greco, E.M.; Gatsonis, C.; Aberle, D. Airflow limitation and histology shift in the national lung screening trial. The nlst-acrin cohort substudy. *Am. J. Respir. Crit. Care Med.* **2015**, *192*, 1060–1067. [[CrossRef](#)]
7. Young, R.P.; Hopkins, R.J.; Christmas, T.; Black, P.N.; Metcalf, P.; Gamble, G.D. Copd prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur. Respir. J.* **2009**, *34*, 380–386. [[CrossRef](#)]
8. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [[CrossRef](#)]
9. Conrads, T.P.; Hood, B.L.; Veenstra, T.D. Sampling and analytical strategies for biomarker discovery using mass spectrometry. *BioTechniques* **2006**, *40*, 799–805. [[CrossRef](#)]
10. Szabo, M.; Hajba, L.; Kun, R.; Guttman, A.; Csanky, E. Proteomic and glycomic markers to differentiate lung adenocarcinoma from copd. *Curr. Med. Chem.* **2020**, *27*, 3302–3313. [[CrossRef](#)]
11. Zmay, T.N.; Zmay, G.S.; Kolovskaya, O.S.; Zukov, R.A.; Petrova, M.M.; Gargaun, A.; Berezovski, M.V.; Kichkailo, A.S. Current and prospective protein biomarkers of lung cancer. *Cancers* **2017**, *9*, 155. [[CrossRef](#)] [[PubMed](#)]
12. Chung, K.; Nishiyama, N.; Yamano, S.; Komatsu, H.; Hanada, S.; Wei, M.; Wanibuchi, H.; Suehiro, S.; Kakehashi, A. Serum agr2 as an early diagnostic and postoperative prognostic biomarker of human lung adenocarcinoma. *Cancer Biomark. Sect. A Dis. Markers* **2011**, *10*, 101–107. [[CrossRef](#)] [[PubMed](#)]
13. Sholl, L.M. Biomarkers in lung adenocarcinoma: A decade of progress. *Arch. Pathol. Lab. Med.* **2015**, *139*, 469–480. [[CrossRef](#)] [[PubMed](#)]
14. Bittner, N.; Ostoros, G.; Geczi, L. New treatment options for lung adenocarcinoma—In view of molecular background. *Pathol. Oncol. Res. POR* **2014**, *20*, 11–25. [[CrossRef](#)] [[PubMed](#)]
15. Liu, Z.; Gao, Y.; Hao, F.; Lou, X.; Zhang, X.; Li, Y.; Wu, D.; Xiao, T.; Yang, L.; Li, Q.; et al. Secretomes are a potential source of molecular targets for cancer therapies and indicate that apoe is a candidate biomarker for lung adenocarcinoma metastasis. *Mol. Biol. Rep.* **2014**, *41*, 7507–7523. [[CrossRef](#)]
16. Li, W.; Zheng, H.; Qin, H.; Liu, G.; Ke, L.; Li, Y.; Li, N.; Zhong, X. Exploration of differentially expressed plasma proteins in patients with lung adenocarcinoma using itraq-coupled 2d lc-ms/ms. *Clin. Respir. J.* **2018**, *12*, 2036–2045. [[CrossRef](#)]
17. Lai, T.; Wu, D.; Chen, M.; Cao, C.; Jing, Z.; Huang, L.; Lv, Y.; Zhao, X.; Lv, Q.; Wang, Y.; et al. Ykl-40 expression in chronic obstructive pulmonary disease: Relation to acute exacerbations and airway remodeling. *Respir. Res.* **2016**, *17*, 31. [[CrossRef](#)]
18. Johansson, S.L.; Roberts, N.B.; Schlosser, A.; Andersen, C.B.; Carlsen, J.; Wulf-Johansson, H.; Saekmose, S.G.; Titlestad, I.L.; Tornoe, I.; Miller, B.; et al. Microfibrillar-associated protein 4: A potential biomarker of chronic obstructive pulmonary disease. *Respir. Med.* **2014**, *108*, 1336–1344. [[CrossRef](#)]
19. Angata, T.; Fujinawa, R.; Kurimoto, A.; Nakajima, K.; Kato, M.; Takamatsu, S.; Korekane, H.; Gao, C.X.; Ohtsubo, K.; Kitazume, S.; et al. Integrated approach toward the discovery of glyco-biomarkers of inflammation-related diseases. *Ann. N. Y. Acad. Sci.* **2012**, *1253*, 159–169. [[CrossRef](#)]
20. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-generation machine learning for biological networks. *Cell* **2018**, *173*, 1581–1592. [[CrossRef](#)]
21. Byvatov, E.; Schneider, G. Support vector machine applications in bioinformatics. *Appl. Bioinform.* **2003**, *2*, 67–77.
22. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
23. Geyer, P.E.; Holdt, L.M.; Teupser, D.; Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **2017**, *13*, 942. [[CrossRef](#)] [[PubMed](#)]
24. Megger, D.A.; Bracht, T.; Meyer, H.E.; Sitek, B. Label-free quantification in clinical proteomics. *Biochim. Biophys. Acta* **2013**, *1834*, 1581–1590. [[CrossRef](#)]
25. Witzke, K.E.; Grosserueschkamp, F.; Jutte, H.; Horn, M.; Roghmann, F.; von Landenberg, N.; Bracht, T.; Kallenbach-Thieltges, A.; Kafferlein, H.; Bruning, T.; et al. Integrated fourier transform infrared imaging and proteomics for identification of a candidate histochemical biomarker in bladder cancer. *Am. J. Pathol.* **2019**, *189*, 619–631. [[CrossRef](#)]
26. Bracht, T.; Schweinsberg, V.; Trippler, M.; Kohl, M.; Ahrens, M.; Padden, J.; Naboulsi, W.; Barkovits, K.; Megger, D.A.; Eisenacher, M.; et al. Analysis of disease-associated protein expression using quantitative proteomics-fibulin-5 is expressed in association with hepatic fibrosis. *J. Proteome Res.* **2015**, *14*, 2278–2286. [[CrossRef](#)]

27. Naboulsi, W.; Megger, D.A.; Bracht, T.; Kohl, M.; Turewicz, M.; Eisenacher, M.; Voss, D.M.; Schlaak, J.F.; Hoffmann, A.C.; Weber, F.; et al. Quantitative tissue proteomics analysis reveals versican as potential biomarker for early-stage hepatocellular carcinoma. *J. Proteome Res.* **2016**, *15*, 38–47. [[CrossRef](#)]
28. Anderson, N.L.; Anderson, N.G. The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteom. MCP* **2002**, *1*, 845–867. [[CrossRef](#)]
29. Niu, L.; Geyer, P.E.; Wewer Albrechtsen, N.J.; Gluud, L.L.; Santos, A.; Doll, S.; Treit, P.V.; Holst, J.J.; Knop, F.K.; Vilsboll, T.; et al. Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. *Mol. Syst. Biol.* **2019**, *15*, e8793. [[CrossRef](#)]
30. Captur, G.; Heywood, W.E.; Coats, C.; Rosmini, S.; Patel, V.; Lopes, L.R.; Collis, R.; Patel, N.; Syrris, P.; Bassett, P.; et al. Identification of a multiplex biomarker panel for hypertrophic cardiomyopathy using quantitative proteomics and machine learning. *Mol. Cell. Proteom. MCP* **2020**, *19*, 114–127. [[CrossRef](#)]
31. Durham, A.L.; Adcock, I.M. The relationship between copd and lung cancer. *Lung Cancer* **2015**, *90*, 121–127. [[CrossRef](#)] [[PubMed](#)]
32. Koshiol, J.; Rotunno, M.; Consonni, D.; Pesatori, A.C.; De Matteis, S.; Goldstein, A.M.; Chaturvedi, A.K.; Wacholder, S.; Landi, M.T.; Lubin, J.H.; et al. Chronic obstructive pulmonary disease and altered risk of lung cancer in a population-based case-control study. *PLoS ONE* **2009**, *4*, e7380. [[CrossRef](#)] [[PubMed](#)]
33. Mouronte-Roibas, C.; Leiro-Fernandez, V.; Ruano-Ravina, A.; Ramos-Hernandez, C.; Casado-Rey, P.; Botana-Rial, M.; Garcia-Rodriguez, E.; Fernandez-Villar, A. Predictive value of a series of inflammatory markers in copd for lung cancer diagnosis: A case-control study. *Respir. Res.* **2019**, *20*, 198. [[CrossRef](#)] [[PubMed](#)]
34. Cho, W.C.; Kwan, C.K.; Yau, S.; So, P.P.; Poon, P.C.; Au, J.S. The role of inflammation in the pathogenesis of lung cancer. *Expert Opin. Ther. Targets* **2011**, *15*, 1127–1137. [[CrossRef](#)]
35. Basile, U.; Gulli, F.; Gragnani, L.; Napodano, C.; Pocino, K.; Rapaccini, G.L.; Mussap, M.; Zignego, A.L. Free light chains: Eclectic multipurpose biomarker. *J. Immunol. Methods* **2017**, *451*, 11–19. [[CrossRef](#)]
36. Braber, S.; Thio, M.; Blokhuis, B.R.; Henricks, P.A.; Koelink, P.J.; Groot Kormelink, T.; Bezemer, G.F.; Kerstjens, H.A.; Postma, D.S.; Garssen, J.; et al. An association between neutrophils and immunoglobulin free light chains in the pathogenesis of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2012**, *185*, 817–824. [[CrossRef](#)]
37. Hu, J.; Boeri, M.; Sozzi, G.; Liu, D.; Marchiano, A.; Roz, L.; Pelosi, G.; Gatter, K.; Pastorino, U.; Pezzella, F. Gene signatures stratify computed tomography screening detected lung cancer in high-risk populations. *EBioMedicine* **2015**, *2*, 831–840. [[CrossRef](#)]
38. Jung, Y.J.; Oh, I.J.; Kim, Y.; Jung, J.H.; Seok, M.; Lee, W.; Park, C.K.; Lim, J.H.; Kim, Y.C.; Kim, W.S.; et al. Clinical validation of a protein biomarker panel for non-small cell lung cancer. *J. Korean Med. Sci.* **2018**, *33*, e342. [[CrossRef](#)]
39. Sanchez-Navarro, A.; Murillo-de-Ozores, A.R.; Perez-Villalva, R.; Linares, N.; Carbajal-Contreras, H.; Flores, M.E.; Gamba, G.; Castaneda-Bueno, M.; Bobadilla, N.A. Transient response of serpin3 during cellular stress. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2022**, *36*, e22190. [[CrossRef](#)]
40. Sanchez-Navarro, A.; Gonzalez-Soria, I.; Caldino-Bohn, R.; Bobadilla, N.A. An integrative view of serpins in health and disease: The contribution of serpin3. *Am. J. Physiol. Cell Physiol.* **2021**, *320*, C106–C118. [[CrossRef](#)]
41. Zhang, Y.; Tian, J.; Qu, C.; Peng, Y.; Lei, J.; Li, K.; Zong, B.; Sun, L.; Liu, S. Overexpression of serpin3 promotes tumor invasion and migration, epithelial-mesenchymal-transition in triple-negative breast cancer cells. *Breast Cancer* **2021**, *28*, 859–873. [[CrossRef](#)] [[PubMed](#)]
42. Jung, Y.J.; Katilius, E.; Ostroff, R.M.; Kim, Y.; Seok, M.; Lee, S.; Jang, S.; Kim, W.S.; Choi, C.M. Development of a protein biomarker panel to detect non-small-cell lung cancer in korea. *Clin. Lung Cancer* **2017**, *18*, e99–e107. [[CrossRef](#)] [[PubMed](#)]
43. Borlak, J.; Langer, F.; Chatterji, B. Serum proteome mapping of egf transgenic mice reveal mechanistic biomarkers of lung cancer precursor lesions with clinical significance for human adenocarcinomas. *Biochim. Biophys. Acta. Mol. Basis Dis.* **2018**, *1864*, 3122–3144. [[CrossRef](#)] [[PubMed](#)]
44. Kim, Y.J.; Gallien, S.; El-Khoury, V.; Goswami, P.; Sertamo, K.; Schlessner, M.; Berchem, G.; Domon, B. Quantification of saa1 and saa2 in lung cancer plasma using the isotype-specific prm assays. *Proteomics* **2015**, *15*, 3116–3125. [[CrossRef](#)] [[PubMed](#)]
45. Sung, H.J.; Ahn, J.M.; Yoon, Y.H.; Rhim, T.Y.; Park, C.S.; Park, J.Y.; Lee, S.Y.; Kim, J.W.; Cho, J.Y. Identification and validation of saa as a potential lung cancer biomarker and its involvement in metastatic pathogenesis of lung cancer. *J. Proteome Res.* **2011**, *10*, 1383–1395. [[CrossRef](#)]
46. Sung, H.J.; Jeon, S.A.; Ahn, J.M.; Seoul, K.J.; Kim, J.Y.; Lee, J.Y.; Yoo, J.S.; Lee, S.Y.; Kim, H.; Cho, J.Y. Large-scale isotype-specific quantification of serum amyloid a 1/2 by multiple reaction monitoring in crude sera. *J. Proteom.* **2012**, *75*, 2170–2180. [[CrossRef](#)]
47. Hughes, C.S.; Moggridge, S.; Muller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **2019**, *14*, 68–85. [[CrossRef](#)]
48. Valikangas, T.; Suomi, T.; Elo, L.L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* **2018**, *19*, 1–11. [[CrossRef](#)]
49. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
50. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300. [[CrossRef](#)]
51. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.