








## Article

# Proteomic Network Analysis of Bronchoalveolar Lavage Fluid in Ex-Smokers to Discover Implicated Protein Targets and Novel Drug Treatments for Chronic Obstructive Pulmonary Disease

Manoj J. Mammen<sup>1,2,\*</sup>, Chengjian Tu<sup>3,4</sup>, Matthew C. Morris<sup>5</sup>, Spencer Richman<sup>5</sup>, William Mangione<sup>2</sup>, Zackary Falls<sup>2</sup>, Jun Qu<sup>3,4</sup>, Gordon Broderick<sup>5</sup>, Sanjay Sethi<sup>6,7</sup> and Ram Samudrala<sup>2,\*</sup>

- <sup>1</sup> Department of Medicine, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA
- <sup>2</sup> Department of Biomedical Informatics, Jacobs School of Medicine and Biological Sciences, State University of New York, Buffalo, NY 14214, USA; w Mangione@buffalo.edu (W.M.); zmfalls@buffalo.edu (Z.F.)
- <sup>3</sup> Department of Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo, NY 14260, USA; ctu3@buffalo.edu (C.T.); junqu@buffalo.edu (J.Q.)
- <sup>4</sup> New York State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, NY 14203, USA
- <sup>5</sup> Center for Clinical Systems Biology, Rochester General Hospital, Rochester, NY 14621, USA; matthew.morris@rochesterregional.org (M.C.M.); scr6045@rit.edu (S.R.); gordon.broderick@rochesterregional.org (G.B.)
- <sup>6</sup> WNY VA Healthcare System, Buffalo, NY 14215, USA; ssethi@buffalo.edu
- <sup>7</sup> Department of Medicine, Jacobs School of Medicine and Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14214, USA
- \* Correspondence: manoj\_mammen@urmc.rochester.edu (M.J.M.); ram@compbio.org (R.S.); Tel.: +1-(585)-275-4161 (M.J.M.); +1-(206)-251-8852 (R.S.)



**Citation:** Mammen, M.J.; Tu, C.; Morris, M.C.; Richman, S.; Mangione, W.; Falls, Z.; Qu, J.; Broderick, G.; Sethi, S.; Samudrala, R. Proteomic Network Analysis of Bronchoalveolar Lavage Fluid in Ex-Smokers to Discover Implicated Protein Targets and Novel Drug Treatments for Chronic Obstructive Pulmonary Disease. *Pharmaceuticals* **2022**, *15*, 566. <https://doi.org/10.3390/10.3390/ph15050566>

Academic Editor: Nektarios Barabutis

Received: 14 March 2022

Accepted: 21 April 2022

Published: 1 May 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:** Bronchoalveolar lavage of the epithelial lining fluid (BALF) can sample the profound changes in the airway lumen milieu prevalent in chronic obstructive pulmonary disease (COPD). We compared the BALF proteome of ex-smokers with moderate COPD who are not in exacerbation status to non-smoking healthy control subjects and applied proteome-scale translational bioinformatics approaches to identify potential therapeutic protein targets and drugs that modulate these proteins for the treatment of COPD. Proteomic profiles of BALF were obtained from (1) never-smoker control subjects with normal lung function (n = 10) or (2) individuals with stable moderate (GOLD stage 2, FEV<sub>1</sub> 50–80% predicted, FEV<sub>1</sub>/FVC < 0.70) COPD who were ex-smokers for at least 1 year (n = 10). After identifying potential crucial hub proteins, drug–proteome interaction signatures were ranked by the computational analysis of novel drug opportunities (CANDO) platform for multiscale therapeutic discovery to identify potentially repurposable drugs. Subsequently, a literature-based knowledge graph was utilized to rank combinations of drugs that most likely ameliorate inflammatory processes. Proteomic network analysis demonstrated that 233 of the >1800 proteins identified in the BALF were significantly differentially expressed in COPD versus control. Functional annotation of the differentially expressed proteins was used to detail canonical pathways containing the differential expressed proteins. Topological network analysis demonstrated that four putative proteins act as central node proteins in COPD. The drugs with the most similar interaction signatures to approved COPD drugs were extracted with the CANDO platform. The drugs identified using CANDO were subsequently analyzed using a knowledge-based technique to determine an optimal two-drug combination that had the most appropriate effect on the central node proteins. Network analysis of the BALF proteome identified critical targets that have critical roles in modulating COPD pathogenesis, for which we identified several drugs that could be repurposed to treat COPD using a multiscale shotgun drug discovery approach.

**Keywords:** proteomics; drug repurposing; translational bioinformatics; interaction signature; bronchoalveolar lavage fluid; chronic obstructive pulmonary disease; label-free quantitation; plasma; serum; biomarker

## 1. Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality and morbidity in the US [1–7]. Additionally, COPD results in millions of hospitalizations in the developing world [1,3,8–10]. The prevalence of cigarette smoking continues to rise in most developing countries worldwide [11–13]. However, only 25–50% of tobacco smokers develop COPD, suggesting only a subset develops an exaggerated inflammatory process that leads to lung destruction [11,12,14]. Bronchoalveolar lavage fluid (BALF) and bronchial samples from ex-smokers reveal active inflammation long after smoking cessation [15,16].

Although structural changes in the airways, parenchyma, and pulmonary vessels are typical in patients with COPD, the lower airways and the alveoli are the initial sites of the inflammatory process [17,18]. The inflammatory process initiated by smoking persists after cessation and is likely exaggerated by autoimmunity and infection [19,20]. Accurate and precise measurement of the molecular mediators in the airways should aid in rigorous analysis of their role in disease.

There has been a keen interest in understanding the genetic determinants of COPD, as the interaction between genes and the environment leads to protein expression, ultimately resulting in either healthy or disease states. However, genomic data alone does not predict protein abundance or activity; proteins are the ultimate participants in integrated biological processes that lead to healthy physiological function or pathology. Proteome-based analysis of bronchoalveolar lavage fluid (BALF) in COPD can identify tissue-specific markers of inflammation that can lead to understanding the mechanisms of COPD progression.

We sought to determine an unbiased proteome-based analysis of BALF in COPD under stable conditions (not in exacerbation status) to identify a broad series of molecules involved in COPD pathogenesis. A label-free proteomics mass spectroscopy method was utilized. The differentially expressed proteins were analyzed using multiple bioinformatics tools to identify critical pathways that were altered in these ex-smoker patients with COPD compared to never-smoker healthy controls and important “hub” proteins implicated in COPD etiology, and identify putative drug candidates that can be repurposed to treat COPD.

The raw proteomic data used in this manuscript were initially detailed in a previously published methodology manuscript using strict criteria (two peptide identification criteria for a protein,  $\geq 1.5$  folds change, and  $p$ -value  $< 0.05$ ) to identify 423 individual proteins with 76 proteins expressed differently between COPD and controls [21]. In this analysis, we adopted a pragmatic approach to the same raw proteomic data (one peptide identification criterion,  $\geq 1.5$  folds change, and  $p$ -value  $< 0.05$ ) that identified 1831 individual proteins and 233 differentially expressed proteins between the two groups. The latter, more practical approach provides important information for biomarker and therapeutic target discovery that may be utilized in future research to discover valuable interventions.

## 2. Results

### 2.1. Study Population Characteristics

Characteristics for subjects included in the BALF study are shown in Table 1, with the only significant differences between the two groups being in tobacco smoke exposure and lung function.

**Table 1.** Clinical parameters of never-smoking healthy subjects and ex-smokers with stable COPD in BALF study.

	Control Subjects n = 10	COPD Subjects (GOLD Stage 2) n = 10	p-Value
Age (years)	63.4 ± 11.7	67.8 ± 8.5	0.15
Sex			0.31
Male	6	7	
Female	4	3	
Race			0.083
Caucasian	8	10	
African-American	2	0	
BMI (kg/m <sup>2</sup> )	28.5 ± 4.2	32 ± 9.7	0.32
Years patient quit smoking	NA	12.9 ± 4.4	
Tobacco smoking, Pack years	NA	56.6 ± 17.2	<0.001
FEV <sub>1</sub> (% predicted)	96.3 ± 14.8	65.9 ± 8.1	<0.001 *
FVC (% predicted)	95.6 ± 13.4	87.6 ± 13.1	0.19
FEV <sub>1</sub> /FVC	77.6 ± 3.8	57.8 ± 8.6	<0.001 *

FEV<sub>1</sub>: forced expiratory volume in 1 s. FVC: forced vital capacity. Years quit: Years subjects quit tobacco smoking. Pack-Years: the average number of packs of cigarettes smoked per week multiplied by the years the subject smoked cigarettes. \*: The significant differences in lung function noted between the control subjects and COPD subjects are expected based on a priori selection of the cohort based on different lung function capacities.

## 2.2. BALF Proteome Characteristics

A total of 1831 unique proteins were identified in the BALF proteome. A total of 233 proteins (>1.5-fold absolute change, *p*-value < 0.05) had a significant differential expression in BALF samples from patients with COPD versus healthy ex-smokers; 138 proteins were decreased in COPD while 95 proteins were increased (Tables S2 and S3).

## 2.3. Manually Curated Pathway Analysis

### 2.3.1. Functional Annotation of Differentially Expressed Proteins and Transcription Factor Interactions

The 233 differentially quantified proteins were characterized by their biological processes, transcription factor interactions, and cellular localization, by employing NIH's DAVID [22,23]. The proteins involved in several biological processes implicated in COPD pathogenesis (total number of proteins, number upregulated, and number downregulated) such as proteolysis [24] (20,4,16), extracellular matrix [25] (13,6,7), cell adhesion [26,27] (11,2,9), cytoskeleton [28] (32,14,18), defense response [29,30] (16, 7,9), cell migration [31] (12,4,8), and oxidation-reduction [32] (11,2,9) were altered in COPD. When examining the lung lining fluid, the largest single group of differentially expressed proteins was associated with the extracellular space (49,30,19) (see Table S4).

Transcription factors (Table S5) associated with the differentially expressed proteins (total number of proteins associated with the transcription factor) included serum response factor-SRF (148), transcription factor 8-AREB6 (166), signal transducer and activator of transcription factor 1-STAT1 (69), zinc finger protein-GFI1 (97), signal transducer and activator of transcription factor 3-STAT3 (101), nuclear factor kappa-light-chain-enhancer of activated B cells-NF-κB (79), CCAAT/enhancer-binding protein β-CEBPB (109), paired box gene 2-PAX2 (113), and activating transcription factor 2-CREBP1 (95).

### 2.3.2. Bioinformatic Pathway Analysis of BALF Proteomic Data

The protein expression datasets were imported into IPA (Ingenuity Systems) and projected onto the relevant biological pathways; processes linked to the differentially expressed proteins were analyzed with the IPA's manually curated knowledge database. Of the 233 differentially expressed proteins, 217 matched the IPA-curated database and were analyzed. Sixteen pathways were noted to have several proteins associated with the differentially expressed BALF dataset (Table S6), including proteins implicated in cellular movement, cellular death and survival, cell morphology, immune cell trafficking, and cell cycle. Figures S1–S4 depict IPA networks of selected pathways with the highest number of differentially expressed proteins.

### 2.4. Computational Drug Prediction

One-hundred-and-thirty out of 233 BALF differentially expressed proteins were identified in the CANDO human protein library. This subset of proteins within the CANDO platform was used to predict 189 putative drug candidates with the most similar protein interaction signatures to the set of known drugs used to treat COPD (Figure 1 and Table S7). Many of the drugs were corticosteroids; however, other putative drugs included tezacaftor [33], a recently developed drug to potentiate sodium channel activity to treat cystic fibrosis; two additional drugs predicted to treat COPD, gemfibrozil [34,35], and pioglitazone [35,36], are drugs currently used to treat hyperlipidemia and diabetes, respectively.

### 2.5. Candidate Key Mediators of COPD Pathology Based on Literature Derived Drug Enrichment

#### 2.5.1. Literature Informed Protein-Protein and Protein-Drug Interaction Network

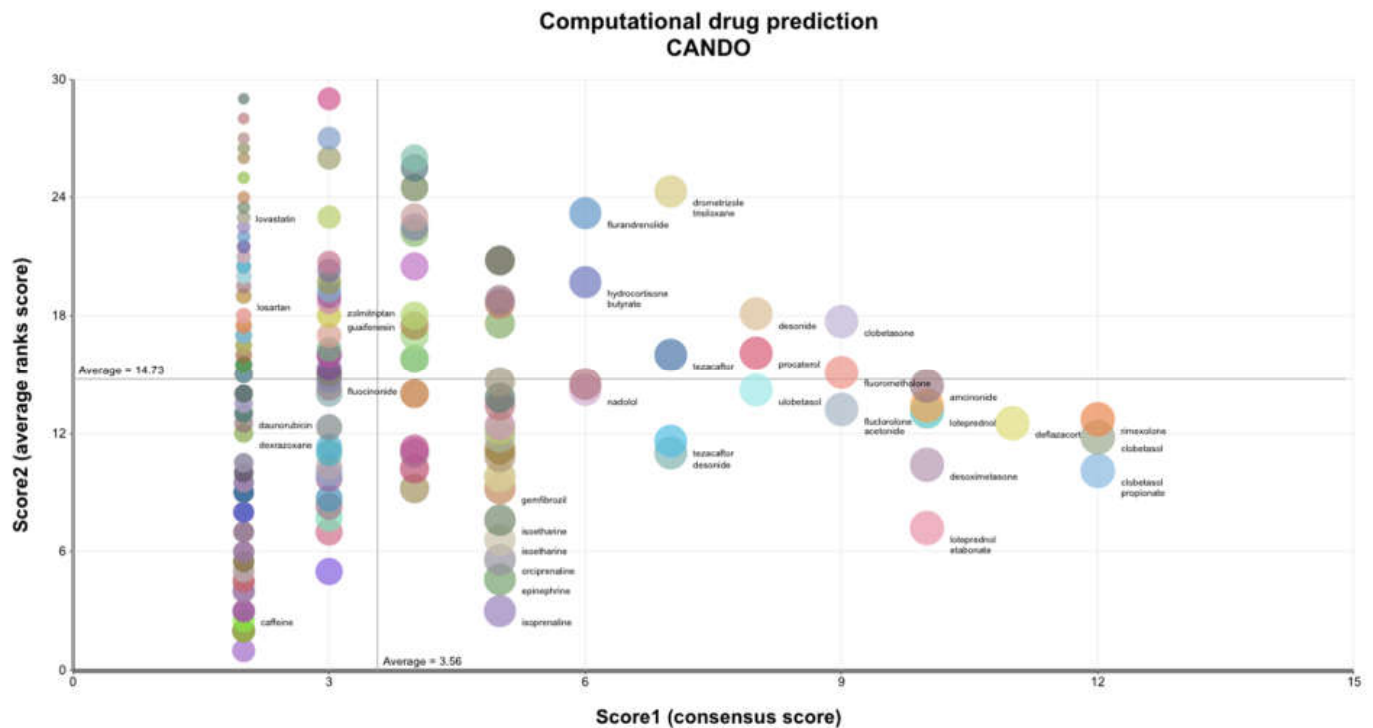
A total of 233 proteins were identified as differentially expressed between COPD patients and healthy controls by mass spectrometry. Of these, 214 were represented in the Elsevier knowledge graph [37], with the remainder comprising specific immunoglobulin chain proteins. A query of the knowledge graph for documented regulatory interactions between these protein entities yielded 206 regulatory edges supported by 807 references (with a median of one reference per edge). One-hundred-and-twelve of the 214 identified proteins could not be connected to the broader network circuit by a documented interaction. The protein entities in this network were then assessed in terms of their importance as mediators of signal transfer based on their betweenness centrality (Figure 2).

#### 2.5.2. Network Topological Analysis

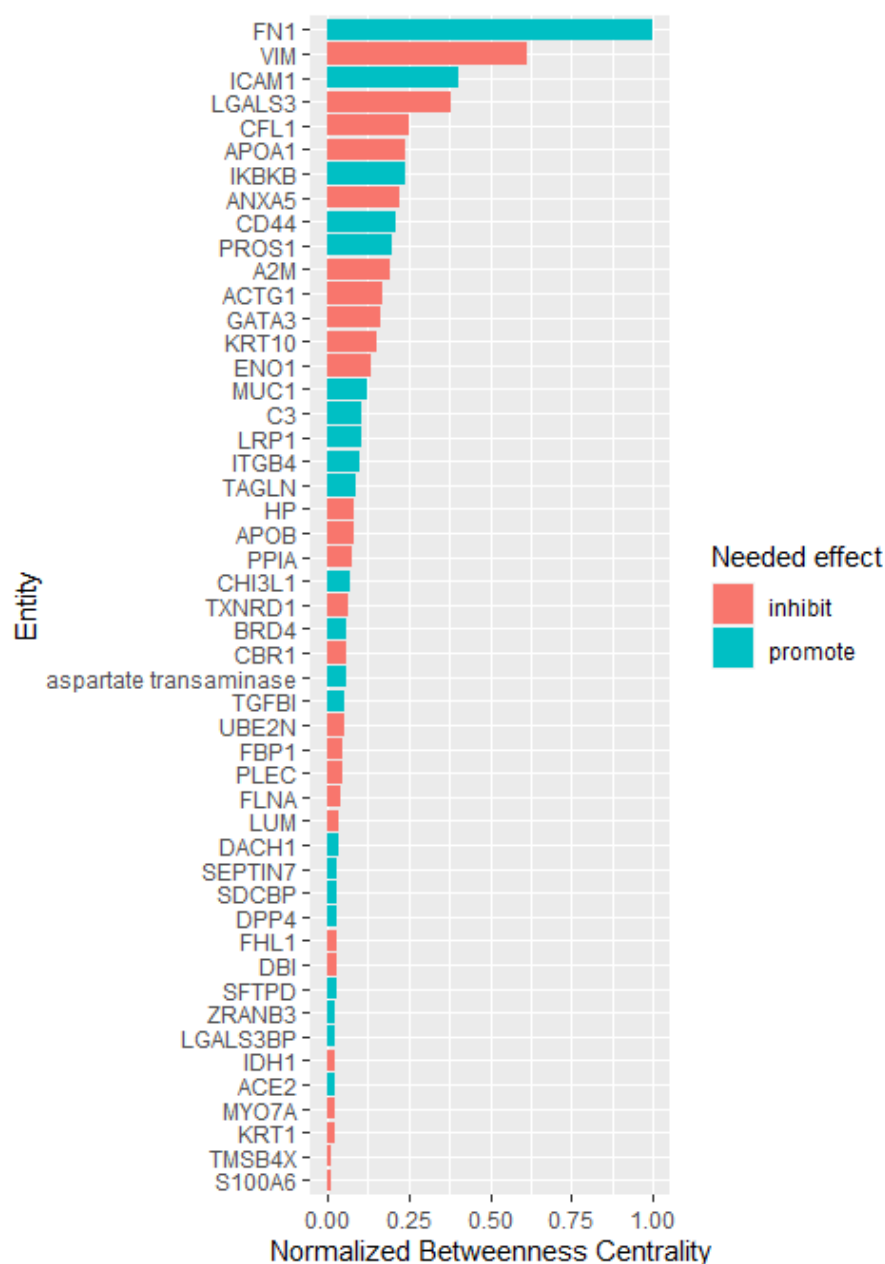
Four nodes representing proteins in the network stood out based on the normalized betweenness centrality values representing a greater than linear baseline increase from the next lower-ranking node: fibronectin, vimentin, intercellular adhesion molecule 1 (ICAM1), and galectin-3. These potentially key signaling mediators had a betweenness centrality of at least 25% of the maximum.

Analysis of the initial data reveals fibronectin and ICAM1 are reduced in COPD patients relative to healthy controls; thus, any candidate therapeutic should target an increase in their activity. The reverse is true for vimentin and galectin-3. We, therefore, sought drugs or combinations of drugs predicted to accomplish the appropriate activation or inhibition of the four most central nodes. Specifically, drugs that will promote central node proteins that were downregulated in the COPD cohort and inhibition of central node proteins that were overabundant in COPD. Therefore, the idealized drug vector constitutes interactions leading to desirable modulation of the central hub protein. CANDO identified 189 distinct drugs (Figure 1, Tables S6 and S7) with relevance for COPD; 39 of these represented in the Elsevier knowledge graph [37] were analyzed for their enrichment of the desired agonist and antagonist effects on the most central entities in the protein regulatory network. Highly enriched drugs or drug pairs are more likely than randomly selected drugs to exert appropriate inhibition or promotion of the most central proteins. Two single drugs (fluocinolone acetonide and dexrazoxane) and 57 two-drug combinations were significantly enriched. Fluocinolone acetonide and dexrazoxane appeared in 54% and 46%

of all significantly enriched two-drug combinations, respectively, far greater than the other drugs appearing in these combinations (Figure 3). Flucinolone acetonide and dexrazoxane are the most enriched two-drug combination leading to an idealized drug vector.

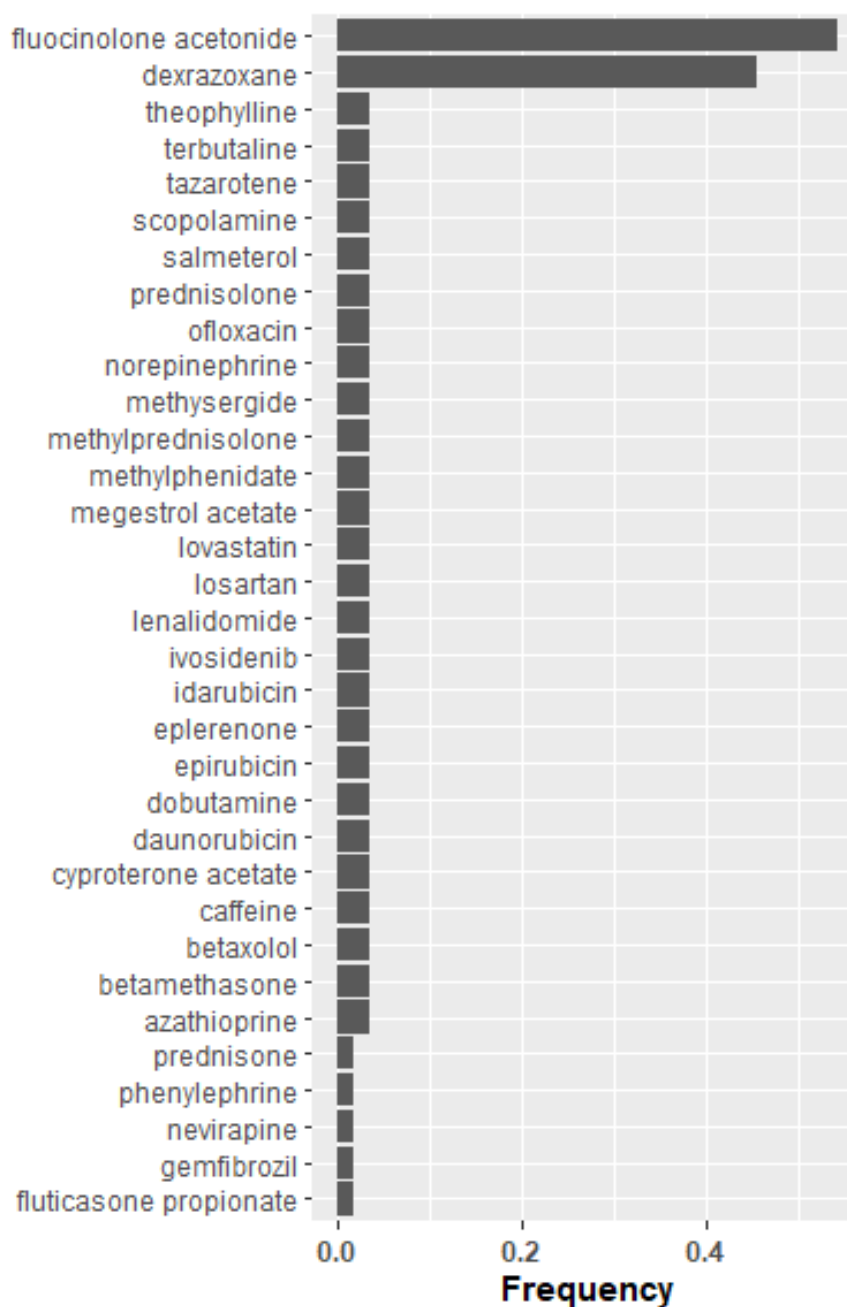


**Figure 1.** Putative drug candidates for treating COPD generated using the CANDO platform. A subset of 130 proteins from the CANDO human protein library was identified from 233 differentially expressed proteins in the BALF. These 130 proteins were utilized to generate BALF-specific interaction signatures for 2450 FDA-approved drugs via our in-house docking protocol BANDOCK (see methods). These drug–proteome interaction signatures were compared to 34 known drugs used to treat COPD to predict 189 of the most similar putative drug candidates. The 189 drugs are represented by colored circles, with the diameter of the circles decreasing with descending overall rank. Drug name labels are depicted for the selection of the 189 drugs shown by the colored circles. The horizontal axis plots the consensus score count or the number of times the particular drug is listed within the top 30 most similar drugs to those known to treat COPD based on interaction signature similarity. The vertical axis plots the average of the cumulative ranks of the consensus scores for the putative drug. The overall rank of a putative drug is determined by initially sorting the drug by the consensus score, as noted above, and then additional sorting by the average rank. Many of the drug candidates were corticosteroids not used to treat COPD; however, other putative drugs included tezacaftor, a drug to potentiate sodium channel activity in the treatment of cystic fibrosis; two additional drugs predicted to treat COPD, gemfibrozil, and pioglitazone, are drugs currently used to treat hyperlipidemia and diabetes, respectively. This analysis indicates that the CANDO platform applied to the BALF proteome can generate putative drug candidates for COPD treatment.



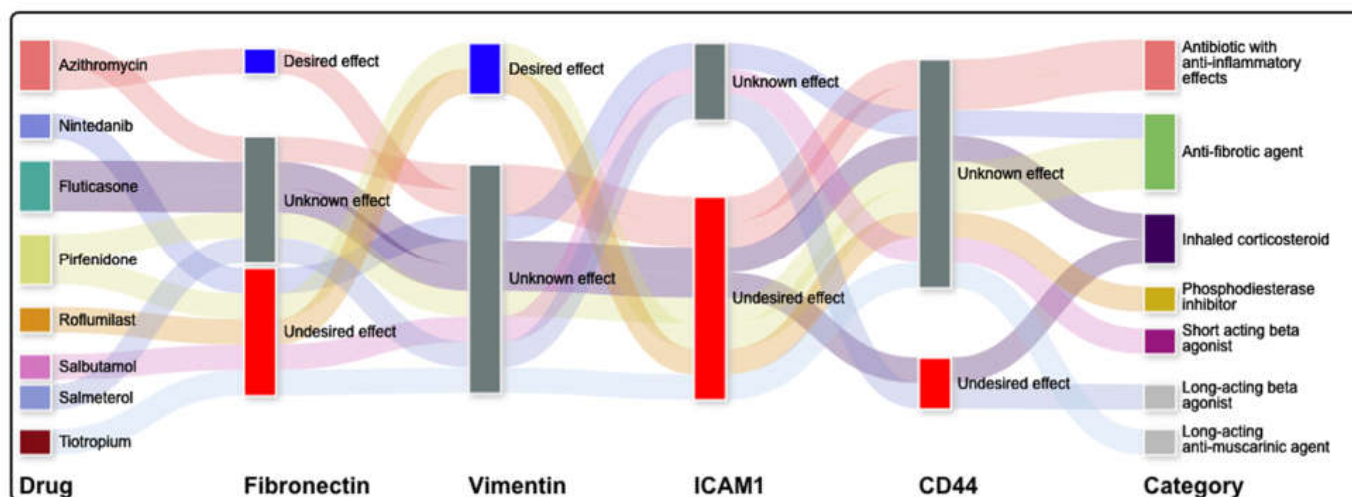
**Figure 2.** BALF network centrality nodes ranked by betweenness centrality. Betweenness centrality quantitatively describes how a node (in this case, a differentially expressed protein in the BALF proteome) mediates the interaction between communities of neighboring nodes in the network. Shown are 44 network entities with betweenness centrality  $> 0.01$ , normalized to the maximum betweenness centrality present in the network. The betweenness centrality scores for all nodes were expressed as fractions of the maximum betweenness centrality present in the network. The (red and blue) colors indicate the needed effect (inhibition/induction) to restore these entities from COPD levels to the normal levels in healthy control subjects. The four nodes with  $\geq 25\%$  of the maximum betweenness centrality (fibronectin) with normalized betweenness centrality values representing a greater than the linear baseline increase from the next lower ranking node are fibronectin, vimentin, intercellular adhesion molecule1 (ICAM1), and galectin-3 (LGALS3). These potential key signaling mediators had a betweenness centrality of at least 25% of the maximum. Topological analysis of the interaction network regulatory interactions documented in the literature suggests that these proteins were central mediators of COPD [38–43]. Colors indicate the needed effect to restore these entities to the normal levels in healthy control subjects.





**Figure 3.** Drug frequency amongst idealized drug combinations predicted to modulate central proteins in COPD. The drugs initially identified by CANDOR, which were predicted to activate or inhibit the four central nodes with the highest maximum betweenness centrality, are listed earlier (Figure 2). Specifically the drugs are predicted to promote central node proteins which were downregulated in the COPD cohort and inhibit central node proteins (identified by network topological graph), which were overabundant in the COPD cohort. The idealized drug vector constitutes interactions leading to desirable modulation of the central hub protein. Representation of individual drug frequency among the 57 significantly enriched two-drug combinations (idealized drug vectors) out of the 39 proteins represented in the Elsevier knowledge graph are listed in descending order. Fluocinolone acetonide and dexrazoxane appeared in 54% and 46% of all significantly enriched two-drug combinations, respectively, which is far greater than other drugs appearing in these combinations. The combination of fluocinolone acetonide and dexrazoxane is the most enriched two-drug combination leading to an idealized drug vector that most likely reverses the protein levels of the four central nodes to levels found in healthy control subjects.

We additionally conducted a targeted query to assess the predicted effects of drugs commonly applied in pulmonary disease treatment on the most central proteins of this regulatory network (Figure 4 and Table S8). While some of these drugs have been documented to have the desired effect on two of the central proteins, fibronectin or vimentin, all have been documented to have the opposite effect on at least one of the most central proteins. Therefore, they were not significantly enriched out of the set of all possible candidate drugs.



**Figure 4.** Commonly used pulmonary drugs and their putative effects on four central node proteins in COPD. A Sankey diagram categorizing putative drugs' likely effects on four central node proteins in COPD is depicted. There are eight drugs on the left of the diagram used to treat different pulmonary diseases, with the corresponding drug classes displayed on the right side of the diagram. The effects of these drugs on four nodes (fibronectin, vimentin, intercellular adhesion molecule1 (ICAM1), and cd44) are detailed in the middle of the diagram, with broad lines connecting the proteins in the right to the putative effect (desired, unknown, and undesired). While some of these have been documented to have the desired effect on fibronectin (promotion) or vimentin (inhibition), all have been reported to have the opposite effect on at least one of the most central proteins. This suggests that using drugs commonly used to treat pulmonary disease, if repurposed for COPD, may have contrary effects on the mediators of the pathways involved in COPD, reinforcing the need to have a more nuanced approach to drug repurposing.

### 3. Discussion

Our investigation of the COPD BALF proteome utilizing novel bioinformatic techniques revealed significant differences in proteins involved in multiple biological processes, including lung-specific mechanisms, protease/anti-protease homeostasis, immunoregulation, and the extracellular matrix. Proteomic profiling of the complex pathways implicated in COPD provides broader physiological exploration not provided by studying one entity at a time. We identified several differentially expressed proteins in COPD versus controls that, based on a review of published literature, have not been previously implicated in COPD etiology. This preliminary analysis illustrates how our BALF proteomic analysis represents a powerful approach to elucidate COPD pathogenesis and identify novel biomarkers.

Employing the bioinformatics tool DAVID and IPA, putative pathway networks were constructed based on the differentially expressed proteins in the BALF proteome that implicated multiple transcription factor pathways and disparate biological processes, such as extracellular space, proteolysis, extracellular matrix, cell adhesion, cytoskeleton, defense response, cell migration, and oxidation reduction.

The CANDO platform identified 189 drug candidates with similar protein interaction signatures based on the BALF proteome compared to known drugs used to treat COPD. However, while most putative drug and protein interactions are likely inhibitors, the



induction or inhibition of a target protein is indeterminable with solely the binding potential between drug and protein pairs.

Topological analysis of the interaction network connecting 233 proteins differentially expressed in COPD through regulatory interactions documented in the literature suggested that ICAM1 [38] and galectin-3 [39] are important central mediators of inflammation. At the same time, both fibronectin [40,41] and vimentin [42] are central mediators of inflammation and fibrogenesis. CD44 [43] is one of the most abundant receptors in mesenchymal stem cells and has been implicated in cell migration in response to lung injury. This body of evidence corroborates the pathway-enrichment analysis results described above and points to fibrosis and innate inflammation as important processes governing the pathogenesis and progression of COPD. A literature knowledge-based query (Elsevier knowledge graph) of drugs with desired drug–target interactions (generated using CANDO) identified putative drugs, such as anti-neoplastic, anti-fibrotic drugs, and inflammation regulators that would restore key central proteins to the levels characteristic of healthy controls. Our results also suggest that currently utilized medications for COPD have disparate effects on the identified central node proteins that are key putative mediators of COPD pathogenesis and progression.

In contrast, the corticosteroid flucinolone acetonide [44,45] and the cardioprotective agent dexrazoxane [46,47] were highly enriched for the desired effects on central network entities, both individually and in combination. Flucinolone acetonide is a stronger potentiator than other corticosteroids of the TGF- $\beta$  pathway [48], which is noted to be dysregulated in COPD [49], and flucinolone acetonide may be more effective than comparable corticosteroids in improved homeostasis in that pathway. Dexrazoxane [46,47] is used to reduce cardiac toxicity associated with anthracycline-based chemotherapy agents by binding to iron and reducing reactive oxygen species; with oxidative stress as a significant factor in COPD pathogenesis [50], antioxidative therapy may be beneficial.

The documented actions of these immunomodulators were predicted here to substantially counteract the observed dysregulation of centrally connected proteins in COPD patients. The relatively high representation of immunomodulators among the candidate agents and the increased centrality of fibrosis-related proteins is consistent with the paradigm of airway remodeling as central to COPD pathology [51]. With additional data, this regulatory circuit could be used as a testbed for computational evaluation of these and other candidate drug effects using network topological methods [52].

### 3.1. Limitations and Strengths

Our approach does have some limitations. The variability in how much BALF is recovered from each aliquot of saline infused in to the lower airway in COPD vs. control subjects is inherent in most BALF proteomic analyses. However, the BALF proteins were normalized to albumin BALF concentrations to account for the variability. The examination of protein levels without accounting for post-translational modifications, such as phosphorylation, may neglect important differences in protein interactions and activity, despite no significant differences in protein levels. Also, the BALF samples were from subjects in the COPD group who were ex-smokers. This exclusion limits the generalizability of our findings, particularly in current smokers, since the acute effects of tobacco smoke were excluded in our study design.

However, we confined our analysis to ex-smokers with moderate COPD to obtain some uniformity of the COPD phenotype and to avoid the acute inflammatory effects of current smoking. Future work on proteomic profiles will inform us of the difference between such profiles in current smokers and different stages of COPD.

### 3.2. Comparison to Previously Published Studies

A sputum proteomics study endeavored to identify COPD severity biomarkers by employing 2D gel electrophoresis and revealed 15 proteins that were significantly differentially expressed between healthy smoker controls and subjects with GOLD stage II;

subsequently, 9 of the 15 candidate proteins were validated with Western Blot. Of the nine candidate proteins validated with Western Blot, seven were statistically significantly different between groups, specifically albumin, alpha-2-HS glycoprotein, transthyretin, PSP94, apolipoprotein A1, lipocalin-1, and PLUNC [53]. Employing quantitative ELISA data normalized for protein content, the investigators identified apolipoprotein A1 and lipocalin-1 as statistically differentially expressed in COPD. Although apolipoprotein A1 and lipocalin-1 were identified in our study of the BALF proteome, the proteins were not significantly differentially expressed, likely due to the differences in expression in the different biocompartments of sputum vs. bronchoalveolar lumen.

A 2D differential gel electrophoresis study and subsequent mass spectroscopy were performed by Ohlmeier et al., which compared healthy smokers, non-smokers, and smokers with GOLD stage II COPD and revealed a different set of 15 proteins that were differentially expressed between the groups [54]. Of these proteins, polymeric immunoglobulin receptor levels in lung tissue and blood between the three groups were correlated with airflow obstruction.

In Lee et al., tumor-free lung tissue harvested from patients with lung cancer resection, when examined via 2D gel electrophoresis/MALDI-TOF-MS, revealed 8 proteins that were upregulated in subjects with COPD compared to non-smokers and 10 significantly differentially expressed proteins between subjects with COPD and smoking subjects without COPD [55]. Two of the identified proteins, matrix metalloprotease 13 (MMP13) and thioredoxin-like 2, were confirmed to be increased in COPD subjects with Western Blot and immunohistochemical staining, with MMP13 localized to the alveolar macrophage and type II pneumocytes and thioredoxin-like 2 found in the bronchial epithelium. Thioredoxin-like 2, which contains thioredoxin, was found in the BALF proteome but not significantly differentially expressed. However, MMP13 was not identified in our BALF study due to differences in study populations and variable biocompartments.

#### 4. Methods

We analyzed the protein quantifications derived from the BALF of subjects with COPD and healthy ex-smoker control subjects via liquid chromatography and mass spectroscopy. We then used pathway analysis tools to identify the relevant cellular pathways associated with differentially expressed proteins quantified from the BALF analysis. We subsequently employed the Computational Analysis of Novel Drug Opportunities (CANDO) platform (<https://github.com/ram-compbio/CANDO> Accessed on 1 December 2021) to identify FDA-approved drugs that could be repurposed to COPD based on their putative interaction with the differentially expressed proteins. Using topological network analysis, we identified putative hub proteins that modulate the cellular pathways associated with COPD. Using the medical literature to predict the repurposed drugs effects on the most important hub protein, we created a refined list of drugs predicted to modulate the cellular pathway to impede COPD pathogenesis and generate proteomic interaction signatures for the compounds.

##### 4.1. Recruitment of Subjects

BALF was obtained in a NHLBI-funded study of innate lung defense in COPD [56]. All procedures received approval from the Institutional Review Board (IRB), Veterans Affairs Western New York Healthcare System (WNY-VA), and strictly adhered to institutional guidelines.

##### 4.2. Ethics Statement

This study is a sub-study of a larger group of patients with COPD and healthy controls to understand biological determinants of exacerbation frequency and was approved by the Institutional Review Boards of the Veterans Affairs Western New York Healthcare System and the University at Buffalo. The participants gave written consent to the study via an IRB-approved consent form.

#### 4.3. Inclusion/Exclusion Criteria

The inclusion criteria and procedures for this study have been described previously and are provided in the supplementary material [55]. After informed consent, 116 volunteers were divided into three groups: (1) healthy non-smokers, (2) ex-smokers with COPD, and (3) active smokers with COPD who underwent bronchoscopy and bronchoalveolar lavage. The methodology for bronchoscopy, lavage, and sample processing is included in the supplementary material.

We selected BALF obtained from 10 ex-smokers with moderate COPD and 10 healthy non-smoking controls for proteomic analysis, respectively. To minimize variability due to the effects of acute smoking and disease severity, we confined this analysis to ex-smokers and moderate stage 2 disease per the Global Obstructive Lung Disease (GOLD) [56] criteria of the forced expiratory volume in 1 s ( $FEV_1$ ); 50–80% predicted with the ratio of  $FEV_1$  over forced vital capacity ( $FEV_1/FVC$ )  $< 0.70$ . All ex-smokers had ceased smoking for at least 1 year.

#### 4.4. Bronchoscopy and BALF Sample Preparation

The research bronchoscopy and BALF sample preparation were performed as described previously [57].

#### 4.5. Protein Identification/Quantification

To investigate the soluble molecules in the epithelial lining fluid that may participate in COPD pathogenesis, unbiased proteomic analysis of BALF commenced without protein depletion or fractionation. Details of the methodology were published [21] and are also provided in the supplementary material.

#### 4.6. Long Gradient Nano-RPLC/Mass Spectrometry

The complete separation of the complex peptide mixture utilized a nano-LC/nanospray setup [21]; the ion-current long gradient approach with mass spectrometry and subsequent protein identification was performed as described in Tu et al. [21,58,59]. All proteins identified with one or more peptide hits, fold change of  $\geq 1.5$ , and  $p$ -value  $< 0.05$  are included as part of the differentially expressed BALF proteome.

#### 4.7. Bioinformatics Analyses

##### 4.7.1. Manually Curated Pathway Analysis

Gene ontology, transcription factors, and expression locations were determined by uploading the protein expression dataset onto a web-based tool, the NIH's Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.7 (<http://david.abcc.ncifcrf.gov/>; Accessed on 21 March 2021) [22,23]. Biological networks were generated with Ingenuity Pathway Analysis (IPA, Ingenuity Systems), a web-based relational database and network generator. Proteins overrepresented in the uploaded datasets in biological networks, canonical pathways, and biological processes were identified.

##### 4.7.2. Literature Informed Protein-Protein and Protein-Drug Interaction Network

In addition to annotating differentially expressed proteins with the manually curated pathways cataloged in IPA, a network of protein–protein interactions was created using known regulatory relationships extracted from published scientific literature using the MedScan text-mining engine [60] as well as protein–drug interactions cataloged in the Reaxsys medicinal chemistry database (Elsevier, Amsterdam). These are embedded in the broader Elsevier Knowledge Graph database [37] and were accessed via the Pathway Studio interface (Elsevier, Amsterdam, The Netherlands) [61].

##### 4.7.3. Shotgun Multiscale Drug Discovery Platform

We used the CANDO platform [62–69] to predict drugs that can be repurposed to treat stable COPD. In CANDO, a compound/drug is potentially repurposable for an indication

when it is found to have similar binding interactions with a specific proteome or library of proteins as a drug with known approval for the indication of interest.

This study calculated the interaction scores between 2450 United States Federal Drug Administration (FDA)-approved drugs from the CANDO version 2.3 compound library and a curated human library of 8385 proteins, including 5316 solved X-ray crystallography structures and 3069 computed protein structures modeled by I-TASSER [70,71]. The interaction scores were calculated using the bioanalytic docking (BANDOCK) protocol in the CANDO, which utilizes predicted binding site information and chemical similarity to determine an interaction score that is a surrogate for the likelihood of interaction between a compound and protein [62]. Binding sites were predicted for all human proteins using COACH [72], which uses the consensus of three complementary methods utilizing structure and sequence information to find similarities to solved structures in the Protein Data Bank (PDB) [73,74]. For each binding site predicted by COACH, a confidence score (Pscore) and an associated co-crystallized ligand are output. The ligand is then compared to the query compound/drug using chemical fingerprinting methods, which enumerates the presence or absence of molecular substructures on the compound/drug. The Sorensen–Dice coefficient [75,76] was used for the protein–ligand and compound/drug fingerprints (Cscore). The BANDOCK interaction score outputted for each compound–protein pair is the product of the Pscore and the Cscore.

For this analysis, we focused on the differentially expressed proteins in the BALF proteome (as described) and drugs used to treat COPD (“MESH:D029424”) (Table S1). We selected proteins in the CANDO human protein library represented in the differentially expressed BALF proteome. We then used the CANDO platform to predict the top drug candidates that could be repurposed to treat COPD based on the compound–proteome interaction signature similarity to drugs currently approved/used to treat stable COPD. The protocol iterates through 34 known drugs used to treat stable COPD, counting the number of times drugs not associated with COPD show up in the top 30 most similar compounds to the known treatments, then outputs the consensus predictions ranked by the number of times each compound appeared across all top 30 lists. The similarity between a given drug and all other drugs in the library is determined by comparing their proteomic interaction signatures using the cosine similarity metric. Compounds with greater similarity scores rank stronger than those with low similarity. Thereby, drugs that were most similar (in terms of interaction signatures) to multiple drugs used to treat COPD will be ranked highest.

#### 4.7.4. Network Topological Analysis

Although not a complete descriptor, the topological location and aspects of the connectivity linking a node to a broader biological network can inform the node’s function in mediating network behavior. Among the measures of a node’s importance or centrality, *betweenness* centrality has been used to describe how a node might serve as an important mediator of information flow in a regulatory network. In this work,  $C_b(n)$  for each node  $n$  of a network was calculated using the Brandes algorithm [77]. The betweenness centrality of a node  $n$  reflects the amount of control that this node exerts over the interaction between communities of neighboring nodes in the network [78] and can be computed as follows:

$$C_b(n) = \sum_{s \neq t \neq n} (\sigma_{s,t}(n)/\sigma_{s,t}) \quad (1)$$

where  $s$  and  $t$  are the source and target nodes in the network different from  $n$ ,  $\sigma_{s,t}$  denotes the number of shortest paths from all  $s$  to all  $t$ , and  $\sigma_{s,t}(n)$  is the number of shortest paths from  $s$  to  $t$  that must pass through node  $n$ . Here, unweighted betweenness centralities were calculated for each node in the literature-informed protein–protein network. The betweenness centrality scores for all nodes were expressed as fractions of the maximum betweenness centrality present in the network. All calculations were conducted in R version 4.0.2 (Vienna, Austria) [79].

#### 4.7.5. Literature Based Drug Enrichment Analysis

Using putative drugs ranked by CANDO and further analyzed via the Elsevier knowledge graph [37], a drug enrichment analysis was performed to predict which drugs can most closely mimic an idealized intervention against the hub proteins identified in the network topological analysis. Drugs are represented as vectors with a length equal to the empirically derived number of protein entities in the network model. Each index value is listed as 0 if there is no interaction between the drug and the corresponding model entity, a 1 if the drug promotes that entity, or a -1 if the drug inhibits that entity. Next, the cosine similarity,  $S_c$ , between each drug vector and the idealized intervention vector is calculated [80]. Cosine similarity is calculated as:

$$S_c(D, M) = \frac{D \cdot M}{\|D\| \|M\|}$$

where  $D$  is the drug vector and  $M$  is the idealized intervention. Higher  $S_c$  indicates a closer match between the drug vector and the idealized vector. An  $S_c$  of 1 means the two vectors are identical, and  $-1$  indicates that the two are exactly opposed. For multidrug combinations, the net-effect of the individual drug vectors is calculated as:

$$\text{sgn} \left( \sum_{i=1}^n \vec{d}_i D_i \right)$$

where  $n$  is the total number of drugs in the combination,  $D_i$  is the vector corresponding to the  $i$ th drug, and  $\text{sgn}$  is the sign function. The cosine similarity of the net-effect vector and idealized vector is then calculated.

The statistical significance of these enrichment scores is determined empirically from an estimated null distribution of cosine similarities. This null distribution uses a set of model-relevant background drugs for which each interacts with at least one entity in the network. All CANDO drugs of interest were included in the background. Empirical  $p$ -values are estimated as

$$\hat{p} = \frac{(r + 1)}{(n + 1)}$$

where  $r$  is the number of null  $S_c$  values greater than the observed  $S_c$ , and  $n$  is the total number of null  $S_c$  values.

#### 4.8. Statistical Analysis

Statistical analysis was performed with SPSS/19. Demographic values were depicted as mean  $\pm$  SEM.

### 5. Conclusions

In summary, our work provides a valuable pipeline for identifying many proteins associated with COPD pathogenesis that illustrate the complexity of the development of this disease and identify putative therapeutic treatment options using cutting-edge bioinformatics approaches. Identifying differentially expressed proteins will form the basis for future mechanistic studies of critical pathways and novel treatment discovery. The validation of our proposed therapeutic approach in animal models and human pilot studies are important next steps.



**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ph15050566/s1>, Figure S1: IPA Network 1: Cellular Movement, Inflammatory Response, Cardiovascular System Development and Function. Figure S2: IPA network 2: cell death and survival, drug metabolism, small molecule biochemistry. Figure S3: IPA network 9: cellular movement, hematological system development and function, immune cell trafficking. Figure S4: IPA network 11: cell cycle, visual system development and function, hair and skin development and function. Figure S5: IPA network putative upstream regulators. Table S1: Drugs commonly prescribed for stable COPD. Curated drugs currently indicated for COPD ("MESH:D029424") from the Comparative Toxicogenomics Database2 and manually curated to include drugs used to treat stable COPD in the United States. Table S2: Unique proteins upregulated in BALF (n = 95), Differentially expressed proteins with at least 1.5x fold change increase in the BALF proteome in COPD versus control cohort samples. Table S3: Unique proteins downregulated in BALF (n = 138). Differentially expressed proteins with at least 1.5x fold change decrease in the BALF proteome in COPD versus control cohort samples. Table S4: See David\_FuncAnnotClustering\_BALF.csv for DAVID functional annotation clustering file. Table S5: Transcription factors associated with binding sites on genes from differentially expressed proteins in BALF. Transcription factors that have association with binding sites on genes from differentially expressed proteins in BALF as noted by DAVID [ref]. Table S6: Top functional networks of differentially expressed molecules in the BALF proteome. The top biological functions associated with molecular pathways imputed with IPA that are significantly associated differentially expressed molecules measured in the BALF proteome. Red represents upregulated, and green represents downregulated proteins. The networks are collections of interconnected molecules assembled by a network algorithm. Each connection represents known relationships between the molecules, found in the Ingenuity Knowledge Base. The score is the degree of relevance of network eligible molecules to the BALF dataset. The score takes into account the number of network eligible molecules in the network and its size, as well as the total number of network eligible molecules analyzed and the total number of molecules in the Ingenuity Knowledge Base that could potentially be included in networks. The network score is based on the hypergeometric distribution and is calculated with the right-tailed Fisher's Exact Test:  $\text{Score} = -\log(\text{Fisher's Exact test result})$ . Focus Molecules are the number of proteins identified in the BALF proteome that is found in the network. Table S7: Computational drug prediction CANDO Table S8: Predicted interactions of drugs treating respiratory diseases and central node proteins.

**Author Contributions:** Conceptualization, M.J.M. and R.S.; methodology, M.J.M., C.T., J.Q., M.C.M., W.M., Z.F., S.R., G.B., S.S. and R.S.; software, M.J.M., M.C.M., W.M., Z.F., S.R., G.B. and R.S.; formal analysis, M.J.M., C.T., J.Q., M.C.M., S.R., G.B., S.S. and R.S.; data curation, M.J.M., C.T. and J.Q.; writing—original draft preparation, M.J.M., C.T., J.Q., M.C.M., W.M., Z.F., S.R., G.B., S.S. and R.S.; writing—review and editing, M.J.M., C.T., J.Q., M.C.M., W.M., Z.F., S.R., G.B., S.S. and R.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** Mammen was supported by a research grant from the University at Buffalo. A National Heart, Lung, and Blood Institute grant to Sethi supported the collection of bronchoalveolar lavage specimens. This work is also supported, in part, by National Institute Health (NIH) awards U54HD071594 (JQ) and by American Heart Association (AHA) award 12SDG9450036 (JQ). Rochester Regional Health also supported this work in conjunction with the US Department of Defense Congressionally Directed Medical Research Programs (CDMRP) under the Peer Reviewed Medical Research Program (PRMRP) award W81XWH1910804 (Broderick—PI; Sethi—Partnering PI, JQ-CoI), as well as by Elsevier AB (Amsterdam). The development of the CANDO platform was supported in part by a NIH Director's Pioneer Award (DP1OD006779), NIH Clinical and Translational Sciences Award (UL1TR001412), NIH Buffalo Research Innovation in Genomic and Healthcare Technology Education Award (T15LM012495), NIH NCATS ASPIRE Design Challenge Award, NIH NCATS ASPIRE Reduction-to-Practice Award, and startup funds from the Department of Biomedical Informatics at the University at Buffalo.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board Institutional Review Boards of the Veterans Affairs Western New York Healthcare System and the University at Buffalo under protocol ID "Innate Immunity in COPD" on February 2007.

**Informed Consent Statement:** Not applicable.



**Data Availability Statement:** The source code and data used to produce the CANDO results and analyses presented in this manuscript are available from Github repository: <https://github.com/ram-compbio/CANDO>.

**Acknowledgments:** The authors would like to thank Catherine Wrona for her technical support.

**Conflicts of Interest:** The authors declare no conflict of interest. The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Defense. To obtain the raw proteomic data on a DVD media, please contact Jun Qu, [junqu@buffalo.edu](mailto:junqu@buffalo.edu).

## Abbreviations

BALF	bronchoalveolar lavage fluid
BANDOCK	bioanalytical docking
CANDO	computational analysis of novel drug opportunities
COPD	chronic obstructive pulmonary disease
DAVID	database for annotation visualization and integrated discovery
GO	gene ontology
IPA	ingenuity pathway analysis
LTQ Orbitrap	linear ion trap combined with an orbitrap analyzer mass spectrometer

## References

- Brown, D.W.; Croft, J.B.; Greenlund, K.J.; Giles, W.H. Deaths From Chronic Obstructive Pulmonary Disease—United States, 2000–2005. *J. Am. Med. Assoc.* **2009**, *301*, 1331.
- Croft, J.B.; Wheaton, A.G.; Liu, Y.; Xu, F.; Lu, H.; Matthews, K.; Cunningham, T.J.; Wang, Y.; Holt, J.B. Urban-Rural County and State Differences in Chronic Obstructive Pulmonary Disease—United States, 2015. *Morb. Mortal. Wkly. Rep.* **2018**, *67*, 205–211. [[CrossRef](#)] [[PubMed](#)]
- GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir. Med.* **2017**, *5*, 691. [[CrossRef](#)]
- US Preventive Services Task Force (USPSTF). Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force recommendation statement. *JAMA* **2016**, *315*, 1372–1377. [[CrossRef](#)]
- From the Global Strategy for the Diagnosis, Management and Prevention of COPD. Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2022. Available online: <http://www.goldcopd.org> (accessed on 1 April 2022).
- Gartman, E.J.; Mulpuru, S.S.; Mammen, M.J.; Alexander, P.E.; Nici, L.; Aaron, S.D.; Ruminjo, J.K.; Thomson, C.C. Summary for Clinicians: Clinical Practice Guideline on Pharmacologic Management of Chronic Obstructive Pulmonary Disease. *Ann. Am. Thorac. Soc.* **2021**, *18*, 11–16. [[CrossRef](#)]
- Nici, L.; Mammen, M.J.; Charbek, E.; Alexander, P.E.; Au, D.H.; Boyd, C.M.; Criner, G.J.; Donaldson, G.C.; Dreher, M.; Fan, V.S.; et al. Pharmacologic Management of Chronic Obstructive Pulmonary Disease. An Official American Thoracic Society Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* **2020**, *201*, e56–e69. [[CrossRef](#)]
- Geitona, M.; Hatzikou, M.; Steiropoulos, P.; Alexopoulos, E.C.; Bouros, D. The cost of COPD exacerbations: A university hospital—Based study in Greece. *Respir. Med.* **2011**, *105*, 402–409. [[CrossRef](#)]
- Hutchinson, A.F.; Brand, C.; Irving, L.; Roberts, C.; Thompson, P.; Campbell, D. Acute care costs of patients admitted for management of chronic obstructive pulmonary disease exacerbations: Contribution of disease severity, infection and chronic heart failure. *Intern. Med. J.* **2010**, *40*, 364–371. [[CrossRef](#)]
- Ciapponi, A.; Alison, L.; Agustina, M.; Demián, G.; Silvana, C.; Edgardo, S. The Epidemiology and Burden of COPD in Latin America and the Caribbean: Systematic Review and Meta-Analysis. *COPD J. Chronic Obstr. Pulm. Dis.* **2013**, *11*, 339–350. [[CrossRef](#)]
- Rennard, S.I.; Vestbo, J. COPD: The dangerous underestimate of 15%. *Lancet* **2006**, *367*, 1216–1219. [[CrossRef](#)]
- Rennard, S.I.; Vestbo, J. Natural Histories of Chronic Obstructive Pulmonary Disease. *Proc. Am. Thorac. Soc.* **2008**, *5*, 878–883. [[CrossRef](#)] [[PubMed](#)]
- Schubert, C. Anti-tobacco efforts going up in smoke. *Nat. Med.* **2006**, *12*, 866. [[CrossRef](#)]
- Mammen, M.; Sethi, S. COPD and the microbiome. *Respirology* **2016**, *21*, 590–599. [[CrossRef](#)] [[PubMed](#)]
- Rutgers, S.R.; Postma, D.S.; Ten Hacken, N.H.; Kauffman, H.F.; van der Mark, T.W.; Koëter, G.H.; Timens, W. Ongoing airway inflammation in patients with COPD who do not currently smoke. *Thorax* **2000**, *55*, 12–18. [[CrossRef](#)]
- Kunz, L.I.Z.; Lapperre, T.S.; Snoeck-Stroband, J.B.; Budulac, S.E.; Timens, W.; Van Wijngaarden, S.; Schrumppf, J.A.; Rabe, K.F.; Postma, D.S.; Sterk, P.J.; et al. Smoking status and anti-inflammatory macrophages in bronchoalveolar lavage and induced sputum in COPD. *Respir. Res.* **2011**, *12*, 34. [[CrossRef](#)]

17. Hogg, J.C. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* **2004**, *364*, 709–721. [[CrossRef](#)]
18. Hogg, J.C.; Chu, F.; Utokaparch, S.; Woods, R.; Elliott, W.M.; Buzatu, L.; Cherniack, R.M.; Rogers, R.M.; Sciruba, F.; Coxson, H.O.; et al. The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* **2004**, *350*, 2645–2653. [[CrossRef](#)]
19. Cosio, M.G.; Saetta, M.; Agusti, A. Immunologic Aspects of Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* **2009**, *360*, 2445–2454. [[CrossRef](#)]
20. Sethi, S. Infection as a comorbidity of COPD. *Eur. Respir. J.* **2010**, *35*, 1209–1215. [[CrossRef](#)]
21. Tu, C.; Mammen, M.J.; Li, J.; Shen, X.; Jiang, X.; Hu, Q.; Wang, J.; Sethi, S.; Qu, J. Large-Scale, Ion-Current-Based Proteomics Investigation of Bronchoalveolar Lavage Fluid in Chronic Obstructive Pulmonary Disease Patients. *J. Proteome Res.* **2013**, *13*, 627–639. [[CrossRef](#)]
22. Da Huang, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
23. Yu, Y.-K.; Da Huang, W.; Bt, S.; Ra, L. Faculty Opinions recommendation of Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13. [[CrossRef](#)]
24. Shapiro, S.D. Proteolysis in the lung. *Eur. Respir. J. Suppl.* **2003**, *44* (Suppl. 44), 30s–32s. [[CrossRef](#)] [[PubMed](#)]
25. Annoni, R.; Lanças, T.; Yukimatsu Tanigawa, R.; de Medeiros Matsushita, M.D.; de Morais Fervezlian, S.D.; Bruno, A.; da Fernando Ferraz Silva, L.; Roughley, P.J.; Battaglia, S.; Dolhnikoff, M.; et al. Extracellular matrix composition in COPD. *Eur. Respir. J.* **2012**, *40*, 1362–1373. [[CrossRef](#)]
26. Riise, G.; Larsson, S.; Löfdahl, C.-G.; Andersson, B. Circulating cell adhesion molecules in bronchial lavage and serum in COPD patients with chronic bronchitis. *Eur. Respir. J.* **1994**, *7*, 1673–1677. [[CrossRef](#)] [[PubMed](#)]
27. Ri, A.; Hagiya, M.; Inoue, T.; Yoneshige, A.; Kimura, R.; Murakami, Y.; Ito, A. Progression of Pulmonary Emphysema and Continued Increase in Ectodomain Shedding of Cell Adhesion Molecule 1 After Cessation of Cigarette Smoke Exposure in Mice. *Front. Cell Dev. Biol.* **2018**, *6*, 52. [[CrossRef](#)]
28. Yang, M.; Kohler, M.; Heyder, T.; Forsslund, H.; Garberg, H.K.; Karimi, R.; Grunewald, J.; Berven, F.S.; Nyren, S.; Sköld, C.M.; et al. Proteomic profiling of lung immune cells reveals dysregulation of phagocytotic pathways in female-dominated molecular COPD phenotype. *Respir. Res.* **2018**, *19*, 1–11. [[CrossRef](#)]
29. Sethi, S. Bacterial Infection and the Pathogenesis of COPD. *Chest* **2000**, *117*, 286S–291S. [[CrossRef](#)]
30. Zuo, L.; He, F.; Sergakis, G.G.; Koozehchian, M.S.; Stimpfl, J.N.; Rong, Y.; Diaz, P.T.; Best, T.M. Interrelated role of cigarette smoking, oxidative stress, and immune response in COPD and corresponding treatments. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **2014**, *307*, L205–L218. [[CrossRef](#)]
31. Stockley, R.A. Neutrophils and the Pathogenesis of COPD. *Chest* **2002**, *121*, 151S–155S. [[CrossRef](#)]
32. Davies, J.C.; Moskowitz, S.M.; Brown, C.; Horsley, A.; Mall, M.A.; McKone, E.F.; Plant, B.J.; Prais, D.; Ramsey, B.W.; Taylor-Cousar, J.L.; et al. VX-659–Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis and One or Two Phe508del Alleles. *N. Engl. J. Med.* **2018**, *379*, 1599–1611. [[CrossRef](#)] [[PubMed](#)]
33. Frick, M.H.; Elo, O.; Haapa, K.; Heinonen, O.P.; Heinsalmi, P.; Helo, P.; Huttunen, J.K.; Kaitaniemi, P.; Koskinen, P.; Manninen, V.; et al. Helsinki Heart Study: Primary-Prevention Trial with Gemfibrozil in Middle-Aged Men with Dyslipidemia. *N. Engl. J. Med.* **1987**, *317*, 1237–1245. [[CrossRef](#)] [[PubMed](#)]
34. Ferraro, R.A.; Leucker, T.; Martin, S.S.; Banach, M.; Jones, S.R.; Toth, P.P. Contemporary Management of Dyslipidemia. *Drugs* **2022**, *82*, 559–576. [[CrossRef](#)] [[PubMed](#)]
35. Eckland, D.; Danhof, M. Clinical pharmacokinetics of pioglitazone. *Exp. Clin. Endocrinol. Diabet.* **2000**, *108* (Suppl. 2), 234–242. [[CrossRef](#)]
36. Kamdar, M.R.; Stanley, C.E.; Carroll, M.; Wogulis, L.; Dowling, W.; Deus, H.F.; Samarasinghe, M. Text Snippets to Corroborate Medical Relations: An Unsupervised Approach using a Knowledge Graph and Embeddings. *AMIA Summits Transl. Sci. Proc.* **2020**, *2020*, 288–297.
37. Walter, R.E.; Wilk, J.B.; Larson, M.; Vasan, R.S.; Keaney, J.; Lipinska, I.; O’Connor, G.; Benjamin, E. Systemic Inflammation and COPD: The Framingham Heart Study. *Chest* **2008**, *133*, 19–25. [[CrossRef](#)]
38. Mammen, M.J.; Sands, M.F.; Abou-Jaoude, E.; Aalinkeel, R.; Reynolds, J.L.; Parikh, N.U.; Sharma, U.; Schwartz, S.A.; Mahajan, S.D. Role of Galectin-3 in the pathophysiology underlying allergic lung inflammation in a tissue inhibitor of metalloproteinases 1 knockout model of murine asthma. *Immunology* **2017**, *153*, 387–396. [[CrossRef](#)]
39. Muro, A.F.; Moretti, F.A.; Moore, B.B.; Yan, M.; Atrasz, R.G.; Wilke, C.A.; Flaherty, K.R.; Martinez, F.J.; Tsui, J.L.; Sheppard, D.; et al. An Essential Role for Fibronectin Extra Type III Domain A in Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **2008**, *177*, 638–645. [[CrossRef](#)]
40. To, W.S.; Midwood, K.S. Plasma and cellular fibronectin: Distinct and independent functions during tissue repair. *Fibrogen. Tissue Repair* **2011**, *4*, 21. [[CrossRef](#)]
41. Li, F.J.; Surolia, R.; Li, H.; Wang, Z.; Kulkarni, T.; Liu, G.; De Andrade, J.A.; Kass, D.J.; Thannickal, V.J.; Duncan, S.R.; et al. Autoimmunity to Vimentin Is Associated with Outcomes of Patients with Idiopathic Pulmonary Fibrosis. *J. Immunol.* **2017**, *199*, 1596–1605. [[CrossRef](#)]
42. Ouhitit, A.; Thouta, R.; Zayed, H.; Gaur, R.L.; Fernando, A.; Rahman, M.; Welsh, D.A. CD44 mediates stem cell mobilization to damaged lung via its novel transcriptional targets, Cortactin and Survivin. *Int. J. Med. Sci.* **2020**, *17*, 103–111. [[CrossRef](#)] [[PubMed](#)]

43. Mills, J.S.; Bowers, A.; Djerassi, C.; Ringold, H.J. Steroids. CXXXVII.1 Synthesis of a New Class of Potent Cortical Hormones. 6 $\alpha$ ,9 $\alpha$ -Difluoro-16 $\alpha$ -hydroxyprednisolone and its Acetonide. *J. Am. Chem. Soc.* **1960**, *82*, 3399–3404. [[CrossRef](#)]
44. Campochiaro, P.A.; Nguyen, Q.D.; Hafiz, G.; Bloom, S.; Brown, D.M.; Busquets, M.; Ciulla, T.; Feiner, L.; Sabates, N.; Billman, K.; et al. Aqueous Levels of Fluocinolone Acetonide after Administration of Fluocinolone Acetonide Inserts or Fluocinolone Acetonide Implants. *Ophthalmology* **2013**, *120*, 583–587. [[CrossRef](#)] [[PubMed](#)]
45. Wiseman, L.R.; Spencer, C.M. Dexrazoxane. A review of its use as a cardioprotective agent in patients receiving anthracycline-based chemotherapy. *Drugs* **1998**, *56*, 385–403. [[CrossRef](#)] [[PubMed](#)]
46. Cvetković, R.S.; Scott, L.J. Dexrazoxane. *Drugs* **2005**, *65*, 1005–1024. [[CrossRef](#)] [[PubMed](#)]
47. Hara, E.S.; Ono, M.; Pham, H.T.; Sonoyama, W.; Kubota, S.; Takigawa, M.; Matsumoto, T.; Young, M.F.; Olsen, B.R.; Kuboki, T. Fluocinolone acetonide is a potent synergistic factor of TGF- $\beta$ 3-associated chondrogenesis of bone marrow-derived mesenchymal stem cells for articular surface regeneration. *J. Bone Miner. Res.* **2015**, *30*, 1585–1596. [[CrossRef](#)]
48. Di Stefano, A.; Sangiorgi, C.; Gnemmi, I.; Casolari, P.; Brun, P.; Ricciardolo, F.L.; Contoli, M.; Papi, A.; Maniscalco, P.; Ruggeri, P.; et al. TGF-beta Signaling Pathways in Different Compartments of the Lower Airways of Patients With Stable COPD. *Chest* **2018**, *153*, 851–862. [[CrossRef](#)]
49. Kirkham, P.A.; Barnes, P.J. Oxidative Stress in COPD. *Chest* **2013**, *144*, 266–273. [[CrossRef](#)]
50. Moisieieva, N.V.; Burya, L.V.; A Kapustianskaya, A.; A Kolenko, I.; A Romyantseva, M.; Shumeiko, O.H. Comprehensive patterns of comorbidity: Copd and depression. *Asp. Treatment. Wiadomości. Lek.* **2018**, *71*, 588–591.
51. Morris, M.C.; Richman, S.; Lyman, C.A.; Qu, J.; Mammen, M.J.; Sethi, S.; Broderick, G. Hacking the Immune Response to Infection in Chronic Obstructive Pulmonary Disease. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)2020, Cincinnati, OH, USA, 26–28 October 2020.
52. Nicholas, B.L.; Skipp, P.; Barton, S.; Singh, D.; Bagmane, D.; Mould, R.; Angco, G.; Ward, J.; Guha-Niyogi, B.; Wilson, S.; et al. Identification of Lipocalin and Apolipoprotein A1 as Biomarkers of Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* **2010**, *181*, 1049–1060. [[CrossRef](#)]
53. Ohlmeier, S.; Mazur, W.; Linja-Aho, A.; Louhelainen, N.; Rönty, M.; Toljamo, T.; Bergmann, U.; Kinnula, V.L. Sputum Proteomics Identifies Elevated PIGR levels in Smokers and Mild-to-Moderate COPD. *J. Proteome Res.* **2011**, *11*, 599–608. [[CrossRef](#)] [[PubMed](#)]
54. Lee, E.J.; In, K.H.; Kim, J.H.; Lee, S.Y.; Shin, C.; Shim, J.J.; Kang, K.H.; Yoo, S.H.; Kim, C.H.; Kim, H.-K.; et al. Proteomic Analysis in Lung Tissue of Smokers and COPD Patients. *Chest* **2009**, *135*, 344–352. [[CrossRef](#)] [[PubMed](#)]
55. Berenson, C.S.; Garlipp, M.A.; Grove, L.J.; Maloney, J.; Sethi, S. Impaired phagocytosis of nontypeable Haemophilus influenzae by human alveolar macrophages in chronic obstructive pulmonary disease. *J. Infect. Dis.* **2006**, *194*, 1375–1384. [[CrossRef](#)] [[PubMed](#)]
56. Singh, D.; Agusti, A.; Anzueto, A.; Barnes, P.J.; Bourbeau, J.; Celli, B.R.; Criner, G.J.; Frith, P.; Halpin, D.M.G.; Han, M.; et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: The GOLD science committee report 2019. *Eur. Respir. J.* **2019**, *53*, 1900164. [[CrossRef](#)]
57. Berenson, C.S.; Wrona, C.T.; Grove, L.J.; Maloney, J.; Garlipp, M.A.; Wallace, P.K.; Stewart, C.C.; Sethi, S. Impaired Alveolar Macrophage Response to Haemophilus Antigens in Chronic Obstructive Lung Disease. *Am. J. Respir. Crit. Care Med.* **2006**, *174*, 31–40. [[CrossRef](#)]
58. Tu, C.; Li, J.; Young, R.; Page, B.J.; Engler, F.; Halfon, M.S.; Canty, J.J.M.; Qu, J. Combinatorial Peptide Ligand Library Treatment Followed by a Dual-Enzyme, Dual-Activation Approach on a Nanoflow Liquid Chromatography/Orbitrap/Electron Transfer Dissociation System for Comprehensive Analysis of Swine Plasma Proteome. *Anal. Chem.* **2011**, *83*, 4802–4813. [[CrossRef](#)]
59. Tu, C.; Li, J.; Bu, Y.; Hangauer, D.; Qu, J. An ion-current-based, comprehensive and reproducible proteomic strategy for comparative characterization of the cellular responses to novel anti-cancer agents in a prostate cell model. *J. Proteom.* **2012**, *77*, 187–201. [[CrossRef](#)]
60. Novichkova, S.; Egorov, S.; Daraselia, N. MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics* **2003**, *19*, 1699–1706. [[CrossRef](#)]
61. Cheadle, C.; Cao, H.; Kalinin, A.; Hodgkinson, J. Advanced literature analysis in a Big Data world. *Ann. N. Y. Acad. Sci.* **2016**, *1387*, 25–33. [[CrossRef](#)]
62. Falls, Z.; Mangione, W.; Schuler, J.; Samudrala, R. Exploration of interaction scoring criteria in the CAND0 platform. *BMC Res. Notes* **2019**, *12*, 1–6. [[CrossRef](#)]
63. Schuler, J.; Samudrala, R. Fingerprinting CAND0: Increased Accuracy with Structure- and Ligand-Based Shotgun Drug Repurposing. *ACS Omega* **2019**, *4*, 17393–17403. [[CrossRef](#)] [[PubMed](#)]
64. Mangione, W.; Falls, Z.; Chopra, G.; Samudrala, R. cando.py: Open Source Software for Predictive Bioanalytics of Large Scale Drug-Protein-Disease Data. *J. Chem. Inf. Model* **2020**, *60*, 4131–4136. [[CrossRef](#)] [[PubMed](#)]
65. Hudson, M.; Samudrala, R. Multiscale Virtual Screening Optimization for Shotgun Drug Repurposing Using the CAND0 Platform. *Molecules* **2021**, *26*, 2581. [[CrossRef](#)] [[PubMed](#)]
66. Schuler, J.; Falls, Z.; Mangione, W.; Hudson, M.L.; Bruggemann, L.; Samudrala, R. Evaluating the performance of drug-repurposing technologies. *Drug Discov. Today* **2021**, *27*, 49–64. [[CrossRef](#)] [[PubMed](#)]
67. Horst, J.; Pieper, U.; Sali, A.; Zhan, L.; Chopra, G.; Samudrala, R.; Featherstone, J. Strategic Protein Target Analysis for Developing Drugs to Stop Dental Caries. *Adv. Dent. Res.* **2012**, *24*, 86–93. [[CrossRef](#)] [[PubMed](#)]
68. Minie, M.; Chopra, G.; Sethi, G.; Horst, J.; White, G.; Roy, A.; Hatti, K.; Samudrala, R. CAND0 and the infinite drug discovery frontier. *Drug Discov. Today* **2014**, *19*, 1353–1363. [[CrossRef](#)]

69. Chopra, G.; Samudrala, R. Multiscale Modelling of Relationships between Protein Classes and Drug Behavior Across all Diseases Using the CANDO Platform. *Mini-Rev. Med. Chem.* **2015**, *15*, 705–717. [[CrossRef](#)]
70. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [[CrossRef](#)]
71. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [[CrossRef](#)]
72. Yang, J.; Roy, A.; Zhang, Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588–2595. [[CrossRef](#)]
73. Dutta, S.; Burkhardt, K.; Young, J.; Swaminathan, G.J.; Matsuura, T.; Henrick, K.; Nakamura, H.; Berman, H.M. Data Deposition and Annotation at the Worldwide Protein Data Bank. *Mol. Biotechnol.* **2008**, *42*, 1–13. [[CrossRef](#)] [[PubMed](#)]
74. Gore, S.; Sanz-García, E.; Hendrickx, P.; Gutmanas, A.; Westbrook, J.D.; Yang, H.; Feng, Z.; Baskaran, K.; Berrisford, J.; Hudson, B.P.; et al. Validation of Structures in the Protein Data Bank. *Structure* **2017**, *25*, 1916–1927. [[CrossRef](#)] [[PubMed](#)]
75. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
76. Gragera, A.; Suppakitpaisarn, V. Semimetric properties of sørensen-dice and tversky indexes. In *International Workshop on Algorithms and Computation*; Springer: Cham, Switzerland, 2016.
77. Brands, U. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **2001**, *25*, 163–177. [[CrossRef](#)]
78. Gyorgy, A. A Practical Step-by-Step Guide for Quantifying Retroactivity in Gene Networks. *Synth. Gene Circuits* **2021**, *2229*, 293–311. [[CrossRef](#)]
79. RC Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2013.
80. Thada, V.; Jaglan, V. Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *Int. J. Innov. Eng. Technol.* **2013**, *2*, 202–205.