Network Signature Techniques for Drug Discovery in Pulmonary Fibrosis

Evangelos Karatzas[1]

[1]Department of Informatics and Telecommunications, University of Athens, 15703, Athens, Greece
vagkaratzas@di.uoa.gr

Abstract. Fibrotic diseases cover a spectrum of systemic and organ specific diseases that affect a large portion of the population, currently without cure. Idiopathic Pulmonary Fibrosis (*IPF*) is an interstitial lung disease and one of the most common and studied fibrotic diseases. Understanding the underlying biological mechanisms and interactions of a disease remains a time consuming and costly task. Computational methodologies that reveal pathway communities can be of great value as they help expedite the process of identifying how perturbations in a single pathway can affect others. Drug Repurposing (DR) is a methodology where already existing drugs are tested against diseases outside their initial spectrum to reduce the high cost of new drug development.

Our 4 main objectives are to (i) identify key differentially expressed genes of fibrotic diseases, (ii) explore the perturbed biological pathways, (iii) suggest repurposed drugs as potential anti-fibrotic candidates for further testing and (iv) identify which fibrotic diseases resemble *IPF* based on common terms, to potentially pursue common regimens.

We analyze transcriptomics datasets to identify key genes implicated in fibrotic diseases. We use these genes as input in DR tools and then propose a novel drug re-ranking methodology via a scoring formula that consolidates standard repurposing scores with structural, functional and side effect scores. Following, we present a pathway analysis and community detection methodology, based on Random Walk theory, where a walker crosses a pathway-to-pathway network under the guidance of a disease-related map. The latter is a gene network that we construct by integrating multi-source information regarding a specific disease. The most frequent trajectories highlight communities of pathways that are expected to be strongly related to the disease under study. By applying our pathway analysis methodology on 9 different fibrotic maladies, we identify various common highlighted pathways as well as unique entries for some of the diseases.

Keywords: Idiopathic Pulmonary Fibrosis, Drug Repurposing, Gene Expression, Biological Pathways, Chemical Structures

1 Dissertation Summary

1.1 Introduction

Fibrotic diseases constitute a group of incurable maladies that are recognized by a fibrotic phenotype and affect a large portion of the population. They cover a spectrum of systemic and organ specific diseases where their corresponding mechanisms escape the homeostatic

Dissertation Advisor: A. Delis, Professor.

signals and due to over-repairing, they cause tissue scarring. This leads to the functional failure of the organ or system. *Idiopathic Pulmonary Fibrosis* (*IPF*) is a rare, incurable disease of the respiratory system during which fibrotic tissue and scars appear in the lungs. *IPF* leads to death within *3-5* years due to the rapid fibrotic progression. According to a 2014 study, article data from *21* countries suggest an incident rate of *3-9* cases per *100,000* (per year) for North America and Europe and lower rates for East Asia (*1.2–3.8* per *100,000*) and South America (*0.4–1.2* per *100,000*) [1]. Recent developments on the field have led to updates in *IPF*'s diagnosis' guidelines. A multidisciplinary committee of *IPF* experts has provided new guidelines for *IPF* diagnosis by combining evidence from high-resolution computed tomography (HRCT) and histopathological patterns of 'usual interstitial pneumonia (UIP)', 'possible UIP' and 'indeterminate for UIP' [2]. The committee also strongly advise against serum biomarker (MMP7, SPD, CCL18, KL6) measurements as an approach to distinguishing *IPF* from other ILDs because of the high false-positive and false-negative result rates.

Despite the improvement of available technologies in the pharmaceutical industry, the cost of commercializing a new drug doubles every 9 years [3]. Designing novel organic compounds in a systematic fashion is a daunting task as it has been estimated that there can be up to 10^{60} molecules with drug-like properties [4]. Drug Repurposing or Drug Repositioning (DR) is the process during which known drugs are applied to different diseases. *In silico* DR studies have been published regarding maladies such as *Alzheimer's Disease* (*AD*), where Siavelis et al. [5] used a bioinformatics pipeline to discover potential inhibitor drugs against *AD*. Bourdakou et al. [6] constructed and used a DR pipeline in the use case of Breast Cancer. More DR studies have been published on Malaria and Tuberculosis [7], on other parasitic diseases such as Trypanosomiasis, Toxoplasmosis, Cryptosporidiosis and Leishmaniasis [8], on Small Cell Lung Cancer [9] as well as on Gastrointestinal Stromal Tumor [10].

The various categories of biological data are known as omics. Multiomics is a biological analysis approach where data consist of multiple "omes". Some of the major omics categories are: (i) genomics which concern structure, function, evolution, mapping, and editing of the genome (DNA), (ii) transcriptomics which concern the transcribed gene expression into mRNA as well as into non-coding RNA, (iii) proteomics which concern the functional proteins translated from mRNA and (iv) metabolomics, which concern metabolites such as sugars, nucleotides, amino acids and lipids.

The goal of the current thesis is to (i) identify key regulatory genes of *IPF* and other fibrotic diseases, (ii) highlight key biological pathways that are involved in the pathogenesis of fibrosis, (iii) propose candidate drugs for further experimental validation, (iv) identify fibrotic diseases common to *IPF* to pursue further common treatments among them and develop tools and methodologies regarding Pathway Analysis and DR.

1.2 DR Pipeline against *IPF*

For the first DR pipeline that we utilize [11], we download *IPF* gene expression datasets from GEO. We perform statistical analysis in order to extract key differentially expressed genes per experiment with the R/Bioconductor package Limma. We select the top-150 and top-150 under-expressed genes (*p-value* < 0.05) (due to the input restrictions of DR tools), of each dataset based on their fold change (FC). We also use the NetWalker tool, as a Random Walk-based approach, to construct a second list with the same number of over- and under-expressed genes.

We use the differentially expressed gene lists from the analyses as input in 3 DR tools (old version of CMap, LINCS (old version) and SPIEDw) and the tools return drug lists sorted by ascending inhibition scores. Negative drug score values suggest inhibiting mode of action against the disease's signature, while positive drug scores hint for inducing action for the disease. We re-rank the candidates by integrating inhibition, structural and functional properties as well as potential side effects in a composite scoring schema via our methodology, termed CoDReS (Composite Drug Repurposing Scoring).

For the pathway analysis section of our initial DR pipeline, we use the gene lists (over- and under-expressed) of each experiment as input to Enrichr's KEGG pathway analysis to discover molecular mechanisms involved in the pathogenesis of *IPF*. Finally, we identify miRNAs that can potentially silence over-expressed genes from our analysis.

1.3 Drug Re-ranking

In our second work, we created a tool based on our initial CoDReS methodology [12]. Currently, CoDReS is hosted in the CING - Bioinformatics Group Servers (C-BIG Servers) (http://bioinformatics.cing.ac.cy/CoDReS/). We implement a scoring formula based on data from online biological repositories that are regularly updated. We calculate a composite score (*CoDReS*) for each drug, as the normalized weighted sum of the initial a-priori score (aS) with a functional (FS) and a structural score (StS) as introduced below:

$$CoDReS_i = \frac{w_{aS} * aS_i + w_{FS} * FS_i + w_{StS} * StS_i}{maxCoDReS}, i = 1, \dots, N\ drugs$$

The weights w_{aS}, w_{FS} and w_{StS} are user-defined parameters that determine the desired influence of each part (a-priori, functional and structural scores respectively) to the final score and have equal default values. The a-priori scores can be uploaded by the user and are automatically normalized in the unit interval [0, 1] by dividing by the absolute maximum a-priori score. The functional score requires the calculation of two different parameters: (i) the Confidence Score, which reflects the gene-disease association and (ii) the Ki, which is an inhibitory constant, measured in nM, and represents the reciprocal of the binding affinity between the inhibitor (drug) and the enzyme (target). The smaller the Ki, the greater the binding affinity. We calculate the FS for each drug as the sum of the products of Confidence Score with the inverse value of Ki, for each gene target of the drug that has been related to the queried disease. Each drug's FS is finally normalized in [0, 1] by dividing with the maximum FS.

$$FS_i = \frac{\sum_{j=1}^{nGenes} ConfidenceScore_j \frac{1}{Ki_j}}{maxFS}$$

The structural score calculates a substance's drug-likeness based on the Lipinski "rules of 5" and Veber's rules. According to the Lipinski rules, in order for a drug to be orally active in humans, it should conform to the following rules: (i) have *≤ 5 hydrogen bond donors*, (ii) have *≤ 10 hydrogen bond acceptors*, (iii) weigh *< 500 Da* and (iv) have an octanol-water partition coefficient *(log P) ≤ 5*. The Veber's rules further require that the chemical substance (v) contains ≤ 10 rotatable bonds and (vi) its polar surface area does not exceed *140 $Å^2$* (angstrom^2). The final StS for each drug is a value within the range [0, 1] calculated in the following way:

$$StS_i = 1 - \frac{numViolations}{6}$$

where "6" is the maximum number of structural rules that a drug might violate.

1.4 Pathway Analysis

Following, our next work concerns a novel Pathway Analysis methodology, termed PathWalks [13]. PathWalks methodology integrates random walks and shortest paths computations to walk on a pathway-to-pathway network under the guidance of a synthetic gene network that we construct by integrating a-priori molecular information related to a disease. The PathWalks methodology exploits two main network components related to a disease of interest which need to be constructed before the execution of the algorithm.

The first component is the multisource information map; this is a synthetic gene-to-gene network which represents integrated information (e.g., gene co-expression, physical interactions, miRNA targets) from biological databases in the form of weighted connections. Mathematically, the gene network is represented as a graph (G_g) and described as $G_g = (V_g, E_g)$, where V_g is the set of nodes (genes) and E_g is the set of connections among nodes. The walker performs random walks on the gene network and the visited nodes indicate the walker's new destination on the PathWalks' second component; the functional connectivity network of biological pathways.

We construct the pathway-to-pathway network (Gp = (Vp, Ep)), by parsing the biological pathways' functional connectivity information from KEGG. Pathways that contain genes already associated with the studied disease, receive higher numeric-value edge scores (i.e., visitation probability). The walker moves on the pathway-to-pathway network according to the instructions given by the map (gene-to-gene network) in order to explore biological pathway relations regarding the disease under examination.

A sorted list of the most visited pathways is generated after a set number of iterations. In order for the algorithm to converge, the two last sorted pathway-visitation lists must have a similarity index above a selected threshold. Finally, the algorithm highlights the most frequently visited edges (i.e., pathway-to-pathway connections) and nodes (pathways), revealing interesting pathway communities, according to the multisource map. In this study, we explore two use case scenarios from different disease settings; *AD* as a neurodegenerative disease and *IPF* as a fibrotic disease.

1.5 DR Pipeline on Multiple Fibrotic Diseases

In our final work, we revisit our initial DR pipeline, refine it and apply it on 9 fibrotic diseases. We analyze gene expression data and present common and unique: (i) genes, (ii) biological pathways and (iii) candidate repurposed drugs among these diseases, in an effort to better understand and potentially treat fibrosis, while focusing around *IPF*.

2 Results and Discussion

2.1 DR Pipeline against *IPF*

For our initial DR pipeline, after analyzing the multiple *IPF* stage-related experiments from datasets GSE10667, GSE24206 and GSE44723, we result into two gene lists per experiment derived from Limma's statistical analysis and from NetWalker's random walk approach. To identify the most important genes related to each stage of the disease, we combine the two gene lists per experiment, creating a unified over-expressed list and a unified under-expressed list per experiment. Finally, we integrate the information from all datasets by building over- and under-expressed gene lists separately, including genes that are common in at least two datasets for each stage of the disease. After we re-rank the candidate-drug lists, we propose niclosamide, lycorine, naltrexone and anisomycin for further experimentation against *IPF*.

As far as biological pathways are concerned, cell communication, extracellular matrix receptor interaction, focal adhesion, cytokine cytokine receptor interaction and colorectal cancer are the 5 pathways that we observe in all stages of *IPF* through our experiments.

In the last part of our initial DR pipeline, we search for microRNAs that are related to fibrotic diseases in HMDD v2.0 (Human microRNA Disease Database version 2.0). We find their gene targets in mirTarBase and compare them with the gene targets of the repurposed drugs from our study. The hsa-miR-208a-5p miRNA targets the CYP1B1 gene that we identify with NetWalker to be over-expressed in *IPF* experiment “stage 2 vs normal” and we suggest that further experiments should be carried out to test any potential anti-fibrotic action.

2.2 CoDReS

In the following section, we discuss the validity of our CoDReS algorithm, where we consider examples disregarding a-priori scores. We choose the top-40 diseases from DisGeNET with the most correlated genes that have at least 20 drug candidates in Malacards. For each disease, we create a mixture list of 200 drugs: *95*% randomly selected from DrugBank and *5*% of the top drugs reported from Malacards repository as developed/used for the selected disease.

After executing CoDReS for each experiment, we count the number of the actual disease-related drugs that we observe in the top-*5*% of the ranked drugs based on their CoDReS along with a *p-value* calculated through a hypergeometric distribution test. We repeat this procedure 100 times for each disease and then calculate the median, maximum, minimum and average *p-value* metrics for each disease.

CoDReS ranked effectively (*median p-value* < *.05*) the input drugs in 35/40 diseases. CoDReS failed to rank drugs correctly in 5 out of 40 diseases. This failure can be partially explained since the top-10 drugs corresponding to most of these diseases contain abstract substances or generic categories such as “Anti-Inflammatory Agents”, “Cytochrome P-450 Enzyme Inhibitors”, “Immunologic Factors” or drugs with close to zero gene targets participating in the disease.

2.3 PathWalks

PathWalks implements shortest path traversing on a functional connectivity network of biological pathways. Due to the network's topology and the assigned edge weights, certain pathway nodes are consistently highlighted in the results. We perform a PathWalks execution with random biological pathway selection at each iteration (without gene map guidance) to identify these topology-favored nodes that are not necessarily highlighted due to their association with each use case disease. For this random-PathWalks experiment, we use our functional connectivity network of biological pathways and assign edge weights equal to the number of common genes between two pathways. We first compare the top-10% ranked pathway lists among the respective *IPF* and *AD* PathWalks and the random PathWalks experiments to identify which pathways are reranked due to direct association with the biological map and which mostly due to the topology. We then compare the top-10% PathWalks results (31 pathways) with the respective top-31 significant results from other pathway analysis tools to evaluate our results.

PathWalks brings 19 pathways to the top of the results of *AD* and 25 of *IPF* due to the integrated biological information rather than due to topology. “Serotonergic synapse” and “Notch signaling” pathways are the first two entries highlighted directly by *AD’s* gene map. “Cytokine-cytokine receptor interaction”, “TGF-beta signaling” and “Chemokine signaling” pathways are the top-3 *IPF* related results with direct biological connection to the disease.

Nevertheless, we do not necessarily consider topology-favored nodes as true negative entries. Topology-favored nodes either contain functional connections with multiple biological pathways (high degree-value) or connect distinct functional subnetworks (high betweenness-value). Therefore, perturbations in the functional connectivity network potentially affects these nodes indirectly. However, we observe that several of the topology-favored pathways decrease in rank for non-relevant diseases. For example, the “Oxidative phosphorylation” pathway is ranked 2nd in the random PathWalks example and 9th in the *AD* use case, but only 162nd in the use case of *IPF*.

To evaluate our findings, we compare our PathWalks results with those derived from pathway analysis tools including GeneTrail3, Enrichr and EnrichNet. We feed as input to these tools the gene nodes of each map. Subsequently, we establish common highlighted pathway entries between PathWalks and the tools in discussion. This exercise partially helps validate our PathWalks-derived results and constitutes a common pathway analysis technique. For example, Glaab and colleagues have successfully used the intersection of the results of the enrichment analysis tools SAM-GS and GAGE while testing for the confidence of their EnrichNet tool’s pathway analysis results. PathWalks also exclusively highlights several biological pathways not necessarily favored by the topology. Furthermore, the key value-added of PathWalks compared to prior pathway analysis approaches, is that it yields functional connections among pathways as well as proposes pathway clusters. In

Validating pathway analysis methodologies is an invariably challenging task since ground truths and gold standards are often unavailable. Yu and colleagues [14] discuss these difficulties and present a model which can evaluate a pathway analysis methodology based on the consistency of its results on smaller subsets of a main gene expression dataset. However, such an approach can only be followed when parsing gene expression datasets. In our case, that entails gathering of multi-omics data from various sources, we choose to validate our PathWalks results by comparing our results with the results from other tools, similar to Glaab’s approach [15]. Furthermore, we identify corroborating bibliographic evidence to further ascertain the effectiveness of PathWalks mechanisms in *AD* and *IPF*. Without doubt, there is no single best approach in pathway analysis or in validating its results. Although common indications provided by several tools offer a baseline for validating results, one should keep in mind that every individual tool contributes its own incremental value-added through its own unique produced outcome(s).

A number of PathWalks results for *IPF* are neither highlighted by the benchmark tools we explore in our analysis nor by the random (no-map) PathWalks execution. The pathway of "Endocytosis", which is directly connected to "Cytokine-cytokine receptor interaction", is ranked 11th but there is little evidence in bibliography associating this pathway with *IPF*. Specifically, Hsu and others show that *IPF* and *Systemic Sclerosis-Pulmonary Fibrosis* share enriched functional groups regarding genes involved in caveolin-mediated endocytosis [16]. Caveolins are a family of plasma membrane proteins which form caves that are involved in receptor-independent endocytosis [17]. In another study, Shi and colleagues suggest a possibility that *IPF* patients may have perturbations in extracellular matrix endocytosis due to caveolin-1 turnover of the fibronectin matrix [18].

Similarly, the "Apelin signaling" pathway, which is directly connected to "MAPK signaling", ranked 23rd and was uniquely produced by PathWalks. Apelin is an endogenous ligand that binds to the G-protein-coupled receptor, is expressed in multiple tissues and organ systems and is implicated in various physiological processes [19]. There is no bibliographic evidence directly associating this pathway with *IPF*. Hence, both "Apelin signaling" and "Endocytosis" pathways should be further explored for potential contribution to the fibrogenesis of *IPF* patients.

2.4 DR Pipeline on Multiple Fibrotic Diseases

In the final results section, we present our DR findings regarding fibrotic diseases. *IPF* has the most common over-expressed genes with Dupuytren's Disease (35) and the most common under-expressed genes with Myelofibrosis (28). Schistosomiasis has the fewest common genes with *IPF* (*3* over- and *3* under-expressed). Through our dataset analysis, we identify the upregulated gene LCN2 associated with *IPF*, Cystic Fibrosis *(CF)*, Schistosomiasis and Systemic Sclerosis (*SSc)* and the under-expressed gene FBLN1 associated with *CF*, Myelofibrosis, Polycystic Kidney Disease and *SSc*. According to the bibliography, even though FBLN1's mRNA levels might be decreased in Chronic Obstructive Pulmonary Disease where small airway fibrosis occurs, on the protein level it might have already been accumulated in the ECM. FBLN1 levels were found increased in serum and bronchoalveolar lavage fluid of asthma patients [20] and in the plasma and lung tissue of *IPF* patients [21]. Based on our results (under-expression of FBLN1) and the bibliography, we observe an association among FBLN1 and fibrotic diseases but only regarding specific tissues (e.g., lung, myocardium). We suggest that further proteomics analyses should be carried out to identify the quantity of the fibulin-1 translated protein in the related fibrotic tissues and its potential involvement in fibrosis.

Following, through the PathWalks runs we identify key disease-specific as well as common pathways between the 9 fibrotic maladies. We observe seven common pathways (favored by the topology) across all nine diseases. These include "Metabolic", "Cancer", "MAPK signaling", "PI3K-Akt signaling", "Non-alcoholic fatty liver disease", "Oxidative phosphorylation" and "Calcium signaling" pathways. We also observe unique pathways highlighted for some of the diseases, with the most interesting ones potentially linking Oral Submucous Fibrosis to myocardial diseases.

We use the key differentially expressed gene lists from our dataset analysis, as input in two signature-based DR tools; CMap and L1000CDS[2]. We then re-rank the drug lists based on our CoDReS tool. Two of the drug candidates returned by CoDReS are common entries among 3 of the 9 fibrotic diseases. Hydrocortisone is selected as an anti-fibrotic drug candidate from the re-ranking procedure for *IPF*, *CF* and *SSc*. Similarly, memantine is highlighted in the use cases of *IPF*, Dupuytren's Disease and Schistosomiasis. In an effort to further screen the repurposed and re-ranked drug candidates, we explore structural similarities among them and drugs that have previously failed in clinical trials against fibrotic diseases. Memantine's action has not been studied against fibrotic diseases but there are hints suggesting further experimentation [22, 23]. On the other hand, low doses of hydrocortisone have shown to attenuate fibrosis especially in the early stages [24, 25]. However, hydrocortisone has high structural similarity with drugs that have previously failed in clinical trials against fibrosis. Even though we suggest prioritizing drugs that are dissimilar to "failed" ones, we do not reject the possibility that even these candidates might succeed in combination with other drugs or at different dosages and/or stages of fibrosis.

In the sequel, we highlight the most promising anti-fibrotic candidates while focusing on *IPF*, by identifying the drugs' gene targets inside key pathway communities. We focus on the targeted *IPF* highlighted pathways that are directly associated with its integrated genetic information map in an effort to explore and distinguish anti-fibrotic candidates among these 89 drugs. Specifically, these are the "Pancreatic secretion", "Protein digestion and absorption" and "Complement and coagulation cascades" pathways. There are 34/121 re-ranked drugs that target at least one gene participating in these pathways. We bibliographically explore the 7 candidates that target at least 2 out of the 3 most important pathways of IPF, namely celecoxib, digoxin, captopril, ibuprofen, staurosporine, nafcillin and wortmannin. Among these, 4 seem most appropriate to further study against fibrosis; captopril, ibuprofen, nafcillin and digoxin.

We observe gene, pathway and repurposed candidate similarities among *IPF* and the rest of the fibrotic diseases. Specifically, *IPF* shares several terms with Dupuytren's Disease having 35 common over-expressed and 16 common under-expressed genes, 2 common key pathways with direct association to the respective genetic information maps and 2 common identified drug candidates. *IPF* and Myelofibrosis share 28 under-expressed genes. *IPF* and IgG4-related Disease share 23 over-expressed genes and 3 key pathways. Finally, *IPF* shares 20 over-expressed genes and 1 key pathway with SSc and 3 drug candidates and 1 key pathway with CF. We suggest that common treatments for *IPF* and the aforementioned diseases, especially Dupuytren's Disease, should be further pursued.

3 Conclusions

During my PhD thesis I studied and analyzed gene expression datasets mainly revolving around *IPF* as well as other fibrotic diseases. My contribution lies in the suggestion of potential unique and shared (i) genes, (ii) biological pathways and (iii) anti-fibrotic drug candidates among *IPF* and other fibrotic diseases. I also designed and developed in silico DR-related tools (CoDReS, ChemBioServer 2.0 [26]) as well as implemented and presented a methodology for pathway community detection (PathWalks).

More specifically, some of the most promising drug candidates that we suggest for further experimentation against fibrosis are: niclosamide, lycorine, naltrexone, anisomycin, captopril and ibuprofen. Niclosamide already has positive results against fibrotic cell lines and pulmonary fibrosis mouse models. We are currently testing niclosamide against myofibroblasts from lungs of *IPF* patients and healthy individuals in the Department of Medicine in Democritus University of Thrace.

We identify "Apelin signaling" and "Endocytosis" pathways as novel indications through our PathWalks methodology that should be further experimentally pursued for their potential contribution to the fibrogenesis of *IPF*.

Finally, we suggest further pursuing of common treatments among *IPF* and the fibrotic diseases that, based on our results, seem more "similar" to *IPF*. We conclude that these diseases are Dupuytren's Disease, *SSc*, *CF* and IgG4-related Disease.

As far as our developed DR tools are concerned, (i) CoDReS concerns the re-ranking and repurposing of drug candidates based on their functional relation to a disease of interest as well as their drugability and (ii) ChemBioServer concerns DR based on structural similarity of substances.

Our *in silico* bioinformatics analyses and tools face certain limitations regarding the validation of results compared to those from wet-lab experiments. Our suggestions constitute indications that require further experimental validation. However, since our input data come from biologically curated databases and our methodologies are scientifically valid, we consider our indications as reliably screened candidates.

4 References

[1] J. Hutchinson, A. Fogarty, R. Hubbard, and T. McKeever, "Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review," *European Respiratory Journal,* vol. 46, pp. 795-806, 2015.

[2] G. Raghu, M. Remy-Jardin, J. L. Myers, L. Richeldi, C. J. Ryerson, D. J. Lederer*, et al.*, "Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline," *American journal of respiratory and critical care medicine,* vol. 198, pp. e44-e68, 2018.

[3] J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency," *Nature reviews Drug discovery,* vol. 11, p. 191, 2012.

[4] P. G. Polishchuk, T. I. Madzhidov, and A. Varnek, "Estimation of the size of drug-like chemical space based on GDB-17 data," *Journal of computer-aided molecular design,* vol. 27, pp. 675-679, 2013.

[5] J. C. Siavelis, M. M. Bourdakou, E. I. Athanasiadis, G. M. Spyrou, and K. S. Nikita, "Bioinformatics methods in drug repurposing for Alzheimer's disease," *Briefings in bioinformatics,* vol. 17, pp. 322-335, 2015.

[6] M. M. Bourdakou, E. I. Athanasiadis, and G. M. Spyrou, "Discovering gene re-ranking efficiency and conserved gene-gene relationships derived from gene co-expression network analysis on breast cancer data," *Scientific reports,* vol. 6, p. 20518, 2016.

[7] A. Nzila, Z. Ma, and K. Chibale, "Drug repositioning in the treatment of malaria and TB," *Future medicinal chemistry,* vol. 3, pp. 1413-1426, 2011.

[8] K. T. Andrews, G. Fisher, and T. S. Skinner-Adams, "Drug repurposing and human parasitic protozoan diseases," *International Journal for Parasitology: Drugs and Drug Resistance,* vol. 4, pp. 95-111, 2014.

[9] N. S. Jahchan, J. T. Dudley, P. K. Mazur, N. Flores, D. Yang, A. Palmerton*, et al.*, "A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors," *Cancer discovery,* vol. 3, pp. 1364-1377, 2013.

[10] Z. Y. Pessetto, S. J. Weir, G. Sethi, M. A. Broward, and A. K. Godwin, "Drug repurposing for gastrointestinal stromal tumor," *Molecular cancer therapeutics,* vol. 12, pp. 1299-1309, 2013.

[11] E. Karatzas, M. Bourdakou, G. Kolios, and G. Spyrou, "Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score," *Scientific reports,* vol. 7, p. 12569, 2017.

[12] E. Karatzas, G. Minadakis, G. Kolios, A. Delis, and G. M. Spyrou, "A Web Tool for Ranking Candidate Drugs Against a Selected Disease Based on a Combination of Functional and Structural Criteria," *Computational and structural biotechnology journal,* vol. 17, pp. 939-945, 2019.

[13] E. Karatzas, M. Zachariou, M. M. Bourdakou, G. Minadakis, A. Oulas, G. Kolios*, et al.*, "PathWalks: Identifying pathway communities using a disease-related map of integrated information," *Bioinformatics*.

[14] C. Yu, H. J. Woo, X. Yu, T. Oyama, A. Wallqvist, and J. Reifman, "A strategy for evaluating pathway analysis methods," *BMC bioinformatics,* vol. 18, p. 453, 2017.

[15] E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, and A. Valencia, "EnrichNet: network-based gene set enrichment analysis," *Bioinformatics,* vol. 28, pp. i451-i457, 2012.

[16] E. Hsu, H. Shi, R. M. Jordan, J. Lyons-Weiler, J. M. Pilewski, and C. A. Feghali-Bostwick, "Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension," *Arthritis & Rheumatism,* vol. 63, pp. 783-794, 2011.

[17] T. M. Williams and M. P. Lisanti, "The caveolin proteins," *Genome biology,* vol. 5, p. 214, 2004.

[18] F. Shi and J. Sottile, "Caveolin-1-dependent β1 integrin endocytosis is a critical regulator of fibronectin turnover," *Journal of cell science,* vol. 121, pp. 2360-2371, 2008.

[19] K. Tatemoto, M. Hosoya, Y. Habata, R. Fujii, T. Kakegawa, M.-X. Zou*, et al.*, "Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor," *Biochemical and biophysical research communications,* vol. 251, pp. 471-476, 1998.

[20] J. Y. Lau, B. G. Oliver, M. Baraket, E. L. Beckett, N. G. Hansbro, L. M. Moir*, et al.*, "Fibulin-1 is increased in asthma–a novel mediator of airway remodeling?," *PLoS One,* vol. 5, 2010.

[21] J. Jaffar, S. Unger, T. J. Corte, M. Keller, P. J. Wolters, L. Richeldi*, et al.*, "Fibulin-1 predicts disease progression in patients with idiopathic pulmonary fibrosis," *Chest,* vol. 146, pp. 1055-1063, 2014.

[22] S. Abbaszadeh, A. Javidmehr, B. Askari, P. M. Janssen, and H. Soraya, "Memantine, an NMDA receptor antagonist, attenuates cardiac remodeling, lipid peroxidation and neutrophil recruitment in heart failure: A cardioprotective agent?," *Biomedicine & Pharmacotherapy,* vol. 108, pp. 1237-1243, 2018.

[23] Y. Li, Y. Liu, X. Peng, W. Liu, F. Zhao, D. Feng*, et al.*, "NMDA receptor antagonist attenuates bleomycin-induced acute lung injury," *PLoS One,* vol. 10, 2015.

[24] J. V. DeSa, "LXXXVII Submucous Fibrosis of the Palate and Cheek," *Annals of Otology, Rhinology & Laryngology,* vol. 66, pp. 1143-1159, 1957.

[25] M. Xu, B. Deng, Y.-l. Chow, Z.-z. Zhao, and B. Hu, "Effects of curcumin in treatment of experimental pulmonary fibrosis: a comparison with hydrocortisone," *Journal of ethnopharmacology,* vol. 112, pp. 292-299, 2007.

[26] E. Karatzas, J. E. Zamora, E. Athanasiadis, D. Dellis, Z. Cournia, and G. M. Spyrou, "ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing," *Bioinformatics,* vol. 36, pp. 2602-2604, 2020.