Drug repurposing through a bioinformatics pipeline applied on IPF and other fibrotic diseases

Evangelos Karatzas[1], Andrea Kakouri[2, 3], George Kolios[4], Alex Delis[1] and George M. Spyrou[2, 3], *

[1]Department of Informatics and Telecommunications, University of Athens, 15703, Athens, Greece,

[2]The Cyprus Institute of Neurology and Genetics, 6 Iroon Avenue, 2371, Nicosia, Cyprus,

[3]The Cyprus School of Molecular Medicine, 6 International Airport Avenue, Nicosia, 2370, Cyprus,

[4]Laboratory of Pharmacology, Department of Medicine, Democritus University of Thrace, Alexandroupolis, Greece.

*Corresponding author. George M. Spyrou, Tel.: +357 22 392852; E-mail: georges@cing.ac.cy

Abstract

Subject: Fibrotic diseases cover a spectrum of systemic and organ-specific maladies that affect a large portion of the population, currently without cure. Common ground among these diseases is the uncontrollable fibrogenesis which is responsible for the accumulated damage in the susceptible tissues. *Idiopathic Pulmonary Fibrosis* (*IPF*), an interstitial lung disease, is one of the most common and studied fibrotic diseases and still remains an active research target.

Objective: We highlight unique and common (i) genes, (ii) biological pathways and (iii) candidate repurposed drugs among nine fibrotic diseases. We bibliographically explore the resulting candidate substances for potential anti-fibrotic mode of action and underline those diseases that appear to be more similar to *IPF*, to be jointly studied for treatments.

Methodology: We identify key genes for the nine fibrotic diseases by analyzing transcriptomics datasets. We construct gene-to-gene networks for each disease and utilize them to explore functional communities of biological pathways. We also use the most significant genes as input in Drug Repurposing (DR) tools and re-rank the resulting candidates according to their structural properties and functional relation to each disease.

Results: We identify seven biological pathways involved in all nine fibrotic diseases as well as pathways unique to some of them. We observe a group of cardiomyopathy-related pathways that are exclusively highlighted for *Oral Submucous Fibrosis* (*OSF*). Based on our DR results, we suggest (i) captopril and (ii) ibuprofen, which seem to slow the progression of fibrotic diseases

according to existing bibliography, as well as (iii) nafcillin and (iv) memantine, which haven't been studied against fibrosis yet, for further wet-lab experimentation. We also suggest digoxin to be tested against *OSF*, since we observe cardiomyopathy-related pathways implicated in *OSF* and there is bibliographic evidence that digoxin may potentially clear myocardial fibrosis. Finally, we observe that *IPF* shares several involved genes, biological pathways and candidate inhibiting-drugs with *Dupuytren's Disease*, *IgG4-related Disease*, *SSc* and *Cystic Fibrosis*. We suggest that treatments among *IPF* and the aforementioned diseases should be jointly pursued.

Keywords: Idiopathic Pulmonary Fibrosis, Drug Repurposing, Drug Re-ranking, Pathway Analysis, Disease Similarity

Introduction

Fibrotic diseases constitute a group of incurable maladies that are recognized by a fibrotic phenotype affecting the respective organs or tissues. The related mechanisms escape the homeostatic signals and due to over-repairing, cause tissue scarring. The plethora and complexity of the underlying perturbations are responsible for the lack of treatments. Even though fibrotic diseases target various biological structures, they share several underlying mechanisms.

One of the most important fibrotic mechanisms is the extracellular matrix (ECM) deposition which is known to drive both the initiation as well as the progression of fibrogenesis (Herrera, *et al.*, 2018). ECM is a three-dimensional network of extracellular macromolecules such as collagen, enzymes and glycoproteins, that provide structural and biochemical support for surrounding cells (Theocharis, *et al.*, 2016). The uncontrolled accumulation of ECM macromolecules is responsible for the replacement of normal with non-functional fibrotic tissue. The transforming growth factor-β (TGF-β) cytokine is a key regulator of ECM, since TGF-β signals, and particularly SMAD proteins which are TGF-β signal transducers, act as transcription factors that induce ECM's expression in mesenchymal cells (Verrecchia and Mauviel, 2007). Other common mechanisms among fibrotic diseases include the (i) bone morphogenic protein (BMP) signaling, (ii) overexpression of connective tissue growth factor (CTGF), (iii) Wnt/β-catenin pathway and (iv) platelet-derived growth factor (PDGF) signaling (Rosenbloom, *et al.*, 2017). In particular, PDGF-A-/PDGFRα signaling loops stimulate fibroblasts to synthesize ECM and release pro-fibrotic mediators (Bonner, 2004).

Idiopathic Pulmonary Fibrosis (*IPF*) is one of the most common and studied fibrotic diseases and remains an active research target. *IPF* is an interstitial lung disease (*ILD*), primarily involving the tissue and space around the air sacs of the lungs. *IPF* is of unknown etiology and due to rapid fibrotic progression, leads to death in about 3-5 years. According to a 2014 study, data from 21 countries present 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) (Hutchinson, *et al.*, 2015). Recent developments have led to updates in the diagnosis guidelines of *IPF*. A multidisciplinary committee of *IPF* experts 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*' (Raghu, *et al.*, 2018). The committee also strongly advises against serum biomarker (MMP7, SPD, CCL18, KL6) measurements as an approach to distinguish *IPF* from other *ILDs* because of the high false-positive and false-negative result rates.

In a previous work we analyzed fibrotic molecular and phenotypic data from biological databases (Malacards and Human Phenotype Ontology) for 14 fibrotic diseases in an initial effort to group fibrotic diseases and identify similarities with *IPF* (Karatzas, *et al.*, 2019). The diseases that we observed being more similar to *IPF* were *Silicosis*, *Cystic Fibrosis* (*CF*), *Systemic Sclerosis* (*SSc*) and *IgG4-related Disease*. Nevertheless, drawing conclusions regarding the similarities of the underlying mechanisms and pathogenesis among diseases is a challenging task.

In this paper we study transcriptomics datasets regarding 9 of the 14 fibrotic diseases from our previous study and identify common and unique (i) genes, (ii) biological pathways and (iii) candidate repurposed drugs while focusing around *IPF*, in an effort to better understand and potentially treat fibrosis. We analyze gene expression datasets from microarray and next generation RNA-Sequencing (RNA-Seq) experiments to identify over- and under-expressed genes related to each disease. We then use the respective gene lists to drive random walks on a functional connectivity network of biological pathways through our pathway analysis methodology, called PathWalks, to ultimately identify key fibrotic pathway communities (Karatzas, et al.). We also use the gene lists as input in computational signature-based Drug Repurposing (DR) tools (Lussier and Chen, 2011) to identify potential therapeutic candidates against fibrosis. Following, we re-rank the repurposed drugs based on their drugability and functional relation to each disease (i.e., gene targets) through our CoDReS tool (Karatzas, *et al.*, 2019). We further screen the drug candidates by prioritizing compounds that are structurally dissimilar to drugs that have already failed in clinical trials against fibrotic diseases. This is justifiable since drugs with similar structures tend to have similar mode of action (Campillos, *et al.*, 2008). Finally, we explore the existing bibliography for evidence that further supports the potential anti-fibrotic action of our proposed drugs. We present a flowchart of analysis pipeline in Figure 1.

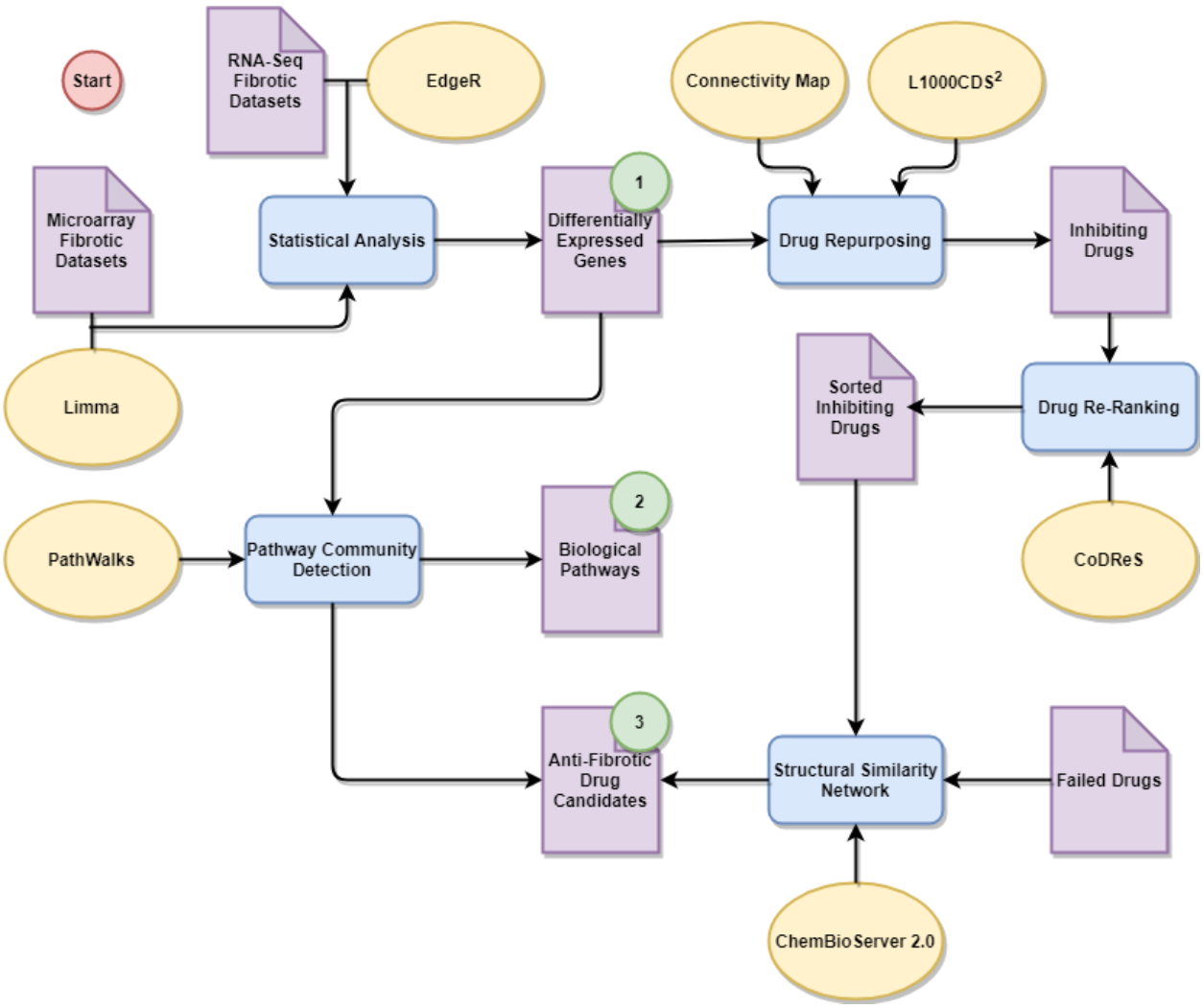

Figure 1: Flowchart of our pipeline regarding pathway analysis and drug repurposing for fibrotic diseases based on transcriptomics data. Purple pages represent data, blue rectangles represent procedures, yellow ellipses represent tools and methodologies and green circles indicate results regarding: (1) genes, (2) biological pathways and (3) inhibiting-drug candidates. We start by analyzing transcriptomics data from several tissues regarding 9 fibrotic diseases to extract key respective gene lists. We then identify common and unique perturbed biological pathways as well as potential therapeutic candidates by exploiting these gene lists.

Materials and Methods

Microarray Dataset Analysis

We download microarray datasets for the 9 fibrotic diseases from Gene Expression Omnibus (GEO) (Edgar, *et al.*, 2002). We analyze the pre-processed *Series Matrix Files* of the data, provided by each experiment. We apply the *normalizeQuantiles* function, from the Limma package (Ritchie, *et al.*, 2015) of the R programming language to non-normalized datasets. We then apply a $\log_2$ transformation to non-$\log_2$ datasets. If a dataset is already $\log_2$ transformed but not normalized, we remove the $\log_2$ transformation, then normalize and finally re-apply the $\log_2$ transformation. After executing the Limma statistical analysis, we use a significance cutoff-threshold (*p-value* <

0.05) on the results. We remove any genes that have dual sign probes (over- and under-expressed at the same time). In the case of duplicate gene entries, we keep the result with the highest significance (i.e., lowest *p-value* score). Finally, we extract lists of the top-*150* over- and top-*150* under-expressed genes per experiment which we then use as input in pathway analysis and DR tools.

Here, we provide some clarifications regarding specific samples of the datasets that we analyze. For the *Myelofibrosis* dataset GSE124281, we discard disease sample GSM3526859 since it was pooled from two patients and also exclude disease samples from patients with *Essential Thrombocythemia*. For the *Oral Submucous Fibrosis* (*OSF*) dataset GSE64216, we exclude the two *Squamous Cell Carcinoma* samples. We also note that in the *CF* dataset GSE15568 patients carry the CF-specific D508 mutated CFTR-allele while in the *CF* dataset GSE40445 patients have a confirmed F508del homozygosity for the CFTR gene. We exclude five *Sjogren's Syndrome* samples from the *IgG4-related Disease* dataset GSE40568. We also exclude seven *Hepatitis B* samples from the *Schistosomiasis* dataset GSE61376. Finally, we exclude the *Non-Specific Interstitial Pneumonia* (*NSIP*) and *IPF-NSIP* samples from the *IPF* dataset GSE110147. Table 1 shows dataset details regarding the related disease, sample tissue and remaining number of samples (total, disease and control) after our exclusions.

Table 1: Dataset information regarding the GEO microarray experiments. We analyze 4 *IPF* microarray datasets where all samples are derived from lung tissue. *CF* samples come from 2 microarray datasets from rectal mucosal epithelia and nasal epithelial cells respectively. *SSc* samples are derived from skin tissue. The 2 *Myelofibrosis* datasets contain bone marrow and bone marrow stromal cell samples respectively. The *OSF* samples are derived from oral buccal mucosa. The *Dupuytren's Disease* samples are derived from hand connective tissue. The *Polycystic Kidney Disease* samples are derived from renal cysts. The *IgG4-related Disease* samples come from labial salivary glands and finally, *Schistosomiasis* samples from liver tissue. Additionally, the number of samples (total, disease and control) that we experimented on, for each dataset entry, is shown on the table.

Dataset ID	Disease	Tissue	#Samples	#Disease	#Control
GSE110147	*IPF*	Lung	33	22	11
GSE53845	*IPF*	Lung	48	40	8
GSE32537	*IPF*	Lung	207	167	50
GSE35145	*IPF*	Lung	8	4	4
GSE15568	*CF*	Rectal mucosal epithelia	29	16	13
GSE40445	*CF*	Nasal epithelial cells	10	5	5
GSE95065	*SSc*	Skin	33	18	15
GSE124281	*Myelofibrosis*	Bone marrow	12	4	8
GSE44426	*Myelofibrosis*	Bone marrow mesenchymal stromal cells	12	6	6
GSE64216	*OSF*	Oral buccal mucosa	6	4	2
GSE75152	*Dupuytren's*	Hand connective tissue	24	12	12
GSE7869	*Polycystic Kidney*	Renal cysts	21	18	3
GSE40568	*IgG4-related Disease*	Labial salivary gland	8	5	3
GSE61376	*Schistosomiasis*	Liver	10	6	4

RNA-Seq Dataset Analysis

We also download RNA-Seq datasets from GEO that are available for *IPF*, *CF* and *SSc*. We analyze the pre-processed *Raw Count Matrix Files* of the data provided by each experiment using R's EdgeR package (Robinson, *et al.*, 2010). We perform normalization on the gene count matrices for RNA composition, by calculating the normalizing factors for the library sizes using the *trimmed mean of M-values* method between each pair of samples. We then test for differential expression among patient and healthy control samples using the *quasi-likelihood F-test (QLF)*, a statistical method based on negative binomial distribution that calculates the natural variation between biological replicates. The *QLF-test* offers a stricter and more reliable error rate control in estimating the uncertainty in gene dispersion (Chen, *et al.*, 2016). We then use a significance cutoff-threshold on the results (*p-value = 0.05*). In agreement with the microarray dataset analysis procedure, we extract lists of the top-*150* over- and top-*150* under-expressed genes per experiment.

Regarding the GSE124548 *CF* dataset, we exclude the post-drug patient samples and compare pre-drug patient and healthy control samples. Similarly, for the *IPF* dataset GSE99621 we exclude the macroscopically unaffected (normal-appearing) patient samples and test for differentially expressed genes between macroscopically affected and healthy controls. Table 2 shows dataset details regarding each related disease, sample tissue and number of samples (total, disease and control) after our exclusions.

Table 2: Dataset information regarding the GEO RNA-Seq experiments. We analyze 2 RNA-Seq datasets containing *IPF* lung samples. We also analyze 1 RNA-Seq dataset from *CF* whole blood samples and 1 RNA-Seq dataset of *SSc* samples from monocyte-derived macrophages. Additionally, the number of samples (total, disease and control) that we experimented on, for each dataset entry, is shown on the table.

Dataset ID	Disease	Tissue	#Samples	#Disease	#Control
GSE99621	*IPF*	Lung	16	8	8
GSE134692	*IPF*	Lung	72	46	26
GSE124548	*CF*	Whole blood	40	20	20
GSE104174	*SSc*	Monocyte-derived macrophages	72	15	57

Pathway Community Detection

We aggregate the key differentially expressed genes (over- and under-expressed) among experiments per disease. We utilize these gene lists in conjunction with our PathWalks methodology to detect biological pathways and their functional communities that are unique or common among fibrotic diseases. PathWalks is a pathway analysis and community detection methodology that guides a walker on pathway-to-pathway network of functional connections, based on an integrated network of genetic information.

We construct 9 gene-to-gene networks (one per fibrotic disease) based on the aforementioned gene lists by using them as input in Cytoscape's (Shannon, *et al.,* 2003) GeneMANIA plug-in (Montojo, *et al.*, 2010). Specifically, we execute GeneMANIA with the following options: (i) *0* resultant genes and attributes (i.e., not adding enriched genes or functional categories), (ii) weight assigned based on query gene and (iii) network types based on the default selections for gene co-expression, physical interaction and pathways. We export the "*normalized max weight*" column of GeneMANIA's results that contains the relation weights between gene pairs. We then aggregate these weight-values among the different types of networks for the same gene pair and finally translate the gene symbols to their respective identifier from the KEGG database (Kanehisa, *et al.*, 2017) to match them with the respective biological pathways they participate in.

The GeneMANIA results present a weight for each participating type of network and a separate weight for each edge. GeneMANIA's execution algorithm consists of two parts; a label propagation function which assigns weights on the edges of the composite network and a ridge regression (regularized linear regression) function which assigns weights to the various types of networks (Mostafavi, *et al.*, 2008). The final composite network is a weighted sum of the individual data sources. The participating networks' weights sum to *100%* and reflect the relevance of each data source for predicting membership in the query list (Warde-Farley, *et al.*, 2010). Edges are weighted relatively to the co-functionality between genes stated in each dataset. The final edge values are weighted sums of the respective normalized weights multiplied by the respective network's weight (Zuberi, *et al*., 2013). We note that each entry of GeneMANIA's "*normalized max weight*" column describes the normalized value for the maximum raw edge weight of a gene pair, among its various edge-entries for the same network (i.e., multiple scores from different databases for the same gene pair and network type).

Following, we create the functional connectivity network of biological pathways based on KEGG. We assign edge scores equal to the number of common genes between two pathways. We use the same network of pathways as a PathWalks component for the 9 different executions (1 per disease) to compare pathway results that are driven by the respective genetic information networks.

The pathway nodes and their functional communities that are highlighted by the PathWalks algorithm tend to favor hubs with high betweenness centrality and strength scores, due to the algorithm's use of shortest paths while traversing the network of pathways. The betweenness centrality metric represents a node's participation in the total shortest paths of a network while the strength metric represents a node's sum of edge weights. We pay special attention to the most traversed pathways that are not necessarily highlighted due to the topology but mainly through the guidance of the genetic information map to identify key pathways that are commonly or uniquely perturbed in fibrotic diseases.

Drug Repurposing

We use the differentially expressed gene lists from each experiment as input in two signature-based DR tools; Connectivity Map (CMap) (Subramanian, *et al.*, 2017) and L1000CDS^2 (Duan, *et al.*, 2016). We use the top-*150* over- and top-*150* under-expressed genes from each experiment as input in both tools since CMap suggests each input gene list should be between *10* and *150* genes for optimal results. We need the resultant substances to inverse the input genes' (disease) signature. To achieve this in CMap, we choose the detailed list results, then select the "*Compound*" choice of the "*Perturbagen type*" subset and sort the results by ascending connectivity score. In L1000CDS^2 we select the "*reverse*" option in the "*Configuration*" settings on the query page. We keep substances with summary-inhibition score less than *-50* (*-100* denotes maximum inhibition) from each CMap experiment and all 50 returned entries from L1000CDS^2.

We then use our drug re-ranking tool CoDReS (Karatzas, et al., 2019) to screen the lists of drug candidates and identify the most promising inhibitors for each fibrotic disease. CoDReS integrates information from biological databases to calculate the functional association and binding affinities among input drugs and queried disease through gene targets as well as calculates the potential drugability of each drug according to the Lipinski and Veber rules (Lipinski, et al., 1997; Veber, et al., 2002). CoDReS constructs structural clusters of the input substances, based on their chemical composition, by utilizing the affinity propagation algorithm (Bodenhofer, *et al.*, 2011) and proposes the top-ranked inhibitors per cluster for further testing against the considered disease. We compile an aggregated list of repurposed drug candidates, derived from both DR tools across all experiments, for each disease and use them as input in CoDReS while selecting the respective disease for each execution.

To further screen the results, we calculate the structural similarities among the short drug lists from CoDReS and substances that have previously failed in clinical trials against fibrotic diseases. To achieve this, we construct a structural similarity network of our re-ranked drugs and the failed drugs of fibrotic diseases as found in repoDB (Brown and Patel, 2017) (last update: July 28, 2017). We use an edge cutoff-threshold of *substance similarity = 0.7*. With this approach, we avoid prioritizing candidates with a lower chance of success for further experimentation, since drugs with similar chemical structures tend to have similar mode of action. We construct the structural similarity network through the ChemBioServer 2.0 web tool (Karatzas, et al., 2020) and visualize it with Cytoscape (Shannon, *et al.*, 2003).

In the sequel, we highlight the most promising anti-fibrotic candidates while focusing on *IPF*, by identifying the gene targets of the re-ranked drugs that participate in key implicated biological pathways. We utilize drug-target information from the DrugBank (Wishart, *et al.*, 2018), DrugCentral (Ursu, *et al.*, 2019) and DGIdb (Cotto, *et al.*, 2018) databases and pathway-gene information from the KEGG pathway database (Kanehisa, *et al.*, 2017). We finally explore the existing bibliography for anti-fibrotic indications concerning our proposed drug candidates to propose the seemingly most appropriate for further *in vitro* and *in vivo* testing.

Results

Key Fibrosis-related Genes

In this section we present key fibrotic gene candidates based on the datasets and methodologies we describe in the relative microarray and RNA-Seq analyses paragraphs. We aggregate the top-150 over- and top-150 under-expressed gene lists among same disease experiments and observe that *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 gene entries with *IPF* (*3* over- and *3* under-expressed). We identify the LCN2 gene being over-expressed in *IPF*, *CF*, *Schistosomiasis* and *SSc* and the FBLN1 gene being under-expressed in *CF*, *Myelofibrosis*, *Polycystic Kidney Disease* and *SSc* and over-expressed in *IgG4-related Disease*.

LCN2 encodes for the lipocalin-2 protein which is associated with neutrophil gelatinase (Kjeldsen, *et al.*, 1993) and is known to limit bacterial growth (Flo, et al., 2004). Takahashi *et al.* study the LCN2 expression in (i) the skin of patients with *SSc*, (ii) bleomycin-treated mice and (iii) Fli1-deficient endothelial cells. Their experiments show that the LCN2 gene is associated with dermal fibrosis in early *Diffuse Cutaneous SSc* cases (some of which are also diagnosed with *ILD*) but not with *ILD* markers such as diffusion lung capacity for carbon monoxide or vital capacity (Takahashi, *et al.*, 2015). Their results also show correlation between LCN2 and progressive skin sclerosis as well as pulmonary vascular involvement that leads to pulmonary arterial hypertension in *SSc*. Furthermore, Nakagawa and colleagues demonstrate increased expression of LCN2, correlated with tubulointerstitial fibrosis and tubular cell injury in patients with *Chronic Kidney Disease* (Nakagawa, *et al.*, 2015). In another study, Kim *et al.* show increased LCN2 levels in urine samples of patients with *Chronic Hepatitis C* accompanied by hepatic fibrosis (Kim, et al., 2010).

On the other hand, contradicting studies show no direct correlation between LCN2 and hepatic fibrosis. Regarding *Nonalcoholic Fatty Liver Disease* (*NAFLD*), Milner *et al.* observe no association between LCN2 and steatosis, lobular inflammation, ballooning or fibrosis even though the expression of LCN2 is significantly elevated in *NAFLD* samples compared to controls (Milner, *et al.*, 2009). Furthermore, Borkham-Kamphorst and others study the expression levels of LCN2 in rat models with acute and chronic liver injury. Their results show correlation of LCN2 to liver damage and resulting inflammatory responses but not to the degree of liver fibrosis (Borkham-Kamphorst, *et al.*, 2011). Our results are in agreement with bibliographic evidence regarding the over-expression of the LCN2 gene in various fibrotic diseases and their respective targeted tissues, such as *SSc* and the affected skin tissue. However, the contradicting findings regarding various liver-associated fibrotic diseases imply that LCN2 might be expressed differently among diseases, even if they share the same fibrotic phenotype.

The FBLN1 gene encodes for the fibulin-1 protein which is an ECM and plasma glycoprotein (Argraves, *et al.*, 1990). Chen *et al.* study the effects of TGF-β1 on the regulation of FBLN1 on primary human airway smooth muscle cells from volunteers with and without *Chronic*

Obstructive Pulmonary Disease (*COPD*) where small airway fibrosis occurs. They show that TGF-β1 causes a decrease in FBLN1 mRNA and soluble FBLN1 while it increases the FBLN1 in ECM due to the sequestration of soluble FBLN1 (Chen, *et al.*, 2013). This fact can partially justify the under-expression of FBLN1 in our transcriptomics analyses (excluding the use case of *IgG4-related Disease*), since gene expression is measured by the level of the corresponding mRNA present in a cell.

Another study suggests that FBLN1 may be involved in the pathogenesis of not only *COPD* but also of asthma and *IPF*. Specifically, they show that the deletion of FBLN1's variant FBLN1C in mouse models inhibits airway and lung remodeling in chronic asthma and lung fibrosis while also protects against *COPD* (Liu, *et al.*, 2016). The FBLN1C variant is known to increase the proliferation of lung fibroblasts in *COPD* and *IPF* patients (Ge, *et al.*, 2015). Regarding *IPF*, Jaffar and colleagues identify increased soluble FBLN1 levels in the serum of patients even though there is no overexpression of FBLN1's deposition in the airway tissue (Jaffar, *et al.*, 2012). Finally, Hansen *et al.* study FBLN1 with regard to myocardial fibrosis. They suggest that, since *Aortic Valve Stenosis* (*AVS*) causes cardiac fibrosis and since they observe elevated expression of the FBLN1 gene in *AVS* samples against controls, the FBLN1 must be expressed as part of the fibrotic process (Hansen, *et al.*, 2013). According to the aforementioned bibliographic evidence, FBLN1 is undeniably implicated in various fibrotic diseases. However, its expression levels in different tissues and its actual role regarding different fibrotic diseases should be further tested.

We show the 2 common-gene participation networks (over/under-expressed) in Figure 2. Table 3 presents the top-3 over- and top-3 under-expressed genes of each experiment based on their log fold change (logFC) values. In extension, Supplementary Tables 1 and 2 list all top common and unique genes (over- and under-expressed respectively), regarding the 9 fibrotic diseases.

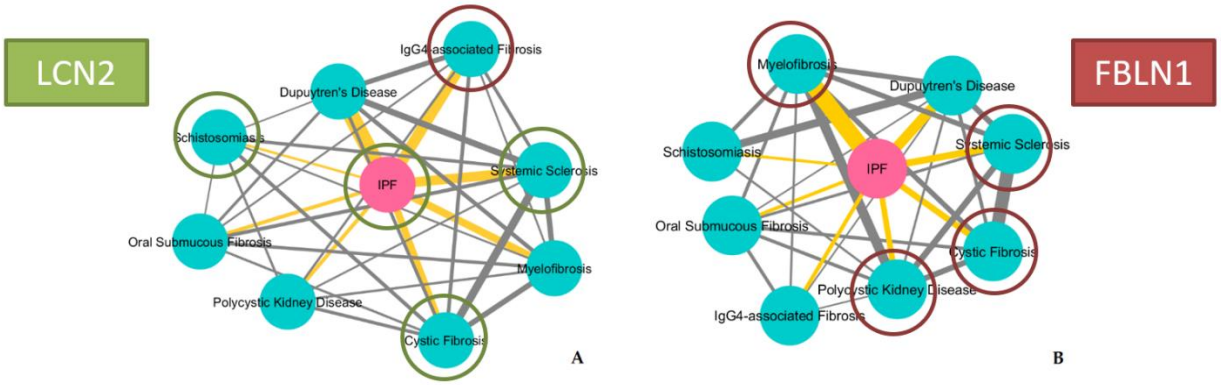

Figure 2: Common-gene participation networks. A. The network edge values are relative to the number of common over-expressed genes between pairs of fibrotic diseases. *IPF* has *35* common over-expressed genes with *Dupuytren's Disease* followed by *23* common with *IgG4-related Disease* and *20* with *SSc*. Our analyses indicate that the LCN2 gene is over-expressed in *IPF, CF, Schistosomiasis* and *SSc*. B. The network edge values are relative to the number of common under-expressed genes between pairs of fibrotic diseases. *IPF* has *28* common under-expressed genes with *Myelofibrosis* followed by *16* with *Dupuytren's Disease* and *10* with *SSc*. Our analyses indicate that the FBLN1 gene is under-expressed in *CF, Myelofibrosis, Polycystic Kidney Disease* and *SSc* but over-expressed in *IgG4-related Disease*. *Schistosomiasis* has only *3* common over- and *3* common under-expressed genes with *IPF*.

Table 3: Top-3 over- and top-3 under-expressed genes of each experiment based on their logFC values. We acquire gene symbols from GEO, from the respective platform-translation files of each experiment.

Dataset ID	Disease Name	Experiment Type	Upregulated Gene	logFC	Downregulated Gene	logFC
GSE110147	*IPF*	Microarray	MMP1	4.08	VTRNA1-1	-5.63
			SPP1	3.96	RNA5SP242	-5.22
			BPIFB1	3.85	SNORD41	-4.33
GSE53845	*IPF*	Microarray	MMP1	4.68	PNMT	-3.97
			UBD	4.22	IL1RL1	-3.40
			MMP7	4.08	DEFA3	-3.26
GSE32537	*IPF*	Microarray	BPIFB1	2.99	IL1R2	-2.81
			ASPN	2.63	MGAM	-2.17
			MMP7	2.56	FCN3	-2.14
GSE35145	*IPF*	Microarray	SFRP2	3.67	IL1R2	-4.13
			MMP7	3.55	S100A12	-3.19
			LOC649923	3.48	S100A8	-3.13
GSE99621	*IPF*	RNA-Seq	TTR	5.77	CHIAP2	-4.90
			OR51E1	5.75	WFDC12	-4.72
			GTF2H4	5.73	PNMT	-4.67
GSE134692	*IPF*	RNA-Seq	ADIPOQ	10.00	DEFA1B	-8.99
			GP1BB	7.89	OLFM4	-7.50
			CST1	6.83	DEFA4	-6.95
GSE15568	*CF*	Microarray	ALDOB	3.08	BRINP3	-2.29
			ISG20	2.39	SNRK	-1.70
			TNFRSF11B	2.10	TLR7	-1.67
GSE40445	*CF*	Microarray	MALAT1	2.60	PROS1	-2.39
			LY6D	2.59	CRISP2	-2.33
			CSF2RB	2.25	FABP6	-2.32
GSE124548	*CF*	RNA-Seq	LOC107987345	4.46	RPL10P6	-2.76
			BCORP1	4.45	RPL10P9	-1.90
			TTTY15	3.91	LOC102723846	-1.80
GSE95065	*SSc*	Microarray	COL4A4	4.68	ZBTB7C	-3.53
			SFRP4	3.16	GSN	-3.22
			CXCL13	2.97	SGCA	-3.12
GSE104174	*SSc*	RNA-Seq	MTND1P23	6.32	LINC00221	-3.65
			HBA2	6.10	DOK5	-2.74
			CRISP3	4.18	KCNJ10	-2.16
GSE124281	*Myelofibrosis*	Microarray	IFI27	3.12	LOC389342	-1.72
			RAP1GAP	2.29	LOC286444	-1.71
			IL1B	2.03	VCAN	-1.66
GSE44426	*Myelofibrosis*	Microarray	OXTR	4.33	HAND2	-7.13
			PTPRB	4.15	CD70	-6.08
			TRPC4	3.95	CHI3L2	-5.96
GSE64216	*OSF*	Microarray	MB	10.12	KRT76	-6.73
			MYLPF	10.03	LOC649923	-6.24
			CKM	9.81	IGLL3	-6.07
GSE75152	*Dupuytren's*	Microarray	THBS4	5.94	LEP	-4.65
			LOC649366	4.97	PLIN	-4.39
			TNC	4.68	PPP1R1A	-4.36
GSE7869		Microarray	UMOD	6.84	SFRP2	-5.97

Dataset ID	Disease Name	Experiment Type	Upregulated Gene	logFC	Downregulated Gene	logFC
	Polycystic Kidney		SLC7A13	6.43	CTHRC1	-5.62
			MIOX	6.41	SFRP4	-5.49
GSE40568	*IgG4-related Disease*	Microarray	ADH1B	3.93	FOS	-5.56
			SPP1	3.73	IGHM	-3.68
			HLA-DRB4	3.23	IFI6	-3.57
GSE61376	*Schistosomiasis*	Microarray	FAM100A	2.85	RPS4Y1	-3.39
			IGFBP2	2.58	GSTM2	-3.33
			PTMS	2.01	GSTM1	-2.89

Identifying Fibrotic Pathway Communities

Following, we utilize the key differentially expressed gene lists per disease, to drive our pathway analysis methodology. The PathWalks algorithm favors pathway-nodes with high betweenness centrality and strength (combination of a node's degree and the sum of its edge-weights) scores. A random walk utilizing shortest paths highlights such nodes. In Figure 3 we point out these biological pathways which are more likely to appear in the PathWalks results due to the topology of the network rather than their direct biological association to each fibrotic disease. Specifically, the depicted orange nodes belong in the top-*5*% of pathways with the highest betweenness centrality. Pink nodes belong in the top-*5*% of pathways with the highest degree. Red nodes belong in both aforementioned groups and yellow nodes depict pathways that do not belong in any of the aforementioned groups but still reach the top-*5*% of the PathWalks results with random pathway selection, due to their high functional connectivity edge values. Since we construct the pathway-to-pathway network based on functional connections, the aforementioned pathways may be sensitive to biological perturbations in most use cases.

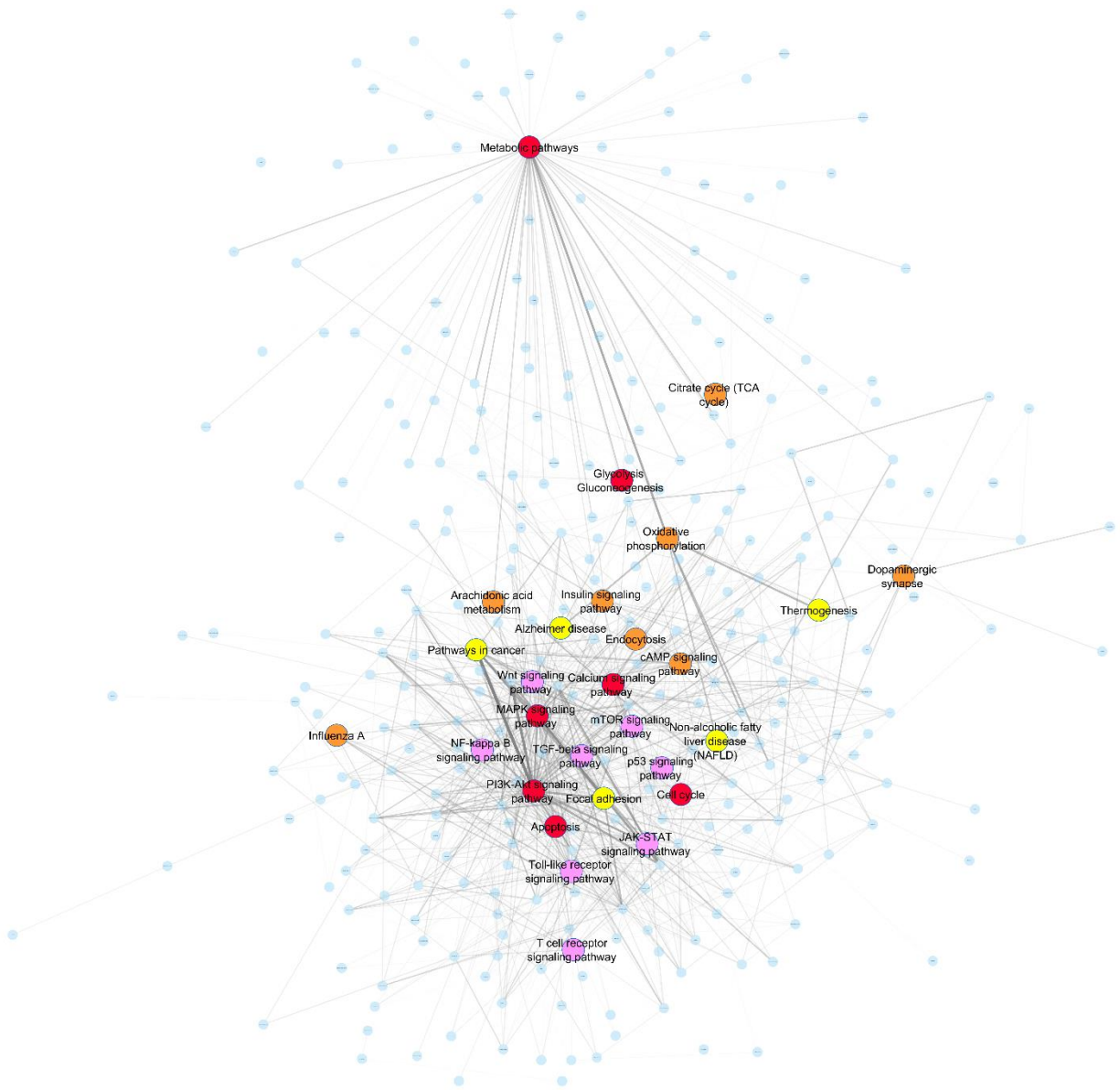

Figure 3: The base pathway-to-pathway network component that we use in the PathWalks execution. An edge indicates a functional connection between two pathways according to KEGG, and its thickness is relative to the number of their common genes. Orange nodes belong in the top-5% of pathways with the highest betweenness centrality. Pink nodes belong in the top-5% of pathways with the highest degree. Red nodes belong in both aforementioned groups. Yellow nodes depict pathways that do not belong in any of the aforementioned groups but still reach the top-5% in the results of a no-map PathWalks execution (i.e., random pathway selections) due to their high functional connectivity edge values. All of these colored nodes are more likely to appear in the results due to their topological characteristics rather than due to direct relation with the genetic information networks of each disease.

We identify key disease-specific as well as common pathways among the 9 fibrotic maladies through the PathWalks executions. In Figure 4 we highlight in green color, nodes that reach the top-5% of the PathWalks results for at least 1 disease and do not belong in the topology-favored groups described in Figure 3. We observe that not every topologically-favored node is

highlighted by PathWalks. Therefore, we believe that even the topologically-favored nodes that are highlighted by the algorithm are not necessarily false-positive entries and should be considered for further experimentation regarding fibrosis. We then focus on the key pathways of all 9 fibrotic diseases and draw convex hulls around their respective highlighted pathways in Figure 5. We observe 7 common pathways across the 9 diseases, all of which are favored by the topology. 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 use cases. The "Cell adhesion molecules" pathway is exclusive for *Myelofibrosis*, the "Antigen processing and presentation" is exclusive for *SSc* and the "IL-17 signaling" is unique for *CF*. "Fat digestion and absorption", "PPAR signaling" and "Glycerolipid metabolism" pathways are unique for *Dupuytren's Disease*. "Influenza A", "RNA transport", "Ribosome biogenesis in eukaryotes" and "Herpes simplex virus 1 infection" pathways are unique for *Schistosomiasis*. Finally, the "Dilated cardiomyopathy", "Cardiac muscle contraction", "Hypertrophic cardiomyopathy", "Adrenergic signaling in cardiomyocytes" and "ECM-receptor interaction" pathways are unique for *OSF*.

We observe that 4 out of the 5 unique pathways of *OSF* are related to cardiomyopathies. A potential connection between *OSF* and cardiomyopathies is missing from the bibliography. However, in their review, Jiang *et al.* highlight drugs from various categories such as steroids, enzymes, cardiovascular drugs, antioxidants, vitamins, and microelements that seem to ameliorate (but not cure) the fibrotic effects of *OSF* (Jiang and Hu, 2009). Specifically, they present cardiovascular drugs that were used in other studies against *OSF* including (i) pentoxifylline (Rajendran, *et al.*, 2006), (ii) buflomedil hydrochloride (Lai, *et al.*, 1995) and (iii) nylidrin hydrochloride (Sharma, *et al.*, 1987) as well as (iv) tea pigment (Li and Tang, 1998) all of which show temporary symptomatic relief. We suggest that a combinatorial treatment including a cardiovascular agent that can target the unique highlighted cardiac-related pathways of the *OSF* disease should be further pursued.

We then present a similarity network based on common pathway terms among the 9 fibrotic diseases in Figure 6. *IPF* does not have any exclusive pathway nodes in the top-*5*% of its PathWalks results. *IPF* shares the most pathways with *IgG4-related Disease* (*14*/*15*) and the fewest with *OSF* (*8*/*15*). More information regarding the unique and common highlighted pathways among the 9 fibrotic diseases can be found in Supplementary Table 3. We present all pathways that were returned in the top-*5*% of the PathWalks results for at least 1 experiment in Table 4. We further note which entries belong to the top-*5*% of pathways with highest betweenness centrality and degree scores or to the top-*5*% results of a random PathWalks execution (without the use of a gene map). We highlight pathways that were returned by the algorithm solely due to their association with the genetic information maps in green color.

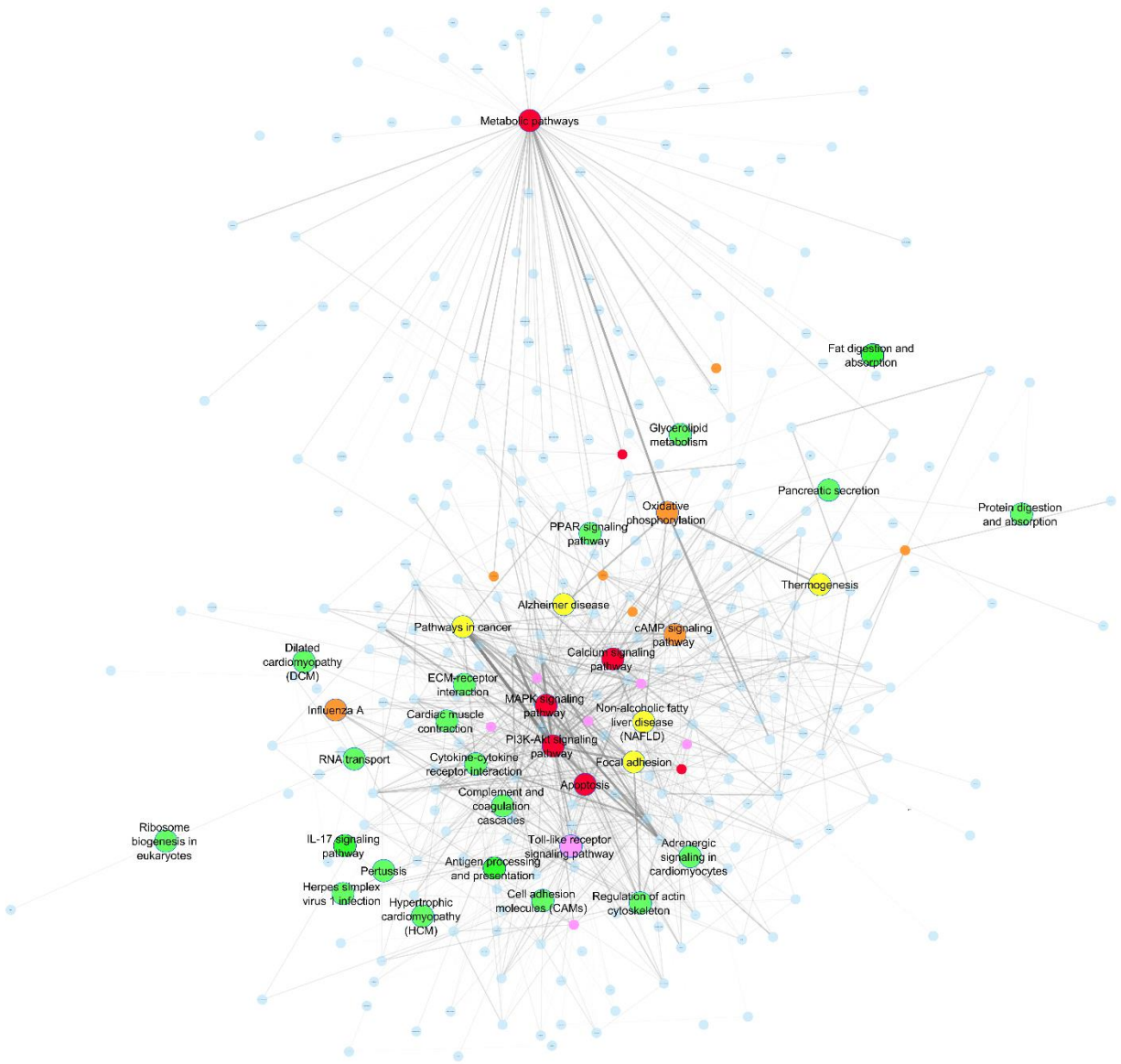

Figure 4: PathWalks results annotated upon the base pathway-to-pathway network of Figure 3. We highlight nodes that are based solely on the guidance of the gene maps in green color. We label all nodes that were returned in the top-5% of any PathWalks experiment. We observe that not every topologically-favored node is highlighted by PathWalks. Therefore, we believe that even the topologically-favored nodes that are highlighted by the algorithm are not necessarily false-positive entries and should be further explored regarding fibrosis.

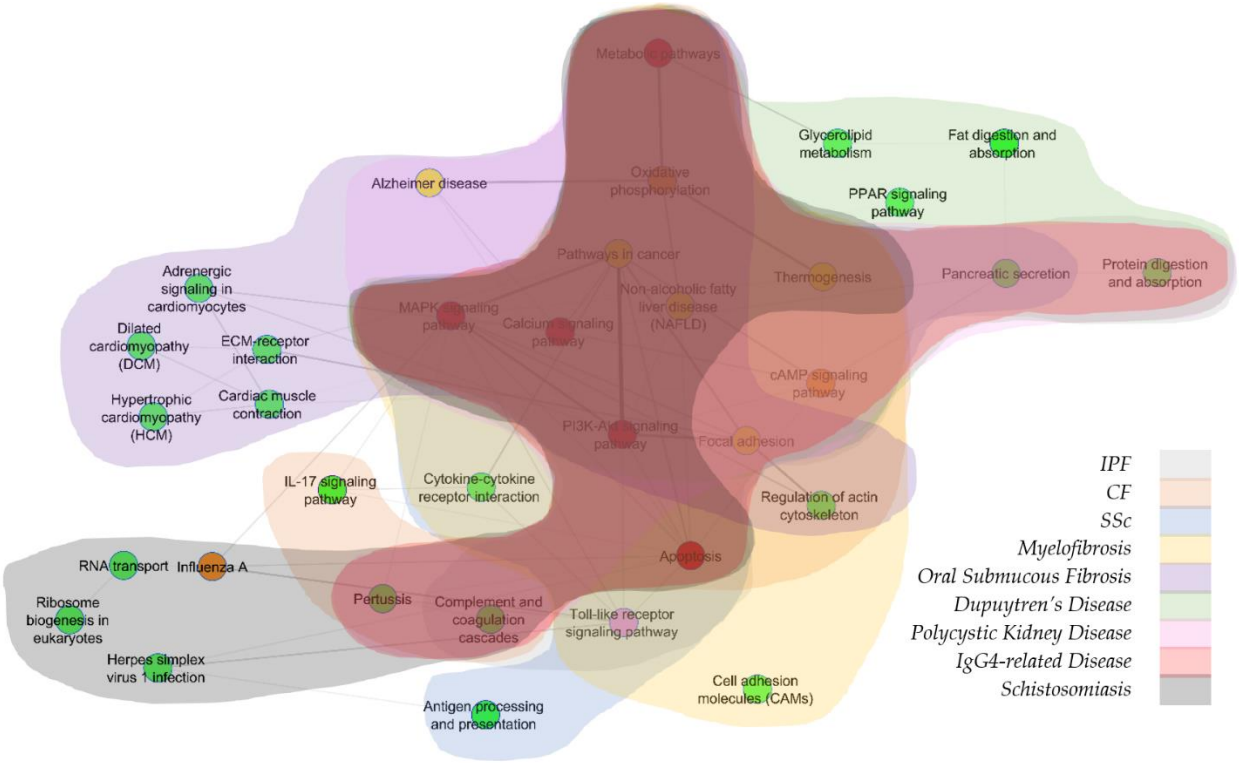

Figure 5: Convex hulls for the 9 fibrotic diseases highlighted pathways from PathWalks. We observe that 7 pathways are common for all diseases including "Metabolic pathways", "Pathways in cancer", "MAPK signaling pathway", "PI3K-Akt signaling pathway", "Non-alcoholic fatty liver disease", "Oxidative phosphorylation" and "Calcium signaling pathway". We also identify unique highlighted pathways for some of the diseases. The "Cell adhesion molecules" pathway which belongs to the yellow convex hull is exclusive for *Myelofibrosis*. The "Antigen processing and presentation" pathway is included in the blue convex hull and is unique for *SSc*. The "IL-17 signaling" pathway is included in the orange convex hull and is unique for *CF*. "Fat digestion and absorption", "PPAR signaling" and "Glycerolipid metabolism" pathways are unique for *Dupuytren's Disease* and belong to the green convex hull. "Influenza A", "RNA transport", "Ribosome biogenesis in eukaryotes" and "Herpes simplex virus 1 infection" pathways belong to the dark gray convex hull and are unique for *Schistosomiasis*. Finally, the purple convex hull contains the "Dilated cardiomyopathy", "Cardiac muscle contraction", "Hypertrophic cardiomyopathy", "Adrenergic signaling in cardiomyocytes" and "ECM-receptor interaction" pathways which are unique for *OSF*. We observe that 4 out of the 5 unique pathways of *OSF* are related to cardiomyopathies hinting for a novel approach of potential treatments since this connection is missing from the bibliography.

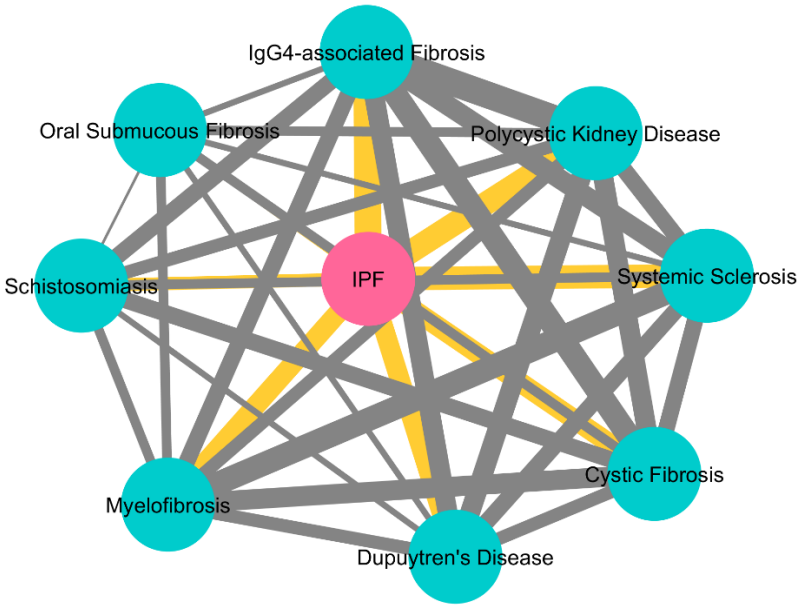

Figure 6: Similarity network among the 9 fibrotic diseases based on their common pathways as indicated by the PathWalks algorithm. *IPF* does not have any unique pathway nodes in the top-5% of its PathWalks results. *IPF* has the most common pathways with *IgG4-related Disease* (14/15) and the fewest with *OSF* (8/15). The edge thickness is relative to the number of common pathways between 2 diseases.

Table 4: Participation table of pathways in the top-5% of betweenness centrality, degree and random PathWalks (without gene map) result nodes. The 1st column of the table contains all pathways that reached the top-5% in at least 1 PathWalks experiment. We draw a check mark for each pathway in the top-5% of betweenness centrality, degree and random PathWalks lists respectively. We paint pathways that are highlighted by the algorithm solely due to their relation with each genetic information map in green.

Top 5% Pathways' Table			
Aggregated List of Top Pathways	**Betweenness Centrality**	**Degree**	**Random PathWalks**
Adrenergic signaling in cardiomyocytes			
Alzheimer disease			✓
Antigen processing and presentation			
Apoptosis	✓	✓	✓
Calcium signaling pathway	✓	✓	✓
cAMP signaling pathway	✓		✓
Cardiac muscle contraction			
Cell adhesion molecules			
Complement and coagulation cascades			
Cytokine-cytokine receptor interaction			
Dilated cardiomyopathy			
ECM-receptor interaction			
Fat digestion and absorption			
Focal adhesion			✓
Glycerolipid metabolism			
Herpes simplex virus 1 infection			
Hypertrophic cardiomyopathy			
IL-17 signaling pathway			
Influenza A	✓		✓
MAPK signaling pathway	✓	✓	✓
Metabolic pathways	✓	✓	✓

Non-alcoholic fatty liver disease			✓
Oxidative phosphorylation	✓		✓
Pancreatic secretion			
Pathways in cancer			✓
Pertussis			
PI3K-Akt signaling pathway	✓	✓	✓
PPAR signaling pathway			
Protein digestion and absorption			
Regulation of actin cytoskeleton			
Ribosome biogenesis in eukaryotes			
RNA transport			
Thermogenesis			✓
Toll-like receptor signaling pathway		✓	✓

Investigating Inhibiting Substances

Based on the methodology described in the Drug Repurposing paragraph of the Materials and Methods section we collect repurposed drug candidates for the 9 respective fibrotic diseases through the CMap and L1000CDS[2] DR tools. We then use our CoDReS re-ranking tool to extract the highest-scoring, structural-representative and inhibiting substances among the repurposed drugs of each disease. For the use case of *IgG4-related Disease* we arbitrarily choose the top-*5* ranked drugs from each DR tool since there is no entry for this disease in CoDReS. We present the re-ranked drug candidates for each disease in Table 6.

Table 5: The re-ranked repurposed drugs for the respective fibrotic diseases. The 1st column contains the disease names while the 2nd column contains the names of the respective maladies in the database of CoDReS. The 3rd column indicates the highest-ranked structural representative inhibitors returned by CoDReS and the 4th column indicates their number. *IgG4-related Disease* is an exception since there is no matching entry in CoDReS. For this use case we present the top-5 ranked drugs from each DR tool (CMap and L1000CDS[2]).

Disease Name	CoDReS Disease Name	Top-Ranked Structural Representatives	#Top Ranked Drugs
IPF	*Idiopathic Pulmonary Fibrosis*	bi-2536, boldine, brazilin, brompheniramine maleate, budesonide, buparlisib, butein, celecoxib, ciglitazone, cinchonine, disulfiram, entinostat, epinephrine, hydrocortisone, icosapent, ipriflavone, memantine, methimazole, mk-2206, pd-0325901, propofol, propranolol, rizatriptan, scriptaid, sorafenib, sunitinib, testosterone, topotecan	28
CF	*Cystic Fibrosis*	acepromazine, apigenin, damnacanthal, delcorine, didanosine, edaravone, entinostat, fasudil, gefitinib, hydrocortisone, mercaptopurine, nafcillin, navitoclax, nilotinib, orantinib, oxybutynin, propofol, quipazine, sb-202190, staurosporine, terbutaline, trametinib, wortmannin, y-27632, ziprasidone	25
SSc	*Systemic Scleroderma*	erlotinib, gsk-1059615, hydrocortisone, l-690330, orphenadrine, radicicol, sb-431542	7
Myelofibrosis	*Myelofibrosis due to another disorder*	alaproclate, anisomycin, brefeldin a, dasatinib, emetine, flurandrenolide, fludroxycortide, fluticasone, l-690330, narciclasine, nortriptyline, pd-0325901, pf-477736, procyclidine, regorafenib, reversine, ruxolitinib, sulpiride, tranylcypromine, trifluoperazine, xanthohumol, zalcitabine	22
OSF	*Oral Submucous Fibrosis*	cefotiam, cephaeline, dephostatin, digoxin, enzastaurin, decafluorobutane, xanthohumol	7
Dupuytren's Disease	*Dupuytren's Disease*	bi-2536, dasatinib, dicycloverine, fostamatinib, ibuprofen, l-690330, memantine, pha-665752, radicicol, triamcinolone, xylazine	11

Polycystic Kidney Disease	*Polycystic Kidney Diseases*	bosutinib, ipsapirone, loreclezole, picotamide, selumetinib, serdemetan, tretinoin, wortmannin	8
IgG4-related Disease	-	BRD-K92317137, brefeldin-a, cerulenin, dexketoprofen, fenpiverinium, gestrinone, importazole, indirubin, parthenolide, rhodomyrtoxin	10
Schistosomiasis	*Schistosomiasis*	1-benzylimidazole, canertinib, captopril, chloroquine, dapsone, dexamethasone, genipin, ibuprofen, iodophenpropit, levocabastine, mdl-28170, memantine, nadolol, parthenolide, phenytoin, propantheline, resorcinol, scriptaid, sunitinib, trametinib, triamterene, vemurafenib	22

There are 2 drug candidates returned by CoDReS that are common among 3 of the 9 fibrotic diseases. Hydrocortisone is highlighted 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*. Hydrocortisone is a corticosteroid and its mode of action against fibrotic diseases is being studied for more than half a century now. It is mostly used for the temporal treatment of *OSF.* More specifically, an old study of DeSa and colleagues reports relief of symptoms in *OSF* patients by injecting the submucosal fibrotic areas with specific dosages of hydrocortisone depending on the stage of the fibrosis (DeSa, 1957). In a more recent study, Singh et al. describe how they have been studying and clinically using optimal doses of a combination of hydrocortisone acetate and hyaluronidase for the medical treatment of *OSF* for 20 years (Singh, *et al.*, 2010).

Hydrocortisone's action has also been studied against *Pulmonary Fibrosis* (*PF*). Xu and others show that hydrocortisone decreases overexpressed fibrotic factors such as TGF-β1, pulmonary type I collagen and inducible nitric oxide synthetase to normal levels in a rat model (Xu, *et al.*, 2007). Regarding *CF*, the study of Tepper *et al.* suggests that the application of intravenous hydrocortisone to the standard treatment of hospitalized infants with *CF*, produces a greater or a more sustained improvement in their lung function (Tepper, et al., 1997). In addition, another old study from L. Zachariae and F. Zachariae demonstrates the effects of hydrocortisone on various fibrotic skin maladies. Specifically, regarding *Dupuytren's Disease* they observe that the institution of hydrocortisone acetate at a mild stage keeps the tissue contracture stationary (Zachariæ and Zachariae, 1957). Regarding *IgG4-related Disease*, Tanabe *et al.* propose a combination of hydrocortisone and thyroxine which result in clinical and laboratorial improvements against the disease (Tanabe, *et al.*, 2006). Finally, Mitani *et al.* study the effects of hydrocortisone in a UV-irradiated mouse model. The UV irradiation causes the appearance of fibroblasts and the accumulation of collagen in the model's lower dermis. They show that hydrocortisone prevents both the fibrosis of lower dermis and the accumulation of ECM components (Mitani, *et al.*, 1999).

Memantine is a low-affinity voltage-dependent uncompetitive antagonist of the N-methyl-D-aspartate receptors found in nerve cells and is used for the treatment of patients with moderate to severe *Alzheimer's Disease*. A direct connection of memantine and fibrotic diseases does not exist in bibliography. However, through a histopathological analysis, Abbaszadeh *et al.* show that a pre-treatment of a heart failure rat model with memantine attenuates myocyte necrosis and fibrosis (Abbaszadeh, *et al.*, 2018). Similarly, Li *et al.* demonstrate that memantine attenuates lung

inflammation in a bleomycin-induced acute lung injury mouse model (Li, *et al.*, 2015). They propose that further studies should be carried out to identify memantine's mode of action regarding both the prevention as well as the treatment of lung fibrosis.

Based on the common re-ranked drug candidates from our analysis we draw another disease similarity network in Figure 8. The edge thickness is relative to the number of common re-ranked drugs between two diseases. We calculate the significance of a disease sharing re-ranked drug candidates with *IPF*. We utilized R's hypergeometric test function *dhyper(x, m, n, k)* where (i) x represents the number of shared CoDReS re-ranked drug candidates between *IPF* and another fibrotic disease, (ii) m represents the number of the intersection between the *IPF* CoDReS drugs (*28*) and the total, unfiltered drug entries that are returned from CMap and L1000CDS[2] across all the experiments of a disease, (iii) n represents the number of all the unfiltered repurposed drugs for the disease, excluding the shared ones with *IPF* and (iv) k represents the number of the CoDReS re-ranked drugs for the disease. For example, *CF* shares *3* (x) CoDReS re-ranked drug candidates with *IPF*, where *26* (m) out of the *28* re-ranked *IPF* drugs exist in the unfiltered outputs from the two DR tools for *CF*, *2448* (n) exist in the *CF* unfiltered DR outputs and were not highlighted for *IPF* through CoDReS and *25* (k) is the number of CoDReS-highlighted *CF* drug candidates. Therefore, the fact that *CF* shares 3, potentially anti-fibrotic, drug candidates with *IPF* is statistically significant ($p\text{-}value \simeq .002$). Based on this test, sharing 2 or 3 re-ranked drug candidates with *IPF* is a statistically significant ($p\text{-}value < .05$) similarity while sharing only 1 drug may happen due to randomness. Detailed information regarding the common and unique highlighted drug candidates among the 9 fibrotic diseases can be found in Supplementary Table 4.

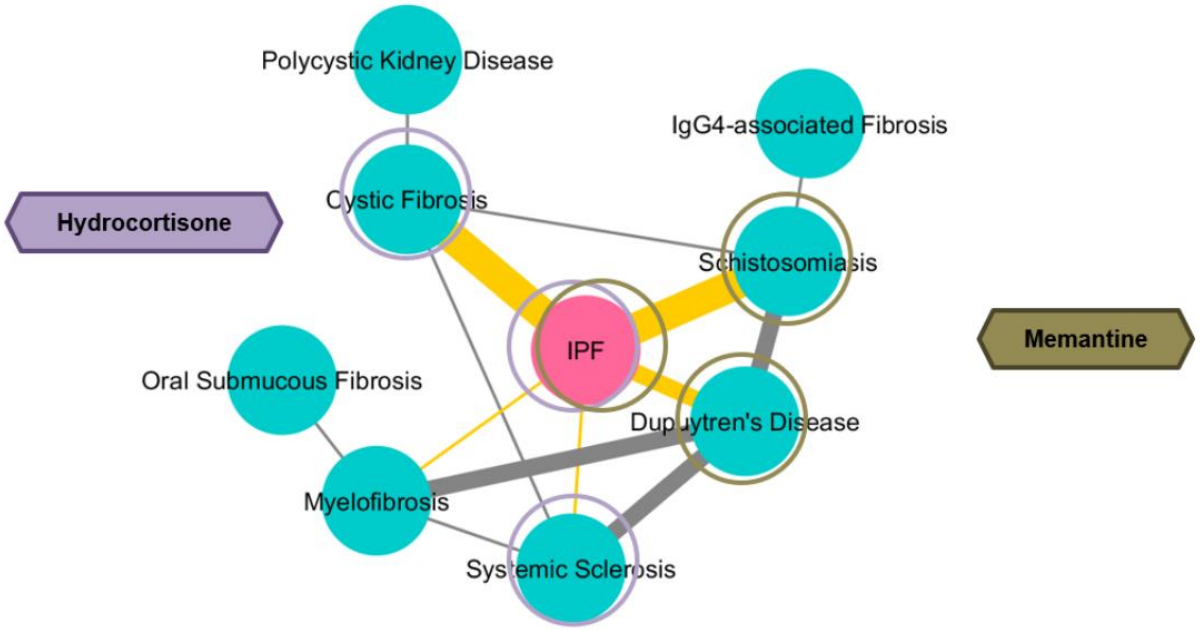

Figure 7: Disease similarity network based on the common re-ranked drug candidates. The edge thickness is relative to the number of common genes between 2 diseases. *IPF* shares 3 drugs with *CF* and *Schistosomiasis* respectively, 2 with *Dupuytren's Disease* and 1 with *SSc* and *Myelofibrosis* respectively. The fact that *IPF* shares 2 or 3 drugs with each of *CF*, *Schistosomiasis* and *Dupuytren's Disease* is statistically significant based on a hypergeometric test. *SSc* and *Myelofibrosis* each share only 1 re-ranked drug with *IPF*, an event that may occur due to randomness. Hydrocortisone is highlighted by the re-ranking procedure for *IPF*, *CF* and *SSc* while 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. The repoDB database contains entries with substances and their clinical trial status against the respective diseases. The 4 different drug statuses annotated in the database are "Approved", "Suspended", "Terminated" and "Withdrawn". We focus on drugs that have previously failed in clinical trials against any of the 9 fibrotic diseases of our study. We consider that drugs with "Suspended", "Terminated" or "Withdrawn" indications have failed in the respective clinical trial. We present the respective names of the 9 fibrotic diseases from repoDB in Table 7. More detailed information regarding drug names, respective fibrotic disease and clinical trial indication combinations can be found in Supplementary Table 5. We note that Alemtuzumab (terminated entry against *SSc*) does not have a chemical structure file in PubChem (Kim, *et al.*, 2016), so we focus our analysis on the *19*-rest, unique "failed" substances.

Table 6: The respective fibrotic disease names in repoDB. More than 1 name is assigned to some of the diseases. There are no matching entries for *OSF* and *IgG4-related Disease* in the database.

Disease Name	repoDB Disease Name
IPF	*Pulmonary Fibrosis*
CF	*Cystic Fibrosis,* *Pulmonary Cystic Fibrosis*

SSc	*Scleroderma,* *Systemic Scleroderma*
Myelofibrosis	*Myelofibrosis,* *Myelofibrosis due to another disorder*
OSF	-
Dupuytren's Disease	*Dupuytren's disease of finger, with contracture*
Polycystic Kidney Disease	*Polycystic Kidney Diseases,* *Polycystic Kidney, Autosomal Dominant*
IgG4-related Disease	-
Schistosomiasis	*Schistosomiasis*

We download the 2-dimensional structural data files (SDF) for the aforementioned *19* failed drugs from PubChem and compile an aggregated SDF file. We then compile another SDF file containing the structures of all unique *121* CoDReS re-ranked drug candidates from this study. We use these 2 SDF files as input in the ChemBioServer "*Attach similar-only nodes to Network*" function. We upload the SDF file of re-ranked drugs as a base network with an edge similarity threshold of *0.5* and the SDF file of failed drugs as a secondary list to infer structural similarity links towards the base network with an edge threshold of *0.7*. We choose the Tanimoto similarity metric and after the execution we export the network created by ChemBioServer. We then remove drug nodes that are dissimilar to the failed drugs and visualize the remaining network in Cytoscape, producing Figure 9. We color the re-ranked drugs gray and the failed drugs green. The edge thickness is relative to the structural similarity between 2 compounds.

We observe that hydrocortisone has high structural similarity with the "failed" drugs iloprost, mycophenolic acid and simvastatin. We suggest giving lower testing priority to the substances depicted in Figure 9 due to their high structural similarity with drugs that have already failed in clinical trials regarding fibrotic diseases. Nevertheless, we do not reject the possibility that even these candidates may act as fibrotic inhibitors in conjunction with other drugs or at different doses and/or stages of fibrosis.

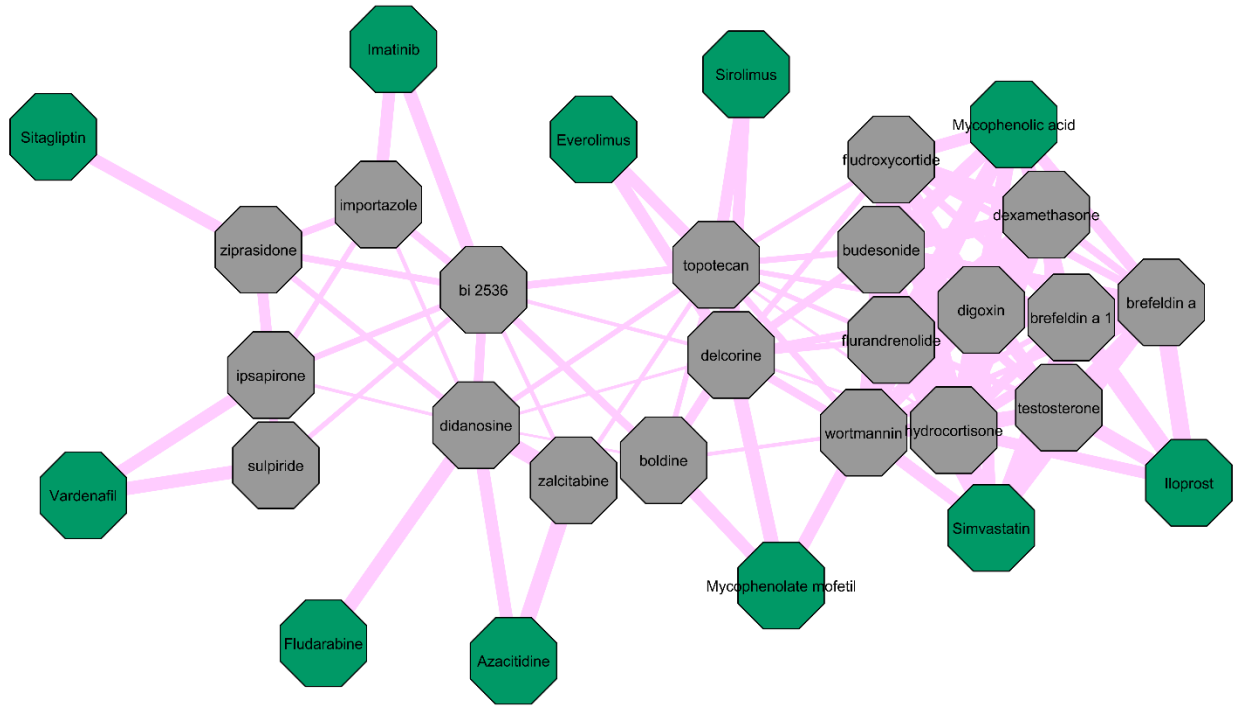

Figure 8: Structural similarity network of the CoDReS re-ranked drug candidates and drugs that have previously failed in clinical trials against fibrotic diseases. Re-ranked drugs that are dissimilar to the "failed" drugs have been removed from the network. We color the re-ranked drugs gray and the "failed" drugs green. The edge size is relative to the structural similarity between 2 compounds. Hydrocortisone, which has been re-ranked for 3 out of the 9 fibrotic diseases, has high structural similarity with iloprost, mycophenolic acid and simvastatin. We suggest giving lower testing priority to the depicted re-ranked substances due to their high similarity with already-failed drugs in clinical trials regarding fibrosis. Nevertheless, we do not reject the possibility that hydrocortisone or any other drug candidate depicted in this figure may act as a fibrotic inhibitor in conjunction with other drugs or at different doses and/or stages of fibrosis.

Highlighting Anti-fibrotic Drug Candidates

Through our analysis pipeline, we conclude in *121* total repurposed and re-ranked drug candidates with potential anti-fibrotic action, *89* of which target at least 1 gene participating in at least 1 of the PathWalks highlighted pathways for the 9 fibrotic diseases. We focus on the drugs that target the *IPF* highlighted pathways that are directly associated with the respective synthetic gene map. Specifically, these 3 pathways are "Pancreatic secretion", "Protein digestion and absorption" and "Complement and coagulation cascades" which we paint green in Figure 10. Figure 10 zooms in the communities of the top-*5*% PathWalks-highlighted pathways for the *IPF* use case and their functional connections. The edge weights are relative to the number of times they were traversed during the PathWalks algorithm. The "Pancreatic secretion" and "Protein digestion and absorption" pathways are highlighted in the use cases of *IPF*, *Dupuytren's Disease*, *Polycystic Kidney Disease* and *IgG4-related Disease*. The "Pancreatic Secretion" pathway is also highlighted in *SSc*. The "Complement and coagulation cascades" pathway is highlighted in the use cases of *IPF*, *CF*, *IgG4-related Disease*, *Polycystic Kidney Disease* and *Schistosomiasis*.

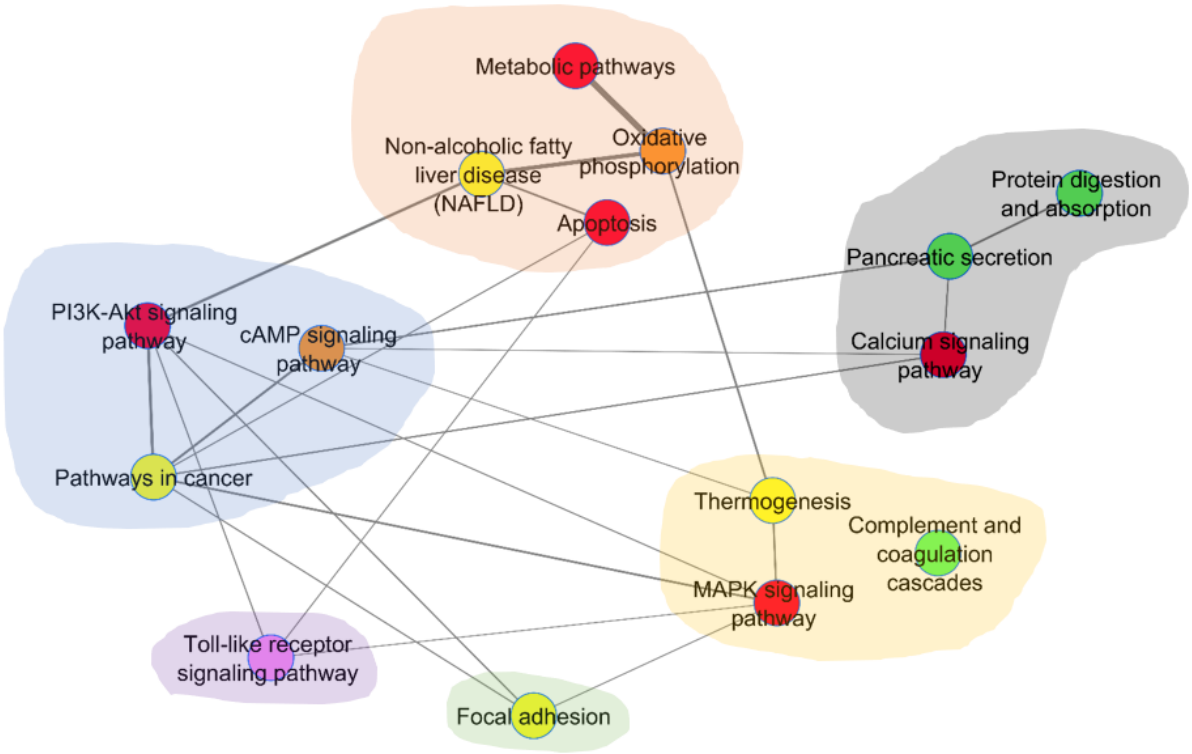

Figure 9: Pathway communities of the top-5% results for *IPF* as calculated by PathWalks. "Pancreatic secretion", "Protein digestion and absorption" and "Complement and coagulation cascades" are the 3 pathways that are highlighted solely due to their direct association with the genetic information network of *IPF* and are colored in green. The edge width is relative to the number of times an edge was traversed during the *IPF* PathWalks execution. The "Pancreatic secretion" and "Protein digestion and absorption" pathways are highlighted in the use cases of *IPF, Dupuytren's Disease, Polycystic Kidney Disease, IgG4-related Disease*. The "Pancreatic Secretion" pathway is also highlighted in *SSc*. The "Complement and coagulation cascades" pathway is highlighted in the use cases of *IPF, CF, IgG4-related Disease, Polycystic Kidney Disease and Schistosomiasis.*

There are *34* out of the *121* total re-ranked drugs that target at least 1 gene participating in the 3 "green" pathways of *IPF* (Figure 10). We present all the combinations of re-ranked drugs, gene targets and biological pathways, where a drug has at least 1 gene target in the highlighted pathways, in Supplementary Table 6. We bibliographically explore the 7 candidates that target at least 2 out of the 3 'green' highlighted pathways of Figure 10 for potential anti-fibrotic mode of action, namely celecoxib, digoxin, captopril, ibuprofen, staurosporine, nafcillin, wortmannin.

Celecoxib is the only candidate targeting all 3 "green" pathways of *IPF*. Specifically, celecoxib is an anti-inflammatory drug that we identify in the CoDReS results for *IPF*. Celecoxib targets the CA2 and COL18A1 genes in "Pancreatic secretion" and "Protein digestion and absorption" pathways respectively and PLAU in "Complement and coagulation cascades". We identify PLAU being over-expressed in the *Dupuytren's Disease* experiments. The anti-fibrotic effects of celecoxib have been widely studied, especially in rat models and mainly regarding liver fibrosis. Many such studies (Chávez, *et al.*, 2010; Ftahy, *et al.*, 2013; Gao, *et al.*, 2013; Paik, *et al.*, 2009; Wen, *et al.*, 2014) highlight the anti-fibrotic potential of celecoxib against hepatic fibrosis. On the other hand,

contradictory studies (Hui, *et al.*, 2006; Liu, *et al.*, 2009) demonstrate pro-fibrotic effects of celecoxib and their results show significant exacerbated liver fibrosis and hepatocellular necrosis in celecoxib-treated rats. However, Yu *et al.* show that celecoxib protects against liver inflammation in early stages of fibrosis even though it has no effect on the actual fibrosis (Yu, *et al.*, 2009). Similarly, Harris *et al.* observe that celecoxib treatment has no effect in the liver fibrosis of mice (Harris, *et al.*, 2018). In a cardiac hypertrophy study, Zhang and colleagues demonstrate that celecoxib treatment prevents collagen accumulation and suppresses CTGF expression in rats (Zhang, *et al.*, 2016). Collagen accumulation and overexpression of CTGF are well-known pro-fibrotic mechanisms. However, because of the contradicting studies, we suggest that other drug candidates from our study should be prioritized instead of celecoxib for further experimentation against fibrosis.

Digoxin is mainly used for the treatment of atrial fibrillation and congestive heart failure (Hallberg, *et al.*, 2007). We identify digoxin as a repurposed drug in the use case of *OSF*. As we mention in the pathway community analysis paragraph, there are 4 unique cardiomyopathy-related pathways involved in *OSF*, a finding that needs further experimental validation. Digoxin targets the "Pancreatic secretion" and "Protein digestion and absorption" through 8 genes; ATP1A1, ATP1A2, ATP1A3, ATP1A4, ATP1B1, ATP1B2, ATP1B3 and FXYD2. ATP1A2 is over-expressed in the use case of *OSF* and under-expressed in *Dupuytren's Disease* while ATP1B1 is over-expressed in *Dupuytren's* and FXYD2 over-expressed in *Polycystic Kidney Disease*. Peckham and colleagues study the effects of digoxin on nasal potential difference in *CF* subjects with perturbed airway epithelium. Their results indicate that digoxin is unlikely to treat *CF* (Peckham, *et al.*, 1995). On the other hand, Moss *et al.* study the effects of digoxin in *CF* patients with heart failure complications. Their results show indications of cardiac process reversion even though it is not certain which treatment among respiratory care, vigorous antibiotic therapy or digoxin, attributed most to the survival (Moss, *et al.*, 1975). Contrary to the previous reports, the study of Coates *et al.* concludes that digoxin neither increases exercise capacity or improves exercising cardiac function in *CF* patients with moderate to severe airway obstruction (Coates, *et al.*, 1982).

Disregarding *CF*, Haller *et al.* observe remarkable clearing of myocardial fibrosis by an affinity-purified digoxin antibody in rats with *Chronic Renal Failure* (Haller, *et al.*, 2012). We observe that digoxin has high structural similarity with the "failed" drugs simvastatin and mycophenolic acid as well as hydrocortisone, as shown in Figure 9. Nevertheless, we suggest that digoxin should be further tested specifically against *OSF* due to the results of our study hinting towards a link between *OSF* and heart diseases.

Captopril is used for the treatment of hypertension and some types of congestive heart failure (Vidt, et al., 1982) and is highlighted from the re-ranking process of our analysis in the use case of *Schistosomiasis*. Captopril targets the F2 and SERPINE1 genes in the "Complement and coagulation cascades" pathway and DPP4 in "Protein digestion and absorption". We observe SERPINE1 being over-expressed in *Myelofibrosis* and *SSc* but under-expressed in *IPF*. There are various studies highlighting captopril's anti-fibrotic mode of action in experimental rat models.

Specifically, regarding lung fibrosis, Ghazi-Khansari's histopathological findings show that captopril improves pulmonary fibrosis in a paraquat-induced lung fibrosis rat model (Ghazi-Khansari, et al., 2007) while Baybutt's study shows that captopril limits fibrosis and interstitial pneumonia in a monocrotaline-induced lung fibrosis rat model (Baybutt, et al., 2007). Jalil and colleagues' study shows attenuation of both interstitial and perivascular fibrosis of myocardium in a renovascular hypertension rat model (Jalil, et al., 1991). Notable reduction of fibrosis has also been observed in rat studies regarding kidney (Cohen, et al., 1996), liver (Ramos, et al., 1994) and colon fibrosis (Wengrower, et al., 2004). On the other hand, there are fewer studies with contradictory results about captopril's action against fibrosis. Tuncer and others do not observe any inhibitory effects of captopril in a rat model with liver fibrosis (Tuncer, et al., 2003) while Okada's team note that captopril significantly enhances myocardial fibrosis in an isoprenaline-induced ventricular fibrosis rat model (Okada, et al., 2010). However, due to the various studies supporting captopril's anti-fibrotic action as well as the indications from our study, we suggest that captopril should be considered for further examination against fibrotic diseases and especially *IPF* due to the various references for its capacity to improve lung fibrosis.

Ibuprofen is an anti-inflammatory agent that is highlighted in the use cases of *Dupuytren's Disease* and *Schistosomiasis*. Ibuprofen targets the PLAT and THBD genes in the "Complement and coagulation cascades" pathway and CFTR in the "Pancreatic secretion" pathway. PLAT is over-expressed in the use case of *SSc*. Carlile *et al.* demonstrate that ibuprofen is a *CF* transmembrane conductance regulator (CFTR) protein corrector and suggest that ibuprofen may be suitable in a *CF* combination therapy (Carlile, *et al.*, 2015). Additionally, Lindstrom *et al.* suggest that high doses of ibuprofen administered based on weight can treat polyposis in children with *CF* (Lindstrom, *et al*., 2007). There are 3 more studies (Konstan, *et al*., 1995; Konstan, *et al*., 2007; Lands, *et al*., 2007) in agreement that high doses of ibuprofen slow the progression of lung disease in *CF* without serious adverse effects. Therefore, ibuprofen should be further experimentally pursued in drug combinations against fibrosis.

Staurosporine targets the PLA2G1B gene in the "Pancreatic secretion" pathway as well as COL1A2 and SLC1A1 in "Protein digestion and absorption". We identify SLC1A1 being over-expressed in *Polycystic Kidney Disease* in our analysis. COL1A2 is under-expressed in *Polycystic Kidney Disease* but over-expressed in the use cases of *IPF*, *IgG4-related* and *Dupuytren's Disease*. We observe that the PLA2G1B gene is under-expressed in 3 out of the 6 total *IPF* experiments and over-expressed in 1 of them. Staurosporine is returned as a repurposed drug for *CF*. However, due to its non-specific selectivity on kinase-binding, staurosporine has a high chance of undesirable side effects (Karaman, *et al.*, 2008; Tanramluk, *et al.*, 2009). Moreover, the study of Lindroos *et al.* shows that staurosporine upregulates PDGFR-α gene expression and its protein levels in pulmonary myofibroblasts of rats (Lindroos, *et al.*, 2001). The overexpression of PDGFR-α is known to induce myofibroblast hyperplasia during pulmonary fibrosis. Therefore, we consider that staurosporine is not a good anti-fibrotic candidate.

We identify nafcillin, which is an antibiotic drug, as a repurposed and re-ranked candidate in the use case of *CF*. Nafcillin targets the F2RL3 gene in the "Complement and coagulation cascades" pathway and PGA5 in "Protein digestion and absorption". F2RL3 is under-expressed in *IPF* experiments. We suggest that nafcillin's mode of action against fibrosis should be explored via cell line and model experiments since there are currently no relevant studies.

Finally, wortmannin is a phosphoinositide 3-kinase inhibitor (Powis, *et al.*, 1994) and is highlighted by CoDReS for the use cases of *CF* and *Polycystic Kidney Disease*. Wortmannin targets the F2R and SLC1A1 genes in the "Complement and coagulation cascades" and "Protein digestion and absorption" pathways respectively. There are a few studies around *SSc* that hint at wortmannin's potential anti-fibrotic mode of action. Shi-Wen *et al.* demonstrate that wortmannin greatly reduces *SSc* patients' fibroblast ability to contract a collagen gel matrix (Shi-Wen, *et al.*, 2004). In a later study, they add that wortmannin blocks the ability of rac1 protein to increase the expression of other profibrotic proteins (Shi-Wen, *et al.*, 2009). Parapuram and colleagues also highlight wortmannin's capacity to reduce the expression of collagen type I, SMA and CCN2 in dermal fibroblasts from PTEN knockout mice (Parapuram, *et al.*, 2011). The study of Zhi *et al.* describes that wortmannin significantly decreases the expression of collagen type I in adult rat cardiac fibroblasts (Zhi, *et al.*, 2013). On the other hand, Wang *et al.* show that wortmannin cancels the effects of cannabinoid receptor type 2 agonist including improving cardiac function, decreasing fibrosis and reducing inflammatory cytokine secretion and oxidative stress in a mouse model with infarcted heart (Wang, *et al.*, 2014). Lastly, Zhang and others also identify that wortmannin abolishes zinc's anti-fibrotic capacity in a rat kidney tubular epithelial cell line (Zhang, *et al.*, 2016). We observe that wortmannin has high structural similarity with the "failed" drugs simvastatin, mycophenolic acid and mycophenolate mofetil, as well as with the already discussed repurposed drugs of our study hydrocortisone and digoxin, as shown in Figure 9. Due to the contradictory results from the bibliography and wortmannin's structural similarity with already "failed" drugs against fibrosis, we do not suggest prioritizing wortmannin in future anti-fibrotic experiments.

Discussion

In this study, we identify genes, biological pathways and potential inhibiting drugs that are associated with fibrotic diseases. We observe the LCN2 gene being differentially over-expressed in 4 out of the 9 studied fibrotic diseases. Bibliographic evidence shows correlation between LCN2 and fibrosis in *SSc*, *Chronic Hepatitis C* and *Chronic Kidney Disease*. However, in other diseases such as NAFLD and cases of liver injury, LCN2 cannot be correlated to the observed fibrosis. More experimentation is needed to better understand the role of LCN2 in fibrotic implications. We also identify FBLN1 being under-expressed in 4 out of the 9 fibrotic diseases and over-expressed in another. According to the bibliography, even though FBLN1's mRNA levels may decrease in *COPD* where small airway fibrosis occurs, on the protein level it is accumulated in the ECM. FBLN1 levels are increased in serum and bronchoalveolar lavage fluid of asthma

patients (Lau, *et al.*, 2010) and in the plasma and lung tissue of *IPF* patients (Jaffar, *et al.*, 2014). Based on our results 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.

Through our PathWalks methodology we identify 7 common biological pathways that are associated with all 9 fibrotic diseases of this study. These include "Metabolic", "Cancer", "MAPK signaling", "PI3K-Akt signaling", "Non-alcoholic fatty liver disease", "Oxidative phosphorylation" and "Calcium signaling" pathways. We do not identify any exclusive fibrotic mechanisms for *IPF* but each of the *Myelofibrosis*, *CF*, *SSc*, *Dupuytren's Disease*, *Schistosomiasis* and *OSF* diseases have unique pathway entries that should be further explored. We observe that the PathWalks results hint evidence for a potential common mechanism between *OSF* and cardiomyopathies. We suggest that combinatorial regimens against *OSF* including cardiovascular agents should be further investigated.

We then use 2 DR tools (CMap and L1000CDS^2) and our drug re-ranking tool (CoDReS) to identify and propose drug candidates with potential anti-fibrotic mode of action. We observe that hydrocortisone and memantine are returned in 3 out of the 9 fibrotic diseases of this study. Memantine's action has not been studied against fibrotic diseases but, as mentioned in the respective Results section, there are hints for its potential anti-fibrotic action that suggest further experimentation (Abbaszadeh, *et al.*, 2018; Li, *et al.*, 2015). As far as hydrocortisone is concerned, low doses have shown to attenuate fibrosis especially in the early stages (DeSa, 1957; Xu, et al., 2007). However, we observe that 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 may be effective in combination with other drugs or at different dosages and/or stages of fibrosis.

We further screen the drug results to present the most promising anti-fibrotic drug candidates. We select substances that target at least 2 out of the 3 PathWalks highlighted biological pathways of *IPF* that are directly associated with the respective synthetic gene map. We explore bibliographic evidence for pro- or anti-fibrotic mode of action regarding these drug candidates. Combining the results of our analysis and the bibliography, we suggest that captopril and ibuprofen should be prioritized for future anti-fibrotic experiments. We also suggest that nafcillin, which hasn't yet been studied against fibrosis, should be considered for *in vitro* and *in vivo* studies. Finally, digoxin should also be further explored, specifically regarding *OSF*.

We observe gene, pathway and candidate drug 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.

Since exploring every possible combination of drug-gene-pathway entries from the results of our analysis is a difficult task, we provide all of our results as Supplementary Material. In Supplementary Tables 1 and 2 we present the top over- and under-expressed genes per disease. In Supplementary Table 3 we present the top-*5*% ranked pathways from the PathWalks algorithm and in Supplementary Table 4 the CoDReS re-ranked drugs that we identify for each fibrotic disease. In Supplementary Table 5 we present the indications for the "failed" drugs of repoDB regarding clinical trials against fibrotic diseases. Finally, in Supplementary Table 6 we link drugs to their gene targets and the respective highlighted pathways. We hope that the results from our study on the gene, pathway and drug levels will be further investigated to extract new knowledge regarding the pathogenesis as well as potential regimens against fibrosis.

Funding

Evangelos Karatzas is a PhD student in the National and Kapodistrian University of Athens. His doctoral thesis is being funded by the IKY (State Scholarships Foundation) scholarship under the Action "Strengthening Human Resources, Education and Lifelong Learning", 2014-2020, co-funded by the European Social Fund (ESF) and the Greek State (MIS 5000432). Andrea Kakouri holds a PhD Student fellow position funded by the European Commission Research Executive Agency Grant BIORISE (No. 669026), under the Spreading Excellence, Widening Participation, Science with and for Society Framework. George M. Spyrou holds the Bioinformatics ERA Chair Position funded by the European Commission Research Executive Agency (REA) Grant BIORISE (Num. 669026), under the Spreading Excellence, Widening Participation, Science with and for Society Framework.

Conflict of Interest: none declared

References

Abbaszadeh, S., *et al.* Memantine, an NMDA receptor antagonist, attenuates cardiac remodeling, lipid peroxidation and neutrophil recruitment in heart failure: A cardioprotective agent? *Biomedicine & Pharmacotherapy* 2018;108:1237-1243.

Argraves, W.S., *et al.* Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *The Journal of cell biology* 1990;111(6):3155-3164.

Baybutt, R.C., *et al.* Effects on cytokines and histology by treatment with the ACE inhibitor captopril and the antioxidant retinoic acid in the monocrotaline model of experimentally induced lung fibrosis. *Current pharmaceutical design* 2007;13(13):1327-1333.

Bodenhofer, U., Kothmeier, A. and Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* 2011;27(17):2463-2464.

Bonner, J.C. Regulation of PDGF and its receptors in fibrotic diseases. *Cytokine & growth factor reviews* 2004;15(4):255-273.
Borkham-Kamphorst, E., Drews, F. and Weiskirchen, R. Induction of lipocalin-2 expression in acute and chronic experimental liver injury moderated by pro-inflammatory cytokines interleukin-1β through nuclear factor-κB activation. *Liver International* 2011;31(5):656-665.
Brown, A.S. and Patel, C.J. A standard database for drug repositioning. *Scientific data* 2017;4:170029.
Campillos, M., *et al.* Drug target identification using side-effect similarity. *Science* 2008;321(5886):263-266.
Carlile, G.W., *et al.* Ibuprofen rescues mutant cystic fibrosis transmembrane conductance regulator trafficking. *Journal of Cystic Fibrosis* 2015;14(1):16-25.
Chávez, E., *et al.* Antifibrotic and fibrolytic properties of celecoxib in liver damage induced by carbon tetrachloride in the rat. *Liver International* 2010;30(7):969-978.
Chen, L., *et al.* Differential regulation of extracellular matrix and soluble fibulin-1 levels by TGF-β1 in airway smooth muscle cells. *PLoS one* 2013;8(6).
Chen, Y., Lun, A.T. and Smyth, G.K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 2016;5.
Coates, A.L., *et al.* The effect of digoxin on exercise capacity and exercising cardiac function in cystic fibrosis. *Chest* 1982;82(5):543-547.
Cohen, E.P., *et al.* Captopril preserves function and ultrastructure in experimental radiation nephropathy. *Laboratory investigation; a journal of technical methods and pathology* 1996;75(3):349-360.
Cotto, K.C., *et al.* DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic acids research* 2018;46(D1):D1068-D1073.
DeSa, J.V. LXXXVII Submucous Fibrosis of the Palate and Cheek. *Annals of Otology, Rhinology & Laryngology* 1957;66(4):1143-1159.
Duan, Q., *et al.* L1000CDS 2: LINCS L1000 characteristic direction signatures search engine. *NPJ systems biology and applications* 2016;2:16015.
Edgar, R., Domrachev, M. and Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* 2002;30(1):207-210.
Flo, T.H., *et al.* Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature* 2004;432(7019):917-921.
Ftahy, M.M., *et al.* Antifibrotic potential of a selective COX-2 inhibitor (celecoxib) on liver fibrosis in rats. *Comparative Clinical Pathology* 2013;22(3):425-430.
Gao, J.-H., *et al.* Celecoxib ameliorates portal hypertension of the cirrhotic rats through the dual inhibitory effects on the intrahepatic fibrosis and angiogenesis. *PloS one* 2013;8(7).
Ge, Q., *et al.* Fibulin1C peptide induces cell attachment and extracellular matrix deposition in lung fibroblasts. *Scientific reports* 2015;5:9496.
Ghazi-Khansari, M., *et al.* Antifibrotic effect of captopril and enalapril on paraquat-induced lung fibrosis in rats. *Journal of Applied Toxicology: An International Journal* 2007;27(4):342-349.
Hallberg, P., *et al.* Digoxin and mortality in atrial fibrillation: a prospective cohort study. *European journal of clinical pharmacology* 2007;63(10):959-971.

Haller, S.T., *et al.* Monoclonal antibody against marinobufagenin reverses cardiac fibrosis in rats with chronic renal failure. *American journal of hypertension* 2012;25(6):690-696.
Hansen, M.L., *et al.* Aortic valve stenosis and atrial fibrillation influence plasma fibulin-1 levels in patients treated with coronary bypass surgery. *Cardiology* 2013;126(3):202-206.
Harris, T.R., *et al.* Celecoxib Does Not Protect against Fibrosis and Inflammation in a Carbon Tetrachloride–Induced Model of Liver Injury. *Molecular pharmacology* 2018;94(2):834-841.
Herrera, J., Henke, C.A. and Bitterman, P.B. Extracellular matrix as a driver of progressive fibrosis. *The Journal of clinical investigation* 2018;128(1):45-53.
Hui, A.Y., *et al.* Effect of celecoxib on experimental liver fibrosis in rat. *Liver International* 2006;26(1):125-136.
Hutchinson, J., *et al.* Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *European Respiratory Journal* 2015;46(3):795-806.
Jaffar, J., *et al.* The Serum Level Of Fibulin-1 Is Elevated In Idiopathic Pulmonary Fibrosis. In, *C103. PATHOGENESIS, BIOMARKERS, AND RISK FACTORS FOR INTERSTITIAL LUNG DISEASE: FROM BENCH TO BEDSIDE*. American Thoracic Society; 2012. p. A5173-A5173.
Jaffar, J., *et al.* Fibulin-1 predicts disease progression in patients with idiopathic pulmonary fibrosis. *Chest* 2014;146(4):1055-1063.
Jalil, J.E., *et al.* Coronary Vascular Remodeling and Myocardial Fibrosis in the Rat with Renovascular Hypertension Response to Captopril. *American journal of hypertension* 1991;4(1_Pt_1):51-55.
Jiang, X. and Hu, J. Drug treatment of oral submucous fibrosis: a review of the literature. *Journal of oral and maxillofacial surgery* 2009;67(7):1510-1515.
Kanehisa, M., *et al.* KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* 2017;45(D1):D353-D361.
Karaman, M.W., *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nature biotechnology* 2008;26(1):127-132.
Karatzas, E., *et al.* Exploring Fibrotic Disease Networks to Identify Common Molecular Mechanisms with IPF. In, *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE; 2019. p. 72-77.
Karatzas, E., *et al.* 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* 2019;17:939-945.
Karatzas, E., *et al.* PathWalks: Identifying pathway communities using a disease-related map of integrated information. *Bioinformatics*.
Karatzas, E., *et al.* ChemBioServer 2.0: An advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing. *Bioinformatics (Oxford, England)* 2020.
Kim, J.-W., *et al.* Increased urinary lipocalin-2 reflects matrix metalloproteinase-9 activity in chronic hepatitis C with hepatic fibrosis. *The Tohoku journal of experimental medicine* 2010;222(4):319-327.
Kim, S., *et al.* PubChem substance and compound databases. *Nucleic acids research* 2016;44(D1):D1202-D1213.

Kjeldsen, L., *et al.* Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *Journal of Biological Chemistry* 1993;268(14):10425-10432.
Konstan, M.W., *et al.* Effect of high-dose ibuprofen in patients with cystic fibrosis. *New England Journal of Medicine* 1995;332(13):848-854.
Konstan, M.W., *et al.* Clinical use of ibuprofen is associated with slower FEV1 decline in children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2007;176(11):1084-1089.
Lai, D., *et al.* Clinical evaluation of different treatment methods for oral submucous fibrosis. A 10-year experience with 150 cases. *Journal of oral pathology & medicine* 1995;24(9):402-406.
Lands, L.C., *et al.* High-dose ibuprofen in cystic fibrosis: Canadian safety and effectiveness trial. *The Journal of pediatrics* 2007;151(3):249-254.
Lau, J.Y., *et al.* Fibulin-1 is increased in asthma–a novel mediator of airway remodeling? *PLoS One* 2010;5(10).
Li, X. and Tang, J. Clinical treatment observation of tea pigment for oral submucous fibrosis. *Hua xi kou qiang yi xue za zhi= Huaxi kouqiang yixue zazhi= West China journal of stomatology* 1998;16(1):50-52.
Li, Y., *et al.* NMDA receptor antagonist attenuates bleomycin-induced acute lung injury. *PLoS One* 2015;10(5).
Lindroos, P.M., *et al.* Regulation of PDGFR-α in rat pulmonary myofibroblasts by staurosporine. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 2001;280(2):L354-L362.
Lindstrom, D.R., *et al.* Ibuprofen therapy and nasal polyposis in cystic fibrosis patients. *Journal of Otolaryngology* 2007;36(5).
Lipinski, C.A., *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* 1997;23(1-3):3-25.
Liu, G., *et al.* Fibulin-1 regulates the pathogenesis of tissue remodeling in respiratory diseases. *JCI insight* 2016;1(9).
Liu, H., Wei, W. and Li, X. Celecoxib exacerbates hepatic fibrosis and induces hepatocellular necrosis in rats treated with porcine serum. *Prostaglandins & other lipid mediators* 2009;88(3-4):63-67.
Lussier, Y.A. and Chen, J.L. The emergence of genome-based drug repositioning. *Science translational medicine* 2011;3(96):96ps35-96ps35.
Milner, K.L., *et al.* Adipocyte fatty acid binding protein levels relate to inflammation and fibrosis in nonalcoholic fatty liver disease. *Hepatology* 2009;49(6):1926-1934.
Mitani, H., *et al.* Alterations of hairless mouse skin exposed to chronic UV irradiation and its prevention by hydrocortisone. *Photochemistry and photobiology* 1999;69(1):41-46.
Montojo, J., *et al.* GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* 2010;26(22):2927-2928.
Moss, A.J., Dooley, R.R. and Mickey, M.R. Cystic fibrosis complicated by heart failure. *Western Journal of Medicine* 1975;122(6):471.
Mostafavi, S., *et al.* GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome biology* 2008;9(1):S4.
Nakagawa, S., *et al.* Molecular markers of tubulointerstitial fibrosis and tubular cell damage in patients with chronic kidney disease. *PLoS One* 2015;10(8).

Okada, M., *et al.* Effects of captopril and telmisartan on matrix metalloproteinase-2 and-9 expressions and development of left ventricular fibrosis induced by isoprenaline in rats. *Biological and Pharmaceutical Bulletin* 2010;33(9):1517-1521.
Paik, Y.-H., *et al.* Celecoxib induces hepatic stellate cell apoptosis through inhibition of Akt activation and suppresses hepatic fibrosis in rats. *Gut* 2009;58(11):1517-1527.
Parapuram, S.K., *et al.* Loss of PTEN expression by dermal fibroblasts causes skin fibrosis. *Journal of investigative dermatology* 2011;131(10):1996-2003.
Peckham, D., *et al.* Effect of oral digoxin, topical ouabain and salbutamol on transepithelial nasal potential difference in patients with cystic fibrosis. *Clinical Science* 1995;89(3):277-284.
Powis, G., *et al.* Wortmannin, a potent and selective inhibitor of phosphatidylinositol-3-kinase. *Cancer research* 1994;54(9):2419-2423.
Raghu, G., *et al.* Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *American journal of respiratory and critical care medicine* 2018;198(5):e44-e68.
Rajendran, R., Rani, V. and Shaikh, S. Pentoxifylline therapy: a new adjunct in the treatment of oral submucous fibrosis. *Indian Journal of dental research* 2006;17(4):190.
Ramos, S.G., *et al.* Captopril reduces collagen and mast cell and eosinophil accumulation in pig serum-induced rat liver fibrosis. *Pathology international* 1994;44(9):655-661.
Ritchie, M.E., *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015;43(7):e47-e47.
Robinson, M.D., McCarthy, D.J. and Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139-140.
Rosenbloom, J., *et al.* Human fibrotic diseases: current challenges in fibrosis research. In, *Fibrosis*. Springer; 2017. p. 1-23.
Shannon, P., *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003;13(11):2498-2504.
Sharma, J., *et al.* Clinical experience with the use of peripheral vasodilator in oral disorders. *International journal of oral and maxillofacial surgery* 1987;16(6):695-699.
Shi-Wen, X., *et al.* Endothelin-1 promotes myofibroblast induction through the ETA receptor via a rac/phosphoinositide 3-kinase/Akt-dependent pathway and is essential for the enhanced contractile phenotype of fibrotic fibroblasts. *Molecular biology of the cell* 2004;15(6):2707-2719.
Shi-Wen, X., *et al.* Rac inhibition reverses the phenotype of fibrotic fibroblasts. *PLoS One* 2009;4(10).
Singh, M., *et al.* Efficacy of hydrocortisone acetate/hyaluronidase vs triamcinolone acetonide/hyaluronidase in the treatment of oral submucous fibrosis. *Indian Journal of Medical Research* 2010;131(5):665.
Subramanian, A., *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;171(6):1437-1452. e1417.
Takahashi, T., *et al.* A possible contribution of lipocalin-2 to the development of dermal fibrosis, pulmonary vascular involvement and renal dysfunction in systemic sclerosis. *British Journal of Dermatology* 2015;173(3):681-689.

Tanabe, T., *et al.* IgG4-associated multifocal systemic fibrosis complicating sclerosing sialadenitis, hypophysitis, and retroperitoneal fibrosis, but lacking pancreatic involvement. *Internal medicine* 2006;45(21):1243-1247.
Tanramluk, D., *et al.* On the origins of enzyme inhibitor selectivity and promiscuity: a case study of protein kinase binding to staurosporine. *Chemical biology & drug design* 2009;74(1):16-24.
Tepper, R.S., *et al.* Lower respiratory illness in infants and young children with cystic fibrosis. *Pediatric pulmonology* 1997;24(1):48-51.
Theocharis, A.D., *et al.* Extracellular matrix structure. *Advanced drug delivery reviews* 2016;97:4-27.
Tuncer, I., *et al.* Anti-fibrogenic effects of captopril and candesartan cilexetil on the hepatic fibrosis development in rat: the effect of AT1-R blocker on the hepatic fibrosis. *Experimental and toxicologic pathology* 2003;55(2-3):159-166.
Ursu, O., *et al.* DrugCentral 2018: an update. *Nucleic acids research* 2019;47(D1):D963-D970.
Veber, D.F., *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry* 2002;45(12):2615-2623.
Verrecchia, F. and Mauviel, A. Transforming growth factor-β and fibrosis. *World journal of gastroenterology: WJG* 2007;13(22):3056.
Vidt, D.G., Bravo, E.L. and Fouad, F.M. Captopril. *New England Journal of Medicine* 1982;306(4):214-219.
Wang, Y., *et al.* Effects of cannabinoid receptor type 2 on endogenous myocardial regeneration by activating cardiac progenitor cells in mouse infarcted heart. *Science China Life Sciences* 2014;57(2):201-208.
Warde-Farley, D., *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research* 2010;38(suppl_2):W214-W220.
Wen, S.L., *et al.* Celecoxib attenuates hepatic cirrhosis through inhibition of epithelial-to-mesenchymal transition of hepatocytes. *Journal of gastroenterology and hepatology* 2014;29(11):1932-1942.
Wengrower, D., *et al.* Prevention of fibrosis in experimental colitis by captopril: the role of TGF-β1. *Inflammatory bowel diseases* 2004;10(5):536-545.
Wishart, D.S., *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 2018;46(D1):D1074-D1082.
Xu, M., *et al.* Effects of curcumin in treatment of experimental pulmonary fibrosis: a comparison with hydrocortisone. *Journal of ethnopharmacology* 2007;112(2):292-299.
Yu, J., *et al.* The anti-inflammatory effect of celecoxib does not prevent liver fibrosis in bile duct-ligated rats. *Liver International* 2009;29(1):25-36.
Zachariæ, L. and Zachariae, F. Hydrocortisone Acetate in the Local Treatment of Fibroses. *Acta Rheumatologica Scandinavica* 1957;3(1-4):76-81.
Zhang, C., *et al.* Celecoxib prevents pressure overload-induced cardiac hypertrophy and dysfunction by inhibiting inflammation, apoptosis and oxidative stress. *Journal of cellular and molecular medicine* 2016;20(1):116-127.
Zhang, X., *et al.* Zinc attenuates tubulointerstitial fibrosis in diabetic nephropathy via inhibition of HIF through PI-3K signaling. *Biological trace element research* 2016;173(2):372-383.

Zhi, H., *et al.* Effects of direct Renin inhibition on myocardial fibrosis and cardiac fibroblast function. *PloS one* 2013;8(12).
Zuberi, K., *et al.* GeneMANIA prediction server 2013 update. *Nucleic acids research* 2013;41(W1):W115-W122.