Dr *AFC*: Drug Repositioning Through Anti-Fibrosis Characteristic

Dingfeng Wu[†], Wenxing Gao[†], Xiaoyi Li[†], Chuan Tian, Na Jiao, Sa Fang, Jing Xiao, Zhifeng Xu, Lixin Zhu, Guoqing Zhang and Ruixin Zhu

Corresponding authors: **Ruixin Zhu (rxzhu@tongji.edu.cn),** Putuo people's Hospital, Department of Bioinformatics, Tongji University, 1239 Siping Road, Shanghai 200092, P.R. China.Tel: 86-21-6598-1041; **Guoqing Zhang (gqzhang@picb.ac.cn),** Bio-Med Big Data Center, Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, P.R.China. Tel: 86-21-5492-0465; **Lixin Zhu (zhulx6@mail.sysu.edu.cn),** Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Department of Colorectal Surgery, the Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510655, P.R. China. Tel: 86-199-46256235

[†] These authors contributed equally to this work.

Abstract

Fibrosis is a key component in the pathogenic mechanism of many diseases. These diseases involving fibrosis may share common mechanisms, therapeutic targets and therefore, common intervention strategies and medicines may be applicable for these diseases. For this reason, deliberately introducing anti-fibrosis characteristics into modelling may lead to more success in drug repositioning. In this study, anti-fibrosis knowledge base was first built by collecting data from multiple resources. Both structural and biological profiles were derived from the knowledge base and used for

Dingfeng Wu is a post-doc fellow at Tongji University, Shanghai, China. His current research interests include machine learning and drug repositioning.
Wenxing Gao, **Xiaoyi Li**, **Sa Fang** and **Jing Xiao** are Master Degree students at Tongji Universtiy, Shanghai, China. They are working on bioinformatics and multi-omics analysis.
Chuan Tian is a scientist at Relay Therapeutics, Cambridge, United States. His current research interests include machine learning and drug repositioning.
Na Jiao is a post-doc fellow at Sun Yat-sen University, Guangzhou, China. Her current research interests include drug repositioning and gut microbiome.
Zhifeng Xu is a junior student at Tongji Universtiy, Shanghai, China. He are working on bioinformatics.
Lixin Zhu is an associate professor at Sun Yat-sen University, Guangzhou, China. His current research interests include drug repositioning and gut microbiome.
Guoqing Zhang is a professor at Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China. His current research interests include machine learning and natural language processing.
Ruixin Zhu is a professor at Tongji University, Shanghai, China. His current research interests include drug repositioning, microbiome and bioinformatics.

constructing machine learning models including Structural Profile Prediction Model (SPPM) and Biological Profile Prediction Model (BPPM). Three external public data sets were employed for validation purpose and further exploration of potential repositioning drugs in wider chemical space. The resulting SPPM and BPPM models achieve area under the receiver operating characteristic curve (AUC) of 0.879 and 0.972 in the training set, and 0.814 and 0.874 in the testing set. Additionally, our results also demonstrate that substantial amount of multi-targeting natural products possess notable anti-fibrosis characteristics and might serve as encouraging candidates in fibrosis treatment and drug repositioning. To leverage our methodology and findings, we developed repositioning prediction platform, Drug Repositioning based on Anti-Fibrosis Characteristic (Dr *AFC*) that is freely accessible via https://www.biosino.org/drafc.

Key words: fibrosis; drug repositioning; web server; knowledge base; natural products

Introduction

Fibrosis is defined as the process of excessive accumulation of fibrous connective tissue in most tissues or organs, where normal cells are replaced by the extracellular matrix (ECM), resulting in disrupted tissue function. In the new era of 21st century, the morbidity and mortality rates of various fibrotic diseases have increased progressively, bringing a huge global health burden. In developed countries, fibroproliferative diseases are responsible for nearly 45% of deaths[1]. One of the well-known fibrotic diseases, idiopathic pulmonary fibrosis(IPF), has a poor prognosis with the 5 year survival rate less than 30% and median survival ranging from 3 to 5 years[2]. The outcomes of IPF patients are even worse than those with many types of cancers [3]. As data obtained by Clinical Practice Research Datalink(CPRD) revealed, the prevalence of IPF patients in board case definitions has doubled from 19.94 per 100,000 patients in 2000 to 38.82 per 100,000 patients in 2012, and a 80% increase in incidence was observed[4]. Another life-threatening fibrotic disease, cardiac fibrosis, is one of the leading factors causing heart failure (HF) [5]. A research from 2008-2014 revealed that in 318 patients with systolic dysfunction, 78% had one type of myocardial fibrosis while 25% had at least 2 types [6].

The polypharmacology of most anti-fibrosis drugs could improve therapeutic efficacy. Recent studies have found that, firstly, fibrosis is the common pathogenic process in most diseases. For example, there are multiple common cellular processes between lung cancer and IPF, including inflammation, cell apoptosis and tissue infiltration [7]. Secondly, fibrosis-related processes have common mechanisms, targets and drugs [8, 9]. A multi-organ fibrosis research discovered a set of 90 common differentially expressed genes across lung, heart, liver and kidney. In the two most active gene networks generated by Ingenuity Pathway Analysis(IPA), these genes play a key role in connective tissue disorders and genetic, skeletal and muscular disorders[10]. Similarly, another multi-organ fibrosis research also obtained a series of 11 metzincin-related differentially expressed genes across heart, lung, liver, kidney

and pancreas including *THBS2*, *TIMP1*, *COL1A2*, *COL3A1*, *HYOU1*, *MMP2* and *MMP7*[11]. Thirdly, fibrosis is a complicated pathological process involving multiple pathways, thus multi-target drugs are appropriate for fibrosis-related diseases[9]. Different pathways interact and counter-interact with each other to establish a "check-and-balance" system, for instance, the core regulators, transforming growth factor-β(TGF-β) and connective tissue growth factor(CTGF) signaling pathways could collaborate to elicit pulmonary and renal fibrosis[12, 13]. In summary, these evidences indicate that anti-fibrosis intervention strategies and medicines may be applicable for more diseases through targeting their common fibrosis-related mechanisms. Therefore, compounds that can more specifically target anti-fibrosis could have greater potential of repositioning and are more applicable for drug repositioning research.

Drug repositioning, or repurposing refers to the "reuse of old drugs", recycling existing drugs for new medical indications. Compared with *de novo* drug discovery, drug repositioning has obvious advantages that it could significantly shorten drug development periods, reduce laboratory cost and minimize potential safety risk. Nowadays, drug repositioning is one of the most efficient strategies in drug development[14]. With the advancement of high-throughput sequencing technology and deep learning, various data-driven computational prediction and analytic models stand out[15, 16], including Similarity Ensemble Approach (SEA)[17] and Connectivity Map(cMAP)[18]. SEA clusters ligands into sets and calculates the similarity scores between ligand sets from ligand topology[17]. cMAP computes the similarity of "signatures" deduced from compound-induced gene profiles to quantify the biological functional relationships between compounds. Moreover, the relationship between compounds and diseases could also be quantified in opposite manner[18]. However, with so many repositioning methods and algorithms have emerged[19-21], there still no attempts hitherto in introducing anti-fibrosis characteristic into drug repositioning strategy.

For the first time, we built the anti-fibrosis knowledge base from anti-fibrosis related research. Based on the knowledge base, two repositioning models, Structural Profile Prediction Model (SPPM) and Biological Profile Prediction Model (BPPM) were constructed with high prediction accuracy. Centered on these two models, we then developed a repositioning computing platform, Drug Repositioning based on Anti-Fibrosis Characteristic (Dr *AFC*), to accelerate the process of exploring repositioning drugs and studying its underlying mechanisms.

Materials and methods

Datasets

Anti-fibrosis knowledge base

Anti-fibrosis related literatures were collected through key word queries “fibrosis AND target” in PubMed from Jan. 1st, 2000 to Oct. 31st, 2019. The compound-target interaction information on “fibrosis” were collected in the CTD[22] from Jan. 1st, 2000 to Oct. 31st, 2019. Anti-fibrosis trials were collected in ClinicalTrials.gov[23] from Jan. 1st, 2000 to Oct. 31st, 2019. Finally, anti-fibrosis treatments, targets and compound-target interactions were extracted and aggregated into the knowledge base.

Model construction

Structural and biological profiles of compounds were collected from DrugBank[24] and cMap, respectively and used for model construction. 2640 approved drugs in DrugBank and 1223 compounds in the anti-fibrosis knowledge base served as the raw data for Structural Profile Prediction Model (SPPM) construction. 6100 biological profiles (gene expression) of 1309 small molecules in cMap served as the raw data for Biological Profile Prediction Model (BPPM).

Case studies

20,263 natural products from TCMID[25], 5968 DrugBank experimental drugs[24] and 5000 random compounds from ChEMBL[26] were collected as external validations and case studies of SPPM. And external biological profiles from GEO database (GSE85871) that contains transcriptomics perturbation profiles of 105 natural products in MCF7 cell line were used for case studies of BPPM.

Methods

Pre-processing of modeling data

In raw chemical structures (from DrugBank approved drugs and the anti-fibrosis knowledge base) and biological profiles (from cMap) data, compounds that appeared in the anti-fibrosis knowledge base were labeled as positive candidates while the rest were labeled as negative candidates. Then, chemical structures were converted into chemical fingerprints (166-bits MACCS keys) for processing chemical information in a fast and convenient way using RDKit[27]. As to biological profiles, Quantile Transformer was used to transform biological profiles into ranking orders to improve the performance of model generalization, and also made datasets from different batches and platforms more comparable.

One-class SVM (nu=0.3) was performed to estimate sample quality, remove outliers and confirm final positive and negative samples. 70% of final samples were used as training set for model selection and super-parameter determination while the remainder as testing set for model validation.

Anti-fibrosis model construction and validation

Four different machine learning algorithms were selected for modeling on training set, including logistic regression, decision tree, random forest and gradient boosting. Among them, method with highest precision and AUC calculated by 5-fold cross-validation was selected for subsequent analysis. Iterative feature elimination (IFE) algorithm was performed to select optimal feature set through one-by-one

feature deletion. Finally, SPPM and BPPM were constructed based on optimal modeling algorithm and feature set, and further validated by testing set.

Drug repositioning mechanism analysis

Network-based inference approaches were wildly used in drug repositioning [20, 21]. Here we infer the potential drug repositioning mechanism through compound-target-disease network. Firstly, based on SPPM and BPPM, the repositioning characteristics of compounds were predicted through their structural or biological profiles, in which compounds with reposition score>0.5 were considered as anti-fibrosis and had repositioning potential. Next, the anti-fibrosis characteristic and potential repositioning mechanisms of these candidates were explored on the basis of compound-target-disease corresponding information in the anti-fibrosis knowledge base. Similar compounds that may interact with same targets and diseases were calculated through Tanimoto similarity of chemical structural fingerprints or Spearman's rank correlation coefficient of biological profiles. Targets and disease information of compounds reported in previous researches were refined from the anti-fibrosis knowledge base to explore anti-fibrosis mechanism of compounds. Finally, the potential mechanisms among compounds in compound-target-disease network displayed in drug repositioning analysis were used to help propose feasible drug repositioning solutions.

Webserver construction of Dr AFC

Dr *AFC* was constructed through PostgreSql database and Django framework. This platform serves as a practical tool for prediction of drug repositioning potential based on compound structures (SPPM) and biological profiles (BPPM) as well as displaying compound-target-disease network of drug repositioning mechanisms. Meanwhile, Dr *AFC* also integrated toolkits such as quantitative estimate of drug-likeness (QED)

from Silicos-it[28], and similarity calculation and structure matching borrowed from RDkit to provide convenient web-based calculations for users.

The overall process is shown in Figure 1.

Results

SPPM and BPPM show high performances for anti-fibrosis prediction

To construct the anti-fibrosis knowledge base, 7058 fibrosis-related references from PubMed, 302 from Comparative Toxicogenomics Database(CTD)[22] and 2664 fibrosis-related trials from ClinicalTrials.gov[23] were collected through text mining. Finally, 1223 anti-fibrosis treatments (containing 902 small molecules), 1067 fibrosis-related targets, 3096 fibrosis-related records from references and 1787 from trials, 1067 anti-fibrosis compound-target interactions were obtained and integrated into anti-fibrosis knowledge base (Figure S1).

In modeling session, 2885 compound structures (from DrugBank approved drugs) [24] and 6100 biological profiles (from cMap) were labeled as positive candidates and negative candidates based on their anti-fibrosis characteristic in the anti-fibrosis knowledge base. After sanity check and outlier removal, 1701 compound structures and 2735 biological profiles were filtered out for model construction (Table S1).

Four different machine learning classifiers were evaluated and compared to choose the most optimal modeling method (Table S2). Gradient boosting was eventually selected according to its highest precision and AUC (Structural profile: Precision=0.737, AUC=0.839, Biological profile: Precision=0.892, AUC=0.912).

In the process of building SPPM and BPPM, we found that even a small number of features could reach certain stability and reasonably good performance (Figure S2, **Figure 2a**). Models based on top 38 features including CHARGE, S and XA(A)A could reach the maximum cross-validation AUC (0.879) in SPPM while top 47 features including RPL30, MRMRPL5 and KPNB1 could reach the maximum cross-validation AUC(0. 972) in BPPM. We discovered that 46 of the top 47 features

in BPPM were connected with fibrosis in CTD inference networks (Figure 2b). Besides, several genes were associated with fibrosis-related indications like retroperitoneal fibrosis, keloids, tissue adhesions and cicatrix.

Finally, SPPM and BPPM were build based on the most optimal modeling method and the selected small feature subset (top 38 features in SPPM and top 47 features in BPPM). In testing set, the average AUC for SPMM reaches 0.814 (Figure 2c) while the average AUC for BPMM reaches 0.874 (Figure 2d).

Case studies

Anti-fibrosis drugs exhibit greater drug repositioning potential

We used SPPM to predict anti-fibrosis drugs from DrugBank experimental drugs and the comparative analysis was performed between the CTD compound-gene interactions of the predicted anti-fibrosis and non-anti-fibrosis drugs. The results show that the anti-fibrosis group accommodates stronger interactions, presumably more genetic effects thus greater repositioning potential (**Figure 3a**).

In Drugbank experimental drugs, multiple drugs with great repositioning potential (Related Genes>500 and Diseases>20, Figure 3b, Table S3) were developed for fibrotic diseases and other diseases. Quercetin was discovered to ameliorate liver fibrosis through regulating macrophage infiltration and polarization, and it could alleviate IPF through fibroblasts apoptosis[29, 30]. Based on our results, we confirm that quercetin interacts with numerous genes and is strongly linked to multiple diseases (Repositioning score=0.856, Related Genes=3938, Diseases=150, Table S3). Another natural compound from turmeric, curcumin (Repositioning score=0.855, Related Genes=903, Diseases=138, Table S3), could also be used for treating multiple fibrotic diseases. It could inhibit fibroblast proliferation and myofibroblast differentiation in IPF[31] while inhibit oxidative stress and exhibit anti-inflammatory effect in liver fibrosis[32]. Apart from fibrosis, curcumin has been applied for osteoarthritis and rheumatoid arthritis treatment [33, 34]. Moreover, other drugs, such

as resveratrol also had great repositioning potential (Repositioning score=0.821, Figure 3b).

Natural compounds are the better repositories for drug repositioning

In order to expand the resources of potential repositioning drugs and further explore the chemical space, we introduced two external molecule sets, natural products from TCMID[25] and random compounds in ChEMBL[26]. SPPM was used to predict the repositioning potential of compounds from both external molecule sets. The results show that there were 35.42%, 77.26% and 37.04% of compounds could be potentially repositioned in DrugBank experimental drugs, TCMID and ChEMBL, respectively. The reserves in natural products from TCMID are significantly higher than others, indicating that natural products are great repositioning repositories and need further researches (Figure 3c).

BPPM was used to discover specific natural products with repositioning potential from gene profiles dataset of 105 natural products (GSE85871). The results show that a total of 66 natural products have anti-fibrosis characteristic and repositioning potential, including ginsenoside Re(Repositioning score=0.979), muscone(Repositioning score=0.974) and cinnamic acid(Repositioning score=0.948) (Table S4). Among them, ginsenoside Re hold the potential to influence HDAC2, HDAC9 and HMGCR and fulfilled anti-fibrosis roles via “inflammation”, “preventing collagen deposition” and “targeting myeloperoxidase” with Drug repositioning mechanism analysis tools in Dr *AFC* (Figure 3d). Ginsenoside Re is the extract of panax ginseng which exhibited protective effects in neural and systematic inflammations through inhibiting the interaction between LPS and TLR4 in macrophages[35]. It was reported to exert anti-fibrosis effect on cardiac fibrosis through down-regulating the expression of p-Smad3, collagen I and reducing the augmentation of collagen fibers[36]. Apart from fibrosis, ginsenoside Re could alleviate inflammation through inhibiting myeloperoxidase activity[37] and decrease

fat accumulation through inhibiting HMGCR and cholesterol biosynthesis[38]. Besides, other ginsenosides, like ginsenoside Rb1, ginsenoside Rc, ginsenoside Rb3, ginsenoside Rb2, ginsenoside Rd and ginsenoside Rg, also exhibit anti-fibrosis characteristic and repositioning potential(Table S4).

Drug Repositioning based on Anti-Fibrosis Characteristic Webserver

Based on SPPM and BPPM, we constructed a computing platform for repositioning research purpose, named Drug Repositioning based on Anti-Fibrosis Characteristic (Dr *AFC*), the main function and workflow of which is shown in **Figure 4**. On Dr *AFC* platform, anti-fibrosis and potential repositioning could be predicted from compound structures or biological profiles. Drug repositioning mechanism analysis could infer the relationships among compounds, fibrosis-related targets and diseases which help understand pathology. Furthermore, drug-likeness estimation, chemical similarity calculation and structure matching were integrated into Dr *AFC* to provide useful information for drug development.

Drug repositioning analysis function

Dr *AFC* allows users to upload compound structures or compound-induced biological profiles for repositioning potential prediction. As shown in Figure 4b, Dr *AFC* accepts SMILES strings of compound structures for SPPM prediction, and accepts gene profiles with row names in Affymetrix U133A probe ID, Entrez ID or gene symbol format for BPPM prediction. Both methods support .txt、.csv or .xlsx files (Figure 4c).

Webserver would perform corresponding prediction analysis automatically based on the uploaded files and display the output on the result page in three aspects (Figure 4d): 1) Basic part includes compound ID, compound name, 2D compound structure(only for SPPM) and SMILES string(only for SPPM). 2) Prediction part includes repositioning scores of anti-fibrosis characteristic and repositioning potential prediction. The repositioning scores ranges from 0 to 1 and higher score indicates

higher potential. If repositioning score≥0.5, the compound would be defined as an anti-fibrosis and potential repositioning compound. 3) Drug repositioning mechanism analysis part. This analysis infers the potential anti-fibrosis and repositioning mechanisms of compound structures or biological profiles users uploaded based on our anti-fibrosis knowledge base. It could provide users potential mechanisms as theoretical foundations for drug repositioning studies.

Other functions

Dr *AFC* also contains drug-likeness estimation, chemical similarity calculation and structure matching tools. Users could upload their compounds in SMILES and perform these additional functions. Drug-likeness estimation could evaluate and score the compound drug-likeness, which ranges from 0 to 1 with higher score indicating higher potential for lead compound. Chemical similarity calculation and structure matching provide convenient ways for users to search compound with similar structures, same structures or substructures, supporting single compound calculation and simultaneous calculation for multiple compounds.

Discussion

Fibrosis is the common mechanism of diseases that attracts global attention. The anti-fibrosis characteristic of a compound could infer the greater repositioning potential it would have. However, the anti-fibrosis characteristic has not been extensively introduced into the realm of drug discovery till now. In this study, we first bridge the gap by developing a platform that can provide intensive information conveniently on drug repositioning based on anti-fibrosis characteristic data, Dr *AFC* (https://www.biosino.org/drafc). This *in silico* platform also provides a highly accurate way to generate data for rational drug design via combining the advanced machine-learning algorithm.

Dr *AFC* was built based on the anti-fibrosis knowledge base, which pioneered the excavation and organization of fibrosis-related studies throughout recent years. Structural profile (SPPM) and biological profile (BPPM) that show extraordinary capabilities in drug repositioning prediction (with AUC 0.814 and 0.874, respectively) were integrated into Dr *AFC*. BPPM show slightly higher performance than SPPM according to the AUC. The possible reason could be that biological profile is more tolerant and could contain information reflecting an overall effect of compound functionally in the body. Biological profile show its advantage in multiple repositioning algorithms previously, such as cMAP[18], L1000CDS^2[39] and MANTRA[40]. Besides, certain therapies without available structure profile like biotech drugs or cocktail therapies could also be studied in repositioning research according to their biological profiles.

In BPPM, 47 biological markers exhibited strong prediction abilities. These genes are directly or indirectly linked to various fibrotic diseases. Interestingly, ribosomal proteins including RPL30, MRPL15, RPL32, RPS3A, RPLP0, RPL7, RPL23A and RPL13A are the main part of these biological markers. Ribosomes serve as significant regulators in immune signaling pathways, tumorigenesis pathways and cardiovascular and metabolic diseases[41, 42]. For example, the expression of RPL30 is negatively correlated with carcinogenesis process in medulloblastoma that usually is accompanied by desmoplasia and could thus serve as a prognosis biomarker[43]. Besides, the over-activation of RNA polymerase in the biogenesis of ribosomes could cause the enhancement of protein synthesis and the decrease of translation accuracy, triggering cancers or exacerbating cancer processes[44]. Furthermore, some biological markers are associated with the spliceosome formation including RBM8A, HNRNPA3, SNRPG and DHX15. Spliceosome is the large molecular machine composed of five snRNA and many proteins, and serves as the catalyzer of pre-RNA introns which are crucial for protein expression and function. It has been reported to

be closely associated with multiple diseases, including cystic fibrosis and pulmonary fibrosis[45, 46].

Based on external molecule sets, natural products are validated to have the strongest anti-fibrosis characteristics and repositioning potential among chemicals from different sources. Natural products provide a wealth of valuable natural resources for modern medicine and are seen as promising and popular candidates for drug repositioning studies[47]. Their privileged scaffolds, structural complexity, abundant stereochemistry and 'metabolite-likeness' are main reasons for the broad-spectrum of biological activities [48, 49]. The multi-targets and synergistic effects of natural products exhibit great advantages in treating diseases undergoing sophisticated mechanisms, such as fibrosis[50]. Our studies show that natural products like ginsenoside have great anti-fibrosis characteristic and repositioning potential and should be top priority when considering repositioned drug discovery. Additionally, the natural products in Drugbank experimental drugs such as quercetin, curcumin and resveratrol, also highlight their strong repositioning capabilities. Therefore, natural products could serve as promising source and the good choice for further drug development and repositioning study.

Conclusion

In summary, based on anti-fibrosis characteristics, we constructed two repositioning models, SPPM and BPPM, which could predict the anti-fibrosis characteristics and repositioning potential from compound structures and compound-induced biological profiles. SPPM and BPPM efficiently utilize the generality of fibrotic diseases, thus greatly increase the success rate of drug repositioning. This study not only established a highly efficient strategy of prediction, but also developed a convenient and user-friendly computing platform, Dr *AFC* (https://www.biosino.org/drafc), for studying fibrosis mechanisms and drug repositioning.

Key Points

- Fibrosis is the common mechanism of diseases which could be applied in drug repositioning.
- We developed a convenient and user-friendly computing platform, Dr *AFC*, for studying fibrosis mechanisms and drug repositioning.
- Dr *AFC* shows high performance on both cross validation and external validation, which demonstrates its potential applications in drug discovery.
- Natural compounds proved to be the better repositories for drug repositioning.

Funding

This work was supported by National Key R&D Program of China [2017YFC0907505, 2016YFC0901904, 2017YFC0908404 to G.Z.]; National Natural Science Foundation of China [81774152 to R.Z., 81770571 to L.Z.]; National Postdoctoral Program for Innovative Talents of China [BX20190393 to N.J.]; China Postdoctoral Science Foundation [2019M651568 to D.W., 2019M663252 to N.J.]; Natural Science Foundation of Shanghai [16ZR1449800 to R.Z.); Science and Technology Service Network Initiative of Chinese Academy of Sciences [Y919C11011 to G.Z.]; and Funds from the University at Buffalo Community of Excellence in Genome, Environment and Microbiome (GEM) [to L.Z.].

Conflict of interests

All the authors have no conflict of interest.

References

1. Wynn TA, Cellular and molecular mechanisms of fibrosis. *J Pathol* 2008;**214**:199-210.
2. Vancheri C and du Bois RM, A progression-free end-point for idiopathic pulmonary fibrosis trials: lessons from cancer. *Eur Respir J* 2013;**41**:262-9.
3. Homma S, Bando M, Azuma A, et al., Japanese guideline for the treatment of idiopathic pulmonary fibrosis. *Respir Investig* 2018;**56**:268-291.

4. Strongman H, Kausar I, and Maher TM, Incidence, Prevalence, and Survival of Patients with Idiopathic Pulmonary Fibrosis in the UK. *Adv Ther* 2018;**35**:724-736.
5. Horn MA and Trafford AW, Aging and the cardiac collagen matrix: Novel mediators of fibrotic remodelling. *J Mol Cell Cardiol* 2016;**93**:175-85.
6. Almehmadi F, Joncas SX, Nevis I, et al., Prevalence of myocardial fibrosis patterns in patients with systolic dysfunction: prognostic significance for the prediction of sudden cardiac arrest or appropriate implantable cardiac defibrillator therapy. *Circ Cardiovasc Imaging* 2014;**7**:593-600.
7. Ballester B, Milara J, and Cortijo J, Idiopathic Pulmonary Fibrosis and Lung Cancer: Mechanisms and Molecular Targets. *Int J Mol Sci* 2019;**20**.
8. Li X, Zhu L, Wang B, et al., Drugs and Targets in Fibrosis. *Front Pharmacol* 2017;**8**:855.
9. Liu J, Sun D, Liu J, et al., FibroAtlas: A Database for the Exploration of Fibrotic Diseases and Their Genes. 2019;**2019**.
10. Wenzke KE, Cantemir-Stone C, Zhang J, et al., Identifying common genes and networks in multi-organ fibrosis. *AMIA Jt Summits Transl Sci Proc* 2012;**2012**:106-15.
11. Rodder S, Scherer A, Korner M, et al., A subset of metzincins and related genes constitutes a marker of human solid organ fibrosis. *Virchows Arch* 2011;**458**:487-96.
12. Meng XM, Nikolic-Paterson DJ, and Lan HY, TGF-beta: the master regulator of fibrosis. *Nat Rev Nephrol* 2016;**12**:325-38.
13. Wang Q, Usinger W, Nichols B, et al., Cooperative interaction of CTGF and TGF-beta in animal models of fibrotic disease. *Fibrogenesis Tissue Repair* 2011;**4**:4.
14. Xue H, Li J, Xie H, et al., Review of drug repositioning approaches and resources. 2018;**14**:1232.
15. Qiu TY, Qiu JX, Feng J, et al., The recent progress in proteochemometric modelling: focusing on target descriptors, cross-term descriptors and application scope. *Briefings in Bioinformatics* 2017;**18**:125-136.
16. Liu L, Tsompana M, Wang Y, et al., Connection Map for Compounds (CMC): A Server for Combinatorial Drug Toxicity and Efficacy Analysis. *Journal of Chemical Information and Modeling* 2016;**56**:1615-1621.
17. Keiser MJ, Roth BL, Armbruster BN, et al., Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;**25**:197-206.
18. Lamb J, Crawford ED, Peck D, et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929-35.
19. Sam E and Athri P, Web-based drug repurposing tools: a survey. *Brief Bioinform* 2019;**20**:299-316.

20. Wu Z, Cheng F, Li J, et al., SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief Bioinform* 2017;**18**:333-347.
21. Peng Y, Wang M, Xu Y, et al., Drug repositioning by prediction of drug's anatomical therapeutic chemical code via network-based inference approaches. *Brief Bioinform* 2020;**00**:1-15.
22. Davis AP, Grondin CJ, Johnson RJ, et al., The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res* 2019;**47**:D948-D954.
23. Zarin DA, Fain KM, Dobbins HD, et al., 10-Year Update on Study Results Submitted to ClinicalTrials.gov. *N Engl J Med* 2019;**381**:1966-1974.
24. Wishart DS, Feunang YD, Guo AC, et al., DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074-D1082.
25. Huang L, Xie D, Yu Y, et al., TCMID 2.0: a comprehensive resource for TCM. 2018;**46**:D1117-D1120.
26. Mendez D, Gaulton A, Bento AP, et al., ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;**47**:D930-D940.
27. Landrum G, RDKit: Open-source cheminformatics. 2006.
28. Silicos-it. Available online: http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/ (accessed on 15 August 2019).
29. Li X, Jin Q, Yao Q, et al., The Flavonoid Quercetin Ameliorates Liver Inflammation and Fibrosis by Regulating Hepatic Macrophages Activation and Polarization in Mice. *Front Pharmacol* 2018;**9**:72.
30. Hohmann MS, Habiel DM, Coelho AL, et al., Quercetin Enhances Ligand-induced Apoptosis in Senescent Idiopathic Pulmonary Fibrosis Fibroblasts and Reduces Lung Fibrosis In Vivo. *Am J Respir Cell Mol Biol* 2019;**60**:28-40.
31. Smith MR, Gangireddy SR, Narala VR, et al., Curcumin inhibits fibrosis-related effects in IPF fibroblasts and in mice following bleomycin-induced lung injury. *Am J Physiol Lung Cell Mol Physiol* 2010;**298**:L616-25.
32. Bruck R, Ashkenazi M, Weiss S, et al., Prevention of liver cirrhosis in rats by curcumin. *Liver Int* 2007;**27**:373-83.
33. Henrotin Y, Priem F, and Mobasheri A, Curcumin: a new paradigm and therapeutic opportunity for the treatment of osteoarthritis: curcumin for osteoarthritis management. *Springerplus* 2013;**2**:56.
34. Chandran B and Goel A, A randomized, pilot study to assess the efficacy and safety of curcumin in patients with active rheumatoid arthritis. *Phytother Res* 2012;**26**:1719-25.
35. Kim JH, Yi YS, Kim MY, et al., Role of ginsenosides, the main active components of Panax ginseng, in inflammatory responses and diseases. *J Ginseng Res* 2017;**41**:435-443.

36. Wang QW, Yu XF, Xu HL, et al., Ginsenoside Re Improves Isoproterenol-Induced Myocardial Fibrosis and Heart Failure in Rats. *Evid Based Complement Alternat Med* 2019;**2019**:3714508.
37. Lee IA, Hyam SR, Jang SE, et al., Ginsenoside Re ameliorates inflammation by inhibiting the binding of lipopolysaccharide to TLR4 on macrophages. *J Agric Food Chem* 2012;**60**:9595-602.
38. Lee S, Lee MS, Kim CT, et al., Ginsenoside Rg3 reduces lipid accumulation with AMP-Activated Protein Kinase (AMPK) activation in HepG2 cells. *Int J Mol Sci* 2012;**13**:5729-39.
39. Duan Q, Reid SP, Clark NR, et al., L1000CDS(2): LINCS L1000 characteristic direction signatures search engine. *NPJ Syst Biol Appl* 2016;**2**.
40. Carrella D, Napolitano F, Rispoli R, et al., Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics* 2014;**30**:1787-8.
41. Zhou X, Liao WJ, Liao JM, et al., Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol* 2015;**7**:92-104.
42. Wang W, Nag S, Zhang X, et al., Ribosomal proteins and human diseases: pathogenesis, molecular mechanisms, and therapeutic implications. *Med Res Rev* 2015;**35**:225-85.
43. De Bortoli M, Castellino RC, Lu XY, et al., Medulloblastoma outcome is adversely associated with overexpression of EEF1D, RPL30, and RPS20 on the long arm of chromosome 8. *BMC Cancer* 2006;**6**:223.
44. Pelletier J, Thomas G, and Volarevic S, Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat Rev Cancer* 2018;**18**:51-63.
45. Box JA, Bunch JT, Tang W, et al., Spliceosomal cleavage generates the 3' end of telomerase RNA. *Nature* 2008;**456**:910-4.
46. Liu X, Jiang Q, Mansfield SG, et al., Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nat Biotechnol* 2002;**20**:47-52.
47. Shen B, A New Golden Age of Natural Products Drug Discovery. *Cell* 2015;**163**:1297-300.
48. Harvey AL, Edrada-Ebel R, and Quinn RJ, The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 2015;**14**:111-29.
49. Zhao C, Ye Z, Ma ZX, et al., A general strategy for diversifying complex natural products to polycyclic scaffolds with medium-sized rings. *Nat Commun* 2019;**10**:4015.
50. Yan H, Huang Z, Bai Q, et al., Natural product andrographolide alleviated APAP-induced liver fibrosis by activating Nrf2 antioxidant pathway. *Toxicology* 2018;**396-397**:1-12.

Figure legends

Figure 1. The schematic of Dr *AFC* construction

Figure 2 Feature selection and model performances

A. Performances of top 30 features through iterative feature elimination in BPPM; **B**. The CTD inference networks of 47 gene features and fibrosis-related diseases; **C**. AUC of SPMM in testing set; **D**. AUC of BPPM in testing set.

Figure 3 Case studies of Dr *AFC*

A. Comparison of the number of genes interacting with compounds predicted as anti-fibrosis and non-anti-fibrosis; **B**. The distribution of related genes, diseases and repositioning score for Drugbank experimental drugs. Compounds with repositioning score>0.5 were considered as anti-fibrosis and had repositioning potential; **C**. The distribution of repositioning scores in different datasets (****: p-value $<10^{-4}$ by two-sided Wilcoxon rank sum test); **D**. Drug repositioning mechanism analysis of ginsenoside Re by Dr *AFC*.

Figure 4. Anti-fibrosis and repositioning computing platform (Dr *AFC*)

A. Dr *AFC* integrated two prediction models, SPPM and BPPM; **B**. SPPM accepts SMILES strings of chemical structures in text or file; **C**. BPPM accepts biological profiles in file; **D**. Repositioning score, label and functional network of compounds were displayed in result; **E**. Drug repositioning mechanism analysis was implemented to infer the drug potential repositioning mechanism through relationships among similar compounds, fibrosis-related targets and diseases.

Supplementary material

Figure S1 Knowledgebase architecture of Dr AFC

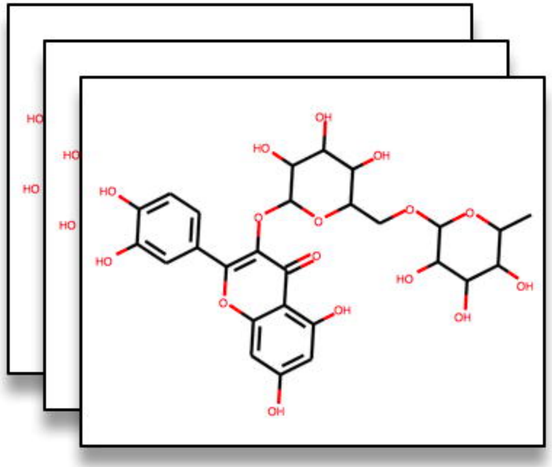
Figure S2 Performances of top 30 features through iterative feature elimination in SPPM

Table S1 The sample size of SPPM and BPPM

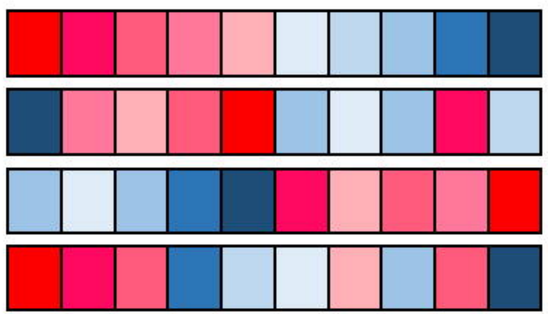
Table S2 Performances of four different machine learning classifiers

Table S3 Drug repositioning prediction in Drugbank Experimental drugs

Table S4 Drug repositioning prediction in natural products(GSE85871)



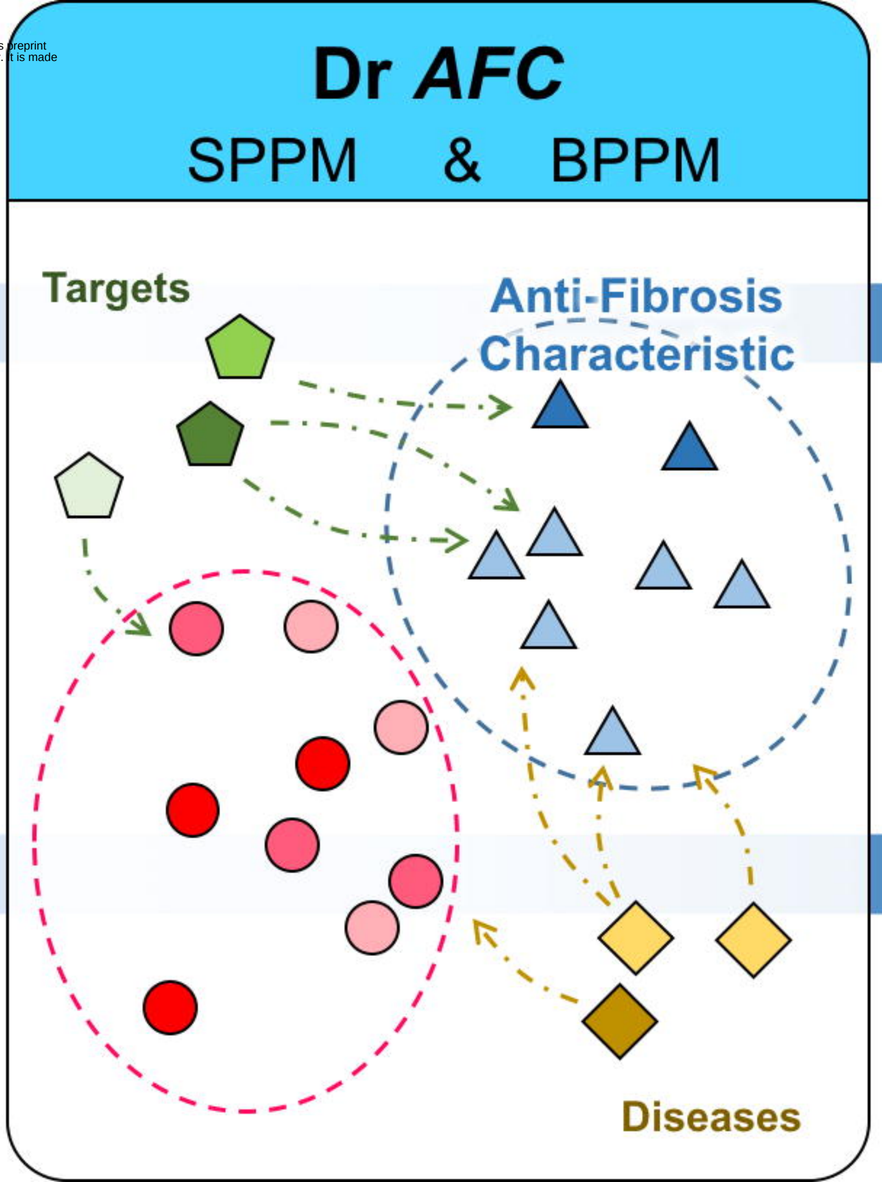
Structural Profile

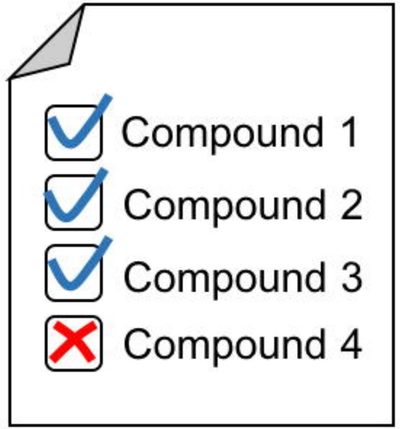


Biological Profile

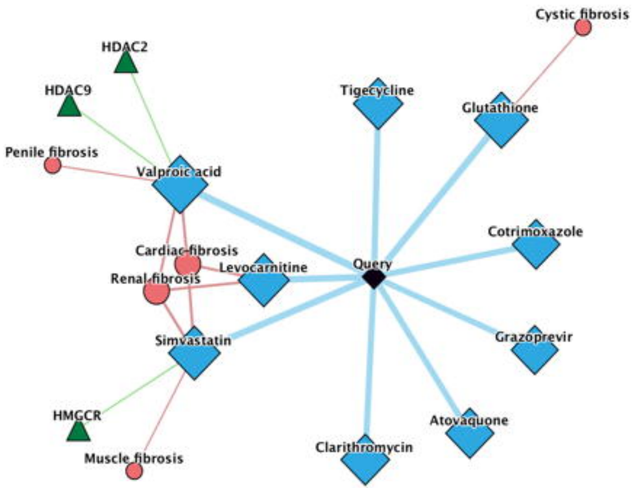
Dr *AFC*

SPPM & BPPM



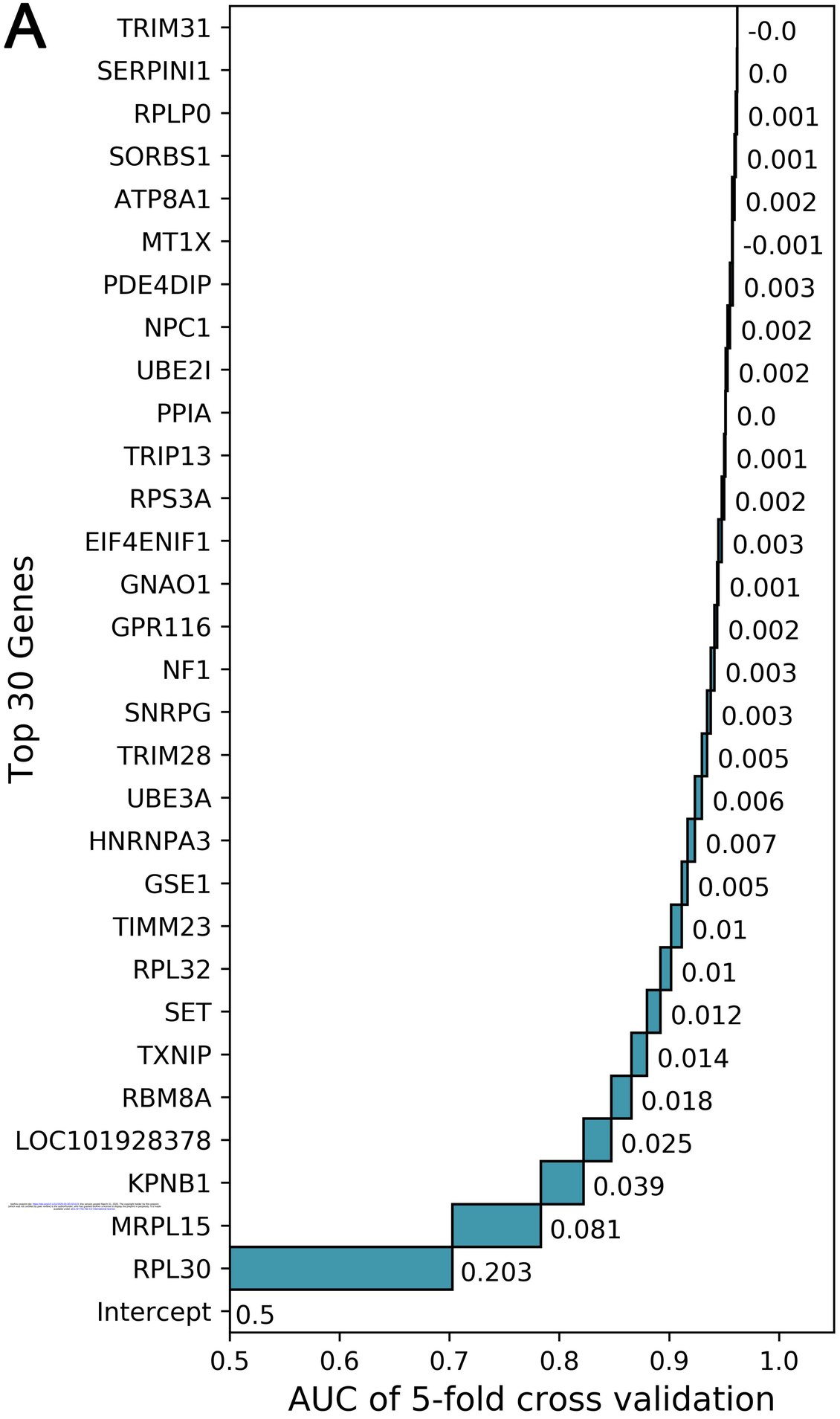


Anti-Fibrosis prediction

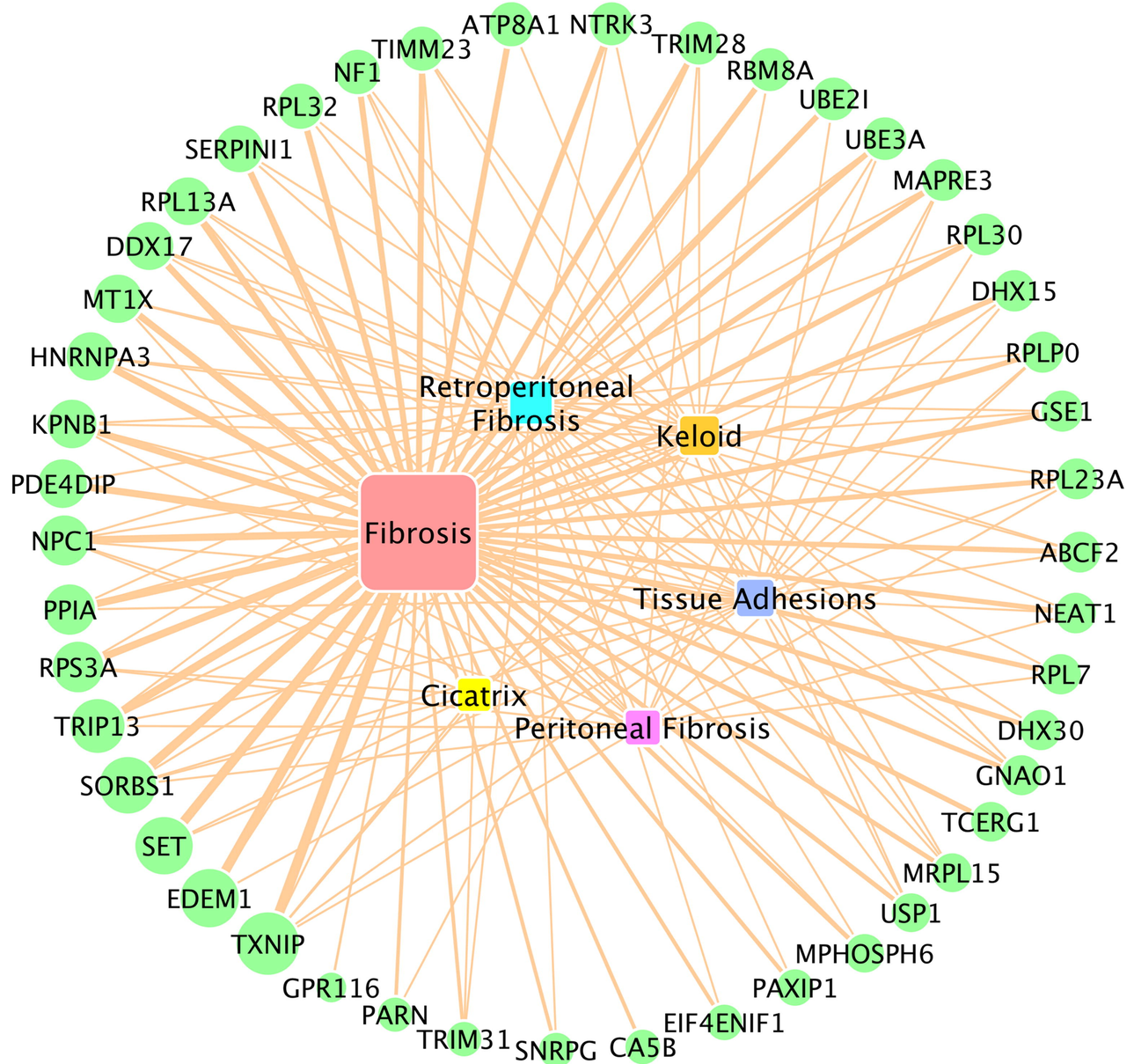


Drug repositioning mechanism

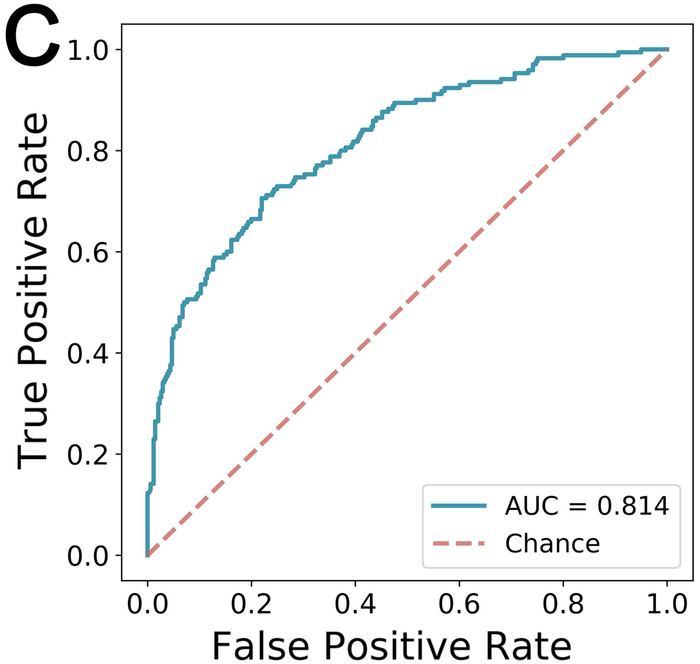
A


Top 30 Genes
TRIM31
SERPINI1
RPLP0
SORBS1
ATP8A1
MT1X
PDE4DIP
NPC1
UBE2I
PPIA
TRIP13
RPS3A
EIF4ENIF1
GNAO1
GPR116
NF1
SNRPG
TRIM28
UBE3A
HNRNPA3
GSE1
TIMM23
RPL32
SET
TXNIP
RBM8A
LOC101928378
KPNB1
MRPL15
RPL30
Intercept
-0.0
0.0
0.001
0.001
0.002
-0.001
0.003
0.002
0.002
0.0
0.001
0.002
0.003
0.001
0.002
0.003
0.003
0.005
0.006
0.007
0.005
0.01
0.01
0.012
0.014
0.018
0.025
0.039
0.081
0.203
0.5
0.5
0.6
0.7
0.8
0.9
1.0
AUC of 5-fold cross validation

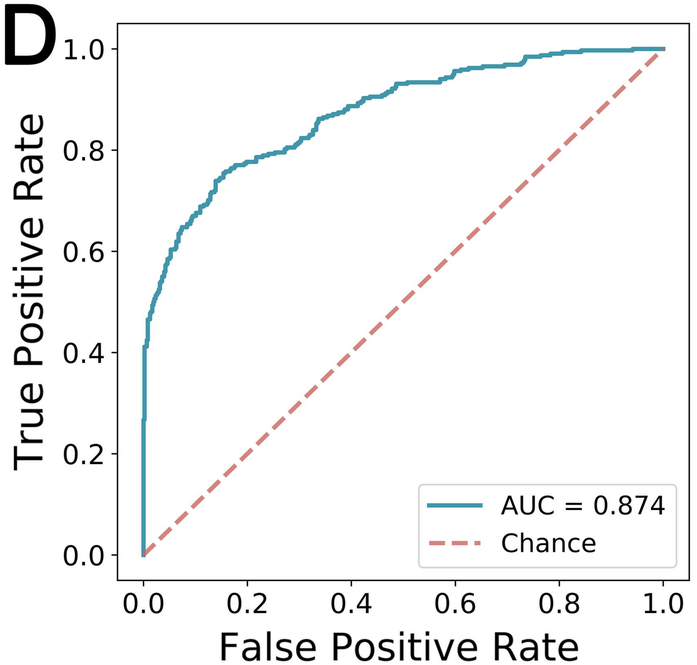
B

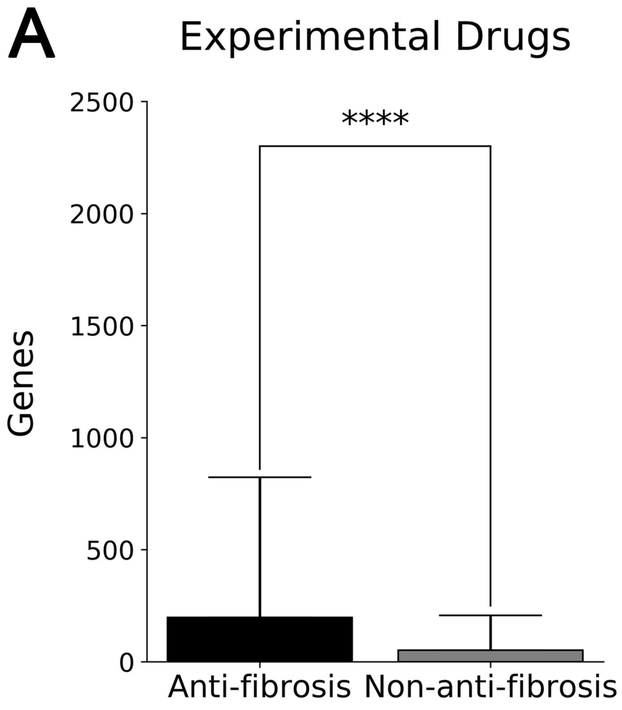

ATP8A1
NTRK3
TRIM28
TIMM23
NF1
RBM8A
UBE2I
RPL32
UBE3A
SERPINI1
MAPRE3
RPL13A
RPL30
DDX17
DHX15
MT1X
RPLP0
HNRNPA3
Retroperitoneal Fibrosis
GSE1
KPNB1
Keloid
RPL23A
PDE4DIP
Fibrosis
ABCF2
NPC1
Tissue Adhesions
NEAT1
PPIA
RPL7
RPS3A
Cicatrix
DHX30
TRIP13
Peritoneal Fibrosis
GNAO1
SORBS1
TCERG1
SET
MRPL15
EDEM1
USP1
TXNIP
MPHOSPH6
GPR116
PAXIP1
PARN
EIF4ENIF1
TRIM31
SNRPG
CA5B

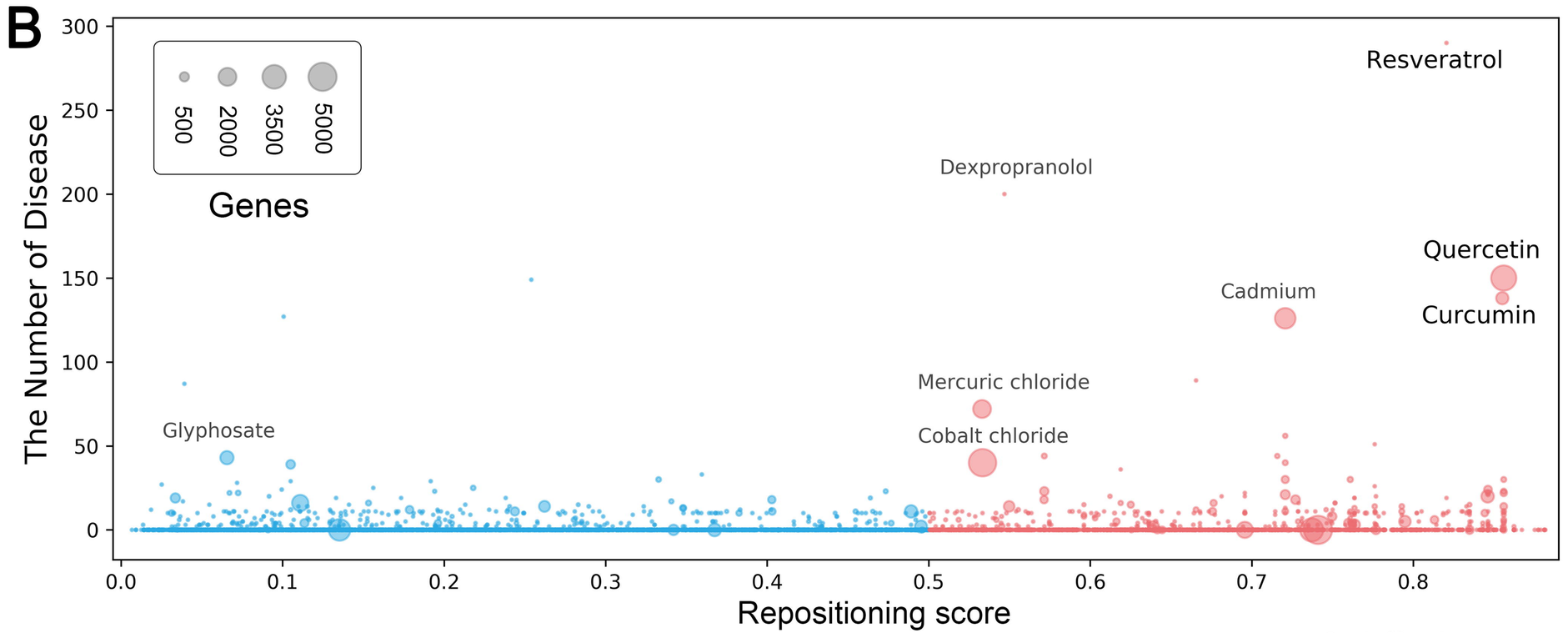
C

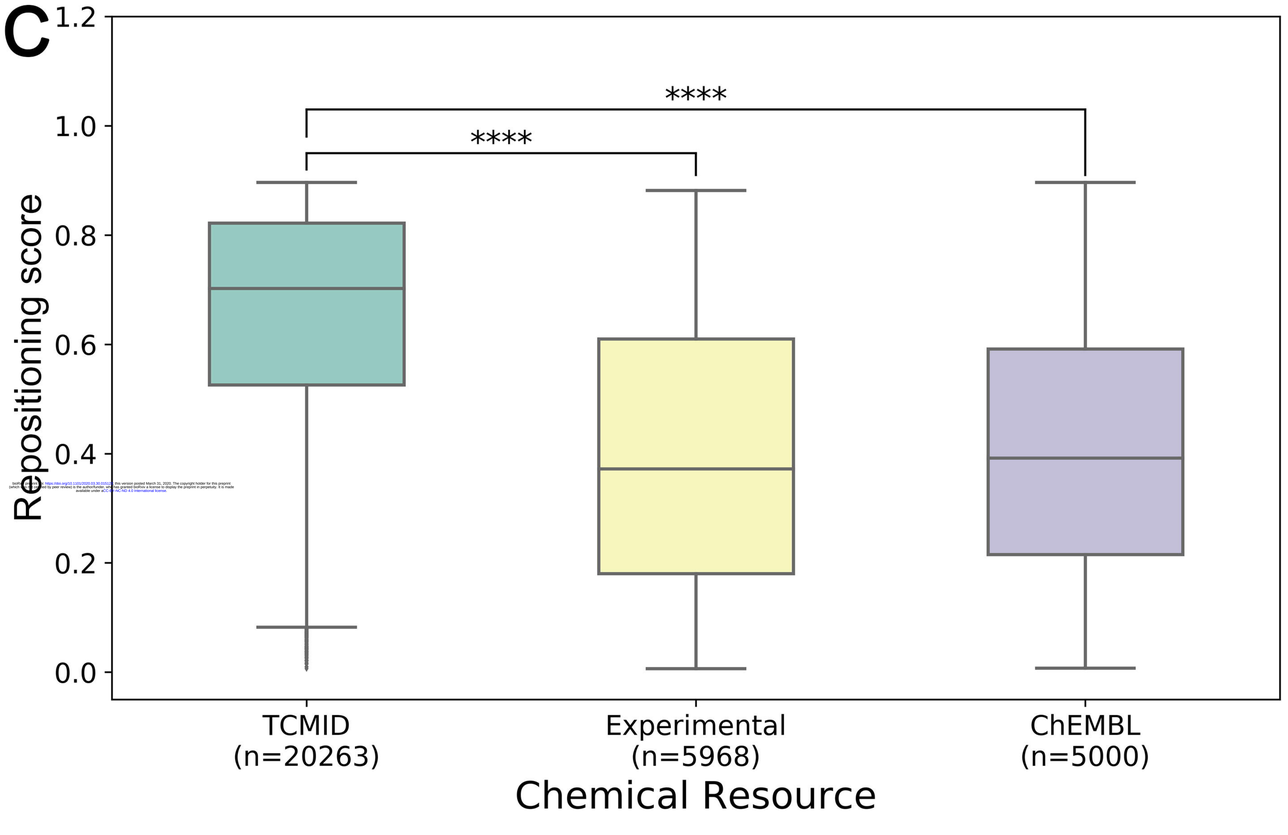

True Positive Rate
False Positive Rate
AUC = 0.814
Chance
0.0
0.2
0.4
0.6
0.8
1.0

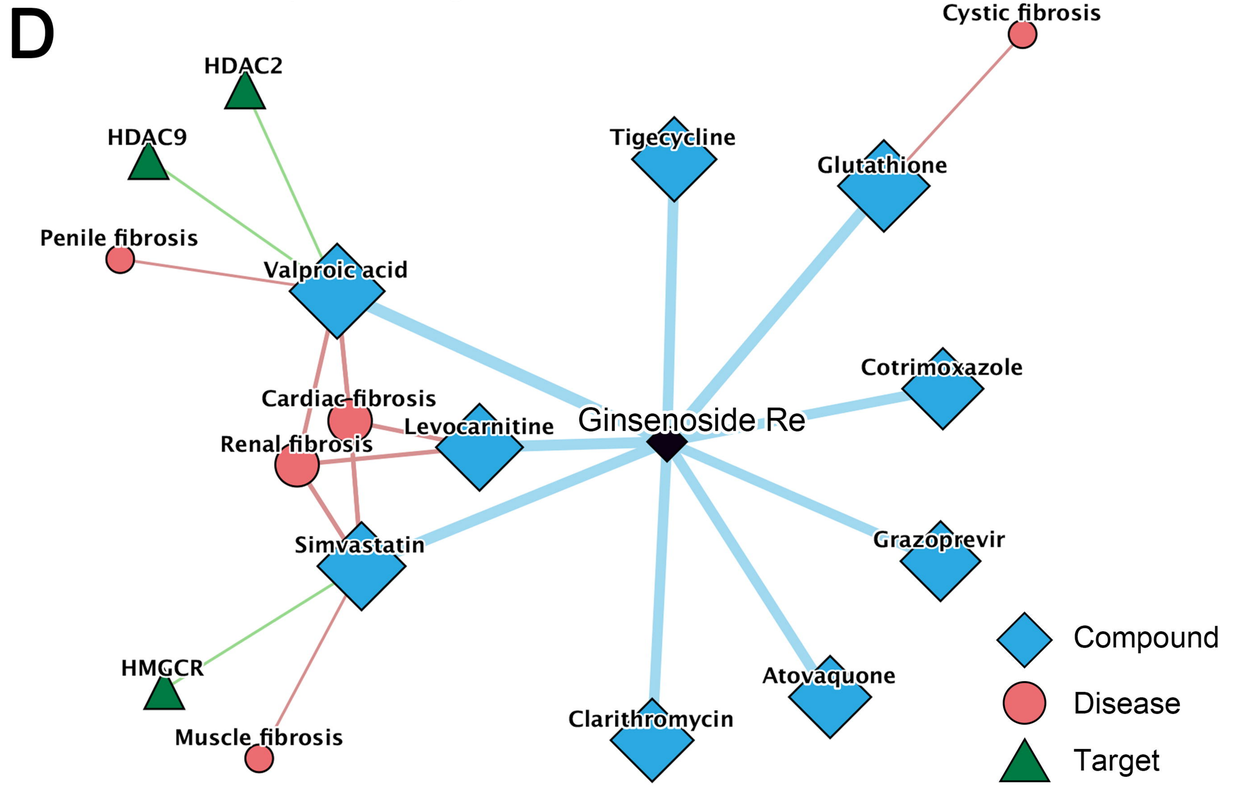
D


True Positive Rate
False Positive Rate
AUC = 0.874
Chance
0.0
0.2
0.4
0.6
0.8
1.0


A
Experimental Drugs
****
Genes
2500
2000
1500
1000
500
0
Anti-fibrosis
Non-anti-fibrosis


B
500
2000
3500
5000
Genes
The Number of Disease
300
250
200
150
100
50
0
Resveratrol
Dexpropranolol
Quercetin
Curcumin
Cadmium
Mercuric chloride
Cobalt chloride
Glyphosate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Repositioning score


C
****
****
Repositioning score
1.2
1.0
0.8
0.6
0.4
0.2
0.0
TCMID
(n=20263)
Experimental
(n=5968)
ChEMBL
(n=5000)
Chemical Resource


D
Cystic fibrosis
HDAC2
HDAC9
Tigecycline
Glutathione
Penile fibrosis
Valproic acid
Cotrimoxazole
Cardiac fibrosis
Levocarnitine
Ginsenoside Re
Renal fibrosis
Simvastatin
Grazoprevir
HMGCR
Atovaquone
Clarithromycin
Muscle fibrosis
Compound
Disease
Target

Dr AFC | Tools | Browse | Download | About | Help | Contact Us

A

Dr AFC: Drug Repositioning based on Anti-Fibrosis Characteristic

Structural Profile Prediction Model

Biological Profile Prediction Model

B

Structural Profile Prediction Model

Upload

A. Paste a List

BGP-15 C1CCN(CC1)CC(CON=C(C2=CN=CC=C2)N)O

B. Choose From a File

浏览... 未选择文件。

Please ensure the List Format is Compound Name + SMILES separated by TAB, and File Format is like example.

Submit Example

STRUCTURE FEATURE (AUC)

C

Biological Profile Prediction Model

Upload

A. Choose From a File

浏览... 未选择文件。

B. Select Signatures

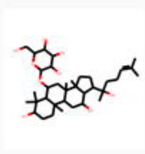
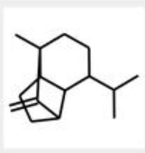
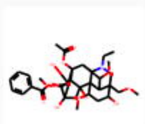
Affymetrix U133A

Please ensure the upload File follows the Rank Order of the example.

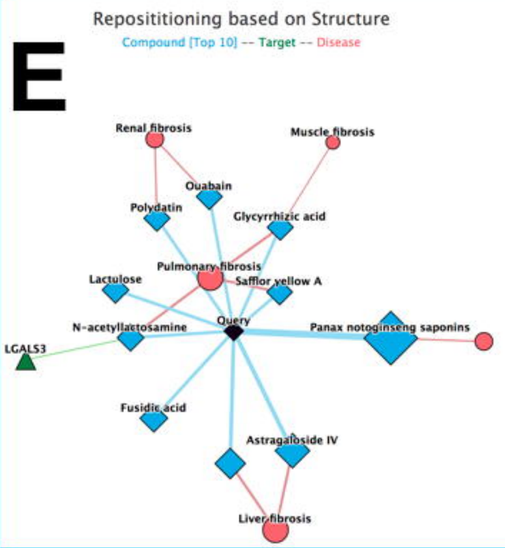
Submit Example

GENE EXPRESSION RANK PROFILE

D

NUMBER	NAME	STRUCTURE	SMILES	SCORE	ANTI-FIBROSIS	MECHANISM
3	ginsenoside rh1		CC(=CCCC(C)(C1CCC2(C1C(CC3C2(CC(C4C3(CCC(C4(C)C)O)C)OC5C(C(C(O5)CO)O)O)O)C)O)C)O)C	0.738	YES	✔
4	(+)-sativen		CC(C)C1CCC2(C3C1C(C2=C)CC3)C	0.433	NO	✖
5	10-aconifine		CCN1CC2(C(CC(C34C2C(C(C31)C5(C6C4(CC(C6OC(=O)C7=CC=CC=C7)(C(C5O)OC)O)O)OC(=O)C)OC)OC)O)COC	0.182	NO	✖

E



Similarity threshold 0.25 Download

Nodes Details

[Type]: Disease

Pulmonary fibrosis

Processes: fibroblasts proliferation, fibroblasts differentiation; prevenet collagen deposition;

Mechanism: nan

1. Wang L, Jin M, Zang B X, et al. Inhibitory effect of safflor yellow on pulmonary fibrosis[J]. Biological and Pharmaceutical Bulletin, 2011, 34(4): 511-516.

2. Zhang D, Liu B, Cao B, et al. Synergistic protection of Schizandrin B and Glycyrrhizic acid against bleomycin-induced pulmonary fibrosis by inhibiting TGF-β1/Smad2 pathways and overexpression of NOX4[J]. International Immunopharmacology, 2017, 48: 67-75.

3. Luo H, Liu B, Zhao L, et al. Galectin-3 mediates pulmonary vascular remodeling in hypoxia-induced pulmonary arterial hypertension[J]. Journal of the American Society of Hypertension, 2017, 11(10): 673-683. e3.