Original Article

Identification of drug compounds for keloids and hypertrophic scars: drug discovery based on text mining and DeepPurpose

Yuyan Pan[1#], Zhiwei Chen[2#], Fazhi Qi[1], Jiaqi Liu[1,3]

[1]Department of Plastic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China; [2]Big Data and Artificial Intelligence Center, Zhongshan Hospital, Fudan University, Shanghai, China; [3]Artificial Intelligence Center for Plastic Surgery and Cutaneous Soft Tissue Cancers, Zhongshan Hospital, Fudan University, Shanghai, China

Contributions: (I) Conception and design: Y Pan; (II) Administrative support: F Qi, J Liu; (III) Provision of study materials or patients: F Qi, J Liu; (IV) Collection and assembly of data: Y Pan, Z Chen; (V) Data analysis and interpretation: Y Pan, Z Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jiaqi Liu, PhD; Fazhi Qi, PhD. Department of Plastic and Reconstructive Surgery, Zhongshan Hospital, Fudan University, 180 Fenglin Rd, Shanghai 200032, China. Email: liujiaqi1213@yahoo.com; qi.fazhi@zs-hospital.sh.cn.

Background: Keloids (KL) and hypertrophic scars (HS) are forms of abnormal cutaneous scarring characterized by excessive deposition of extracellular matrix and fibroblast proliferation. Currently, the efficacy of drug therapies for KL and HS is limited. The present study aimed to investigate new drug therapies for KL and HS by using computational methods.

Methods: Text mining and GeneCodis were used to mine genes closely related to KL and HS. Protein-protein interaction analysis was performed using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and Cytoscape. The selection of drugs targeting the genes closely related to KL and HS was carried out using Pharmaprojects. Drug-target interaction prediction was performed using DeepPurpose, through which candidate drugs with the highest predicted binding affinity were finally obtained.

Results: Our analysis using text mining identified 69 KL- and HS-related genes. Gene enrichment analysis generated 25 genes, representing 7 pathways and 130 targeting drugs. DeepPurpose recommended 14 drugs as the final drug list, including 2 phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) inhibitors, 10 prostaglandin-endoperoxide synthase 2 (PTGS2) inhibitors and 2 vascular endothelial growth factor A (VEGFA) antagonists.

Conclusions: Drug discovery using *in silico* text mining and DeepPurpose may be a powerful and effective way to identify drugs targeting the genes related to KL and HS.

Keywords: Keloids (KL); hypertrophic scars (HS); text mining; DeepPurpose; drug-target interaction; drug therapy

Submitted Dec 15, 2020. Accepted for publication Jan 29, 2021.
doi: 10.21037/atm-21-218
View this article at: http://dx.doi.org/10.21037/atm-21-218

Introduction

Keloids (KL) and hypertrophic scars (HS) are fibroproliferative disorders caused by abnormal wound healing following dermal injury. These scars form due to fibroblast proliferation and are characterized by excessive collagen accumulation (1). There is great variation in the epidemiology of KL and HS depending on the study population; for instance, the incidence among African and Hispanic populations ranges from 4.5–16%, compared with only 0.09% in England (2). Aside from the unpleasant symptoms of HS, such as itching, pain, erythema, and functional damage, its unsightly appearance can cause psychological pain for patients, affecting their quality of life (3).

Currently, treatments for KL and HS include drug injections, surgical excision, laser therapy, radiotherapy, pressure therapy, and cryotherapy. However, corticosteroid

injections can produce side effects such as skin atrophy and telangiectasia. Furthermore, the rate of recurrence among keloid patients treated with surgical excision combined with radiotherapy has been reported to be 21%, with none in craniofacial locations (4). Other therapies may also cause side effects and have unsatisfactory effectiveness (5). However, the molecular mechanism underlying scar formation still needs to be elucidated, and successful treatment of KL and HS remains a challenge.

It takes more than 10 years to discover and develop a new drug, at an average cost exceeding 2.6 billion US dollars. However, new therapeutic purposes for existing drugs may be discovered through drug repositioning (6,7). Drug-target interactions (DTIs) measure the binding affinity of drug molecules to protein targets (8). Therefore, computational methods that can obtain knowledge about the interaction between compounds and target proteins are important in drug research and discovery (R&D). Computer simulation methods can speed up the drug research and development process by systematically prioritizing the most effective compounds. Recently, deep learning (DL) technology has been demonstrated to have the potential to predict compound–protein interactions on a large scale by learning from limited data, and it has been successfully applied in the R&D of new drugs, in which it significantly shortened the time and cost (9,10).

Our previous studies demonstrated that drug discovery using *in silico* text mining and pathway analysis tools may be a method to explore candidate drugs targeting the genes and pathways associated with certain diseases. In this study, we utilized DeepPurpose, a powerful Python toolkit, which presented the most likely drug candidates based on our previous work. DeepPurpose processes the input target amino acid sequences and candidate drug codes by feeding the data into multiple latest deep learning models pre-trained on DAVIS, BindingDB-Kd, and kinase inhibitor bioactivity (KIBA) datasets (11-13). The prediction results are then integrated by DeepPurpose to generate a ranked list, with the drug candidates with the highest predicted binding scores positioned at the top. The top-ranked drug candidates are considered to possess the potential for experimental verification.

DeepPurpose presents the DTI model as an encoder-decoder framework to predict drug-target interactions. Taking the simplified molecular-input line-entry system (SMILES) format of the drug and the target amino acid sequence pair as input, DeepPurpose outputs the score of the binding affinity between the drug and the molecule. For drug molecules, DeepPurpose provides 8 encoders: Morgan, PubChem, Daylight, RDKit 2D, convolutional neural network (CNN), convolutional recurrent neural network (CNN+RNN), Transformer encoders, and Message-Passing Neural Network (MPNN). For protein targets, DeepPurpose provides 7 encoders: amino acid composition (AAC), PseACC, Conjoint Triad, Quasi Sequence, CNN, CNN+RNN, and Transformer (14).

In this study, we investigated new drug therapies for KL and HS by employing computational methods. First, we performed text mining, biological process and pathway analysis, and protein-protein interaction (PPI) analysis to explore the target genes and pathways highly relevant to KL and HS. DTI analysis was then performed to obtain candidate drugs. Finally, DeepPurpose was used to predict the interaction of candidate drugs and gene targets, and the drugs with the highest predicted binding affinity from a ranked list were obtained.

We present the following article in accordance with the MDAR checklist (available at http://dx.doi.org/10.21037/atm-21-218).

Methods

Text mining

In this study, text mining, a process in which high-quality information is derived from biological literature, was performed using pubmed2ensembl (http://pubmed2ensembl.ls.manchester.ac.uk/) (15). The following terms were used as search input: “keloid”, “hypertrophic scar”, “hyperplastic scar”, and “scar hypertrophy”. We chose “Homo sapiens” as the species dataset, then selected “Ensembl Gene ID” and “Associated Gene Name” under GENE. “Search for PubMed IDs” and “filter on Entrez: PMID” drop-down menus were chosen in the search of every query. The intersection of the 4 derived gene lists was used for the next step. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Biological process and pathway analysis

GeneCodis (http://genecodis.cnb.csic.es/) was used to perform enrichment analysis on genes closely related to KL and HS (16). First, the genes identified through text mining were subjected to Gene Ontology (GO) biological process analysis. The most significantly enriched genes in biological processes were selected for Kyoto Encyclopedia

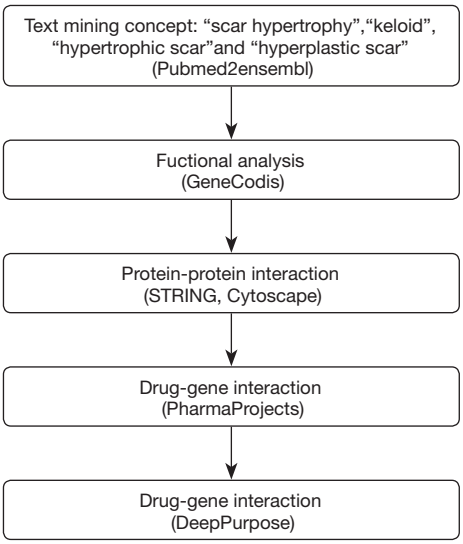

Figure 1 Overall data mining process. Text mining and GeneCodis were used to identify genes related to keloids and hypertrophic scars (KL and HS). Protein-protein interaction analysis was performed in STRING and Cytoscape. Drugs targeting the genes highly related to KL and HS were selected using Pharmaprojects. Based on the drug-target interaction analysis by DeepPurpose, candidate drugs with highest predicted binding affinity were finally derived.

of Genes and Genomes (KEGG) pathway analysis. The most significantly enriched KEGG pathways were selected, and genes associated with the selected pathways were used for further analysis.

Protein-protein interaction network

We used the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (http://string-db.org) to construct a protein-protein interaction (PPI) network in order to visualize the genes from the previous step (17). The genes were input under the “Multiple proteins” menu, and “Homo sapiens” was selected as the species dataset. To obtain the genes with strong interactions, we set a high confidence score of 0.700, and the PPI network of the target genes was generated. Then, the CentiScape plugin in Cytoscape was used to determine the centrality parameters of the PPI network (18). “Degree” and “Betweenness” were chosen as the parameters for the selection of key genes in this study. Degree represents the total number of edges incident to the node, and betweenness refers to the number of shortest paths through the node.

Drug-gene interactions

Drugs targeting the genes highly related to KL and HS were searched for using Pharmaprojects (https://pharmaintelligence.informa.com/) (19). Each gene query returned a drug list detailing the global status, disease, mechanism of action, delivery route, target, chemical structure (SMILES format), and other information about drugs. Drugs with “launched”, “phase I/II/III clinical trial”, “pre-registration”, or “registered” as the global status were screened out, and those with the delivery route of “oral” or “oral, swallowed” were also excluded. These criteria allowed us to obtain candidate drugs with targeting ability, quick onset of action, and few side effects. Drugs derived from the DTI analysis may be candidates for KL and HS treatment.

DeepPurpose

In order to utilize DeepPurpose, we first translated the target proteins into amino acid sequences and the potential drugs into SMILES fingerprints. Taking the sequences and fingerprints as input, we used the pre-trained models provided by DeepPurpose to predict the binding affinity between each paired drug molecule and protein target of interest. As DeepPurpose provides 15 pre-trained models, we predicted the binding affinity score for each pre-trained model individually and screened the potential drug-target interaction by setting appropriate thresholds. We validated the results using the validation set we collected. We also calculated aggregated binding affinity scores with the aggregation schema proposed by DeepPurpose. Finally, the differences in the predicted binding affinity scores obtained using single models and aggregate models were analyzed.

Statistical analysis

Statistical analyses were carried out using machine learning algorithm in DeepPurpose.

Results

Results of text mining, biological process, and KEGG pathway analysis

Through the data mining process described in *Figure 1*, 135 genes relating to “scar hypertrophy”, “keloid”, “hypertrophic scar”, and “hyperplastic scar” were found. After deleting the duplicates, we were left with 69 genes (*Figure 2*). In the analysis of enriched GO biological process

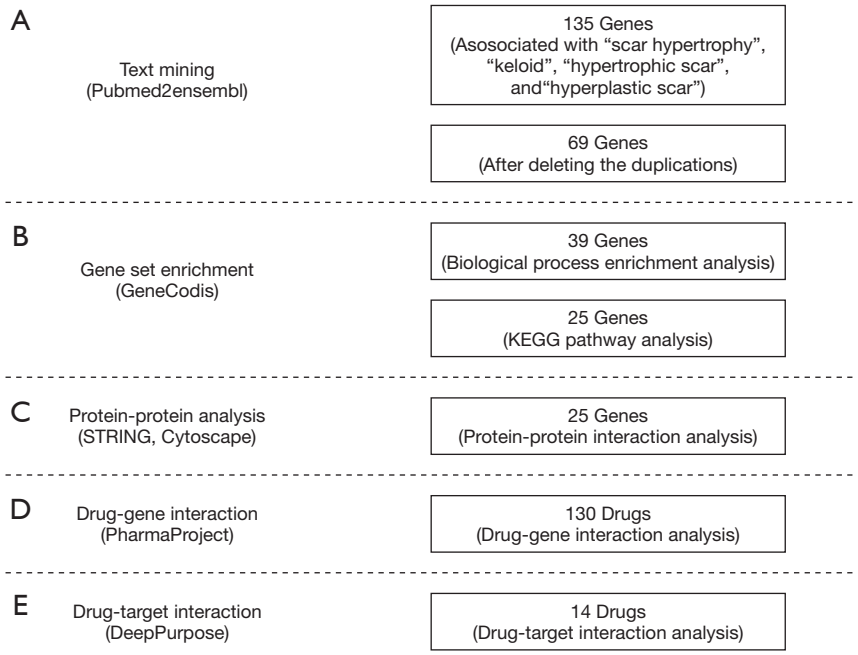

Figure 2 Summary of data mining results. (A) Text mining: 135 genes were found to be associated with "scar hypertrophy", "keloid", "hypertrophic scar", and "hyperplastic scar" using pubmed2ensembl. Sixty-nine genes remained after deletion of the duplicates. (B) Gene set enrichment: GeneCodis biological processes and pathway analysis generated 39 and 25 genes, respectively. (C) Protein-protein interaction analysis was performed using STRING and Cytoscape. (D) Drug-gene interaction: 130 targeting drugs were selected by Pharmaprojects. (E) Drug-target interaction: the 14 candidate drugs with highest predicted binding affinity were finally derived.

annotations, the P value cutoff (P=1.00e-11) was set to select the most enriched biological processes relevant to the pathology of KL and HS, which resulted in 7 sets of annotations containing 39 genes (*Table 1*). The 5 most enriched biological process annotations were: "positive regulation of epithelial to mesenchymal transition" (P=1.41E-13), the "transforming growth factor beta receptor signaling pathway" (P=2.67E-13), the "cytokine-mediated signaling pathway" (P=4.17E-13), "wound healing" (P=4.42E-12), and "pathway-restricted SMAD protein phosphorylation" (P=1.54E-11). For the KEGG pathway analysis, the P value cutoff was set to P=1.00e-14, which resulted in 25 genes in 7 pathways above the cutoff (*Table 2*). The top 3 most enriched biological process annotations were: the "AGE-RAGE signaling pathway in diabetic complications" (P=1.71E-21), "pathways in cancer" (P=5.43E-16), and the "TGF-beta signaling pathway" (P=8.08E-16).

Results of PPI network analysis

The PPIs of the 25 target genes were analyzed using the STRING database (*Figure 3*). Data from STRING were then input into Cytoscape to generate the PPI network (*Figure 4*). In CentiScaPe, the average values of the 2 important centrality parameters, degree and betweenness, were 10.00 and 15.44, respectively. The final gene list included "*CDKN1B*", "*VEGFA*", "*TNF*", "*TGFBR1*", "*TGFBR2*", "*TGFB1*", "*TGFB2*", "*TGFB3*", "*STAT3*", "*PIK3CA*", "*MMP2*", "*SMAD2*", "*SMAD3*", "*IL6*", "*IL6R*", "*FN1*", "*COL1A1*", "*COL1A2*", "*TP53*", "*SP1*", "*PTGS2*", "*MMP9*", "*HGF*", "*FGF2*", and "*FGF7*".

Results of drug-gene interaction analysis

A total of 130 drugs targeting the final gene list were initially selected as possible treatments for KL and HS. These drugs included 30 vascular endothelial growth factor A (VEGFA) receptor antagonists, 27 prostaglandin-endoperoxide synthase 2 (PTGS2) inhibitors, 15 tumor necrosis factor alpha (TNF-α) antagonists, 14 transforming growth factor beta 1 (TGF-β1) antagonists, 8 hepatocyte growth factor (HGF) receptor agonists, 8 interleukin (IL)-6 antagonists, 7 IL-6 receptor (IL-6R) antagonists,

Table 1 Summary of biological process gene set enrichment analysis

Process	Genes in query set	Corrected hypergeometric P value	Genes
Positive regulation of epithelial to mesenchymal transition	10	1.41E-13	*TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1I1, TGFB1, SMAD3, SMAD2, IL6, COL1A1*
Transforming growth factor beta receptor signaling pathway	12	2.67E-13	*TP53, TGFBR3, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, SMAD7, SMAD6, SMAD3, SMAD2, COL1A2*
Cytokine-mediated signaling pathway	16	4.17E-13	*VEGFA, TP53, TNFRSF1B, TNF, TGFB1, STAT3, PTGS2, PIK3CA, MMP9, MMP2, IL6R, IL6, HGF, FN1, FGF2, COL1A2*
Wound healing	11	4.42E-12	*POSTN, TGFBR2, TGFBR1, TGFB3, TGFB2, SMAD3, SMAD2, TNC, FN1, FGF2, COL1A1*
Pathway-restricted SMAD protein phosphorylation	5	1.54E-11	*TGFBR3, TGFBR2, TGFBR1, TGFB1, SMAD7*
Negative regulation of cell population proliferation	17	1.72E-11	*CDKN1B, TP73, TP53, TIMP2, TGFBR2, TGFB3, TGFB2, TGFB1I1, TGFB1, STAT3, SOD2, PTGS2, SMAD6, SMAD3, SMAD2, IL6, DPT*
Positive regulation of pri-miRNA transcription by RNA polymerase II	8	4.52E-11	*TP53, TNF, TGFB2, TGFB1, STAT3, SMAD6, SMAD3, FGF2*

The most significantly enriched biological processes relevant to the pathology of keloids and hypertrophic scars above the P value cutoff (P=1.00E-11) were selected. The analysis of enriched biological processes resulted in 7 sets of annotations containing 39 genes. TGFBR2, transforming growth factor beta receptor 2; TGFBR1, transforming growth factor beta receptor 1; TGFB3, transforming growth factor beta 3; TGFB2, transforming growth factor beta 2; TGFB1I1, transforming growth factor beta 1 included transcript 1; TGFB1, transforming growth factor beta 1; SMAD3, mothers against decapentaplegic homolog 3; SMAD2, mothers against decapentaplegic homolog 2; IL6, interleukin 6; COL1A1, collagen type I alpha 1; TP53, tumor protein 53; TGFBR3, transforming growth factor beta receptor 3; SMAD7, mothers against decapentaplegic homolog 7; SMAD6, mothers against decapentaplegic homolog 6; COL1A2, collagen type I alpha 2; VEGFA, vascular endothelial growth factor A; TNFRSF1B, tumor necrosis factor receptor superfamily member 1B; TNF, tumor necrosis factor; STAT3, signal transducer and activator of transcription 3; PTGS2, prostaglandin-endoperoxide synthase 2; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; MMP9, matrix metalloprotein 9; MMP2, matrix metalloprotein 2; IL6R, interleukin 6; HGF, hematopoietic growth factor; FN1, fibronectin 1; FGF2, fibroblast growth factor 2; POSTN, periostin; TNC, tenascin C; CDKN1B, cyclindependent kinase inhibitor 1B; TP73, tumor protein 73; TIMP2, metallopeptidase inhibitor 2; SOD2, superoxide dismutase 2; DPT, dermatopontin.

5 fibroblast growth factor (FGF2) agonists, 5 TGF-β1 antagonists, 5 PI3 kinase inhibitors, 4 STAT 3 inhibitors, 1 matrix metalloproteinase-9 (MMP-9) inhibitor and 1 TGF-β3 antagonist.

Results of DeepPurpose analysis

DeepPurpose requires drug molecules to be in the SMILES format, so 34 pharmaceutical compounds with SMILES structure were selected for DeepPurpose analysis. Subsequently, each pre-trained model in DeepPurpose generated a ranked list showing the predicted binding affinity between the drugs and molecules (*Table 3*). A threshold of pKd ≥7.0 was used for models based on the DAVIS and the BindingDB datasets, while for models based on the KIBA dataset, the threshold was set to 12.1.

For the generation of the final outcomes, DeepPurpose proposed 3 aggregation schemas—the mean, max, and average of the max and mean—to combine the predictions from different models. We applied these schemas separately on the models trained on the same dataset, which gave us 9 additional ranked lists of binding affinity scores. The chosen thresholds were also used to screen potential drug-target pairs (*Table 4*). The final drug list consisted of 14 drugs, including 2 PI3K inhibitors, 10 PTGS2 inhibitors, and 2 VEGFA antagonists (*Table 5*).

Discussion

Keloids (KL) and hypertrophic scars (HS) are common

Table 2 Summary of Kyoto Encyclopedia of Genes and Genomes (KEGG) process gene set enrichment analysis

Process	Genes in query set	Corrected hypergeometric P value	Genes
AGE-RAGE signaling pathway in diabetic complications	17	1.71E-21	*CDKN1B, VEGFA, TNF, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, STAT3, PIK3CA, MMP2, SMAD3, SMAD2, IL6, FN1, COL1A2, COL1A1*
Pathways in cancer	22	5.43E-16	*CDKN1B, VEGFA, TP53, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, STAT3, SP1, PTGS2, PIK3CA, MMP9, MMP2, SMAD3, SMAD2, IL6R, IL6, HGF, FN1, FGF7, FGF2*
TGF-beta signaling pathway	8	8.08E-16	*TNF, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, SMAD3, SMAD2*
FoxO signaling pathway	10	8.53E-16	*CDKN1B, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, STAT3, PIK3CA, SMAD3, IL6*
Cytokine-cytokine receptor interaction	7	8.85E-16	*TNF, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, IL6*
Hippo signaling pathway	7	8.85E-16	*TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, SMAD3, SMAD2*
Cellular senescence	9	6.52E-15	*TP53, TGFBR2, TGFBR1, TGFB3, TGFB2, TGFB1, PIK3CA, SMAD3, SMAD2*

The most significantly enriched KEGG pathways relevant to the pathology keloids and hypertrophic scars above the P value cutoff (P=1.00E-14) were selected. The analysis of enriched pathway annotations resulted in 7 sets of annotations containing 25 genes. VEGFA, vascular endothelial growth factor A; CDKN1B, cyclindependent kinase inhibitor 1B; TGFBR2, transforming growth factor beta receptor 2; TGFBR1, transforming growth factor beta receptor 1; TGFB3, transforming growth factor beta 3; TGFB2, transforming growth factor beta 2; TGFB1, transforming growth factor beta 1; SMAD3, mothers against decapentaplegic homolog 3; SMAD2, mothers against decapentaplegic homolog 2; IL6, interleukin 6; COL1A1, collagen type I alpha 1; TP53, tumor protein 53; COL1A2, collagen type I alpha 2; TNF, tumor necrosis factor; STAT3, signal transducer and activator of transcription 3; PTGS2, prostaglandin-endoperoxide synthase 2; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; MMP9, matrix metalloprotein 9; MMP2, matrix metalloprotein 2; IL6R, interleukin 6; HGF, hematopoietic growth factor; FN1, fibronectin 1; FGF7, fibroblast growth factor 7; FGF2, fibroblast growth factor 2; SP1, specificity protein 1.

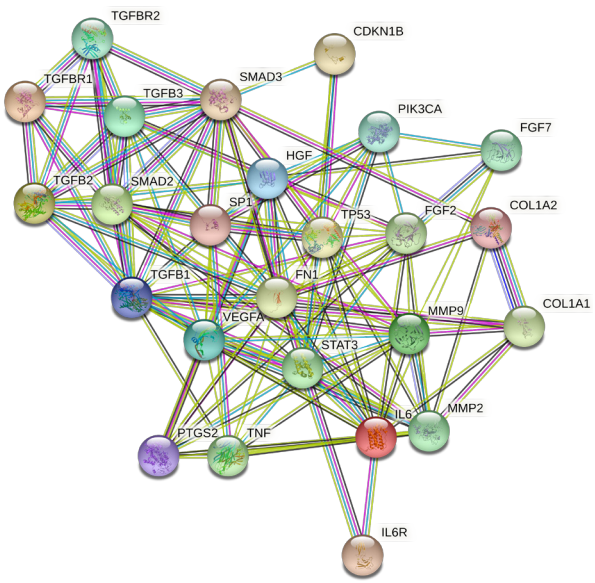

Figure 3 The protein-protein interaction (confidence score, 0.700) network of the 25 targeted genes, generated using STRING. Network nodes represent proteins, and edges represent protein-protein interactions.

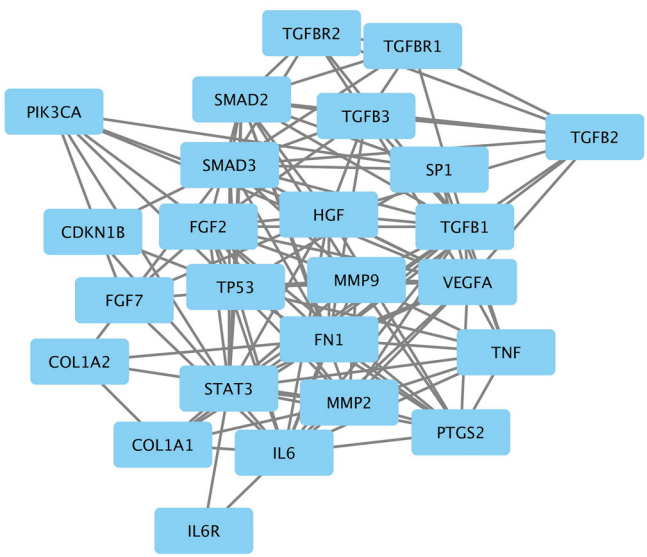

Figure 4 The protein-protein interaction network of the 25 targeted genes, generated by Cytoscape. Network nodes represent proteins and edges represent protein-protein interactions.

Table 3 Identification of drug candidates for keloids and hypertrophic scars by DeepPurpose

Drug name	Target gene	DeepDTA_DAVIS	Morgan_CNN_DAVIS	MPNN_CNN_DAVIS	Daylight_AAC_DAVIS	Morgan_AAC_DAVIS	CNN_CNN_BindingDB	Morgan_CNN_BindingDB	MPNN_CNN_BindingDB	Transformer_CNN_BindingDB	Daylight_AAC_BindingDB	Morgan_AAC_BindingDB	Morgan_CNN_KIBA	MPNN_CNN_KIBA	Daylight_AAC_KIBA	Morgan_AAC_KIBA
NPC-18	*FGF2*	5.161	5.098	3.924	5.178	5.090	6.293	5.319	5.502	5.653	4.939	4.334	10.362	10.634	11.309	10.667
Refanalin	*HGF*	5.123	5.069	5.771	5.457	5.078	6.758	6.732	5.979	5.092	5.313	5.272	11.300	11.622	11.239	11.542
BEBT-908	*PI3KCA*	4.917	5.069	4.915	5.116	5.078	6.618	6.742	5.378	6.928	5.031	5.167	11.333	11.684	11.491	11.601
Bimiralisib	*PI3KCA*	4.901	5.029	5.658	5.839	5.072	5.659	6.621	5.100	7.383	6.480	5.938	11.325	11.352	10.260	11.632
SF-1126	*PI3KCA*	4.937	5.179	4.039	5.124	5.070	7.876	5.287	5.175	6.824	5.225	4.814	11.259	11.396	11.310	11.516
Copanlisib	*PI3KCA*	4.939	5.050	4.728	5.838	5.119	5.673	5.768	5.280	3.866	6.377	4.947	11.472	11.389	11.335	11.668
(S)-flurbiprofen	*PTGS2*	5.308	5.059	6.092	5.548	5.033	5.429	4.548	5.517	3.799	4.418	4.082	11.338	11.754	11.207	11.566
Aceclofenac	*PTGS2*	5.634	5.241	4.632	5.497	5.229	5.202	5.612	4.963	3.689	5.208	5.014	11.463	11.157	11.113	11.477
Azapropazone	*PTGS2*	5.181	5.094	5.942	5.375	5.129	6.863	6.341	5.289	3.799	5.376	4.553	11.776	11.252	11.368	11.604
Betamethasone dipropionate/salicyclic acid	*PTGS2*	5.153	5.806	7.020	5.436	5.060	8.067	7.918	7.324	3.799	5.419	5.444	11.018	12.529	11.270	11.422
Bromfenac	*PTGS2*	5.174	5.053	5.858	5.043	5.074	6.585	5.800	5.066	4.242	4.695	4.903	11.480	11.388	11.285	11.399
Celecoxib	*PTGS2*	5.289	5.076	6.357	5.016	5.033	5.084	6.014	4.969	6.531	5.210	4.747	11.465	11.312	11.286	11.321
Dexketoprofen	*PTGS2*	5.184	5.091	6.088	5.130	5.069	5.545	4.112	5.609	3.799	5.024	3.944	11.347	11.663	11.233	11.488
Diclofenac epolamine	*PTGS2*	5.239	5.132	3.307	5.286	5.092	5.784	6.407	5.109	7.009	5.771	5.169	11.407	11.634	10.917	11.472
Etofenamate	*PTGS2*	5.574	5.052	6.527	5.216	5.043	6.103	6.045	5.140	4.776	4.150	4.642	11.443	11.251	11.397	11.501
Flurbiprofen	*PTGS2*	5.364	5.059	6.344	5.548	5.033	5.872	4.548	5.461	4.562	4.418	4.082	11.338	11.633	11.207	11.566
HTX-011	*PTGS2*	5.383	5.279	5.613	5.481	5.124	6.339	7.277	5.510	3.799	5.310	5.219	11.291	11.583	11.513	11.309
Nimesulide-hyaluronic acid bioconjugate	*PTGS2*	4.995	5.464	5.703	5.200	5.051	6.902	5.590	6.004	5.795	5.230	4.902	10.223	10.043	10.550	10.303
Indometacin	*PTGS2*	5.335	5.071	5.597	5.101	5.049	5.759	6.049	5.473	5.971	4.975	5.308	11.369	11.793	11.304	11.479
Ketorolac	*PTGS2*	5.258	5.055	6.579	5.373	5.071	5.606	5.764	5.121	4.200	5.345	4.413	11.362	11.627	10.589	11.610
Laflunimus	*PTGS2*	5.997	5.049	7.715	5.059	5.120	5.244	6.186	5.413	4.836	4.836	4.921	11.323	11.867	11.786	11.576
Lornoxicam	*PTGS2*	5.334	5.050	4.397	5.386	5.144	6.038	7.186	5.414	3.799	5.013	5.035	10.835	11.205	11.317	11.342
Meloxicam	*PTGS2*	5.607	5.336	5.085	5.652	5.162	6.356	6.506	5.541	3.799	5.367	5.137	11.406	11.651	12.266	11.429
Mesalazine	*PTGS2*	5.493	5.064	4.270	5.037	5.046	4.942	4.258	5.154	5.168	4.759	3.833	11.451	11.464	12.795	11.612
Paracetamol	*PTGS2*	5.281	5.055	4.664	5.017	5.040	4.514	4.963	4.868	4.389	4.844	3.839	11.466	10.973	10.310	11.547
Parecoxib sodium	*PTGS2*	5.459	5.075	6.006	5.546	5.065	5.725	6.621	5.259	5.257	5.542	5.223	11.377	12.262	11.409	11.446
Piroxicam	*PTGS2*	5.419	5.094	5.296	5.860	5.073	5.875	7.150	5.430	3.799	4.993	5.060	10.850	11.647	11.291	10.654
Propacetamol	*PTGS2*	5.032	5.076	5.631	5.068	5.068	5.126	5.429	5.812	4.151	4.378	3.828	11.358	11.474	10.403	11.508
Tiemonium + noramidopyrine	*PTGS2*	5.482	5.833	6.306	5.240	5.101	5.779	6.801	5.674	7.074	4.321	4.897	11.355	11.574	10.611	11.525
Yakuban Tape	*PTGS2*	5.647	5.059	6.478	5.548	5.033	5.874	4.548	5.458	6.738	4.418	4.082	11.338	11.699	11.207	11.566
Pirfenidone	*TGFB1*	4.981	5.049	3.321	5.096	5.070	3.512	4.036	5.272	4.435	4.614	3.839	11.357	10.911	10.654	11.499
Tranilast	*TGFB1*	4.959	5.006	4.952	5.030	5.076	4.427	5.046	5.460	3.377	4.808	3.785	11.874	11.635	11.705	11.692
Pegaptanib octasodium	*VEGFA*	5.034	5.079	3.592	5.260	5.019	7.667	3.645	4.551	7.120	5.893	5.013	11.223	10.153	11.431	11.280
Sunitinib malate	*VEGFA*	5.177	5.611	5.053	5.088	5.154	7.087	6.890	6.101	5.963	5.008	4.993	12.449	11.718	11.862	12.183

DeepPurpose generated a ranked list demonstrating the predicted binding affinity between drugs and target genes. A threshold of pKd ≥7.0 was used for models based on the DAVIS and BindingDB datasets, while the threshold was set to 12.1 for models based on the KIBA dataset. The significant values based on the criteria are in bold. KIBA, kinase inhibitor bioactivity; CNN, convolutional neural network; MPNN, message-passing neural network; AAC, amino acid composition; FGF2, fibroblast growth factor 2; HGF, hematopoietic growth factor; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; PTGS2, prostaglandin-endoperoxide synthase 2; TGFB1, transforming growth factor beta 1; VEGFA, vascular endothelial growth factor A.

Table 4 Identification of drug candidates for keloids and hypertrophic scars by aggregated models

Drug name	Target gene	AVE_DAVIS	MAX_ DAVIS	AVE_MAX_ DAVIS	AVE_ BindingDB	MAX_ BindingDB	AVE_MAX_ BindingDB	AVE_ KIBA	MAX_ KIBA	AVE_MAX_ KIBA
NPC-18	*FGF2*	4.9	5.2	5.0	5.3	6.3	5.8	10.7	11.3	11.0
Refanalin	*HGF*	5.3	5.8	5.5	5.9	6.8	6.3	11.4	11.6	11.5
BEBT-908	*PI3KCA*	5.0	5.1	5.1	6.0	6.9	6.5	11.5	11.7	11.6
Bimiralisib	*PI3KCA*	5.3	5.8	5.6	6.2	7.4	6.8	11.1	11.6	11.4
SF-1126	*PI3KCA*	4.9	5.2	5.0	5.9	7.9	6.9	11.4	11.5	11.4
Copanlisib	*PI3KCA*	5.1	5.8	5.5	5.3	6.4	5.8	11.5	11.7	11.6
(S)-flurbiprofen	*PTGS2*	5.4	6.1	5.8	4.6	5.5	5.1	11.5	11.8	11.6
Aceclofenac	*PTGS2*	5.2	5.6	5.4	4.9	5.6	5.3	11.3	11.5	11.4
Azapropazone	*PTGS2*	5.3	5.9	5.6	5.4	6.9	6.1	11.5	11.8	11.6
Betamethasone dipropionate/salicyclic acid	*PTGS2*	5.7	7.0	6.4	6.3	8.1	7.2	11.6	12.5	12.0
Bromfenac	*PTGS2*	5.2	5.9	5.5	5.2	6.6	5.9	11.4	11.5	11.4
Celecoxib	*PTGS2*	5.4	6.4	5.9	5.4	6.5	6.0	11.3	11.5	11.4
Dexketoprofen	*PTGS2*	5.3	6.1	5.7	4.7	5.6	5.1	11.4	11.7	11.5
Diclofenac epolamine	*PTGS2*	4.8	5.3	5.0	5.9	7.0	6.4	11.4	11.6	11.5
Etofenamate	*PTGS2*	5.5	6.5	6.0	5.1	6.1	5.6	11.4	11.5	11.4
Flurbiprofen	*PTGS2*	5.5	6.3	5.9	4.8	5.9	5.3	11.4	11.6	11.5
HTX-011	*PTGS2*	5.4	5.6	5.5	5.6	7.3	6.4	11.4	11.6	11.5
Nimesulide-hyaluronic acid bioconjugate	*PTGS2*	5.3	5.7	5.5	5.7	6.9	6.3	10.3	10.5	10.4
Indometacin	*PTGS2*	5.2	5.6	5.4	5.6	6.0	5.8	11.5	11.8	11.6
Ketorolac	*PTGS2*	5.5	6.6	6.0	5.1	5.8	5.4	11.3	11.6	11.5
Laflunimus	*PTGS2*	5.8	7.7	6.8	5.2	6.2	5.7	11.6	11.9	11.8
Lornoxicam	*PTGS2*	5.1	5.4	5.2	5.4	7.2	6.3	11.2	11.3	11.3
Meloxicam	*PTGS2*	5.4	5.7	5.5	5.5	6.5	6.0	11.7	12.3	12.0
Mesalazine	*PTGS2*	5.0	5.5	5.2	4.7	5.2	4.9	11.8	12.8	12.3
Paracetamol	*PTGS2*	5.0	5.3	5.1	4.6	5.0	4.8	11.1	11.5	11.3
Parecoxib sodium	*PTGS2*	5.4	6.0	5.7	5.6	6.6	6.1	11.6	12.3	11.9
Piroxicam	*PTGS2*	5.3	5.9	5.6	5.4	7.2	6.3	11.1	11.6	11.4
Propacetamol	*PTGS2*	5.2	5.6	5.4	4.8	5.8	5.3	11.2	11.5	11.3
Tiemonium + noramidopyrine	*PTGS2*	5.6	6.3	5.9	5.8	7.1	6.4	11.3	11.6	11.4
Yakuban Tape	*PTGS2*	5.6	6.5	6.0	5.2	6.7	6.0	11.5	11.7	11.6
Pirfenidone	*TGFB1*	4.7	5.1	4.9	4.3	5.3	4.8	11.1	11.5	11.3
Tranilast	*TGFB1*	5.0	5.1	5.0	4.5	5.5	5.0	11.7	11.9	11.8
Pegaptanib octasodium	*VEGFA*	4.8	5.3	5.0	5.6	7.7	6.7	11.0	11.4	11.2
Sunitinib malate	*VEGFA*	5.2	5.6	5.4	6.0	7.1	6.5	12.1	12.4	12.3

Aggregated models generated a ranked list demonstrating the predicted binding affinity between the drug and the target gene. A threshold of pKd ≥7.0 was used for models based on the DAVIS and the BindingDB datasets, while the threshold was set to 12.1 for models based on KIBA dataset. The significant values based on the criteria were in bold. KIBA, kinase inhibitor bioactivity; AVE, average; MAX, maximum; FGF2, fibroblast growth factor 2; HGF, hematopoietic growth factor; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; PTGS2, prostaglandin-endoperoxide synthase 2; TGFB1, transforming growth factor beta 1; VEGFA, vascular endothelial growth factor A.

Table 5 Candidate drugs targeting genes relevant to keloids and hypertrophic scars

Drug name	Target gene	The highest PKd	Model	Disease
Mesalazine	*PTGS2*	12.795	Daylight_AAC_KIBA	Colitis, ulcerative
Betamethasone dipropionate/salicyclic acid	*PTGS2*	12.529	MPNN_CNN_KIBA	Eczema; inflammatory disease
Sunitinib malate	*VEGFA*	12.449	Morgan_CNN_KIBA	Macular degeneration, age-related, wet; edema, macular, diabetic; retinal vein occlusion
Meloxicam	*PTGS2*	12.266	Daylight_AAC_KIBA	Ankylosing spondylitis, rheumatoid arthritis
Parecoxib sodium	*PTGS2*	12.262	MPNN_CNN_KIBA	Pain, post-operative
SF-1126	*PI3KCA*	7.876	CNN_CNN_BindingDB	Cancer, liver; cancer, myeloma; cancer, neuroblastoma; cancer, solid
Laflunimus	*PTGS2*	7.715	MPNN_CNN_DAVIS	Pain, post-operative; pain, neuropathic, general; spinal cord injury
Pegaptanib octasodium	*VEGFA*	7.667	CNN_CNN_BindingDB	Macular degeneration, age-related, wet; edema, macular, diabetic
Bimiralisib	*PI3KCA*	7.383	Transformer_CNN_BindingDB	Cancer, breast; cancer, CNS; cancer, head and neck; cancer, leukemia, chronic lymphocytic; cancer, lymphoma; cancer, solid; cancer, head and neck; cancer, lymphoma, T-cell, cutaneous; cancer, skin, unspecified; dermatological disease
HTX-011	*PTGS2*	7.277	Morgan_CNN_BindingDB	Pain, postoperative
Lornoxicam	*PTGS2*	7.186	Morgan_CNN_BindingDB	Arthritis, osteo; arthritis, rheumatoid; pain, musculoskeletal; pain, postoperative
Piroxicam	*PTGS2*	7.150	Morgan_CNN_BindingDB	Arthritis, rheumatoid
Tiemonium + noramidopyrine	*PTGS2*	7.074	Transformer_CNN_BindingDB	Gastrointestinal disease; muscle spasm; pain, nociceptive, general
Diclofenac epolamine	*PTGS2*	7.009	Transformer_CNN_BindingDB	Inflammatory disease; pain, musculoskeletal

The final list consisted of 14 drugs which met the criteria of pKd ≥7.0 for models based on DAVIS and BindingDB datasets and pKd ≥12.1 for models based on KIBA dataset. The diseases targeted by the drugs are listed in the table. KIBA, kinase inhibitor bioactivity; CNN, convolutional neural network; MPNN, message-passing neural network; AAC, amino acid composition; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; PTGS2, prostaglandin-endoperoxide synthase 2; VEGFA, vascular endothelial growth factor A.

dermal fibroproliferative disorders, which place a burden on the health of individuals worldwide. However, the pathogeneses of KL and HS have not been elucidated, and current therapeutic approaches have limited effectiveness. Through gene set enrichment analysis, this study identified 25 genes closely related to the pathology of KL and HS, and a list of 14 drugs targeting 3 of the key genes was compiled using DeepPurpose. Potential drugs can be divided into PI3K inhibitors, PTGS2 inhibitors and VEGFA antagonists.

Prostaglandin-endoperoxide synthase 2 encoded by the PTGS2 gene, also known as cyclooxygenase-2 (COX-2), is the rate-limiting enzyme of prostaglandin biosynthesis (20). The involvement of COX-2 in the pathogeneses of scar lesions has been evidenced. Studies have demonstrated that COX-2 is significantly overexpressed in KL and HS tissues, while down-regulation of COX-2 may reduce KL and HS formation (21-24). After tissue injury, COX-derived prostaglandin E2 (PGE2) promotes the recruitment of inflammatory cells, which release TGF-β or platelet-derived growth factors; thereby, extracellular matrix and fibroblast activation is enhanced, leading to fibroblast proliferation and collagen production (25). The reduction of KL and HS formation in patients using nonsteroidal anti-inflammatory

drugs and COX-2 inhibitors has suggested that COX-2 inhibitors may serve as a therapeutic strategy for KL and HS, which is consistent with our findings. Diprosalic, one of the PTGS2 inhibitors found to hold promise in this study, is a combination of betamethasone dipropionate and salicylic acid. It is currently used to treat psoriasis and inflammatory diseases like dermatitis and eczema, as well as to manage subacute and chronic hyperkeratotic and dry dermatoses that are responsive to corticosteroid therapy (26,27). Other COX-2 inhibitors include meloxicam, lornoxicam, piroxicam, mesalazine, parecoxib sodium, HTX-011, tiemonium noramidopyrine, and diclofenac epolamine, the indications for which are postoperative pain and arthritis. These drugs may represent promising treatments for KL and HS.

The involvement of the phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K)/protein kinase B (Akt)/mammalian target of rapamycin (mTOR) signaling pathway in the pathogeneses of KL and HS has been reported previously. Activation of the PI3K/Akt/mTOR pathway has been demonstrated to enhance the inflammation, angiogenesis, and deposition of extracellular matrix components in scar formation; thus, it is considered to be related to several fibrous diseases (28). CUDC-907, a dual inhibitor of the PI3K/Akt/mTOR pathway and histone deacetylase (HDAC), was found to reverse the pathological phenotype of KL fibroblasts (29). In this study, we found 2 PI3K inhibitors to have potential as drug therapies. Bimiralisib is a dual inhibitor of PI3K and the mammalian target of rapamycin (mTOR). It has been identified as a clinical candidate with potential antineoplastic activity, including in malignant lymphomas, primary central nervous system lymphoma (PCNSL), head and neck squamous cell carcinoma (HNSCC), advanced solid tumors, and metastatic breast cancer (30,31). Another PI3K inhibitor, SF-1126, which selectively inhibits all PI3K class IA isoforms as well as DNA-dependent protein kinase (DNA-PK) and mTOR, is the focus of current phase I clinical trials for chronic lymphocytic leukemia and advanced or metastatic solid tumors (32). In a phase I clinical trial, this drug showed considerable efficacy against B-cell malignancies and solid tumors with no dose-limiting toxicities or hepatotoxicities (33). However, the incorporation of novel PI3K inhibitors into treatment strategies for KL and HS still requires further experimental research and long-term trials to ascertain their tolerability, efficacy, and safety.

VEGF (or VEGFA, the most abundant VEGF isoform) has been implicated as a crucial participant in pathological wound healing (34). Multiple studies on KL and HS have reported an association of high VEGFA levels with scar formation (35-38). Furthermore, there is experimental evidence that VEGF inhibition may be an approach to reducing deposition of scar tissue (37,39-41). In this study, we identified 2 VEGF antagonists as potential drugs to treat KL and HS. Sunitinib malate, a dual inhibitor of VEGF and PDGF receptors, is a lead injectable sustained-release candidate used in the treatment of wet age-related macular degeneration (AMD) (42). It is also under development for the treatment of diabetic macular edema and retinal vein occlusion (43). Meanwhile, pegaptanib octasodium, a pegylated oligonucleotide aptamer, is a direct inhibitor of VEGF that is used as an anticancer agent and in AMD. However, clinical testing to determine whether VEGF inhibition is an effective anti-scarring strategy will need to be performed.

In this study, we used DeepPurpose to predict the interactions of candidate drugs and gene targets in order to select the drugs with the highest predicted binding scores. In the knowledge of the relevance between candidate drugs and target genes, the identification of interactions between them became our major objective. The potential of machine learning models to predict the binding affinity between new drug-target pairs has been demonstrated in various studies. Bagherian *et al.* (44) briefly reviewed drug-target interaction prediction by machine learning models. Recently, machine learning methods have been used to search for cures for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (45-47), which has given direction for the promotion of new drug discovery. DeepPurpose, the toolkit we used in the current study, is built on the basis of an encoder–decoder framework. The encoders are generated from novel machine-learning approaches for drug-target interaction prediction to extract features from candidate drugs and target genes, while the decoder is a multi-layer perceptron that uses the extracted features to compute the binding affinity scores. With the 15 pre-trained models and 3 aggregation schemas provided by DeepPurpose, we finally obtained 24 different ranked lists of binding affinity score predictions. Though, we selected all potential drugs that meet the threshold criteria under each model, further analysis of pros and cons of the models may give a better guidance in drug screening with larger datasets. We built a validation set to evaluate these models. For each pair in the validation set, we collect the kinase dissociation constant (K_d) and transformed it to logspace(pK_d) as $pK_d = -\log_{10}\left(\frac{K_d}{10^9}\right)$, which is used as the dependent variable in the models

Table 6 MSE for different models on different datasets

Dataset	Model								
	CNN_CNN	Morgan_CNN	MPNN_CNN	Daylight_AAC	Morgan_AAC	Transformer_CNN	AVE	MAX	AVE_MAX
DAVIS	5.5	5.3	5.2	4.8	5.4	–	5.1	4.4	4.6
BindingDB	3.4	5.1	4.8	5.0	6.5	6.7	4.7	3.5	3.8

Three out of five models (DeepDTA, Morgan_CNN, MPNN_CNN) have smaller MSE when trained on BindingDB than on DAVIS dataset. CNN_CNN model has the smallest MSE, which shows that aggregated models may not always have a better performance though proposed by DeepPurpose's oneline models. CNN, convolutional neural network; MPNN, message-passing neural network; AAC, amino acid composition; AVE, average; MAX, maximum; MSE, mean square error.

trained on DAVIS and BindingDB datasets. The mean squared error (MSE) of each model was calculated, and the results are shown in *Table 6*.

The results paved the way to obtaining the best drug-target pair. Firstly, the MSEs showed that models trained on larger datasets outperformed those trained on smaller datasets. Three out of 5 models (DeepDTA, Morgan_CNN, MPNN_CNN, and Morgan_ACC) had a smaller MSE when trained on the BindingDB dataset than on the DAVIS dataset. This is often the case for machine learning models: those trained on a larger dataset have better generalizability, since the larger the training set is, the greater opportunity is for the model to learn global patterns. Moreover, by comparing the MSE of the single models and the aggregated models, we found that aggregated models do not always outperform single models, especially when aggregation is applied to models with a considerable variance in performance. However, for models trained on the DAVIS dataset, aggregated models performed better. The model with mean schema had a smaller MSE than most single models, while models with the max and the average of the mean and max schemas outperformed even the best single model. With the BindingDB dataset, however, the aggregated models did not perform as well as the best single model but did outperform most of the single models. This implies that although the use of aggregation schema can, to a certain extent, reduce the limitation and bias of single models, it can also introduce additional errors by aggregating the results of poor models.

Conclusions

Our study has demonstrated that drug discovery using *in silico* text mining and DeepPurpose may be a powerful and effective way to find drugs targeting the genes related to KL and HS. Therefore, our study could provide a theoretical basis for the development of novel targeted therapies for KL and HS.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (grant No. 81671915).

Footnote

Reporting Checklist: The authors have completed the MDAR checklist. Available at http://dx.doi.org/10.21037/atm-21-218

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-21-218). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Ogawa R. Mechanobiology of scarring. Wound Repair Regen 2011;19 Suppl 1:s2-9.
2. Marneros AG, Norris JE, Olsen BR, Reichenberger E. Clinical genetics of familial keloids. Arch Dermatol 2001;137:1429-34.
3. Chiang RS, Borovikova AA, King K, et al. Current concepts related to hypertrophic scarring in burn injuries. Wound Repair Regen 2016;24:466-77.
4. Kim S, Choi TH, Liu W, et al. Update on scar management: guidelines for treating Asian patients. Plast Reconstr Surg 2013;132:1580-9.
5. Lee HJ, Jang YJ. Recent Understandings of Biology, Prophylaxis and Treatment Strategies for Hypertrophic Scars and Keloids. Int J Mol Sci 2018;19:711.
6. Moosavinasab S, Patterson J, Strouse R, et al. 'RE:fine drugs': an interactive dashboard to access drug repurposing opportunities. Database (Oxford) 2016;2016:baw083.
7. Mullard A. New drugs cost US$2.6 billion to develop. Nature Reviews Drug Discovery 2014;13:877.
8. Mayr A, Klambauer G, Unterthiner T, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chem Sci 2018;9:5441-51.
9. Smalley E. AI-powered drug discovery captures pharma interest. Nat Biotechnol 2017;35:604-5.
10. Fleming N. How artificial intelligence is changing drug discovery. Nature 2018;557:S55-7.
11. Liu T, Lin Y, Wen X, et al. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 2007;35:D198-201.
12. Tang J, Szwajda A, Shakyawar S, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J Chem Inf Model 2014;54:735-43.
13. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol 2011;29:1046-51.
14. Huang K, Fu T, Glass L, et al. DeepPurpose: a Deep Learning Library for Drug-Target Interaction Prediction and Applications to Repurposing and Screening. 2020.
15. Baran J, Gerner M, Haeussler M, et al. pubmed2ensembl: a resource for mining the biological literature on genes. PLoS One 2011;6:e24716.
16. Carmona-Saez P, Chagoyen M, Tirado F, et al. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 2007;8:R3.
17. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
18. Su G, Morris JH, Demchak B, et al. Biological network exploration with Cytoscape 3. Curr Protoc Bioinformatics 2014;47:8.13.1-24.
19. Jardim DL, Groves ES, Breitfeld PP, et al. Factors associated with failure of oncology drugs in late-stage clinical development: A systematic review. Cancer Treat Rev 2017;52:12-21.
20. Sobolewski C, Cerella C, Dicato M, et al. The role of cyclooxygenase-2 in cell proliferation and cell death in human malignancies. Int J Cell Biol 2010;2010:215158.
21. Abdou AG, Maraee A, Saif H. Immunohistochemical evaluation of COX-1 and COX-2 expression in keloid and hypertrophic scar. Am J Dermatopathol 2014;36:311-7.
22. Rossiello L, D'andrea F, Grella R, et al. Differential expression of cyclooxygenases in hypertrophic scar and keloid tissues. Wound Repair Regen 2009;17:750-7.
23. Louw L. The keloid phenomenon: progress toward a solution. Clin Anat 2007;20:3-14.
24. Wilgus TA, Vodovotz Y, Vittadini E, et al. Reduction of scar formation in full-thickness wounds with topical celecoxib treatment. Wound Repair Regen 2003;11:25-34.
25. Stratton R, Shiwen X. Role of prostaglandins in fibroblast activation and fibrosis. J Cell Commun Signal 2010;4:75-77.
26. Shou M, Galinada W, Wei Y, et al. Development and validation of a stability-indicating HPLC method for simultaneous determination of salicylic acid, betamethasone dipropionate and their related compounds in Diprosalic Lotion. J Pharm Biomed Anal 2009;50:356-61.
27. Guenther LC. Fixed-dose combination therapy for psoriasis. Am J Clin Dermatol 2004;5:71-7.
28. Wong VW, You F, Januszyk M, et al. Transcriptional profiling of rapamycin-treated fibroblasts from hypertrophic and keloid scars. Ann Plast Surg 2014;72:711-9.
29. Tu T, Huang J, Lin M, et al. CUDC-907 reverses pathological phenotype of keloid fibroblasts in vitro and in vivo via dual inhibition of PI3K/Akt/mTOR signaling and HDAC2. Int J Mol Med 2019;44:1789-800.
30. Tarantelli C, Gaudio E, Arribas A, et al. PQR309 Is a Novel Dual PI3K/mTOR Inhibitor with Preclinical Antitumor Activity in Lymphomas as a Single Agent and in Combination Therapy. Clin Cancer Res 2018;24:120-9.
31. Beaufils F, Cmiljanovic N, Cmiljanovic V, et al. 5-(4,6-Dimorpholino-1,3,5-triazin-2-yl)-4-(trifluoromethyl)pyridin-2-amine (PQR309), a Potent,

Brain-Penetrant, Orally Bioavailable, Pan-Class I PI3K/mTOR Inhibitor as Clinical Candidate in Oncology. J Med Chem 2017;60:7524-38.
32. Qin AC, Li Y, Zhou LN, et al. Dual PI3K-BRD4 Inhibitor SF1126 Inhibits Colorectal Cancer Cell Growth in Vitro and in Vivo. Cell Physiol Biochem 2019;52:758-68.
33. Mahadevan D, Chiorean E, Harris W, et al. Phase I pharmacokinetic and pharmacodynamic study of the pan-PI3K/mTORC vascular targeted pro-drug SF1126 in patients with advanced solid tumours and B-cell malignancies. Eur J Cancer 2012;48:3319-27.
34. Le AD, Zhang Q, Wu Y, et al. Elevated vascular endothelial growth factor in keloids: relevance to tissue fibrosis. Cells Tissues Organs 2004;176:87-94.
35. Gira AK, Brown LF, Washington CV, et al. Keloids demonstrate high-level epidermal expression of vascular endothelial growth factor. J Am Acad Dermatol 2004;50:850-3.
36. van der Veer WM, Niessen FB, Ferreira JA, et al. Time course of the angiogenic response during normotrophic and hypertrophic scar formation in humans. Wound Repair Regen 2011;19:292-301.
37. Mogili NS, Krishnaswamy VR, Jayaraman M, et al. Altered angiogenic balance in keloids: a key to therapeutic intervention. Transl Res 2012;159:182-9.
38. Ong CT, Khoo YT, Tan EK, et al. Epithelial-mesenchymal interactions in keloid pathogenesis modulate vascular endothelial growth factor expression and secretion. J Pathol 2007;211:95-108.
39. Wang J, Chen H, Shankowsky HA, et al. Improved scar in postburn patients following interferon-alpha2b treatment is associated with decreased angiogenesis mediated by vascular endothelial cell growth factor. J Interferon Cytokine Res 2008;28:423-34.
40. Salem A, Assaf M, Helmy A, et al. Role of vascular endothelial growth factor in keloids: a clinicopathologic study. Int J Dermatol 2009;48:1071-7.
41. Wu WS, Wang FS, Yang KD, et al. Dexamethasone induction of keloid regression through effective suppression of VEGF expression and keloid fibroblast proliferation. J Invest Dermatol 2006;126:1264-71.
42. A Depot Formulation of Sunitinib Malate (GB-102) in Subjects With Neovascular (Wet) Age-related Macular Degeneration. Available online: https://clinicaltrials.gov/show/NCT03249740
43. GrayBug. Available online: https://www.graybug.vision/our-technologies-and-pipeline/#pipeline
44. Bagherian M, Sabeti E, Wang K, et al. Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. Brief Bioinform 2021;22:247-69.
45. Beck BR, Shin B, Choi Y, et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 2020;18:784-90.
46. Zhang H, Saravanan KM, Yang Y, et al. Deep Learning Based Drug Screening for Novel Coronavirus 2019-nCov. Interdiscip Sci 2020;12:368-76.
47. Nand M, Maiti P, Joshi T, et al. Virtual screening of anti-HIV1 compounds against SARS-CoV-2: machine learning modeling, chemoinformatics and molecular dynamics simulation based analysis. Sci Rep 2020;10:20397.

(English Language Editor: J. Reynolds)

Cite this article as: Pan Y, Chen Z, Qi F, Liu J. Identification of drug compounds for keloids and hypertrophic scars: drug discovery based on text mining and DeepPurpose. Ann Transl Med 2021;9(4):347. doi: 10.21037/atm-21-218