

厚生労働科学研究費補助金（政策科学総合研究事業）
（臨床研究等 ICT 基盤構築・人工知能実装研究事業）
分担研究報告書

課題名 : 新薬創出を加速する症例データベースの構築・拡充/創薬ターゲット推定アルゴリズムの開発
研究分担者名 : 山西 芳裕
国立大学法人九州工業大学大学院 情報工学研究院 生命化学情報工学研究系 教授

研究要旨

IPF の臨床情報と多階層オミックス情報を統合解析するための機械学習手法をスパースモデリングの枠組みで開発した。開発手法を本事業で収集した臨床情報と多階層オミックス情報に適用し、IPF の創薬標的や疾患サブタイプのバイオマーカー探索を行った。これまでに開発したツール化合物探索 AI を遺伝子やタンパク質に関する様々なオミックスデータに適用し、IPF の有望な創薬標的候補を探索して候補を絞り込むことができた。

A. 研究目的

医薬ビッグデータに基づいて、対象疾患の創薬標的候補の「確からしさ」の検証実験に利用可能なツール化合物を探索するインシリコ手法を開発する。様々な疾患に関するマルチオミックスデータや分子ネットワークデータ、既承認薬、開発中止化合物、合成化合物、天然化合物など大規模な化合物の構造データや実験データを収集する。疾患データと化合物データの融合解析を行う統計手法や、多様なオミックス関連データを有効活用して化合物を効率的にスクリーニングできる機械学習の手法を開発する。最終的に、特発性肺線維症に対して見出された創薬標的分子候補を制御するツール化合物の候補をインシリコ予測する。

B. 研究方法

- ① IPF の臨床情報と多階層オミックス情報を統合解析するための機械学習手法をスパースモデリングの枠組みで開発する。
- ② 上記の手法を本事業で収集した臨床情報と多階層オミックス情報に適用し、IPF の創薬標的や疾患サブタイプのバイオマーカー探索を行う。
- ③ これまでに開発したツール化合物探索 AI を遺伝子やタンパク質に関する様々なオミックスデータに適用し、IPF の有望な創薬標的候補を探索する。

（倫理面への配慮）

本研究は、医薬基盤・健康・栄養研究所において倫理審査、承認を得た後、ヒトゲノム・遺伝子解析研究に関する倫理指針及び、人を対象とする医学系研究に関する倫理指針に従って遂行した。

C. 研究結果

- ① IPF の臨床情報と多階層オミックス情報を統合解析するための機械学習手法をスパースモデリングの枠組みで開発する。

大阪大学提供 IPF 関連データの解析を行った。特に、今期は、新しく提供していただいた確定診断の情報を使って解析を行った。

確定診断毎の患者数

diagnosis	n
UIP	126
pro-UIP	80
indeterminate-UIP	53
NSIP	122
OP	10
HP	10
PPFE	5
others	154
HC	39
non-diagnosis	3
	602

合計 602人の患者データ

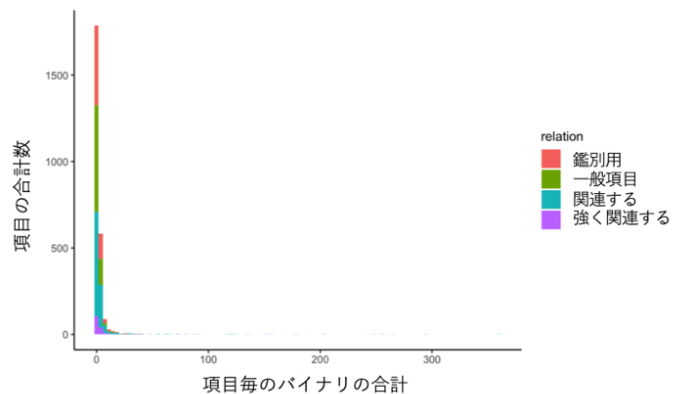
血清エクソソームのプロテオームデータ

2416 UniProt

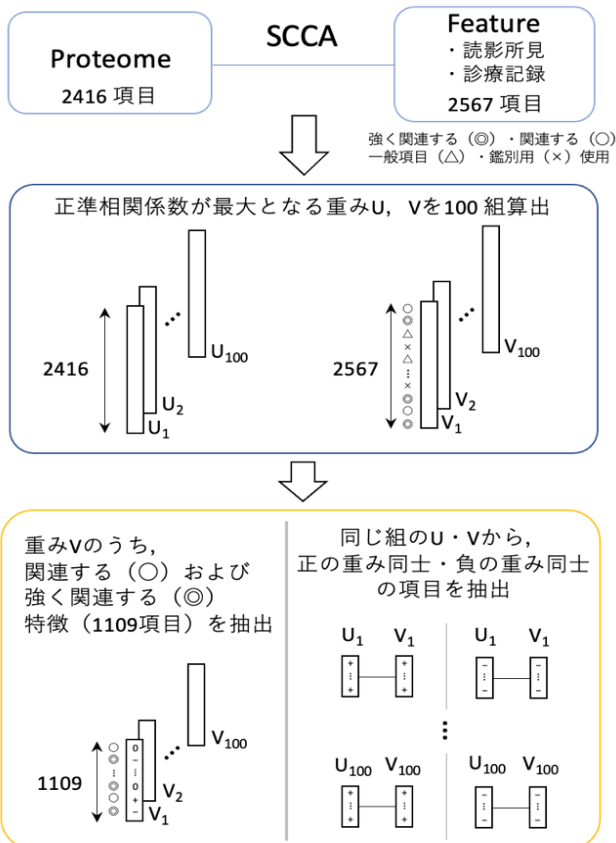
読影所見・診療記録 (バイナリ)

2567 項目

◎ 強く関連する	欧州呼吸器学会、米国胸部疾患学会、日本呼吸器学会において、診断アルゴリズムに存在、ガイドラインに記載あり	169 項目
○ 関連する	呼吸器疾患全般、炎症性疾患、急性増悪などにも関連する (ただし、読影所見については、診断の要素となる所見)	940 項目
△ 一般項目	年齢、性別、併存疾患などにより変動する一般検査項目 (読影所見については、他臓器や様々な疾患に関係する所見)	804 項目
× 鑑別用	IPFでないことを判定する、癌の合併を知る、他の疾患であることを判別する、他の疾患を考える所見 (他の特異性間質性肺炎、感染性肺炎、リウマチ性肺炎など)	654 項目

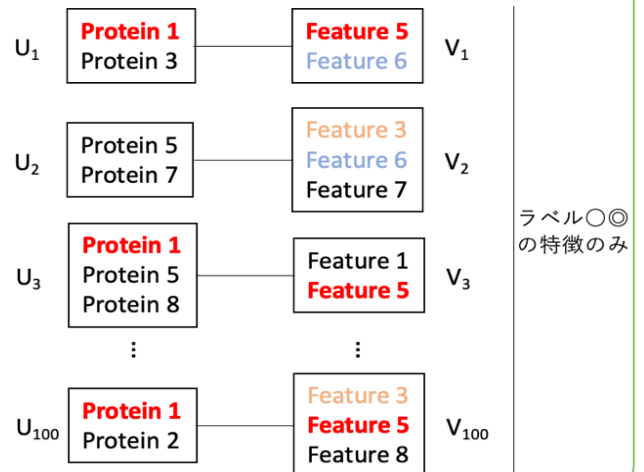


タンパク質と読影所見および診療記録間のスパース正準相関分析 (SCCA) で IPF と関連のあるタンパク質を予測することを試みた。



『タンパク質とIPFに関連する特徴に真に相関がある。』

IPFに関連する特徴に重みがある時、
相関しているタンパク質にも重みがある
と想定し、特徴とタンパク質を抽出



選択された特徴の重み V が最大もしくは最小の組に含まれるタンパク質の重み U を抽出

関連する (○) および強く関連する (◎)
読影所見や診療記録と相関を示す
タンパク質を選択した。

10

タンパク質と特徴を用いた SCCA で選択された IPF に関連する特徴とタンパク質（一部）を以下に示す。

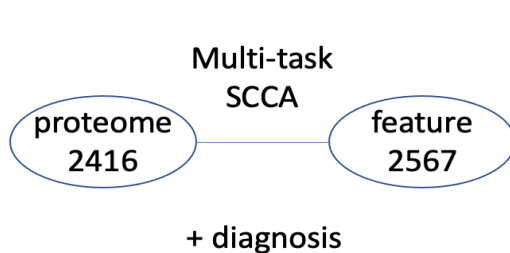
正の重みを示した特徴とタンパク質の組み合わせ

max_V * U	feature	name	relation	uniprot	gene_symbol	function.
0.02	c259	すりガラス影:p:S1 右肺	○	P20160	AZU1	Azurocidin; Cationic antimicrobial protein CAP37; Heparin-bin
0.01	c259	すりガラス影:p:S1 右肺	○	Q14764	MVP; LRP	Major vault protein; Lung resistance-related protein
0.01	c2683	気管支拡張:p:両側肺 末梢	○	P02730	SLC4A1; AE1; DI; Band 3 anion transport protein; Anion exchange protein 1; Sol	
0.01	c2683	気管支拡張:p:両側肺 末梢	○	P11171	EPB41; E41P	Protein 4.1; 4.1R; Band 4.1; EPB4.1; Erythrocyte membrane pro
0.01	c259	すりガラス影:p:S1 右肺	○	Q8IVF2	AHNAK2; C14orf7 Protein AHNAK2	
0.01	c3066	炎症後変化:s:S1 左肺	○	P09493	TPM1; C15orf13; Tropomyosin alpha-1 chain; Alpha-tropomyosin; Tropomyosin	
0.01	c3066	炎症後変化:s:S1 左肺	○	P09493	TPM1; C15orf13; Tropomyosin alpha-1 chain; Alpha-tropomyosin; Tropomyosin	
0.01	c3075	炎症後変化:s:S2 左肺	○	P09493	TPM1; C15orf13; Tropomyosin alpha-1 chain; Alpha-tropomyosin; Tropomyosin	
0.01	c3075	炎症後変化:s:S2 左肺	○	P09493	TPM1; C15orf13; Tropomyosin alpha-1 chain; Alpha-tropomyosin; Tropomyosin	
0.01	c3066	炎症後変化:s:S1 左肺	○	Q16181	SEPTIN7; CDC10; Septin-7; CDC10 protein homolog	
0.01	c3075	炎症後変化:s:S2 左肺	○	Q16181	SEPTIN7; CDC10; Septin-7; CDC10 protein homolog	
0.01	c2683	気管支拡張:p:両側肺 末梢	○	Q00013	MPP1; DXS552E; 55 kDa erythrocyte membrane protein; Membrane protein, pal	
0.01	c295	すりガラス影:p:S6 下葉 両側肺 胸膜	◎	Q6RW13	AGTRAP; ATRAP Type-1 angiotensin II receptor-associated protein; AT1 recept	
0.01	c6085	陰影:p:中葉 右肺 末梢	○	Q6RW13	AGTRAP; ATRAP Type-1 angiotensin II receptor-associated protein; AT1 recept	
0.01	c6097	陰影:p:右肺 底部	○	Q6RW13	AGTRAP; ATRAP Type-1 angiotensin II receptor-associated protein; AT1 recept	
0.01	s13	喫煙歴-N	○	P62318	SNRPD3	Small nuclear ribonucleoprotein Sm D3; snRNP core protein D
0.01	s13	喫煙歴-N	○	O43242	PSMD3	26S proteasome non-ATPase regulatory subunit 3; 26S protea
0.01	c5637	蜂巣肺:s:両側肺 胸膜側	◎	Q8IWB7	WDFY1; FENS1; KWD repeat and FYVE domain-containing protein 1; FYVE dom	
0.01	c294	すりガラス影:s:S6 上葉 右肺	○	P46776	RPL27A	60S ribosomal protein L27a; Large ribosomal subunit protein u
0.01	c3120	炎症後変化:s:S6 上葉 右肺	○	P46776	RPL27A	60S ribosomal protein L27a; Large ribosomal subunit protein u

負の重みを示した特徴とタンパク質の組み合わせ

max_V * U	feature	name	relation	uniprot	gene_symbol	function.
0.05	s8	発熱	○	Q92496	CFHR4; CFHL4; FHR4	Complement factor H-related protein 4
0.04	c295	すりガラス影:p:S6 下葉 両側肺 胸膜	◎	P34910	EVI2B; EVDB	Protein EVI2B; Ecotropic viral integration site 2B protein homolog
0.04	c6085	陰影:p:中葉 右肺 末梢	○	P34910	EVI2B; EVDB	Protein EVI2B; Ecotropic viral integration site 2B protein homolog
0.04	c6097	陰影:p:右肺 底部	○	P34910	EVI2B; EVDB	Protein EVI2B; Ecotropic viral integration site 2B protein homolog
0.03	s8	発熱	○	Q03591	CFHR1; CFHL1; CFHL1; CF	Complement factor H-related protein 1; H factor-like protein 1; H36
0.03	s8	発熱	○	Q99832	CCT7; CCTH; NIP7-1	T-complex protein 1 subunit eta; CCT-eta; HIV-1 Nef-interacting pro
0.02	s8	発熱	○	P19105	MYL12A; MLCB; MRLC3; Myosin regulatory light chain 12A; Epididymis secretory protein Li 24	
0.02	s8	発熱	○	O14950	MYL12B; MRLC2; MYLC2	Myosin regulatory light chain 12B; MLC-2A; Myosin regulatory light c
0.02	s8	発熱	○	Q8IZJ3	CPAMD8; KIAA1283	C3 and PZP-like alpha-2-macroglobulin domain-containing protein 8
0.02	c493	すりガラス影:p:両側肺 周囲 気管支	○	Q92496	CFHR4; CFHL4; FHR4	Complement factor H-related protein 4
0.02	c1402	嚢胞:p:に沿った領域 下葉 両側肺 背側 胸膜	◎	Q92496	CFHR4; CFHL4; FHR4	Complement factor H-related protein 4
0.02	c2677	気管支拡張:p:両側肺 周囲 気管支	○	Q92496	CFHR4; CFHL4; FHR4	Complement factor H-related protein 4
0.02	s8	発熱	○	P08603	CFH; HF; HF1; HF2	Complement factor H; H factor 1
0.02	s8	発熱	○	Q8TAQ9	SUN3; SUNC1	SUN domain-containing protein 3; Sad1/unc-84 domain-containing
0.02	c5796	透過性低下:s:下葉	○	Q9Y411	MYO5A; MYH12	Unconventional myosin-Va; Dilute myosin heavy chain, non-muscle;
0.02	c5856	間質性肺炎増悪:s:	○	Q9Y411	MYO5A; MYH12	Unconventional myosin-Va; Dilute myosin heavy chain, non-muscle;
0.01	c5796	透過性低下:s:下葉	○	P00751	CFB; BF; BFD	Complement factor B; C3/C5 convertase; Glycine-rich beta glycopro
0.01	c5856	間質性肺炎増悪:s:	○	P00751	CFB; BF; BFD	Complement factor B; C3/C5 convertase; Glycine-rich beta glycopro
0.01	s8	発熱	○	Q8WUX1	SLC38A5; JM24; SN2; SN	Sodium-coupled neutral amino acid transporter 5; Solute carrier fam
0.01	c493	すりガラス影:p:両側肺 周囲 気管支	○	Q03591	CFHR1; CFHL1; CFHL1; CF	Complement factor H-related protein 1; H factor-like protein 1; H36

最後に、確定診断の情報を加味してクラス特異的なタンパク質の抽出を試みた。そのために、確定診断のクラスごとに分類し、かつクラス内の患者間で相関のある分子を抽出するための Multi-task SCCA 法を開発した。IPF 患者群で重みがある分子を選んだところ、IPF との関連が報告されている SFTPC, MUC1 (= KL-6), SFTPA, MUC5B などの分子を、抽出することができた。



UIP	126
pro-UIP	80
indeterminate-UIP	53
NSIP	122
OP	10
HP	10
PPFE	5
others	154
HC	39
non-diagnosis	3
	602

確定診断ごとのタンパク質と特徴の間の相関を統計学的に算出する。

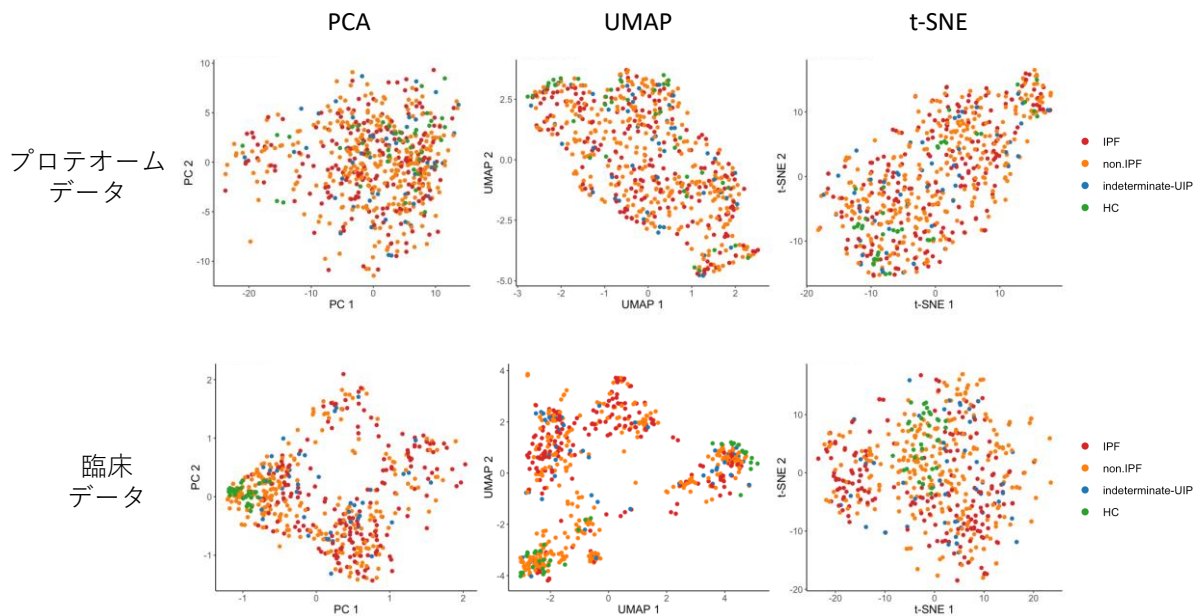
json_uniprot	gene_name	function.	uniprot	pubmed	UIP	pro-UIP	indeterminate-UIP	NSIP	OP	HP	PPFE	others	HC	non-diagnosis
P11686-2	SFTPC; SFTP2	Pulmonary surfactant-associated protein C; Pulmonary P11686			28	0.00063003	0	0	0	0	0	0	0.00053143	0
P15941;P15941	MUC1; PUM	Mucin-1; Breast carcinoma-associated antigen DF3; CiP15941			68	2.61E-05	0.00925768	0	0	0	0	0	0	0
Q8IWL1;Q8IWL1	SFTPA1; COLECA4	Pulmonary surfactant-associated protein A1; 35 kDa pI Q8IWL2			8	0.00123089	0	0	0	0	0	0	0	0
Q8IWL1;Q8IWL1	SFTPA2; COLECA5	Pulmonary surfactant-associated protein A2; 35 kDa pI Q8IWL1			12	0.00123089	0	0	0	0	0	0	0	0
json_uniprot	gene_name	function.	uniprot	pubmed	UIP	pro-UIP	indeterminate-UIP	NSIP	OP	HP	PPFE	others	HC	non-diagnosis
Q9HC84	MUC5B; MUC5	Mucin-5B; Cervical mucin; High molecular weight Q9HC84			97	-1.275E-05	0	0	0	0	0	0	0	0

② 上記の手法を本事業で収集した臨床情報と多階層オミックス情報に適用し、IPF の創薬標的や疾患サブタイプのバイオマーカー探索を行う。

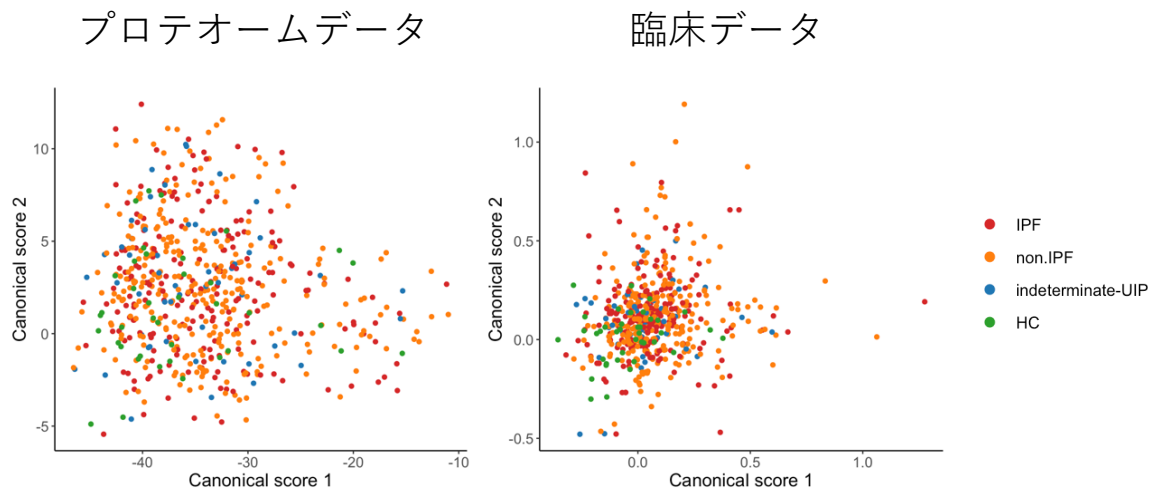
大阪大学コホート（以下、阪大コホート）の確定診断情報を元に、患者を IPF, non. IPF, indeterminate-UIP, 健常者（以下、HC）の 4 グループにまとめ直した。

確定診断情報	グループ
UIP	IPF
pro-UIP	
indeterminate-UIP	
NSIP	non.IPF
OP	
HP	
PPFE	
others	
HC	HC

各グループのプロテオームデータと臨床データから、教師なし学習である PCA, UMAP, t-SNE で次元削減を行なった。各手法で得られた第一成分と第二成分から、各グループを分類することは困難なことが示唆された。

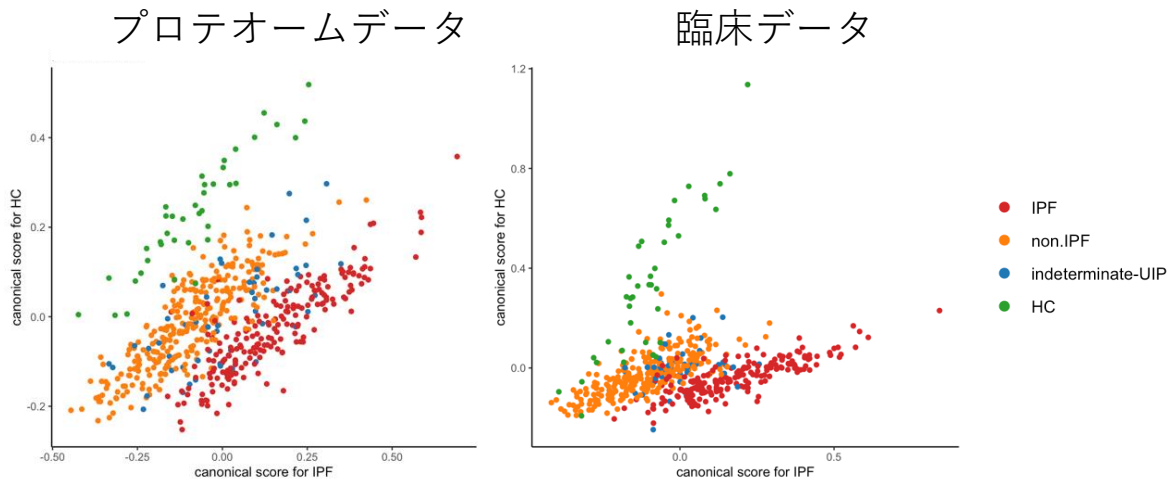


また、教師なし学習であるスパース正準相関分析（以下、SCCA）で、各患者のプロテオームデータおよび臨床データ間で相関を示す特徴量を調べた。第一正準スコアと第二正準スコアから、SCCA では各グループを分類するような特徴量を選択できないことが示唆された。



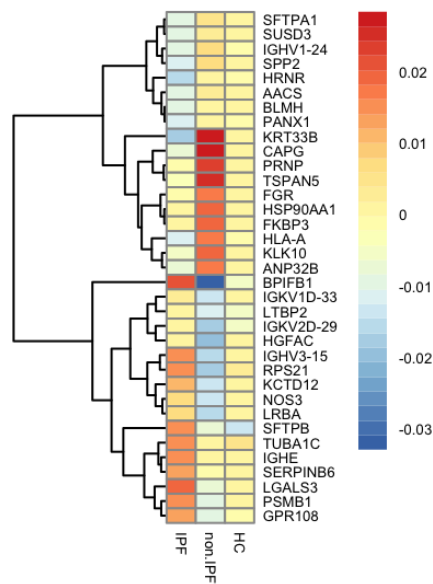
教師なし学習の手法では各グループを分類するような特徴を抽出するのは困難であったため、教師あり学習であるマルチクラススパース正準相関分析（以下 MSCCA）で各グループを分類可能な特徴量の選択を試み

た。IPF と non-IPF を明確に分類することを目的とするため、両グループの患者が混在すると考えられる indeterminate-UIP を除外して行なった。IPF と HC の重みから算出した正準スコアから、プロテオームデータおよび臨床データで、各グループの分類に寄与する特徴が選択されたことが示唆された。また、得られた IPF と HC の重みから indeterminate-UIP の正準得点を算出し、その他 3 グループと比較すると、IPF もしくは non-IPF の特徴を持つ患者が混在していることが示唆された。

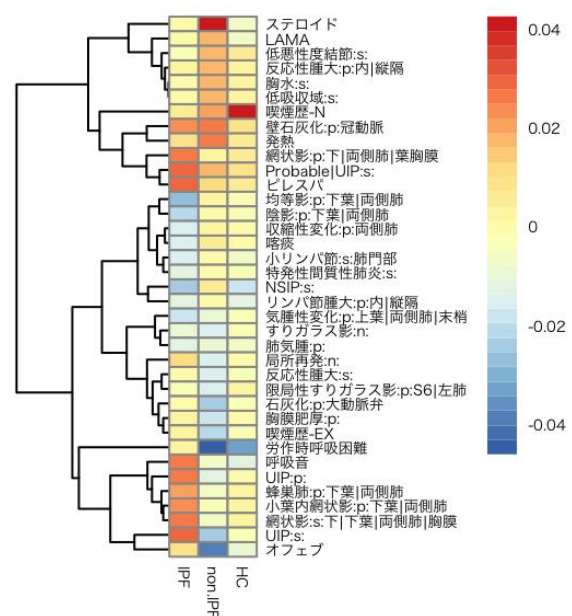


IPF グループと non-IPF グループの分類に強く寄与するタンパク質や臨床項目を調べるため、MSCCA の結果得られた IPF および non-IPF グループの重み上位および下位 10 位までのタンパク質と臨床項目に対する各グループの重みをヒートマップで示した。IPF や non-IPF グループの上位および下位のタンパク質には正負が逆の重みか、もしくは、一方のグループには大きな重みがあり、他方のグループには 0 に近い重みがついていた。

タンパク質の重み



臨床項目の重み



以上のことから、MSCCA によって IPF と non-IPF を判別可能なバイオマーカー候補が選択されたことが示唆された。

③ これまでに開発したツール化合物探索 AI を遺伝子やタンパク質に関する様々なオミクスデータに適用し、IPF の有望な創薬標的候補を探索する。

ヒト細胞における創薬標的分子の摂動応答プロファイルと疾患特異的な遺伝子発現プロファイルを用いた創薬標的分子の予測手法ターゲットリポジショニングの研究開発を引き続き行なった。創薬標的分子をコードする遺伝子のノックダウン (KD) プロファイルが標的分子を阻害した薬の機能を反映し、同様に遺伝子の

過剰発現 (OE) プロファイルが標的分子を活性化した薬の機能を反映していると仮定して、阻害標的分子と活性化標的分子を分けて予測した。疾患特異的な遺伝子発現パターンと、ヒト細胞における遺伝子振動応答の遺伝子発現パターンとの間の逆相関を計算する方法は、精度はあまり高くなかった。部分的に既知の創薬標的分子の情報を学習し、疾患間の類似性を考慮して、創薬ターゲット分子の予測を行うマルチタスク学習アルゴリズムは、疾患の既知の阻害標的分子や活性化標的分子を上手く予測することができた。SNP ベースの方法よりも高い精度を確認できた。

これまでに開発してきた臨床投薬データ、薬剤応答トランスクリプトームデータ、グラフ畳み込みニューラルネットワーク、ランダムフォレスト回帰モデルに基づくインシリコ手法による IPF の創薬標的予測結果を統合して、予測スコアが最も高かった生体分子 (受容体 X) を IPF の創薬標的候補として提示した。その実験検証を加藤先生のグループと共同研究で行ってきた。その結果、SFTPC ノックインマウスとブレオマイシンモデルマウスの 2 種類の IPF 動物モデルで、予測結果の妥当性を確認できた。

詳しい作用メカニズムを調べるため、ブレオマイシン肺線維症モデルマウスに対する受容体 X のアンタゴニスト投与時のトランスクリプトーム解析を行った。加藤グループがマウス実験、山西グループがデータ解析を担当した。

ブレオマイシン投与の影響を評価した。コントロール群は無処理、処理群はブレオマイシン投与後 CMC 連日投与した。IL17 など炎症系のパスウェイが活性化していた。

BLM マウスへの受容体 X のアンタゴニスト投与の効果を評価した。コントロール群は、BLM 投与後 CMC 連日投与、処理群は BLM 投与後受容体 X のアンタゴニスト連日投与した。コントロールと比較して受容体 X のアンタゴニスト投与肺の遺伝子発現解析では、炎症系パスウェイの抑制を認めた。以上のことから、受容体 X のアンタゴニストが抗炎症作用により肺線維化を抑制する可能性が示唆された。

D. 考察

① IPF の臨床情報と多階層オミックス情報を統合解析するための機械学習手法をスパースモデリングの枠組みで開発する。

IPF クラス内の患者間で相関のある分子を抽出するための Multi-task SCCA 法を臨床情報とプロテオーム情報に適用したところ、IPF 患者群で重みがある分子を選んだところ、IPF との関連が報告されている SFTPC, MUC1 (≒ KL-6), SFTPA, MUC5B などの分子を、抽出することができ、開発手法は特徴抽出能力が高いことが示唆された。

② 上記の手法を本事業で収集した臨床情報と多階層オミックス情報に適用し、IPF の創薬標的や疾患サブタイプのバイオマーカー探索を行う。

マルチクラススパース正準相関分析で各グループを分類可能な特徴量の選択を試みた結果、プロテオームデータおよび臨床データで、各グループの分類に寄与する特徴が選択されたことが示唆された。また、得られた IPF と HC の重みから indeterminate-UIP の正準得点を基に、その他 3 グループと比較した結果、IPF もしくは non-IPF の特徴を持つ患者が混在していることが示唆された。

③ これまでに開発したツール化合物探索 AI を遺伝子やタンパク質に関する様々なオミックスデータに適用し、IPF の有望な創薬標的候補を探索する。

開発手法で IPF の有望な創薬標的候補として受容体 X を最終候補として絞り込み、IPF モデルマウスを用いた実験で予測結果の妥当性が示唆された。BLM マウスへの受容体 X のアンタゴニスト投与の効果を評価した結果、コントロールと比較して受容体 X のアンタゴニスト投与肺の遺伝子発現解析では、炎症系パスウェイの抑制を認めた。以上のことから、受容体 X のアンタゴニストが抗炎症作用により肺線維化を抑制する可能性が示唆された。

E. 結論

IPF の臨床情報と多階層オミックス情報を統合解析するための機械学習手法をスパースモデリングの枠組みで開発した。開発手法を本事業で収集した臨床情報と多階層オミックス情報に適用し、IPF の創薬標的や疾患サブタイプのバイオマーカー探索を行った。これまでに開発したツール化合物探索 AI を遺伝子やタンパク質に関する様々なオミックスデータに適用し、IPF の有望な創薬標的候補を探索して候補を絞り込むことができた。

F. 研究発表

1. 論文発表

- 1) Iwata, M., Kosai, K., Ono, Y., Oki, S., Mimori, K., and Yamanishi, Y., "Regulome-based characterization of drug activity across the human diseasome", npj Systems Biology and Applications, 8:44, 2022. Nov. 17, 2022 (Springer-Nature), DOI: 10.1038/s41540-022-00255-4
- 2) Namba, S., Iwata, M., and Yamanishi, Y., "From drug repositioning to target repositioning: prediction of therapeutic targets using genetically perturbed transcriptomic signatures", Bioinformatics, 38:i68-i76, 2022. Sep. 18, 2022 (Oxford University Press), DOI: 10.1093/bioinformatics/btac240

2. 学会発表

なし

G. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし