

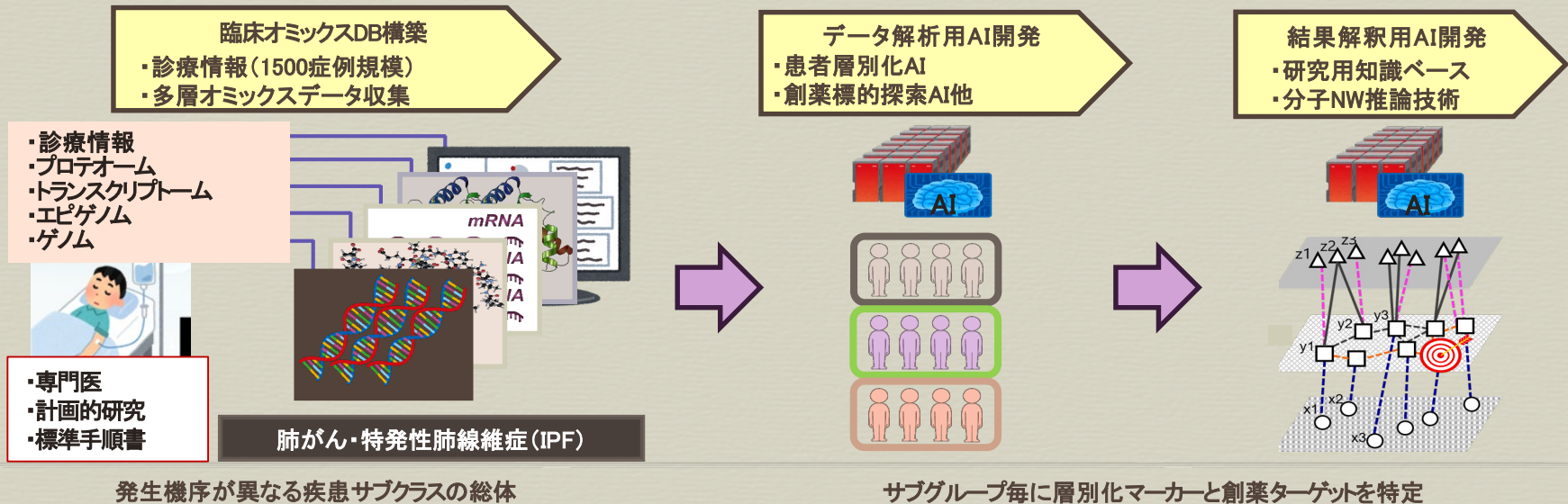
「新薬創出を加速する症例データベースの構築・拡充/創薬  
ターゲットの推定アルゴリズムの開発」  
(PRISM創薬AI)

医薬基盤・健康・栄養研究所

夏目やよい

# PRISM創薬AIでの取り組み

- ◆ 診療情報とオミックスデータを紐づけた超高次元ビッグデータの解析にAIを導入
- ◆ 医薬健栄研を中心に、情報科学、基礎医学、臨床医学、生物学等の様々な領域のトップ研究者を結集  
→ 目標を共有し、役割を分担する本格的なプロジェクト型研究開発
- ◆ 高度専門病院による精緻な医療データ収集 → 日本の強み





# 創薬AI開発に向けた 問題点の洗い出し

## 創薬研究者からの意見:

- ・現在の創薬のボトルネックは創薬標的探索。実験動物に効いてヒトに効かない薬を開発してしまっている。
- ・創薬において患者層別化が重要。同じ診断名であっても適した創薬標的は患者によって違う可能性がある。

→ヒトのデータを用いて、診断名の情報を使うことなく患者層別化をすることができれば  
現在の創薬における問題を解決できるのでは。(問題点の洗い出し)

## 情報科学研究者からの意見:

- ・利用できる人のデータは診療情報とオミックスデータ。この二つは異質(診療情報は様々なデータタイプの混合。オミックスデータは測定値の行列)であり、両方を用いた解析には工夫が必要。
- ・患者層別化=患者のサブグループへの分割。診断名の情報を使わずに似た患者を同じサブグループに入れる事を考える。ただし、診療情報やオミックスデータには、疾患の特徴と直接関連のない項目も多く含まれているため従来法によるサブグループ分割では患者層別化につながらない。

→データタイプに依存せず、異質のデータ二つの間で関連づけられる項目だけを抽出したい。  
抽出した項目を使って患者層別化をできるようなAIを新たに開発する必要がある。(AI開発タスクへの落とし込み)

# 患者層別化AIにできること

オミックスデータ

|     | 分子α   | 分子β    | 分子γ | 分子δ   |     |
|-----|-------|--------|-----|-------|-----|
| 患者A | 12.5  | 2352.4 | 0.3 | 522.9 |     |
| 患者B | 32.7  | 2876.1 | 0.3 | 472.7 | ... |
| 患者C | 176.2 | 3186.3 | 0.4 | 751.6 |     |
| ⋮   | ⋮     | ⋮      | ⋮   | ⋮     |     |

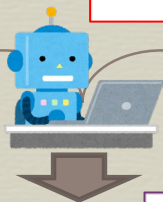
構造化された診療情報

|     | 年齢 | 性別 | 予後 | 治療法X |     |
|-----|----|----|----|------|-----|
| 患者A | 78 | 男  | 悪い | 効く   |     |
| 患者B | 88 | 女  | 悪い | 効かない | ... |
| 患者C | 65 | 男  | 良い | 効かない |     |
| ⋮   | ⋮  | ⋮  | ⋮  | ⋮    |     |

診療情報: 構造化とデータ変換  
オミックスデータ: 測定値のまま

|     | 年齢 | 性別 | 予後 | 治療法X |     |
|-----|----|----|----|------|-----|
| 患者A | 78 | 1  | 1  | 1    |     |
| 患者B | 88 | 0  | 1  | 0    | ... |
| 患者C | 65 | 1  | 0  | 0    |     |
| ⋮   | ⋮  | ⋮  | ⋮  | ⋮    |     |

□ 枠のデータを  
患者層別化AIに入力



出力例①

|     |      |
|-----|------|
| 分子A | HIGH |
| 分子B | HIGH |
| 分子C | LOW  |

出力例①の解釈:  
分子A,Bの値が高く  
分子Cの値が低い患者は  
予後が悪く治療法Xが効く  
傾向がある

|      |      |
|------|------|
| 予後   | LOW  |
| 治療法X | HIGH |

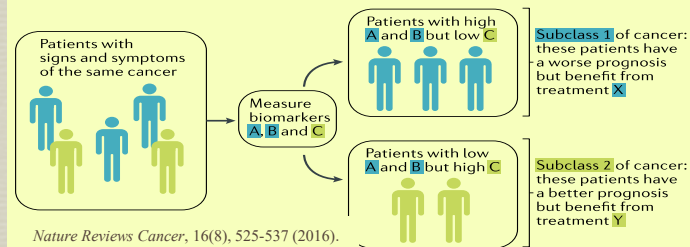
出力例②

|     |      |
|-----|------|
| 分子A | LOW  |
| 分子B | LOW  |
| 分子C | HIGH |

出力例②の解釈:  
分子A,Bの値が低く  
分子Cの値が高い患者は  
予後が良く治療法Yが効く  
傾向がある

|      |      |
|------|------|
| 予後   | HIGH |
| 治療法Y | HIGH |

従来の患者層別化: バイオマーカー



出力に含まれる分子A, B, Cは  
患者層別化バイオマーカーの  
有力な候補となる。

→ データ駆動的な患者層別化

分子A, B, Cの増減パターンから  
体内で起きている分子レベルの  
応答を推論することで、創薬標的  
候補を得ることができる。

→ 創薬標的探索

# その他の創薬AI

➤ 患者層別化AIを含む25種の新規AIを完成 (Web applicationとして公開)

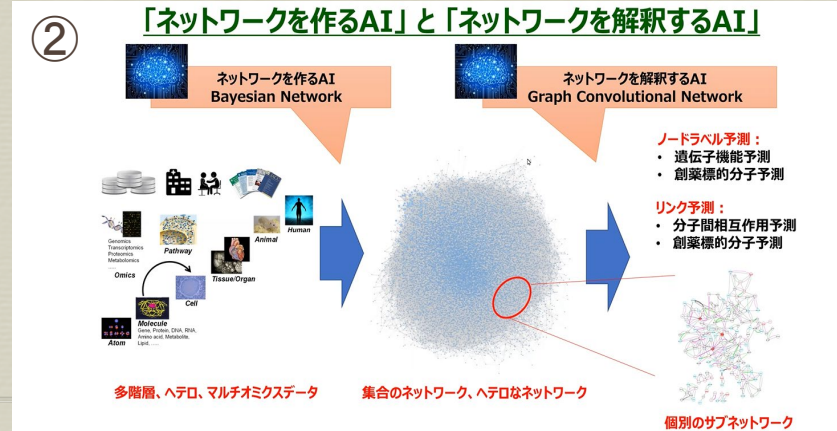
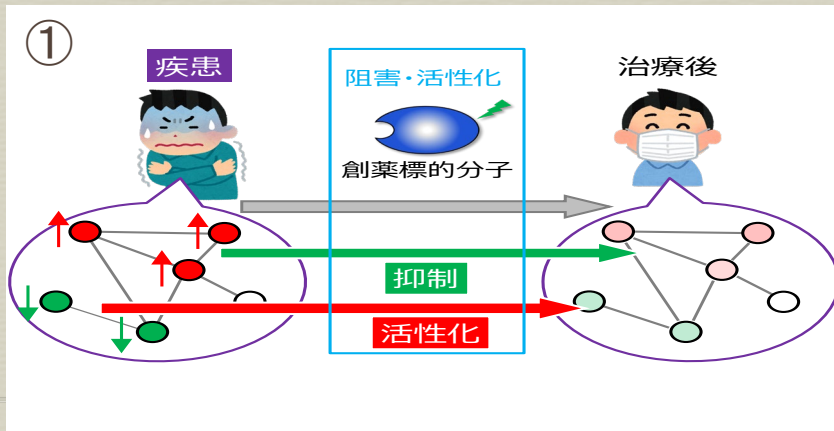
## ① 医薬ビッグデータの統合によって創薬標的探索をするAI

これまで得られている様々なデータを利用し、タンパク質の挙動や遺伝子摂動プロファイルから疾患の創薬ターゲット分子を予測する。

## ② 診療情報や分子の関係性を表すネットワークを作るAI、それを解釈するAI

患者のオミクスデータや診療情報から、分子や診療情報の項目の関係性を表すネットワークを作成する。

IPF患者と健常者のネットワークを比較することによって、疾患に特徴的な生体分子を探索する。





# 事業成果共有プラットフォーム 峰



## ■ 峰について

「峰」は、官民研究開発投資拡大プログラム(Public/Private R&D Investment Strategic Expansion Program：PRISM)の枠組みに基づいた「新薬創出を加速する人工知能の開発」省庁連携研究プロジェクトの成果（収集した医療データや開発した解析プログラム・予測モデル）を公開し、手軽に体験していただく場として公開しているプラットフォームです。

当該事業の対象疾患である特発性肺線維症(IPF)と肺がんにおいて、新規創薬ターゲット（医薬品が作用するタンパク質等の生体内分子）をデータ駆動的に見出すことを目的として、これまでに医療データや既存知識を収集しました。また、これらのデータから創薬ターゲット探索を行うための新規解析プログラム、医療テキストや学術論文から医学・生物学分野の専門用語を自動抽出する自然言語処理プログラムなどの開発を行ってきました。

詳しいプログラムとデータについてはプログラム&データの紹介を、利用については利用方法&申し込み方法を参照していただくようお願いいたします。

峰で公開しているプログラム・予測モデルとデータは、当該事業に参画している下記機関より提供されました。今後も順次追加予定です。

以下、50音順  
医薬基盤・健康・栄養研究所  
九州工業大学  
京都大学  
国立がん研究センター  
産業技術総合研究所  
奈良先端科学技術大学院大学  
理化学研究所




















## お知らせ

2023年3月10日 | お知らせ  
利用申請の受け付け

2022年3月29日 | お知らせ  
ウェブサイト公開しました。

[お知らせ一覧](#)

## Garuda Gadgets

|  |  |  |   |
|--|--|--|---|
| <br>kGCN Network<br>by NIBIOHN    | <br>Multiomics_Anal<br>by NIBIOHN | <br>SFMEDM<br>by NIBIOHN          | <br>NamedEntityRec<br>by NIBIOHN |
| <br>EntityLinker<br>by NIBIOHN    | <br>TargetMine<br>by NIBIOHN      | <br>SubsetBinder<br>by NIBIOHN    | <br>MoEnc<br>by NIBIOHN          |
| <br>JaMIE<br>by NIBIOHN           | <br>Modified_Diet_N<br>by NIBIOHN | <br>Vanishing Ranki<br>by NIBIOHN | <br>RPPA<br>by NIBIOHN           |
| <br>Semantic Search<br>by NIBIOHN | <br>HeaRT<br>by NIBIOHN           | <br>PathoGN<br>by NIBIOHN         | <br>INGOR<br>by NIBIOHN          |
| <br>INGOR ECv<br>by NIBIOHN       | <br>INGOR RC<br>by NIBIOHN        | <br>INGOR Network<br>by NIBIOHN   |   |

# PRISM創薬AI 成功の鍵

---

## ・明確なゴール

事業として目指すゴールを「データ駆動的な創薬標的探索」と具体的に設定。  
それにより、専門が異なる研究者間で「何を実施しなくてはならないか」を  
お互いの専門領域に落とし込みながら議論・提案することができた。

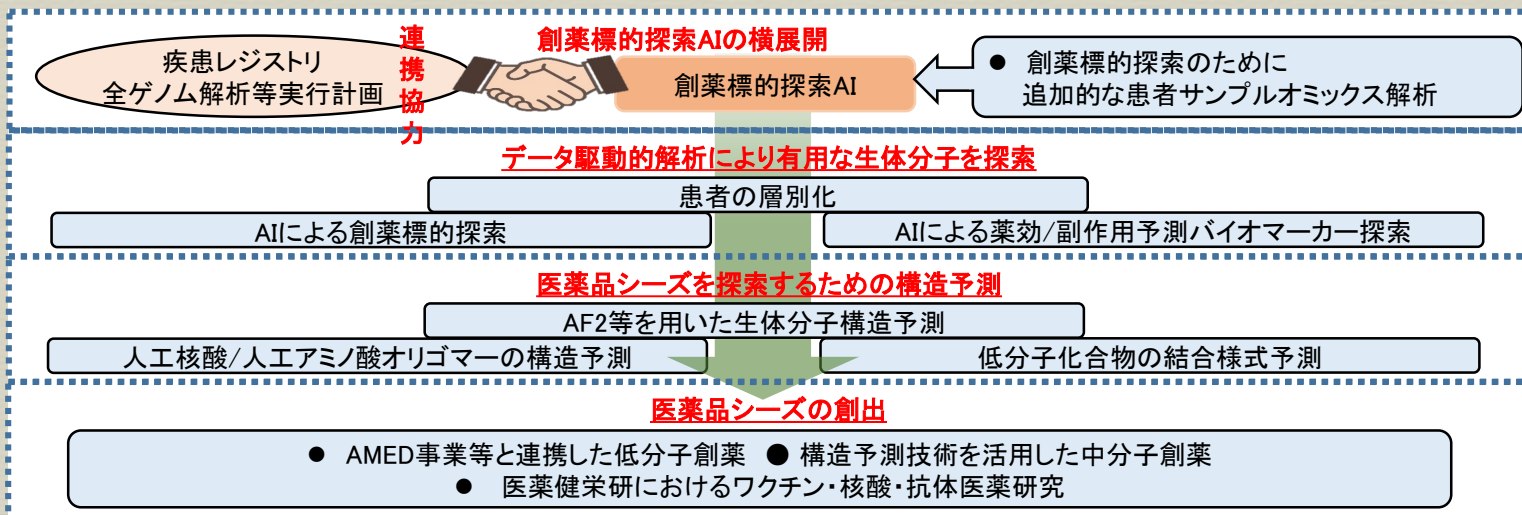
## ・異なる専門を有した研究者による有機的な連携

PRISM創薬AIでは多くの研究機関が参画していたが、専門が異なっているがゆえに  
全体として「寄せ集めになっていないか」という観点で事業開始当初より  
研究者間の有機的な連携が重要視されていた。

創薬における問題点と必要なタスクの洗い出しを実施し、各タスクに最適な研究機関がアサインされる  
ことにより、全体として各研究機関の存在意義が明確になると同時に全体として調和が生まれた。

# PRISM創薬AIのその後

- 肺がん、IPF以外の疾病にも創薬標的候補探索手法を横展開
  - ⇨ 全ゲノム解析等実行計画やAMEDで実施しているGAPFREE事業等との連携
- 創薬標的を特定し、さらにAIを用いて医薬品シーズを創出
  - ⇨ 当所で培ってきた技術の他、理研、京大及びLINCとの富岳を利用した共同研究やAMED事業等と連携





# ポストPRISM創薬AIの課題

- ・ゲノム情報や診療情報といったデータをAI解析に適したビッグデータとして扱うためには、複数の施設・事業間での連携が必要となる。しかし、初めからそのような連携が想定されている場合を除いては患者からの同意取得内容と齟齬がないかどうかの確認や、セキュアなデータ共有手段の確保、関係者のインセンティブなどの各種確認・調整が必要となる。
- ・PRISM創薬AIでは診療情報、オミックスデータともに1時点のみでデータ収集を行なったが、薬の奏功や疾患の増悪など創薬において重要な情報を得るためには時系列(複数の時点)データの収集が望ましい。しかし、そのためにはまとまった予算を安定して得られる状況である必要がある。

# 今後の課題

- ・日本の創薬AI研究において、データの取得の負担が大きい。

## 診療情報:

AIフレンドリーな状態で蓄積されていない

情報セキュリティや個人情報保護の観点で利活用のハードルが高い

## オミックスデータ:

特にゲノム情報については利活用のハードルが高い

データ収集の負担が費用面で大きい

- ・上記のような医療ビッグデータを扱うことができる人材育成が急務  
必要なのはIT人材というよりも、医療ビッグデータに関するドメイン知識を有していて  
且つそれに合わせて適切なデータ解析を行うことができる人材。
- ・創薬AIの開発とデータ解析によって得られた出力が創薬につながるかどうかは  
その生物学的妥当性の検証と創薬の種としての有望性の検証を実施できるかどうか依存する。  
アカデミア創薬やスタートアップの支援により、創薬に向けた円滑な橋渡しが必要。



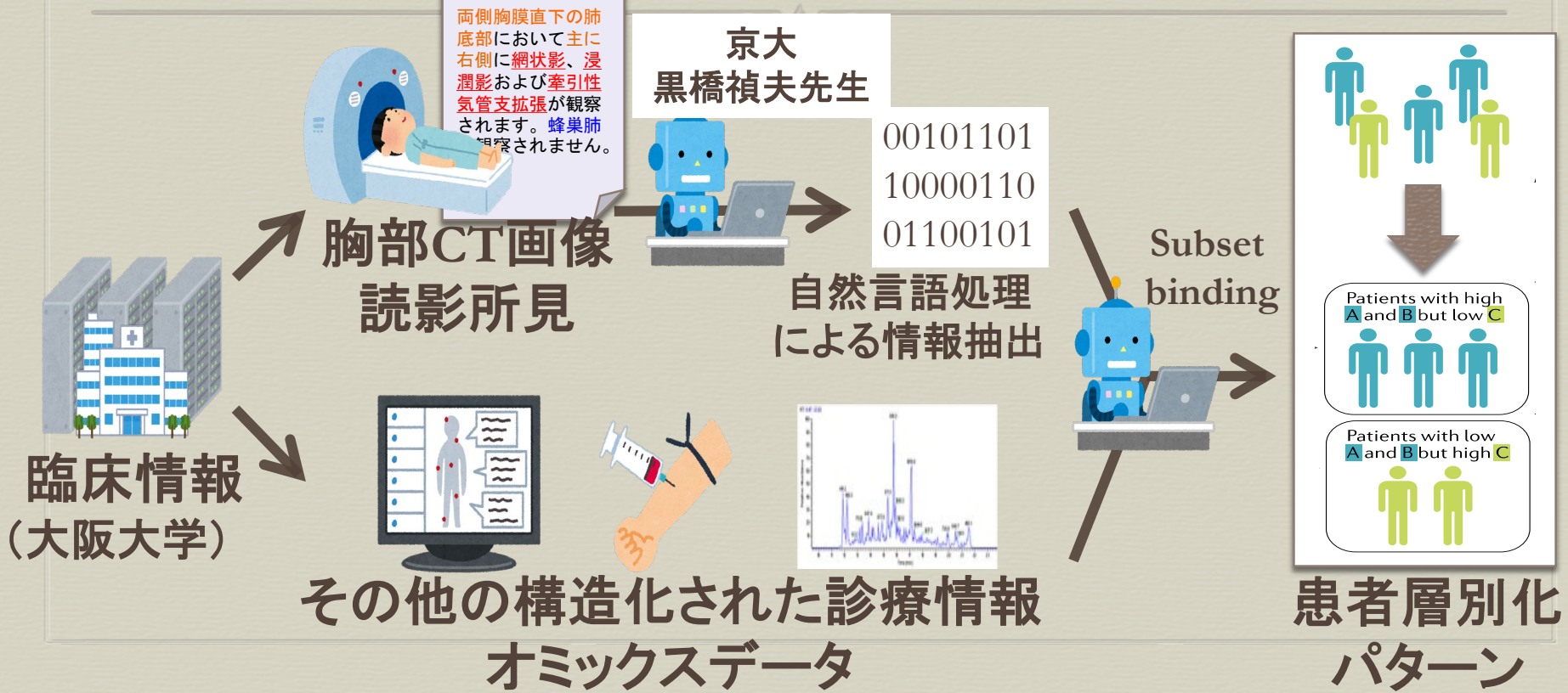
# Appendix



# 患者層別化AIの 妥当性検証プロセス

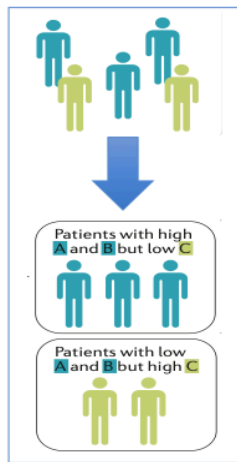
- ① 診療情報、オミックスデータを模した人工データを作成する。  
この人工データの中には、「生体分子AとBの値が高い患者は、診療項目XとYの値が高い」といったパターンを挿入しておく。  
開発した患者層別化AIを用いて、このパターンを検出できるかどうかを確認する。
- ② 患者層別化AIを用いて、実データ(診療情報、オミックスデータ)の解析を行う。次に、診療項目と紐づけられた生体分子の上流解析を行い、疾患に關与する制御因子やパスウェイを推定する。
- ③ 検証実験を実施し、②で見出された制御因子やパスウェイの關与を確認する。

# 大阪大学コホート 臨床情報を用いた創薬標的探索



# Subset bindingが提示した分子の解釈と検証

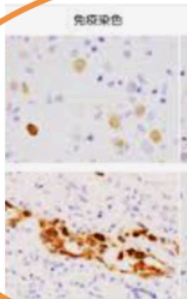
20種類のIPF関連タンパク質  
(7種類のコアタンパク質) を検出



患者層別化  
パターン

エクソソーム内  
プロテオームデータ  
からの生体分子群

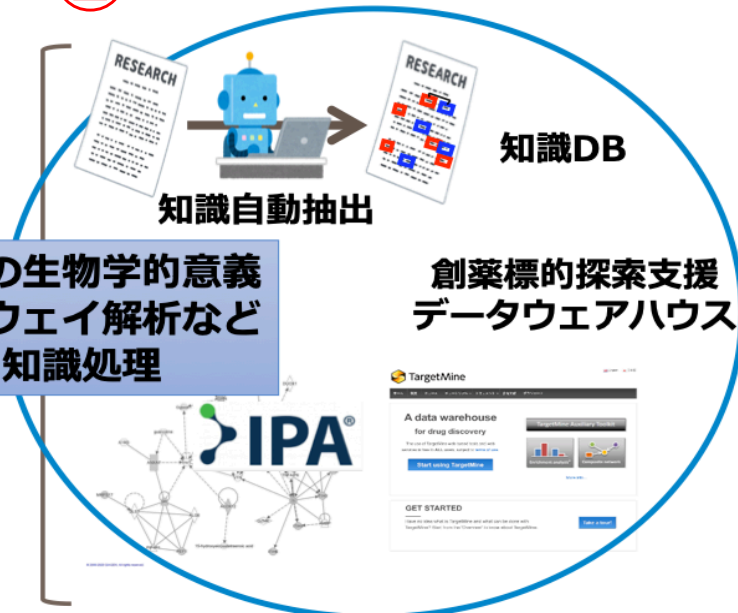
Protein A  
Protein B  
Protein C



肺の病変部位での  
発現を確認する

分子の生物学的意義  
パスウェイ解析など  
知識処理

②

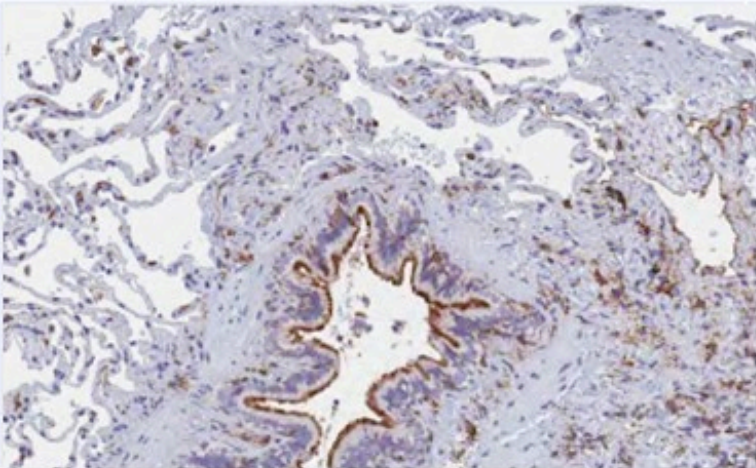
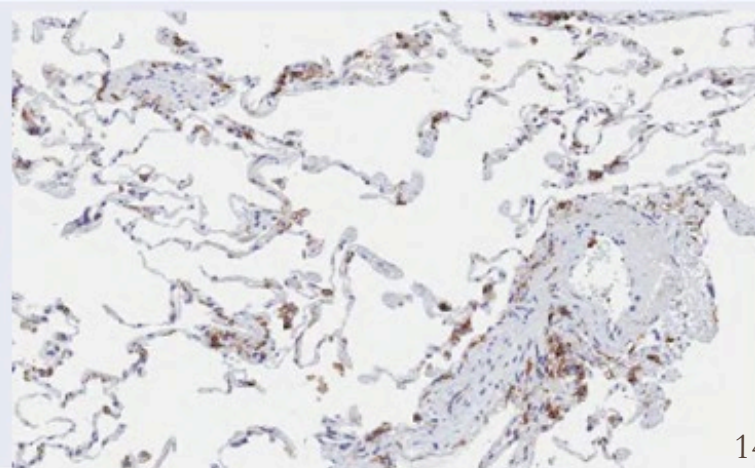


③

代謝物質の変動  
ヒト細胞試験系での確認



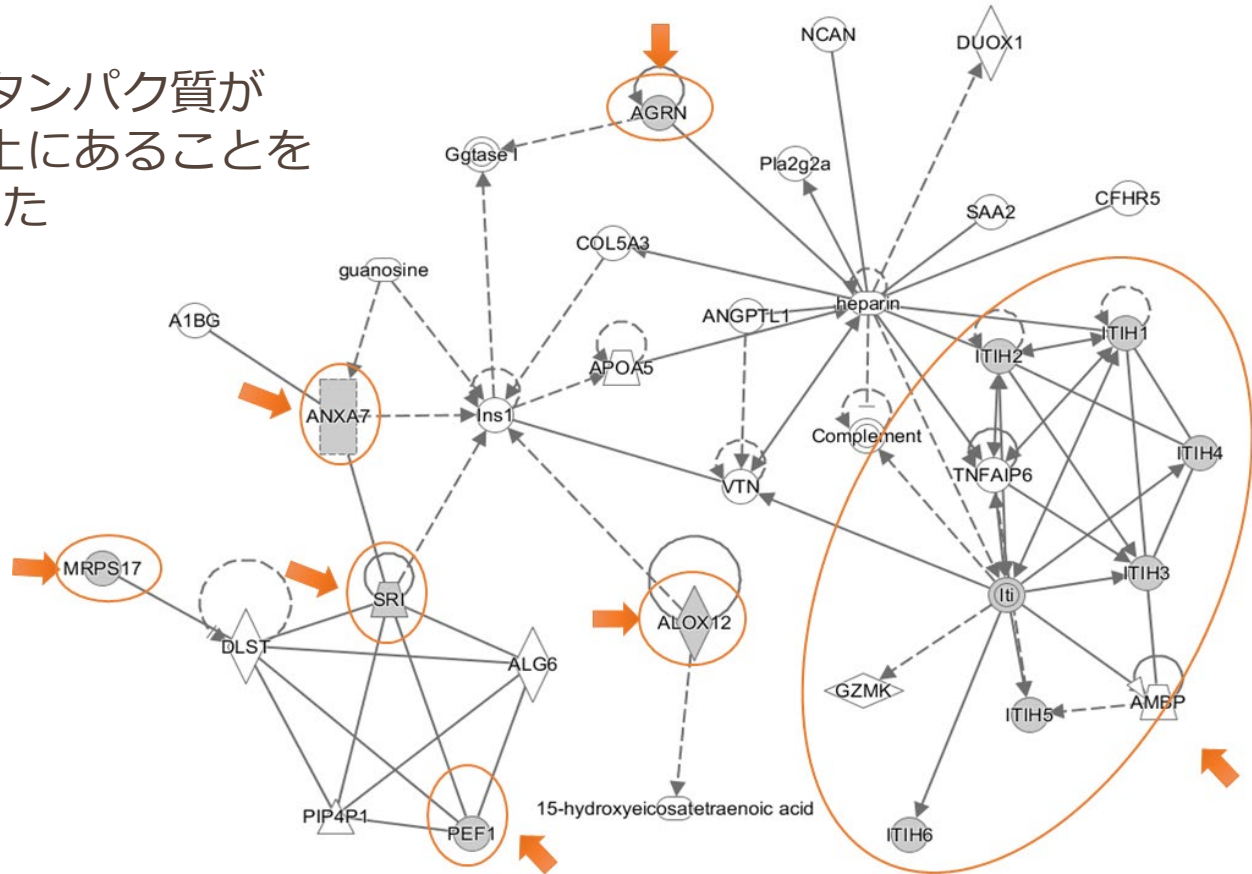
# 見出されたタンパク質の線維化部位における発現亢進

|                                 | 線維化部位   | 正常部位   |
|---------------------------------|---|--|
| 線維化<br>マッソント<br>リクローム<br>染色(MT) | +   | -  |
| 抗LYN<br>抗体による<br>免疫染色           |  |  |

# コアタンパク質の発現を制御する上流因子の探索

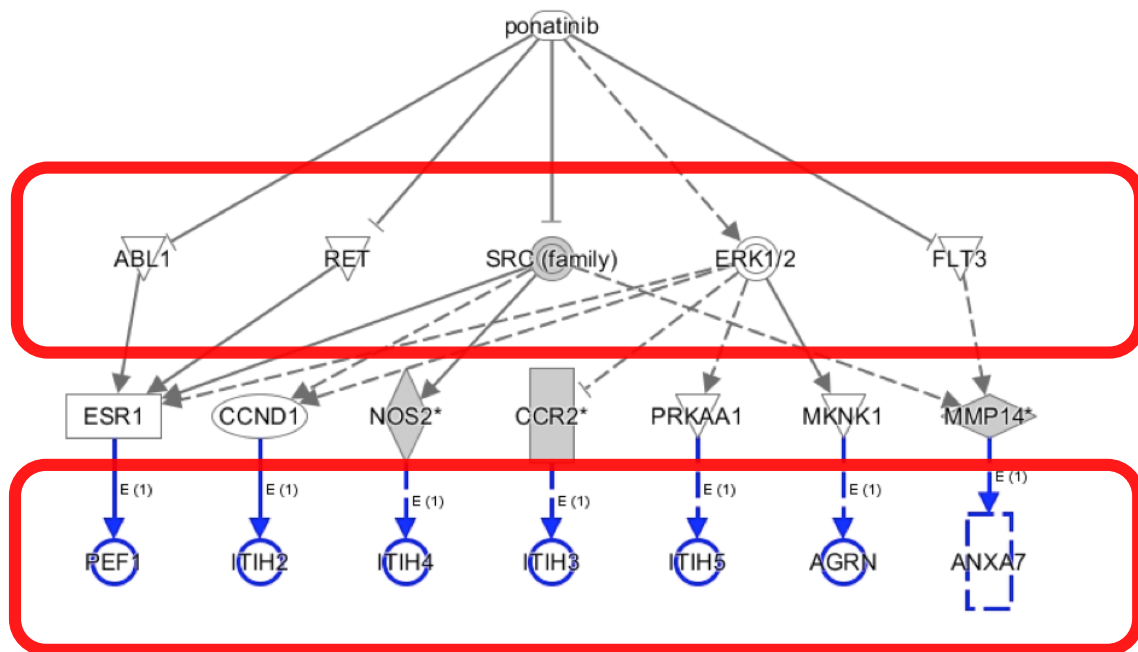
Network 1 : pplPAcore - 2020-09-13 03:35 □□ : pplPAcore : pplPAcore - 2020-09-13 03:35 □□

見出されたコアタンパク質が  
同一ネットワーク上にあることを  
見出した



# 創薬標的候補の提示とdrug repositioning

上流の分子群をponatinibが制御することを見出した



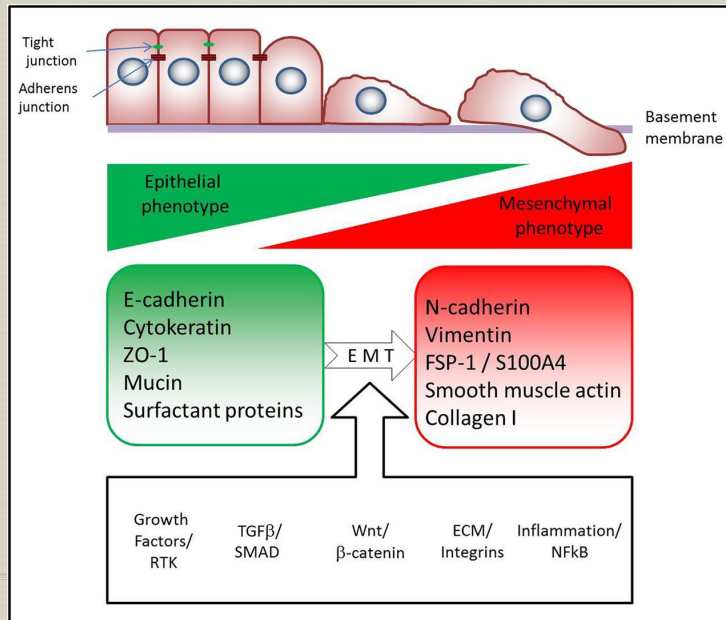
IPF関連タンパク質を  
制御する因子

患者層別化AIが検出した  
IPF関連タンパク質  
(コアタンパク質)



# EMT試験系

## EMT: Epithelial Mesenchymal Transition (上皮細胞の間葉系細胞への形質転換)



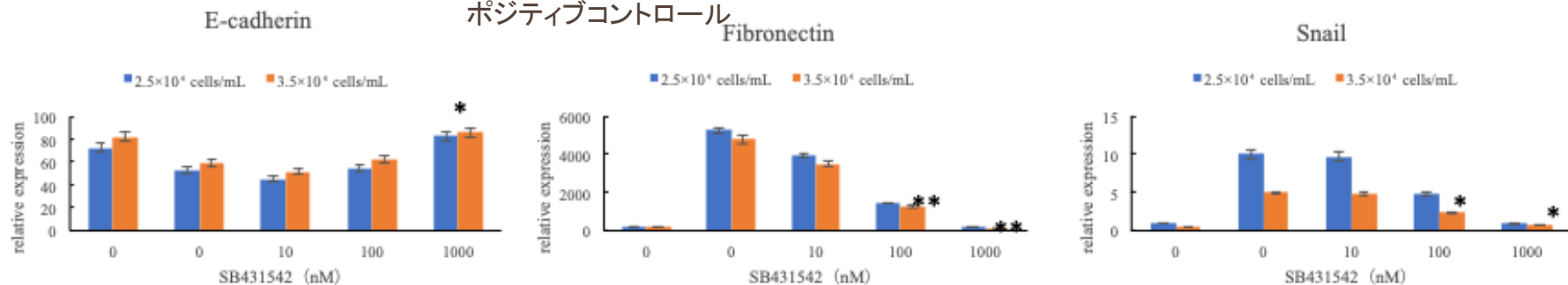
IPFにおける肺線維化の機序に  
EMTの重要性が示唆されている

→in vitroでEMT試験系を樹立し、  
見出したタンパク質が線維化に  
どのように寄与しているかを確認

# ponatinibは濃度依存的にEMTを抑制した

## SB431542 + TGF- $\beta$ (3 ng/mL)

ポジティブコントロール



## Ponatinib + TGF- $\beta$ (3 ng/mL)

