

第4回 ABCIUGウェビナー

2022年10月14日(金) 14:00~15:00

診療情報と生体分子の測定データから 薬の標的を探し出すために

医薬基盤・健康・栄養研究所

AI健康・医薬研究センター

夏目やよい

まず始めに自己紹介

- 2003年 東京大学 農学部 生命化学専修卒業
- 2005年 東京大学大学院 農学生命科学研究科 応用生命化学専攻修士課程修了
- 2008年 東京大学大学院 農学生命科学研究科 応用生命化学専攻博士課程修了
- 2008年 京都大学 化学研究所附属バイオインフォマティクスセンター 特定研究員
- 2010年 科学技術振興機構 さきがけ 専任研究員
- 2015年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト プロジェクト研究員
- 2016年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト 研究員
- 2018年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト サブプロジェクトリーダー
- 2021年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト プロジェクトリーダー (現職)
- 2022年 徳島大学 先端酵素学研究所 特任教授 (現職)

まず始めに自己紹介

2003年 東京大学 農学部 生命化学専修卒業

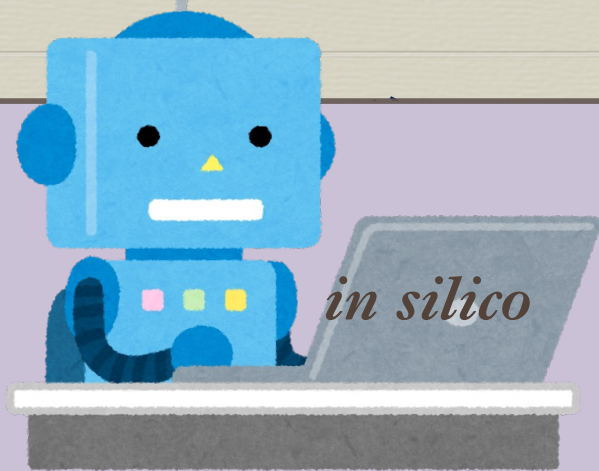
2005年 東京大学大学院 農学生命科学研究科 応用生命化学専攻修士課程修了

2008年 東京大学大学院 農学生命科学研究科 応用生命化学専攻博士課程修了



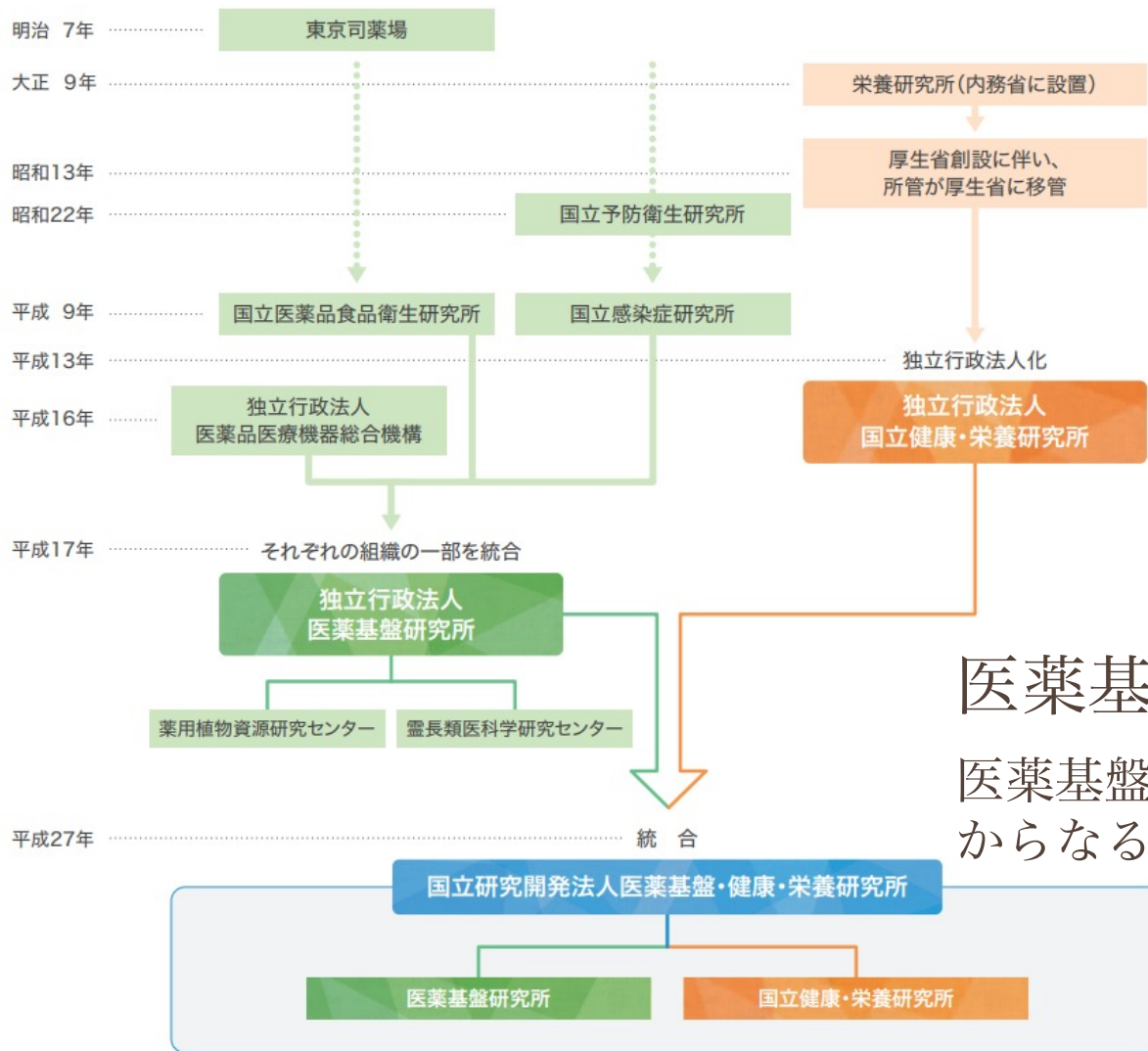
専門

食品科学
毒性学
分子生物学



専門 バイオインフォマティクス 機械学習

- 2008年 京都大学 化学研究所附属バイオインフォマティクスセンター 特定研究員
2010年 科学技術振興機構 さきがけ 専任研究員
2015年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト プロジェクト研究員
2016年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト 研究員
2018年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト サブプロジェクトリーダー
2021年 医薬基盤・健康・栄養研究所
バイオインフォマティクスプロジェクト プロジェクトリーダー (現職)
2022年 徳島大学 先端酵素学研究所 特任教授 (現職)



医薬基盤・健康・栄養研究所：
 医薬基盤研究所と国立健康・栄養研究所
 からなる比較的新しい研究所です

AI健康・医薬研究センターの ミッション

AI

2019年設立

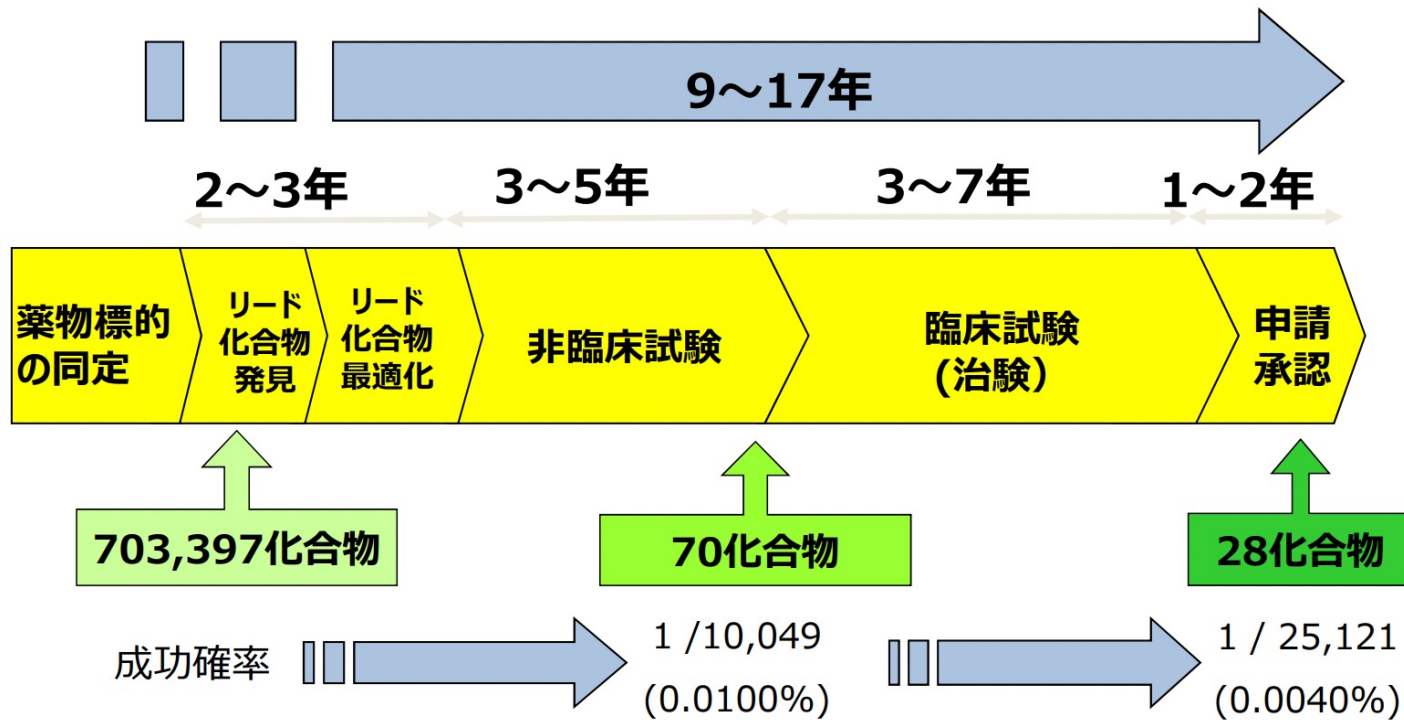
創薬の効率化と、コホート研究を進化させるAI解析技術

創薬・疾患研究から健康・栄養研究にまたがるAI基盤プラットフォームの構築を推進する「AI健康・医薬研究センター」を擁しています。医薬品開発においては創薬標的探索のためのAI開発、薬物動態・毒性予測の統合モデリング、次世代低分子医薬品を創生するプラットフォームの構築を進め、「ヘルス・メディカル連携研究センター」ではヒトを対象にした研究と基礎研究との融合によって得た腸内細菌と生活習慣データの統合解析やデジタルヘルスケアをターゲットにした実績を積み上げています。

医薬品開発に
AIを導入したいのは
なぜか？

医薬品開発に要する期間と成功確率

- 医薬品の開発には10年以上の時間と数百億～数千億円規模の費用が必要。
- 成功確率は年々低下（10年前:1/1.6万→現在:1/2.5万）し、難易度が上昇。



出典：日本製薬工業協会調べ（2011～2015年度）

医薬品開発における課題

- ◆ 以下の理由により医薬品開発の生産性が低下している。
 - 新薬開発の対象が、原因未解明の複雑な疾患にシフト
 - 新薬開発の成功率の低下（20年前:1/1.3万→現在:1/2.3万）
 - 研究開発費の高騰

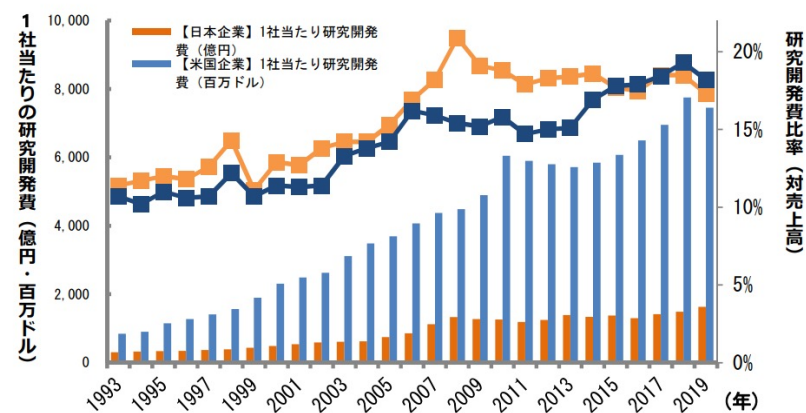
⇒AIの活用による費用・期間の削減・短縮や成功率の向上が期待できる。

■ 新薬開発の成功率（累積成功率）

	2000~2004	2005~2009	2010~2014	2015~2019
前臨床試験開始	1 : 2,158	1 : 3,213	1 : 3,748	1 : 3,740
臨床試験開始	1 : 3,653	1 : 8,698	1 : 9,622	1 : 10,301
承認取得 (自社)	1 : 12,888	1 : 31,064	1 : 24,553	1 : 22,749
承認取得数 (自社)	36	21	29	24

出典:日本製薬工業協会「DATA BOOK 2021」をもとに厚生労働省作成

■ 研究開発費及び研究開発費比率（対売上高）の推移

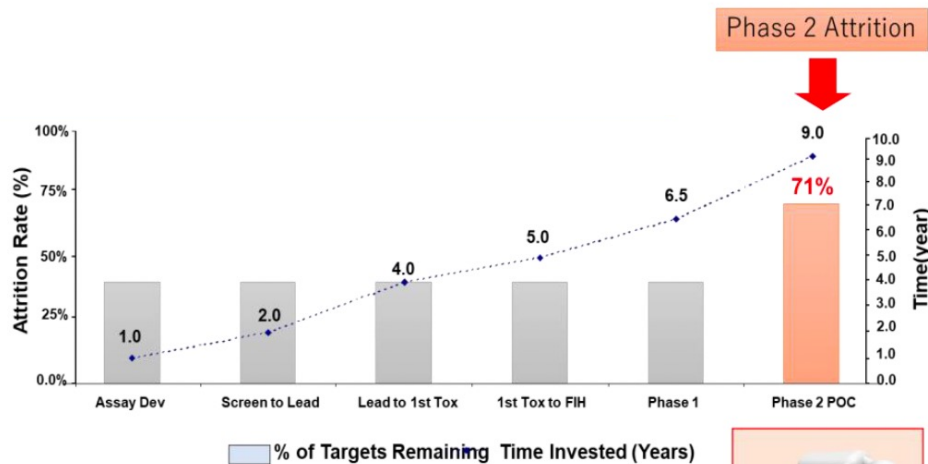


出所: SPEEDA(株式会社ユーザベース)、有価証券報告書、アナニュアルレポート
 出典: 日本製薬工業協会 DATA BOOK2021をもとに医薬産業政策研究所にて作成

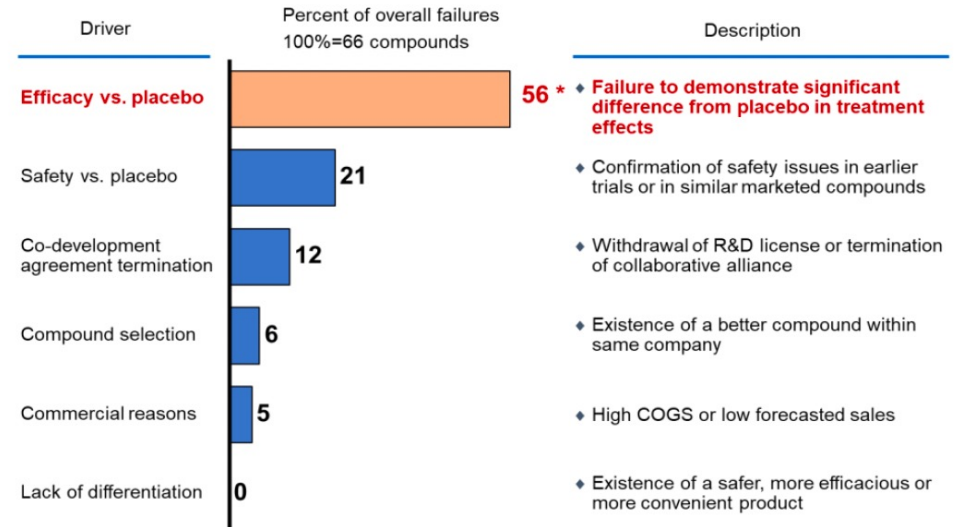
AI創薬の現状

- 多くの製薬企業はIT企業との提携やAI導入を進めている
- 問題は解決したのか？→No
- AIを導入しやすいステップ、しにくいステップがある
- 高精度で実施できるステップ、改善が必要なステップがある
- AIを導入しにくく、かつ改善が必要なステップとは？

医薬品開発の最大の隘路



Assumptions: 1st 5 stages have ~ 60% success rate each, and Phase 2 has ~ 30% success rate. Times adapted from various sources.



* Only 1 indication RIP out of 37 indication RIPs by efficacy vs. placebo had an established MOA
Source: EvaluatePharma, Pharmaprojects; Factiva, Literature Search; team analysis

Phase-2でのPOC取得（Proof of Concept）の失敗率が最も高い

POC取得失敗の最多の原因は「有意な薬効が得られなかった」

臨床 第II相以降に
薬効を実証できない

||

実験動物で認められた薬効が
ヒトでは認められない

新薬開発の初期段階からヒトの情報を基に
創薬標的を探索することはできないか？

創薬標的探索に AIを導入したい



AIの導入が進んでいない：

- ・ AI解析に必要なとなるデータ（診療情報、分子情報）の取得が困難で研究が進んでいない

改善が必要：

- ・ 臨床試験Phase IIでの失敗率の高さの原因

官民研究開発投資拡大プログラム

Public/Private R&D Investment Strategic Expansion Program: PRISM (プリズム)

新薬創出を加速する人工知能の開発



ホーム

PRISMについて

事業紹介

研究体制

アクセス・お問い合わせ

関連リンク

Status: Analysis
CTscan
T1
T2
FLAIR
T1 contrast
新薬創出を加速する人工知能の開発

<https://www.nibiohn.go.jp/prism/>

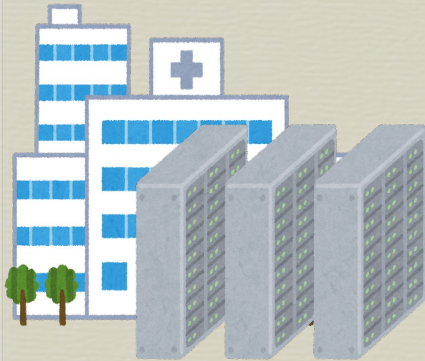
「新薬創出を加速する 人工知能の開発」

これが全てではない！

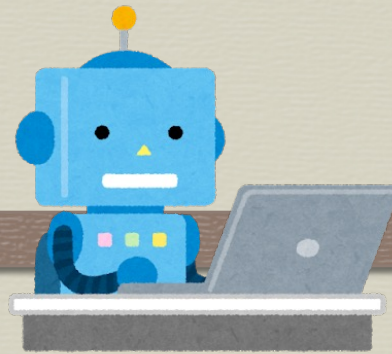
必要な **データ** と **解析技術** が揃って初めて
データ解析ができる

全体概要と対象疾患

肺がん
特発性肺線維症



臨床情報DB構築

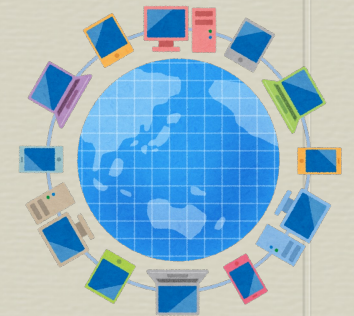


AIによる解析

臨床情報からの
創薬標的提示



創薬標的提示



成果の公開

使用するデータと目的に合わせた
多岐に渡るアルゴリズム開発

オープンプラットフォーム
構築

本事業で進める4つのプロセス

- ①臨床情報の収集とデータベースの構築
- ②臨床情報の利活用を可能とするAIの開発
- ③臨床情報からの創薬標的提示
- ④オープンプラットフォームによる事業成果公開

①臨床情報の収集とデータベースの構築

対象疾患：肺がん および 特発性肺線維症 (IPF)

国立がん研究センター
National Cancer Center Japan

国立がん研究センターは社会と協働し、
全ての国民に最適ながん医療を提供します

NIBIOHN 国立研究開発法人
医薬基盤・健康・栄養研究所
National Institutes of
Biomedical Innovation, Health and Nutrition

お問い合わせ ・ アクセス ・ リンク ・
文字サイズ 中 大 特大 サイト内

研究所紹介 研究と活動 情報公開 関連情報

National Institutes of
Biomedical Innovation,
Health and Nutrition

医薬基盤研究所 (NIBIO) 記事一覧

2020年7月9日 **New**
お知らせ 令和2年度 希少疾病用医薬品・希少疾病用医療機器・希少疾病用再生医療等製品試験研究助成金の交付品目のお知らせ

国立健康・栄養研究所 (NIHN) 記事一覧

2020年3月31日
お知らせ 健康日本21 (第二次) 分析評価事業のページが更新されました

2020年2月16日

①臨床情報の収集とデータベースの構築

対象疾患：肺がん および 特発性肺線維症 (IPF)

国立がん研究センター
浜本隆二先生

世界最大規模の
肺がんDB

本発表

世界初の
IPF DB



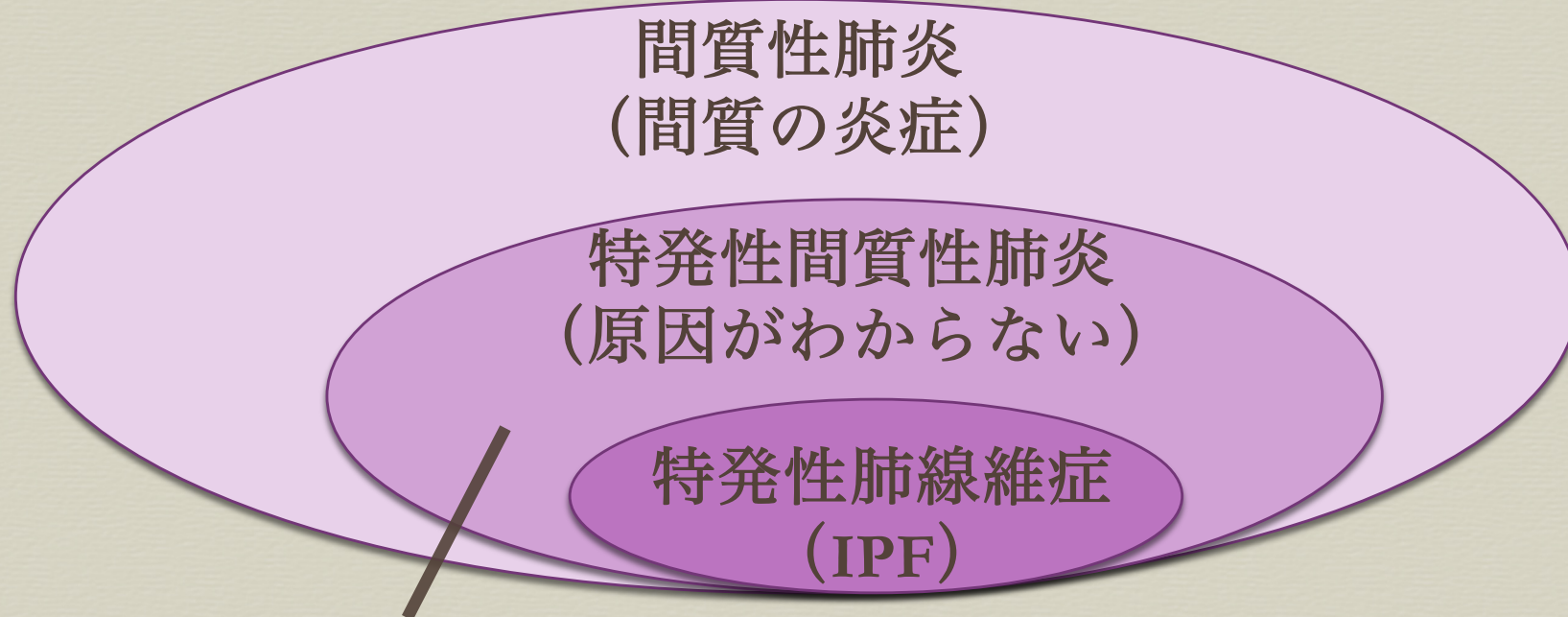
「新薬創出を加速する人工知能の開発」

対象疾患：特発性肺線維症（IPF）

「原因がわからない」

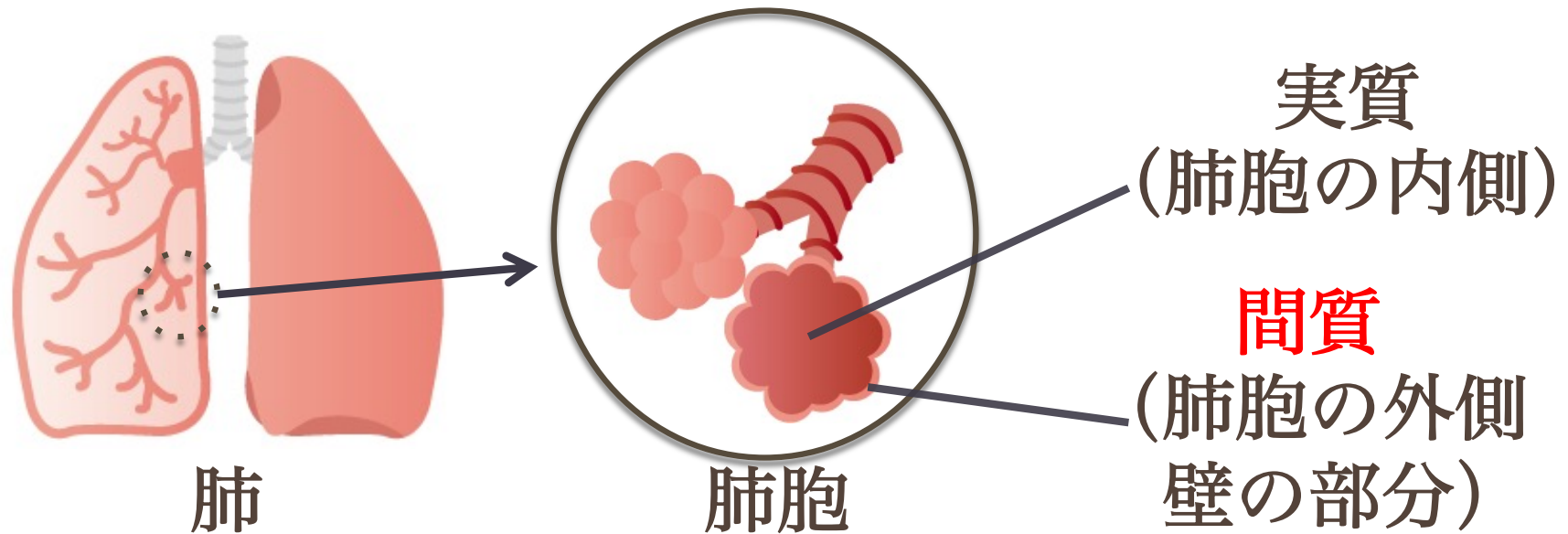
「肺が線維化する病気」

IPFを含む間質性肺炎



国の指定難病

間質性肺炎



間質性肺炎：間質で起こる炎症

特発性間質性肺炎の分類

IPFは①頻度が高い、②治療が難しい、③予後不良
→IPFか、そうでないかの判断が重要



CT画像による
画像診断



肺生検による
病理診断
(必要に応じて)

特発性間質性肺炎

5年生存率は20-40%

特発性間質性肺炎は7種類に分類

特発性肺線維症 (IPF)

非特異性間質性肺炎

特発性器質化肺炎

剥離性間質性肺炎

呼吸細気管支炎を伴う間質性肺疾患

リンパ球性間質性肺炎

急性間質性肺炎

間質性肺炎の データ駆動的**患者層別化**



間質性肺炎DB
診療情報＋生体分子測定データ

データ駆動的に

- IPF患者を判別
できるだろうか？
- IPF患者を更に細分化
できるだろうか？
(e.g. 薬の奏功)

IPFを含む間質性肺炎の 臨床情報収集拠点

①大阪大学コホート→データ解析へ

大阪大学大学院医学系研究科 呼吸器・免疫アレルギー内科

熊ノ郷 淳先生

②神奈川県立循環器呼吸器病センターコホート

神奈川県立循環器呼吸器病センター

小倉高志先生

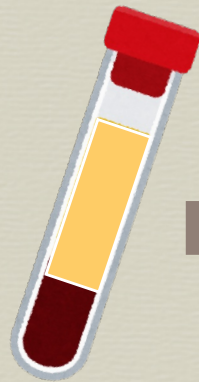
①大阪大学コホート

対象者：

①IPFを含む間質性肺炎患者

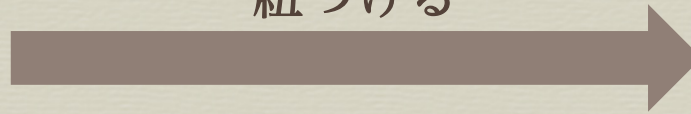
②検査の結果、器質的な呼吸器疾患を有さないと診断された方々（「健常者」として扱う）

①大阪大学コホート



血清

採血日に最も近い
日程における情報と
紐づける



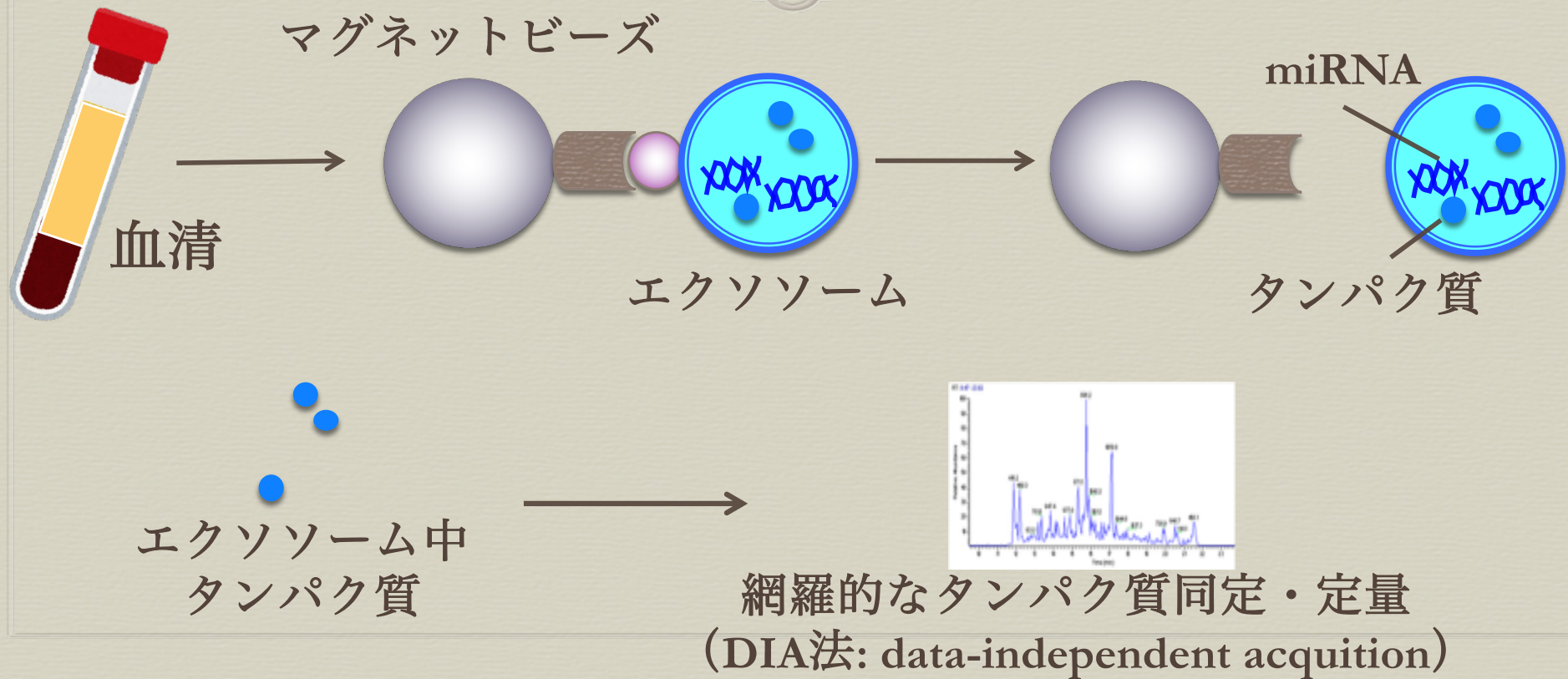
診療記録
CT画像＋読影所見
血液検査値
患者基本情報
初診時間診票



大阪大学医学部呼吸器・免疫内科
バイオバンク
武田吉人先生

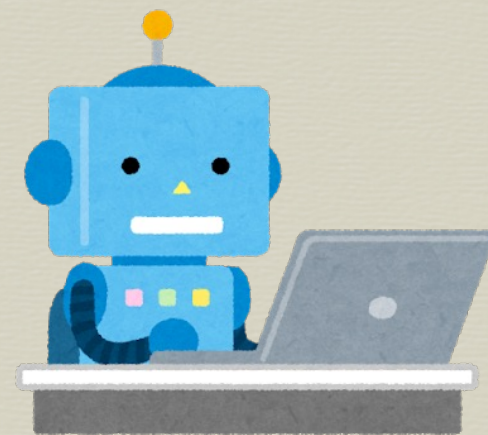
大阪大学医学部附属病院
医療情報部
松村泰志先生
武田理宏先生

オミックスデータ： 血清中エクソソーム プロテオーム解析



大阪大学コホートで収集された臨床情報

- 血清エクソソームプロテオームデータ **602件**
- 診療記録＋患者基本情報 **594件**
- 血液検査値 **594件**
- 読影所見 **538件**
- 初診時間診票 **141件**
- CT画像 **323件**



AIによる創薬標的探索へ

本事業で進める4つのプロセス

- ①臨床情報の収集とデータベースの構築
- ②臨床情報の利活用を可能とするAIの開発
- ③臨床情報からの創薬標的提示
- ④オープンプラットフォームによる事業成果公開

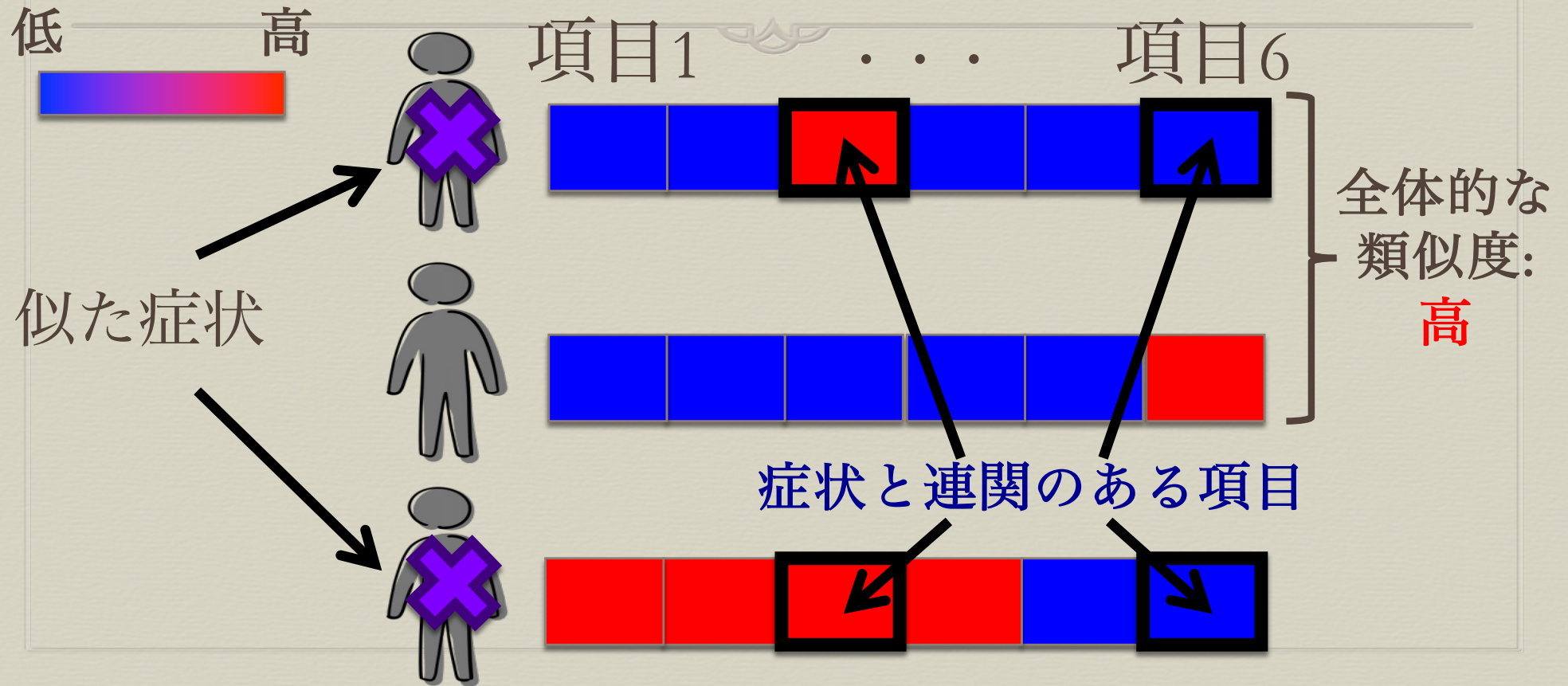
患者層別化と創薬標的探索

対象疾患の患者

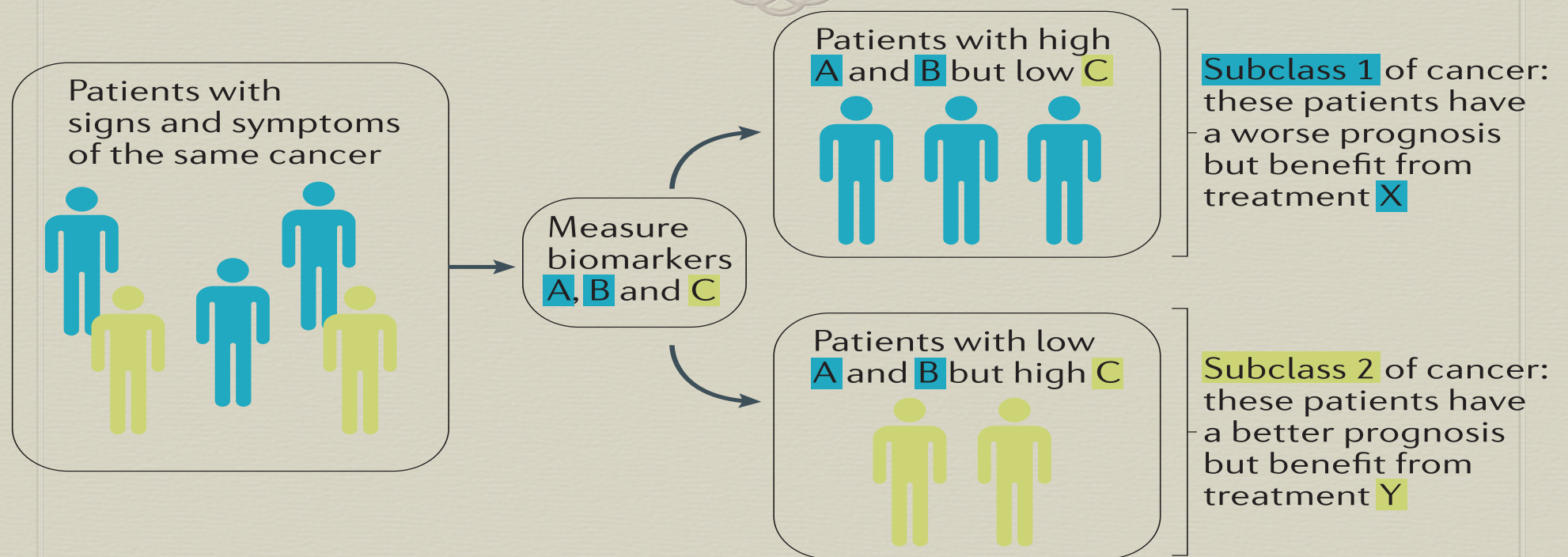
患者数が比較的少ない
疾患サブグループに分割

疾患サブグループに対して
「期待通りに効く薬」の開発を目指す

クラスタリングの欠点

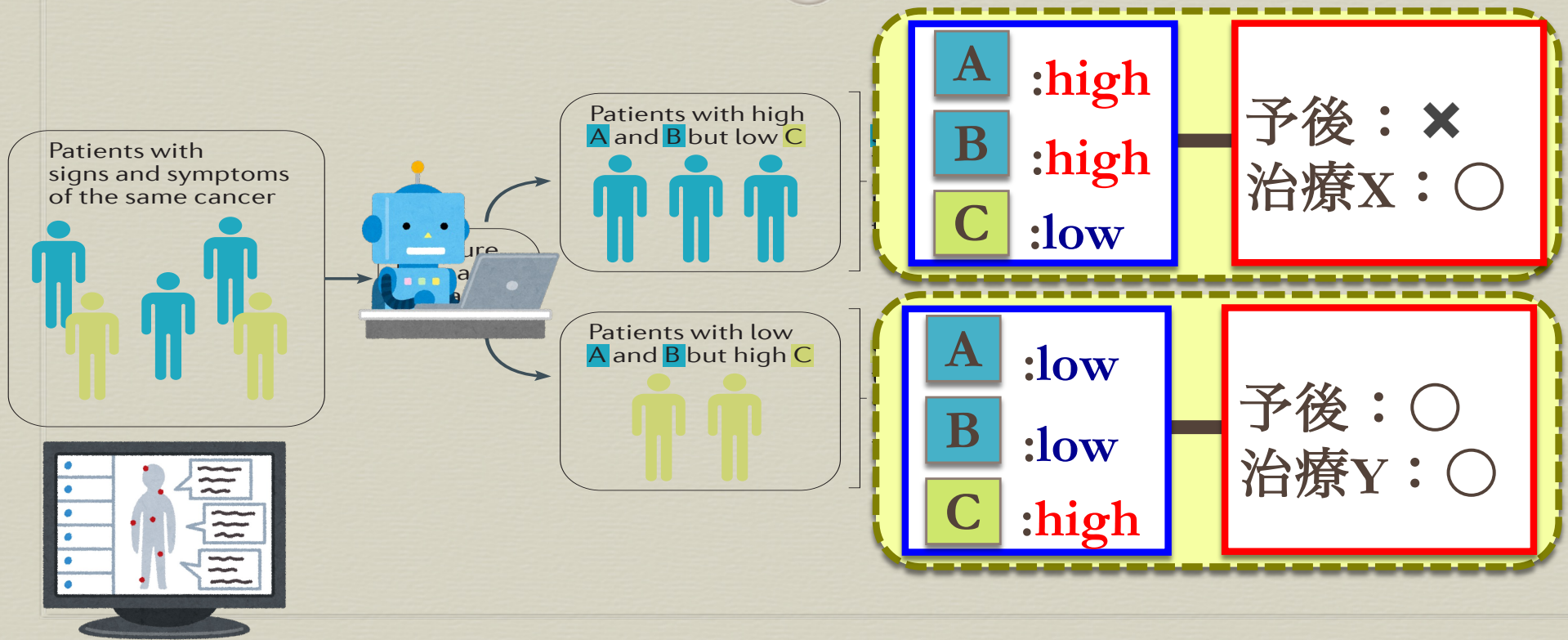


患者層別化（従来法）



Nature Reviews Cancer, 16(8), 525-537 (2016).

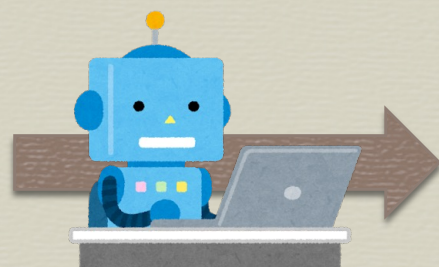
患者層別化 (新規法)



あらゆるデータを層別化するsubset binding (理研AIP 上田修功先生との共同研究)

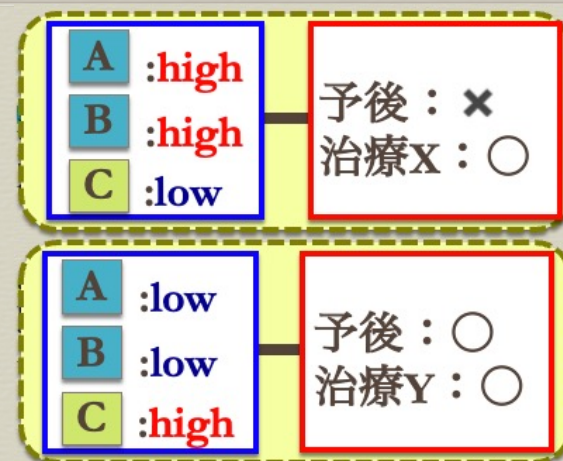
例) IPF臨床情報 (大阪大学)

血中タンパク質データ
診療情報



層別化

AI



入力 :

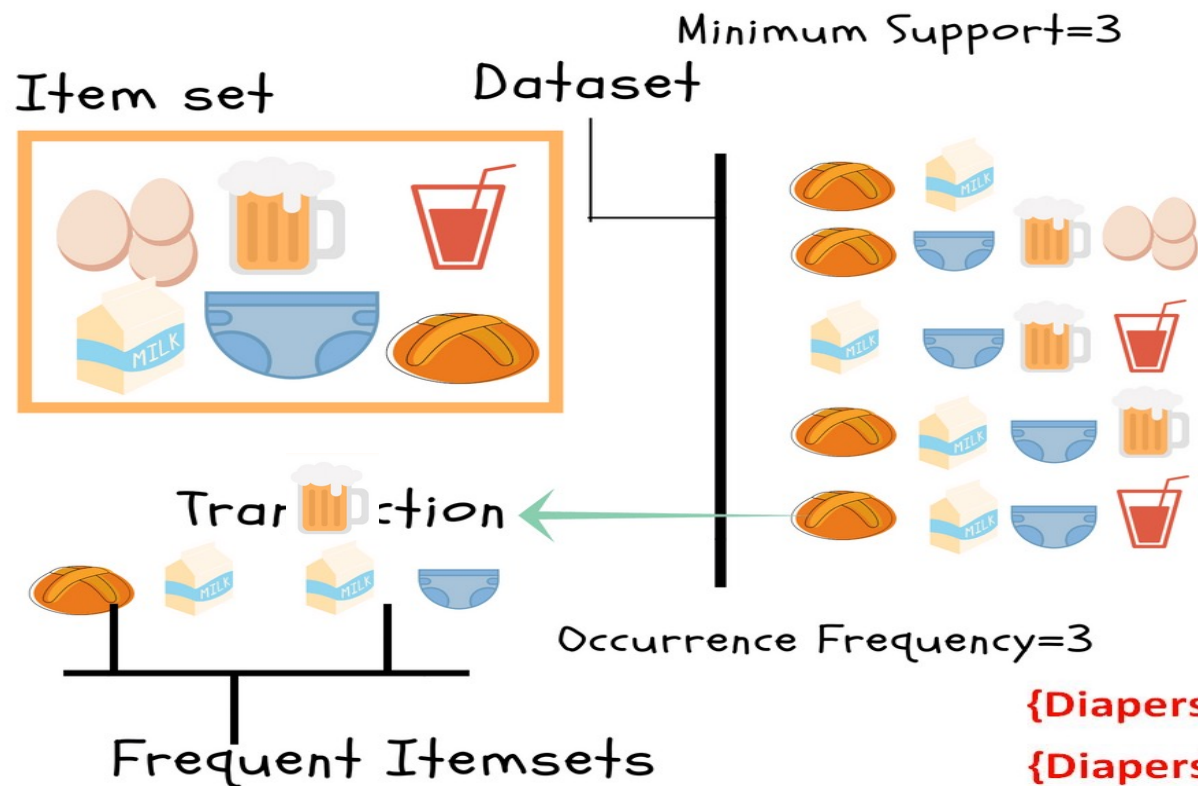
二種類のデータ

出力 :

層別化の条件

層別化された患者を血液から見つけ出せる
層別化された患者に対する創薬標的探索のヒントを得られる

BASIS of the algorithm (association rule mining)



Market basket transactions

ID	Items
1	{Bread, Milk}
2	{Bread, Diapers , Beer , Eggs}
3	{Milk, Diapers , Beer , Cola}
4	{Bread, Milk, Diapers , Beer }
5	{Bread, Milk, Diapers, Cola}
...	...

{Diapers, Beer}

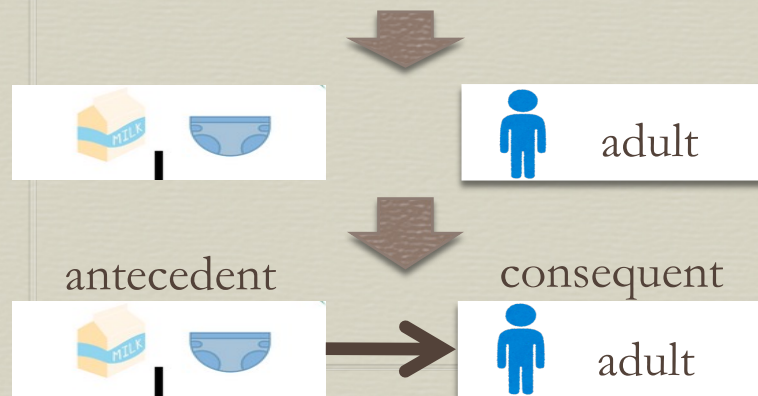
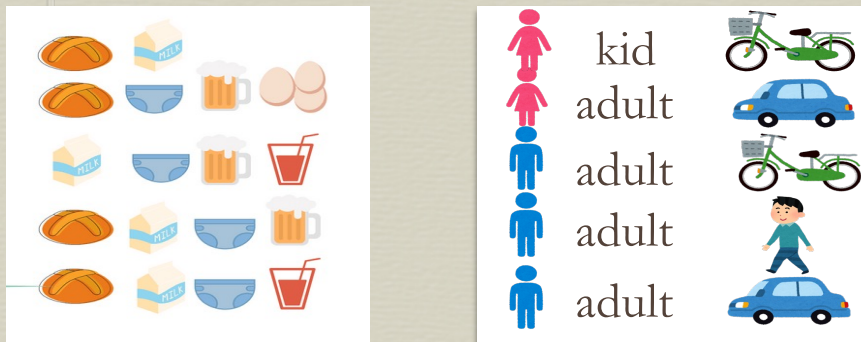
Example of a frequent itemset

{Diapers} → {Beer}

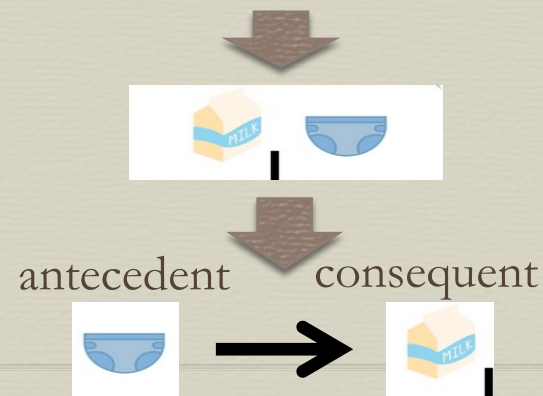
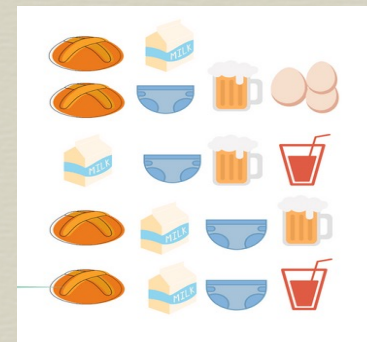
Example of an association rule

Novelty of this algorithm

OUR ALGORITHM



CANONICAL

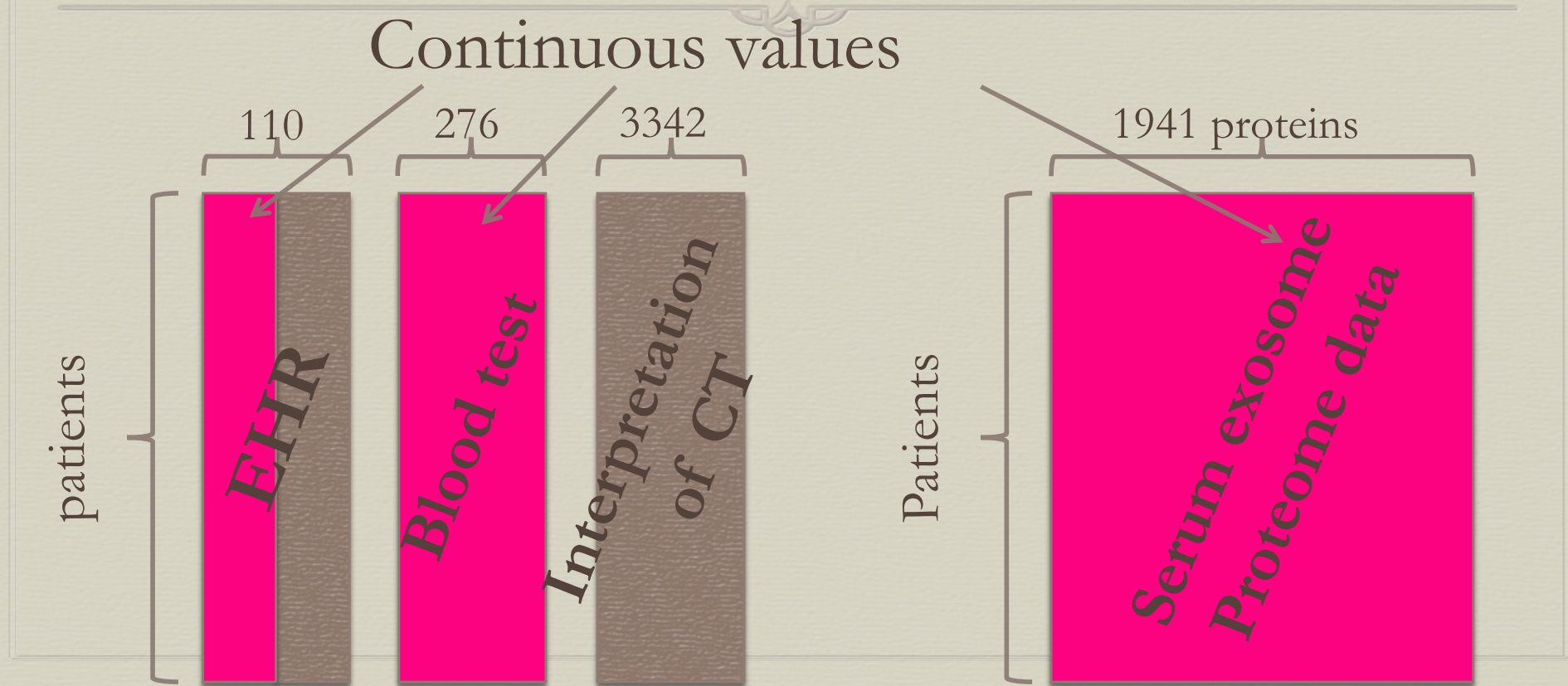


dataset

Frequent
itemsets

Association
rules

Issue: how to handle continuous values



Fuzzy association rule mining

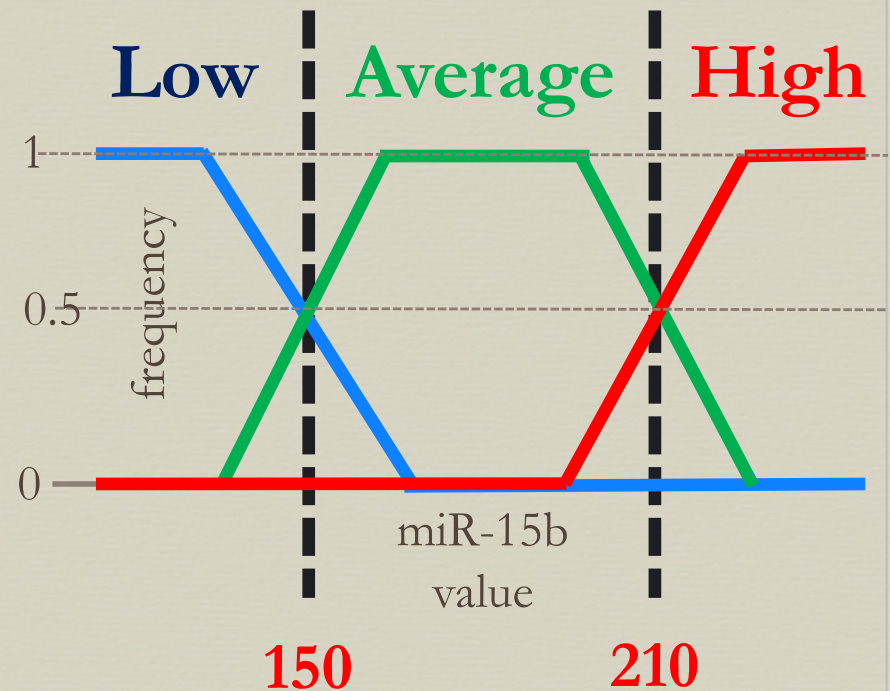
PatientID	miR-15b
1	125.2
2	173.3
3	219.5
4	202.8
	⋮
	⋮

...

PatientID	miR-15b Low	miR-15b Average	miR-15b High
1	1	0	0
2	0	1	0
3	0	0.3	0.7
4	0	0.6	0.4

...

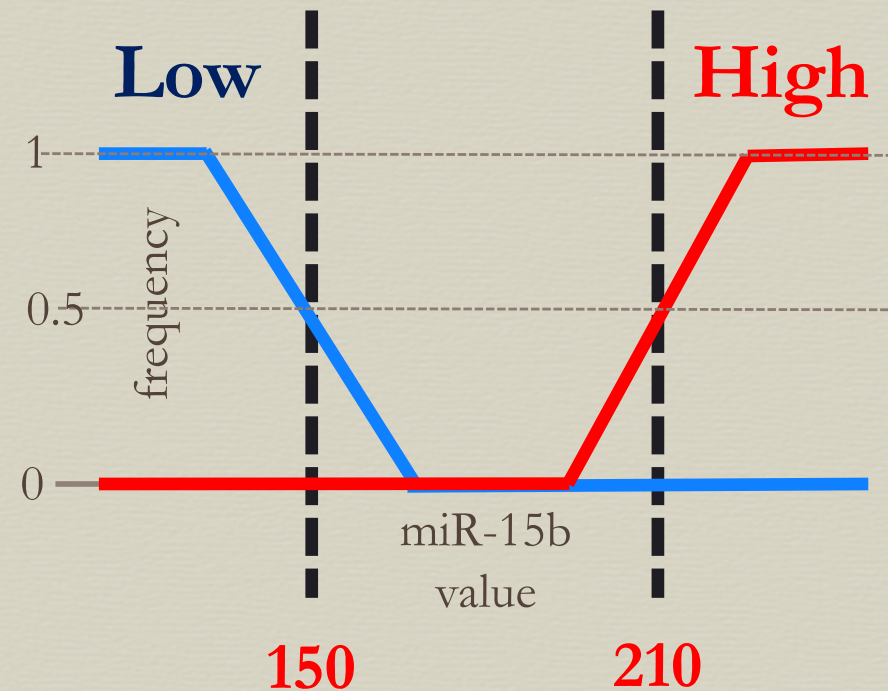
⋮
⋮
⋮



Fuzzy association rule mining

PatientID	miR-15b
1	125.2
2	173.3
3	219.5
4	202.8
	⋮
	⋮

PatientID	miR-15b Low	miR-15b Average	miR-15b High
1	1	0	0
2	0	0	0
3	0	0.3	0.7
4	0	0.6	0.4
	⋮	⋮	⋮



長所



- 2つの異なるデータから、それらを紐づける項目を出力する
(入力データは臨床情報である必要がなく汎用性が高い)
- データ駆動的に層別化をすることができる
(事前知識を必要としないので解析したいデータさえあれば使用可能)
- 離散値でも連続値でもそれらの混合でも解析可能

短所

- ・ 項目数が大きくなるとメモリ負荷が大きくなる

→ 生体分子を測定したデータは項目数が大きいことが多い！

今回解析したデータでは、タンパク質の種類は約2000
これは生体分子測定データとしては少ない方

研究現場の裏話・・・

- ・ 自前のサーバーでメモリ不足、計算が止まってしまう
 - アルゴリズムを改良するか、マシンを変えるか？
 - 最終目標は「創薬標的のデータ駆動的探索が可能だ」という Proof of Concept」
 - 一定の期間で成果を挙げなくてはならない
 - 後者の選択肢に決定

ABCIはポイント制、利用申請をすればすぐに使うことができるので
利用量の様子を見ながらデータ解析を試してみることができた

本事業で進める4つのプロセス

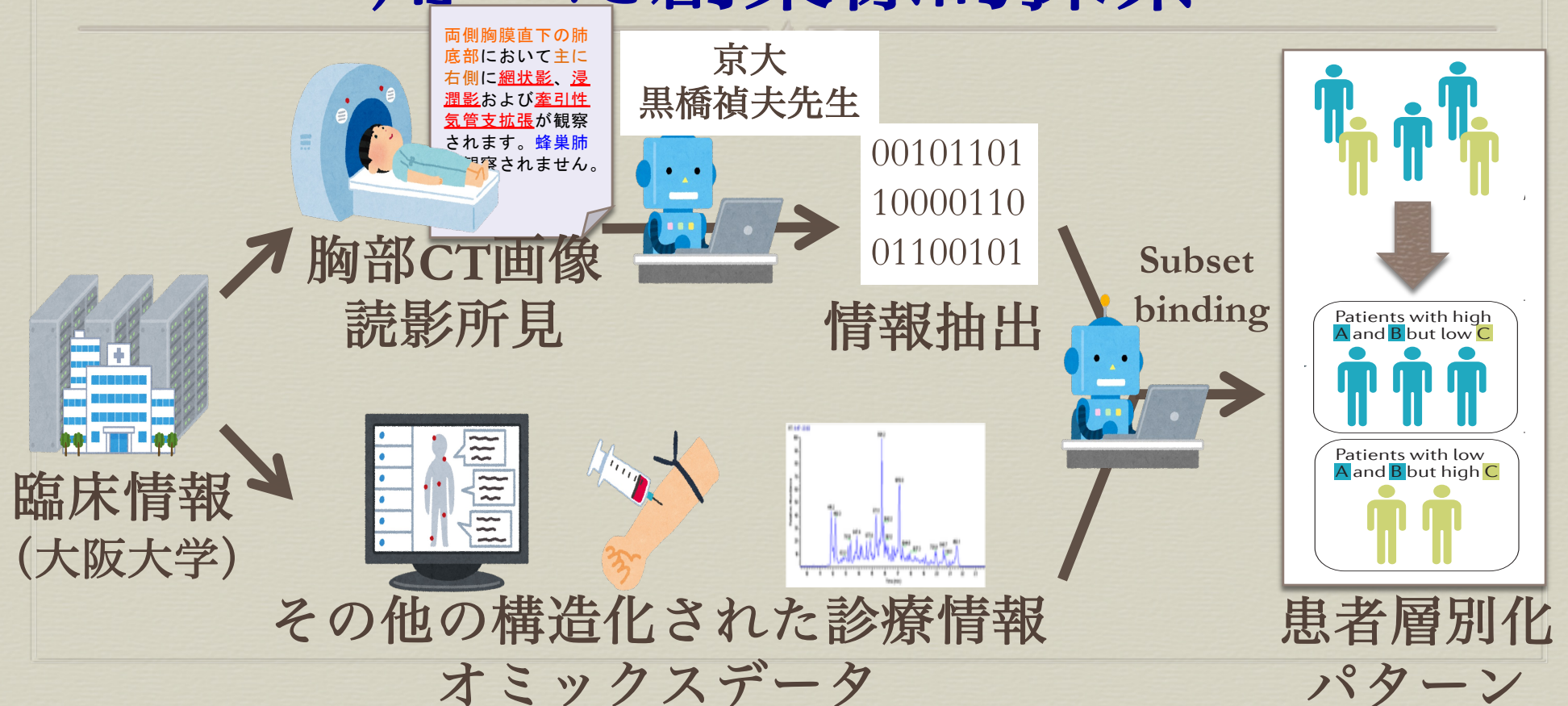
①臨床情報の収集とデータベースの構築

②臨床情報の利活用を可能とするAIの開発

③臨床情報からの創薬標的提示

④オープンプラットフォームによる事業成果公開

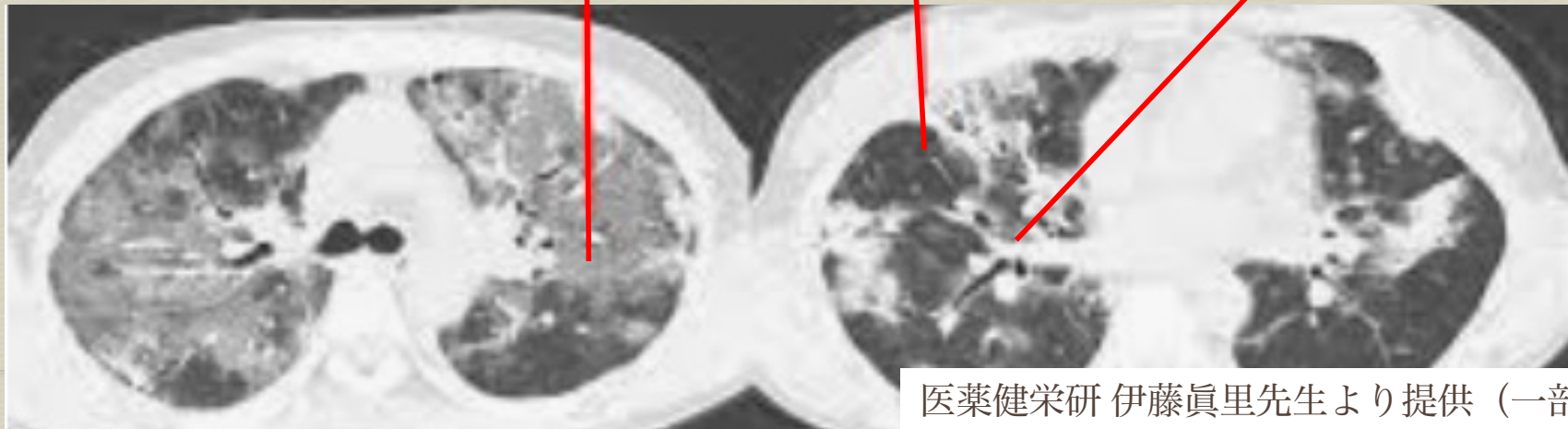
大阪大学コホート 臨床情報を用いた創薬標的探索



解析に用いた診療情報

両側胸膜直下の肺底部において主に右側に網状影、浸潤影および牽引性気管支拡張が観察されます。蜂巢肺は観察されません。Possible UIPパターンを疑います。

斑点状のすりガラス影が右肺に点在しています。一連の間質性肺炎に関連する変化に矛盾はありません。間質性肺炎の疑いがあります。



医薬健栄研 伊藤真里先生より提供（一部改変）

両側胸膜直下の肺底部において主に右側に網状影、浸潤影および牽引性気管支拡張が観察されます。蜂巢肺は観察されません。Possible UIPパターンを疑います。

斑点状のすりガラス影が右肺に点在しています。一連の間質性肺炎に関連する変化に矛盾はありません。間質性肺炎の疑いがあります。

右肺
両側胸膜直下

網状影
浸潤影
牽引性気管支拡張

Positive (あり)

Negative (なし)

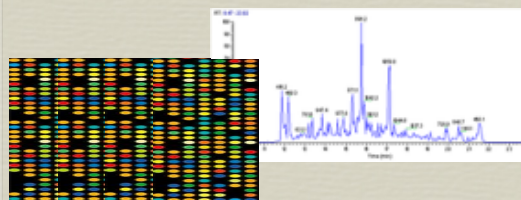
Suspect (疑わしい)

京都大学 黒橋禎夫先生

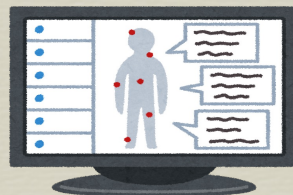
医薬健栄研 伊藤真里先生より提供 (一部改変)

患者層別化AIのワークフロー

オミックス



診療情報



入力

二種類の行列
(合計602サンプル)



出力

患者層別化
の条件

分子A :high

分子B :high

分子C :low

診療項目x :low
(または0)

診療項目y :high
(または1)

実データでの出力例

タンパク質 A **high**
タンパク質 B **high**
タンパク質 C **high**

これらのタンパク質量が
高い傾向の患者さんは

[胸水:n:]
[蜂巣肺:p:下葉|両側肺]
[牽引性気管支拡張:p:下葉|両側肺]
[網状影:p:下葉|両側肺]

IPF!

CT画像の読影所見に
これらのキーワードが
含まれる傾向がある

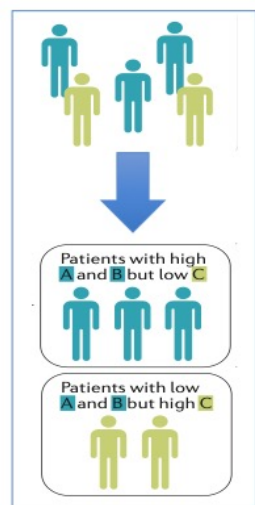
患者層別化に有用なタンパク質群の検出

患者層別化AIが提示した タンパク質の特徴

- これらのタンパク質は線維化との関連が示唆される一方、IPFとの関連は報告なし。
- KOマウスのフェノタイプに肺の炎症、間質性肺炎があるタンパク質が含まれる。
- これらのタンパク質は全て、あるパスウェイに関連。
- 抗線維化薬ニンテダニブの標的分子など既知の線維化メカニズムとの関連性も検出。

Subset Bindingにより提案された分子の解釈と検証

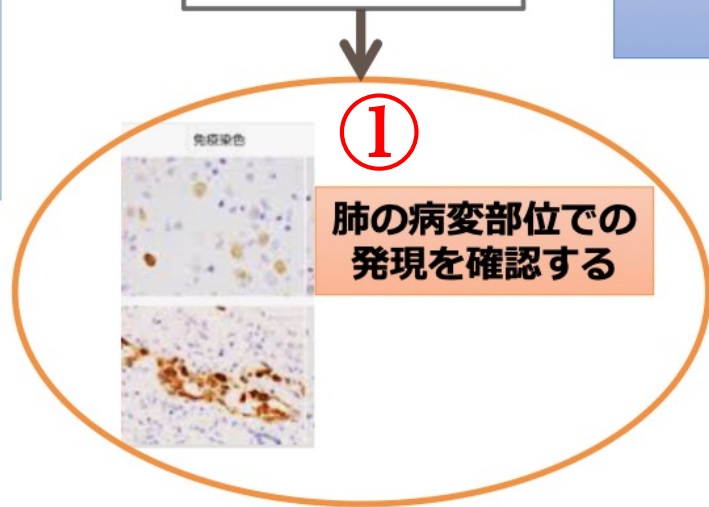
20種類のIPF関連タンパク質
(7種類のコアタンパク質)を検出



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



②

分子の生物学的意義
パスウェイ解析など
知識処理

知識自動抽出

知識DB

創薬標的探索支援
データウェアハウス

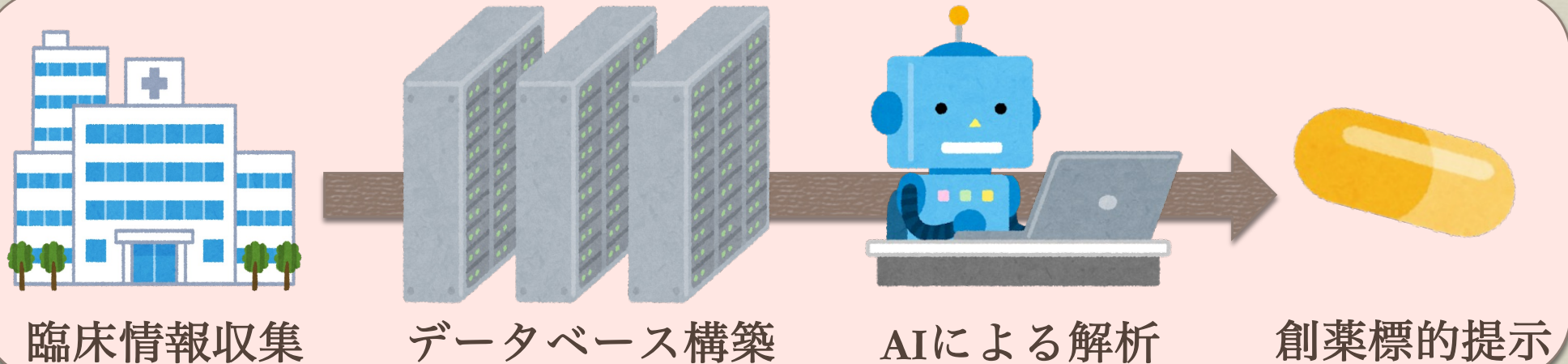


③

代謝物質の変動
ヒト細胞試験系での確認

今後の展望

このスキームはあらゆる疾患に対して適用可能



本事業により、臨床情報から創薬標的探索が可能であることが示された
→ 対象疾患を更に拡大、先制医療・個別化医療の実践へ

ABCI UG活動に期待する事

- ・ ABCIは使いやすい！

- 工学、情報科学などが専門でない研究者・技術者の利用も増えるかも？

- ありがちな問題を解決する手段をユーザー間で共有できると嬉しい

謝辞

理化学研究所

革新知能統合研究センター

上田修功

大阪大学大学院医学系研究科

熊ノ郷淳

武田吉人

京都大学大学院情報学研究科

知能情報学専攻

黒橋禎夫

大阪大学医学部附属病院

医療情報部

松村泰志

武田理宏

医薬基盤・健康・栄養研究所

プロテオームリサーチプロジェクト

足立淳

朝長毅

奈良先端科学技術大学院大学

先端科学技術研究科

荒牧英治

謝辞

医薬基盤・健康・栄養研究所 AI健康・医薬研究センター

水口賢司
伊藤真里
黒田正孝
Yi-An Chen
樋口千洋
深川明子
鎌田英世

進藤順紀
中村恵宣
景山浩充
川井享代
森田隆嗣

長尾知生子
藤原大
野島陽水
山田美樹