

# 創薬ターゲット探索プラットフォームの構築

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

官民研究開発投資拡大プログラム (PRISM)

「AI技術領域」

令和2年度成果

令和3年3月

厚生労働省・文部科学省

# 資料1 「創薬ターゲット探索プラットフォームの構築」の概要

アドオン額(令和2年度):683,250千円(厚生労働省・文部科学省)

元施策・有/PRISM事業・継続予定

## 課題と目標

- (課題) 動物実験頼りで創薬ターゲットを選定する従来の研究スキームが、医薬品開発の最大のボトルネックとなっている第2相試験 (Phase 2) での成功率低迷の要因となっているため、製薬産業界では、患者情報とAIを活用して患者を層別化し、データ駆動的に適切な創薬ターゲットを探索する新たなプラットフォームの開発が新薬創出のゲームチェンジャーとして切望されている。
- (目標) 特発性肺線維症 (IPF) 及び肺がんを対象に、診療情報と網羅的生体分子情報から患者を層別化し、創薬ターゲットをデータ駆動的に探索するプラットフォームの構築に必要な多岐にわたるAI群を開発する。さらに、IPF及び肺がんの新たな創薬ターゲットをそれぞれ1個ずつ特定し、企業導出等により産業界での実創薬に展開する。

## 「創薬ターゲット探索プラットフォームの構築」の概要

- 元施策：  
厚生労働省（医薬基盤・健康・栄養研究所）では、平成29年度より、「新薬創出を加速する人工知能の開発」事業において、IPFを対象に、診療情報と網羅的生体分子情報を収納した疾患統合データベースの構築と新たな創薬ターゲット探索手法の開発を開始した。
- PRISMで実施する理由：
  - ・ 日本が強みを持つ創薬領域での研究開発を強化するプロジェクトであり、創薬ターゲット探索に焦点をあてたAI開発は世界に先駆けた取り組みである。
  - ・ 本施策により、医薬品開発コストの大幅な削減と個別化医療の進展がもたらされ、薬剤費及び医療費の低減が期待される。
  - ・ 有効な治療薬がない肺がん、IPFの新たな創薬ターゲットの特定により、患者や医療従事者の負担低減をもたらす革新的新薬の創出が期待される。
  - ・ 事業成果のオープンプラットフォーム化や導出等による企業と公的機関（公立病院、大学、国研等）とのAI開発等での連携強化を目標としている。
- テーマの全体像：  
IPFと肺がんを対象に、日本トップクラスの医療機関との連携のもと、質の高い診療情報と網羅的生体分子情報をそれぞれ1,500症例以上収集し、世界最大規模の疾患統合データベースを構築する。さらに、これらのデータや科学論文、既存データベース等を活用して、データ駆動的に患者を層別化し、創薬ターゲットを探索する世界最先端のAI群と知識ベースを開発する。医薬健栄研、理研、JST、大学・国研など15機関が参画。

## 出口戦略

- ・ 事業成果を広く健康・医療領域の研究開発に展開するため、疾患データベース、アルゴリズム、研究スキーム等をオープンプラットフォーム化する。オープンプラットフォーム化にあたっては、関係各所との調整のうえ、商用利用の有料化等の制度も含め、持続的かつ自立化が可能な事業運営を目指す。
- ・ オープンプラットフォーム化になじまない創薬ターゲット及び本事業の実施過程で派生的に創出された技術・成果物については民間企業への導出を図る。

## 民間研究開発投資誘発効果等

- 民間投資誘発効果：令和元年度 画像解析AIの導出（経済効果50億円：導入企業による試算）
- 民間からの貢献額：平成30年度から令和2年度の3年間で14億円相当。
  - ・ 平成30年度（内訳）共同研究等29社：共同研究費（3億円）、人件費（2.0億円）、資機材（約0.2億円相当）。
  - ・ 令和元年度（内訳）共同研究等32社（製薬企業20社を含む）：共同研究費（1.7億円）、人件費（2.4億円）、資機材（約0.2億円相当）
  - ・ 令和2年度（内訳）共同研究等44社（製薬企業24社を含む）：共同研究費（1.0億円）、人件費（3.4億円）、資機材（約0.1億円相当）

# 資料2 「創薬ターゲット探索プラットフォームの構築」の概要

## テーマの全体像・説明

本テーマでは、新薬開発の最大のボトルネックであるPhase 2での成功確率を上げるため、AIや大規模知識ベース等の最先端情報科学を実装した革新的な“データ駆動型の創薬標的探索プラットフォーム”を構築する。

## PRISMによる加速・拡充

- 厚労省、文科省が密接な連携体制を構築し、
- 理研AIPセンター及びJST-AIPネットワークラボでのAI等基盤技術並びに産総研での知識抽出技術及び知識発見を行うためのバイオインフォマティクス技術などを利用することにより研究開発を加速する。
- 対象疾患として難病であるIPFに加えて、がんの死因第1位の肺がんも対象として拡充する。

## 「創薬ターゲット探索プラットフォームの構築」の概要

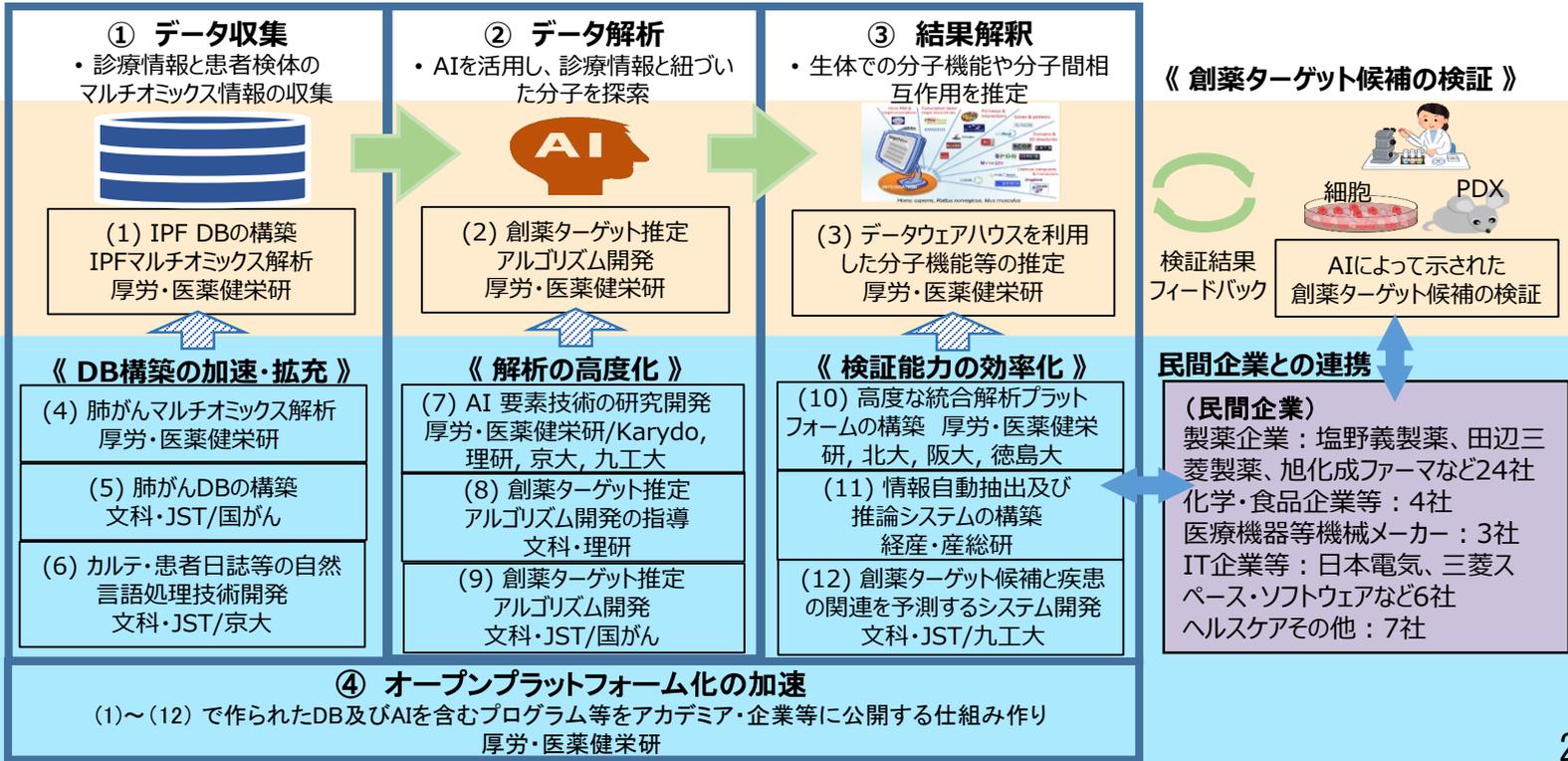
本施策では、① 診療情報及びマルチオミクス情報を収集し、それらを構造化した疾患統合データベースを構築する。② 疾患統合データベース及びAIを活用して患者を層別化し、診療情報と密接に関連した生体分子を探索する。③ AIにより探索された生体分子を既存知識を利用して、その分子機能や分子間相互作用を推定することで、創薬ターゲット候補を決める。この技術を民間企業との連携を通じて社会実装させることで民間研究開発投資を促進する。

### 厚労省 元施策

「新薬創出を加速する人工知能の開発」  
《令和2年度予算》  
300百万円

### PRISMによる加速・拡充

「創薬ターゲット探索プラットフォームの構築」  
《令和2年度予算》  
厚労省：513百万円  
文科省：170百万円



# 資料3 「創薬ターゲット探索プラットフォームの構築」の目標達成状況

○施策全体の目標

データ駆動的に患者を層別化し、創薬ターゲットを探索するAI群の開発及びIPFと肺がんの新たな創薬ターゲットをそれぞれ1個ずつ特定し、共同研究や企業導出等により産業界での実創薬に展開する。

事業名等 (※個別に目標を設定している場合)	当年度目標	目標の達成状況
<p>① (データ収集) 特発性肺線維症(IPF)及び肺がん患者1500例の診療情報及びマルチオミクス測定データを収集し、世界最大規模の疾患統合データベースを構築する。</p>	<p>① (データ収集)</p> <ul style="list-style-type: none"> <li>IPFデータベース構築：300症例の追加</li> <li>肺がんデータベース構築：オミクスデータ欠損値補完</li> </ul>	<ul style="list-style-type: none"> <li>IPF：コロナ感染拡大の影響により当初目標とした肺組織取得症例数を300から150に下方修正した。状況の好転が見込めないため、収集対象をコロナ禍の影響を受けない血液サンプル取得症例に変更し、391症例を収集。累積数901症例。</li> <li>肺がん：目標症例1,500には昨年度到達済みのため、今年度はpan-negative肺がんについて欠損値を補充した（全ゲノム、RNA-seq、エピゲノム）。</li> </ul>
<p>② (データ解析) ③ (結果解釈) 創薬ターゲットの探索に必要なAI及びデータウェアハウス/知識ベースを開発し、それらの性能を実証する。</p>	<p>② (データ解析) ③ (結果解釈)</p> <ul style="list-style-type: none"> <li>13種類のIPF・肺がんのターゲット探索用AIの開発及び性能実証</li> </ul>	<ul style="list-style-type: none"> <li>解析研究に参加する7機関が、IPF及び肺がんの詳細なデータセットを共有し、各機関が重複することなくAI開発に関わる事業目標達成に相補的に貢献する計画である。肺がん及びIPFデータの共有完了。</li> <li>肺がん、IPFのオミクスデータからデータ駆動的に患者を層別化し、創薬ターゲットを探索する5種のAI開発に成功。複数個の肺がん及びIPF創薬ターゲット候補を捕捉し、検証研究を開始。2種のAI開発継続中。</li> <li>論文情報や公共データベースを利用し、オミクスデータから創薬ターゲットを予測する5種のAI開発を完了。</li> <li>読影所見等の臨床/患者テキストからの医療情報抽出システムのプロトタイプ開発を完了。</li> </ul>
<p>④ (オープンプラットフォーム) 事業成果を広く健康・医療領域の研究開発に展開するため、疾患データベース、アルゴリズム、開発スキーム等をオープンプラットフォーム化する。</p>	<p>④ (オープンプラットフォーム)</p> <ul style="list-style-type: none"> <li>先行事例等の調査</li> <li>制度設計</li> </ul>	<ul style="list-style-type: none"> <li>国内外での先行事例及び企業25社の参加意欲等の一次ヒアリングを完了。調査結果をもとにオープンプラットフォームの国内運用の基本構想決定。</li> <li>要配慮個人情報に関する企業ユーザーの懸念に対応するため、臨床情報各項目の個人特定リスク評価と匿名加工方法の検討を完了。</li> <li>ソフトウェアやプログラム言語を問わずAIやデータベースを相互連結させる成果共用基盤システム（GARUDAベース）の設計とプロトタイプ構築を完了。患者層別化AIを基盤上に搭載し、性能確認を完了。</li> <li>オープンプラットフォーム計画の認知度アップのため一流誌への論文発表や情報発信を推進。</li> </ul>

## 資料4 「創薬ターゲット探索プラットフォームの構築」の成果

### ○ データ収集

- IPFは、391症例（目標300）の診療情報とオミックス情報（全ゲノム・DNAメチル化解析・RNA-seq）を追加。累計901症例
- 肺がん症例数は昨年度に世界最大規模（1596症例）を達成済み。今年度は欠損値補填として全ゲノム解析（50症例）、エピゲノム（メチル化、ChIP-seq）解析（250症例）の追加解析を完了した。

### ○ データ解析

- 高次元・小標本データに適用可能なIPF患者層別化AIを完成し、IPFの特徴と紐づけられたタンパク質複数個をデータ駆動的に検出。プロテオームデータと紐づいた診療情報をAIで解析し、データ駆動的に層別化とターゲット探索を行うことができることを示した。
- このことは難病のように既存知識が蓄積されていなくても診療情報と紐づいたオミックスデータを収集すれば患者層別化が可能であることを意味する重要な成果。複数の異なる専門性を有した研究者が連携することで成功（阪大、基盤研、理研、京大情/JST、産総研）。層別化AIは創薬ターゲット及び層別化バイオマーカーの探索のゲームチェンジャーとして期待できることから特許を出願し、一流国際誌に投稿中。
- 肺がん解析プラットフォームの構築をほぼ終了。複数の機械学習・深層学習技術を用いた4種のAIを開発し、マルチオミックスデータから肺がんの新規創薬ターゲット/層別化マーカーとなる遺伝子複数個の特定に成功。ターゲット/バイオマーカー候補の検証研究を開始国がん/JSTを中心に、京大情/JST、基盤研が連携。肺がんオミックスデータを解析するAIについて、成果をまとめた論文を国際誌に投稿した。
- 既存データベースや論文情報を利用し、機械学習や深層学習等の複数の手法を用いて、マルチオミックスデータから創薬ターゲットを予測する新手法5個の開発を完了し、複数の創薬ターゲットを特定（九工大/JST）。新たに2個の新手法を開発中（京大医、Karydo TherapeutiX）。
- 読影所見等の臨床/患者テキストからの医療情報抽出システム：肺がんでは9割以上の高い認識精度を得た。システムのプロトタイプを完成させた。

### ○ 結果解釈

- 基盤研が保有するデータウェアハウス（TargetMine）に3種の公共・市販データベースを追加統合し、計画していた7種のDB追加による機能拡張を完了。IPFの特徴と紐づけられた複数個のタンパク質から新たな創薬標的の絞り込みに活用中。
- 英語文献からの情報自動抽出技術を実装したウェブアプリケーションのプロトタイプ開発を完了。
- 情報自動抽出システムは計画を達成したが、精度向上の改良を継続。抽出した情報から疾患ネットワーク構築し、疾患パスウェイを推論する技術開発を継続。
- オミックス解析結果の解釈精緻化を目的に、各種臓器及び血中エクソソームの糖鎖解析及びIPF肺組織シングルセルオミックス解析を実施中。

### ○ オープンプラットフォーム

- オープンプラットフォームの想定ユーザー（25社）の参加意欲に加え、国内外で先行して類似プラットフォームを運営している機関などの運営状況に関する先行事例調査を完了した。調査結果をもとにオープンプラットフォームの国内運用の基本構想を決定。
- 上記調査の過程で明らかになった要配慮個人情報に関する企業ユーザーの懸念に対応するため、OPFに搭載する臨床情報各項目の個人特定リスク評価と匿名加工方法の検討を完了。
- 生命科学領域でデータベースやAI等の協調利用等のグローバル運用の実績が豊富なSystem Biology Institute（SBI）との連携により、ソフトウェアやプログラム言語を問わずAIやデータベースを相互連結させることができるGARUDAプラットフォーム（SBI）をベースとした成果共用基盤システム（ハードウェア・ソフトウェア）の設計およびプロトタイプ構築を完了。本事業成果の一つである患者層別化AIを基盤上に搭載し、システムの性能確認を完了。
- オープンプラットフォーム計画の認知度アップのため事業成果の一流国際誌への投稿やマスコミに向けた情報発信（令和元年度成果報告会、学会セッション（IIBMP2020）、論文5報、学会発表27件、マスコミ対応4件）

# 資料5 「創薬ターゲット探索プラットフォームの構築」の民間からの貢献及び出口の実績

○民間からの貢献額：平成30年度から令和2年度までの3年間で14億円相当

- ① 平成30年度（内訳）共同研究等29社：共同研究費（3億円）、人件費（2.0億円）、資機材（約0.2億円相当）
- ② 令和元年度（内訳）共同研究等32社（製薬企業20社を含む）：共同研究費（1.7億円）、人件費（2.4億円）、資機材（約0.2億円相当）
- ③ 令和2年度（内訳）共同研究等44社（製薬企業24社を含む）：共同研究費（1.0億円）、人件費（3.4億円）、資機材（約0.1億円相当）

## 令和2年度当初見込み

## 令和2年度実績

新薬創出を加速する症例データベースの構築・拡充（厚労省）

- ① 共同研究 人件費（4.5人年（約15百万円） / 21社）  
資機材（計算機、ソフトウェア等）（約10百万円相当/2社）
- ② 共同研究 1件 人件費（約45百万円）
- ③ 共同研究 遺伝子発現解析等 2件 研究費（6百万円）
- ④ 共同研究 人件費（1.9人年（19百万円））

- ① 共同研究 人件費（2.8人年（約20百万円） / 22社）  
資機材（計算機、ソフトウェア等）（約10百万円相当/2社）
- ② 共同研究 1件 人件費（約60百万円）
- ③ 共同研究 知識ネットワーク解析1件、遺伝子発現解析等 3件 研究費（12.8百万円）、人件費（2人年（約19百万円））
- ④ 共同研究 3件 研究費（27百万円）、人件費（5人年（36百万円））

創薬ターゲット推定アルゴリズムの開発（文科省）

- ・ 共同研究等9件（製薬企業、IT企業等）  
言語処理（読影所見の自動生成及び学術論文からの副作用情報の自動抽出）、肺がん創薬因子を同定する新手法開発、自動アノテーションツールの開発、化合物の標的分子、薬効予測 等
- 人件費 8人年（約80,000千円）/3件  
共同研究費 6件（約150,000千円）

- ① 自然言語処理に関する共同研究 10件 研究費（17百万円）、人件費（58百万円）
- ② 創薬関連の共同研究 1件 研究費（1百万円）、人件費（10百万円）
- ③ インフォマティクスに関する共同研究 6件 研究費（27百万円）、人件費（44百万円）
- ④ がん関係の共同研究費 4件 研究費（10百万円）、人件費（90百万円）

○出口戦略

- ・ 優れた創薬ターゲットを提供することで、民間製薬企業の医薬品開発を喚起し、研究開発投資誘発効果が期待できる。（承認薬1剤当たりの研究開発費約3,000億円）
- ・ 医療・ヘルスケア領域全般に展開できる成果を提供することで、民間企業の研究開発投資誘発効果が期待できる  
例）医療情報システム市場：電子カルテや医用画像関連システム等の医療情報システム全体の国内市場規模は約6,000億円、電子カルテだけで約4,000億円と言われており（JAHISの調査研究）、本事業の成果導出により研究開発投資効果が期待できる。

## 令和2年度当初見込み

## 令和2年度実績

- ・ 民間企業からの共同研究等に関する相談・調整等：10社  
（創薬ターゲット探索、薬効予測、読影所見の自動生成及び副作用情報の自動抽出等の自然言語処理技術など）

- ・ 民間企業からの共同研究等に関する相談・調整等：10社  
（創薬ターゲット探索、薬効予測、読影所見の自動生成及び副作用情報の自動抽出等の自然言語処理技術など）