

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

令和元年度成果

令和2年7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：911,089千円(厚生労働省・文部科学省)
元施策・有/PRISM事業・継続予定

課題・目標

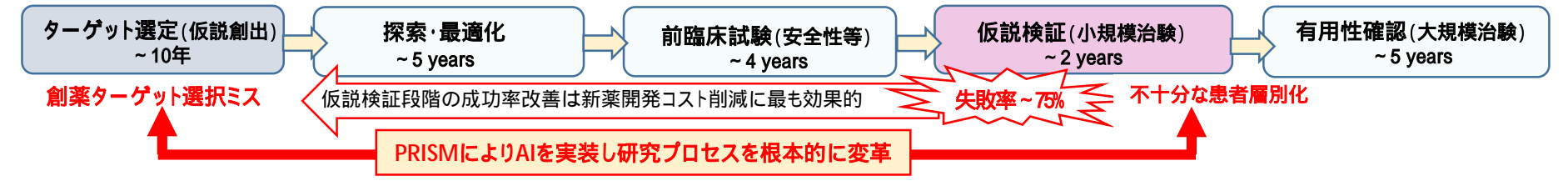
- 課題：創薬ターゲット／患者層別化バイオマーカーのAI探索に必要な要素技術の開発及び性能実証に利用する正確な患者情報に紐付された網羅的生体分子情報（マルチオミックス情報）が十分に蓄積されておらず、アルゴリズム等の要素技術開発に必要なノウハウ・リソースも不足している。
- 目標：① 特発性肺線維症（IPF）及び肺がん患者1500例の臨床情報及びマルチオミックス解析データを収集し、疾患統合データベースを構築する。
創薬ターゲット／患者層別化バイオマーカーの探索に必要な14種類のAI及び知識ベースを開発し、それらの性能を実証する。
IPFと肺がんの新たな創薬ターゲットをそれぞれ1個ずつ特定し、共同研究や企業導出等により産業界での実創薬に展開する。
事業成果を広く健康・医療領域の研究開発に展開するために、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。

「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、「創薬ターゲットの枯渇問題」を克服すべく、「新薬創出を加速する人工知能の開発」事業（運営費交付金）において、AIを活用した新たな創薬ターゲット探索手法の確立を目的に、正確な診療情報に紐づけされた網羅的生体分子情報を収納した疾患統合データベース（DB）の構築及びアルゴリズムの開発を特発性肺線維症（IPF）を対象疾患として開始した。（令和元年度：286百万円）

PRISMで実施する理由：
 ・優れた創薬ターゲットを提供することで、医薬品開発における民間製薬企業の研究開発投資誘発効果が期待できる。
 ・本施策により創出された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。
 ・優れた創薬ターゲットは医薬品開発の成功確率を高めることで医薬品開発経費を効率化し、それを通じて薬剤費の低減（財政支出の効率化）に資することが期待できる。

テーマの全体像：
近年、新薬開発の対象が発生機序が解明されていない難病及び希少疾患にシフトしているため、従来の研究手法では患者で行う仮説検証試験（Phase 2）で有効性を示す新たな創薬ターゲットの特定が極めて困難となっている。「創薬ターゲットの枯渇問題」を解消し、新薬創出を大幅に加速する「ゲームチェンジャー」として、人工知能（AI）技術を活用した臨床データ解析により創薬ターゲット及びプレジジョンメディスンを実現しうる患者層別化バイオマーカーを同定する新しいアプローチの開発に大きな期待が寄せられている。



出口戦略

- ・本事業で創出される「創薬ターゲット探索のための基盤」について商用利用では有料とする仕組みを構築し、もって基盤の維持を図り医薬品の研究開発投資を喚起する。
- ・本事業の実施過程で派生的に作られる技術・成果物を医療システム等に応用するために民間企業への導出を図る。
- ・本事業により創出される解析アルゴリズムは様々なヘルスケア関連のビッグデータ解析に広く活用可能であることから、LINC、GARUDA等の先行プラットフォームとの相互連携等を図り、広く産業界による商用二次利用（有料）を促進することによって自立可能な事業運営を目指す。

民間研究開発投資誘発効果等

優れた創薬ターゲットを提供することで、医薬品開発における民間製薬企業の研究開発投資（承認薬1剤当たりの研究開発費 約3,000億円）が期待でき、本施策により創出された派生的成果物を医療・ヘルスケア領域に提供することで、医療情報システム全体で約6,000億円と言われている（JAHISの調査研究）国内市場での民間企業の研究開発投資誘発効果が期待できる。令和元年度に企業導出したAI搭載ITプラットフォームの経済効果は、worldwideで約50億円という試算が企業側から出ている。民間からの貢献額：令和元年度実績として民間企業32社（製薬企業：20社、化学・食品企業等：7社、IT企業等：5社）と総額4.3億円相当の共同研究を行った。（内訳）：人件費（約2.4億円）、資機材（約0.2億円相当）、共同研究費（約1.7億円）

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、(データ収集)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DBを構築する。(データ解析)疾患統合DB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。(結果解釈)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。さらに、(オープンプラットフォーム化の加速)事業で得られた成果について、アカデミア・企業等に公開する仕組み作りを行う。

・PRISMによる加速・拡充

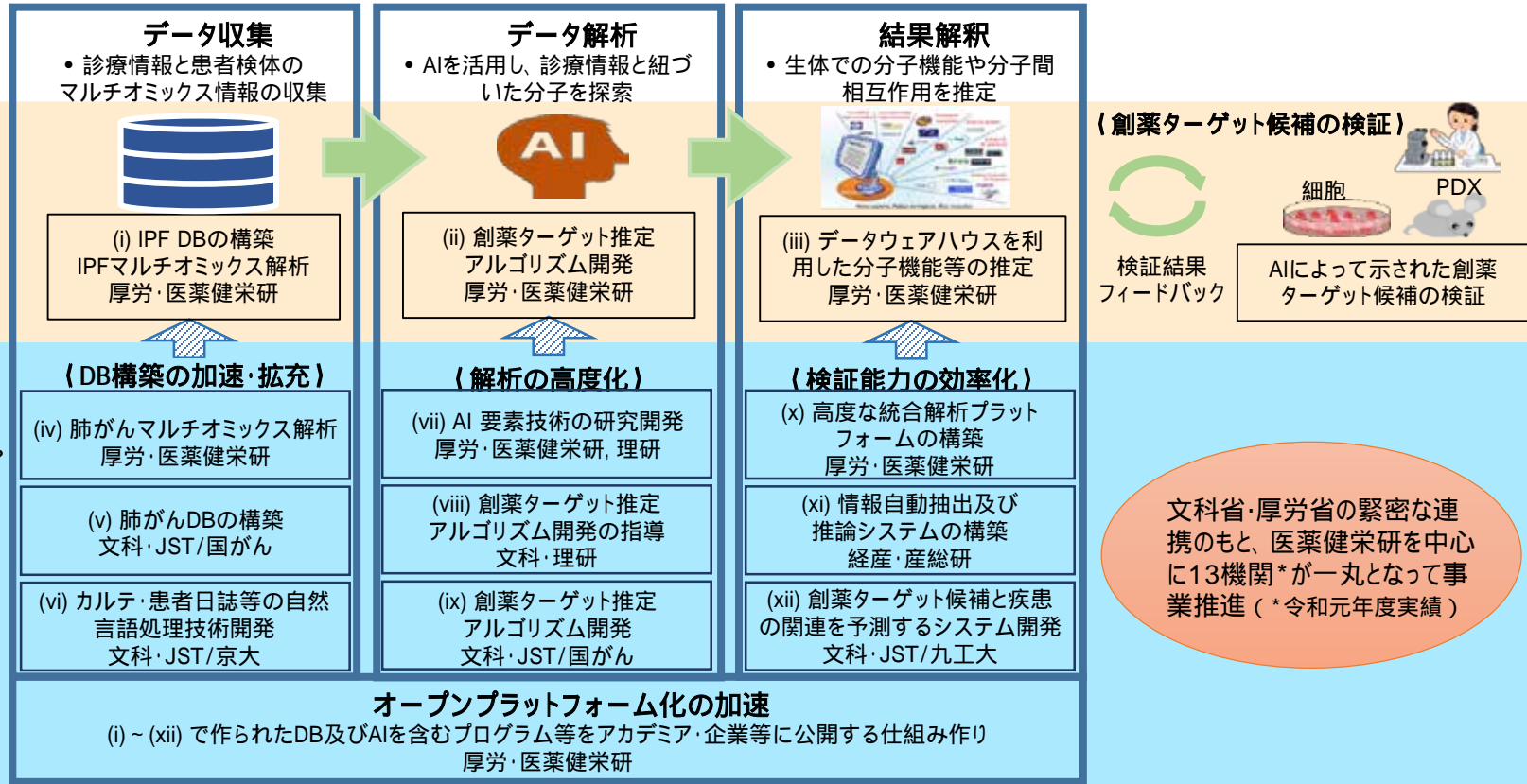
厚労省、文科省が密接な連携体制を構築し、理研AIPセンターでのAI等基盤技術、JST-AIPネットワークラボでの AI技術を用いたマルチオミクス解析基盤技術、バイオマーカー予測法、知識に基づく頑健で高精度な構造的言語処理技術、既承認薬の新しい適応可能疾患を予測する機械学習アルゴリズム、並びに産総研での知識抽出技術及び知識発見を行うためのバイオインフォマティクス技術などの研究成果を、アドオンとして創薬ターゲットの探索に向けたAI開発に発展・展開することで、社会実装や民間企業(製薬企業、医療情報システム会社等)への技術移転等の加速が期待できる。

厚労省 元施策

「新薬創出を加速する人工知能の開発」
(令和元年度予算額) 286百万円

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」
(令和元年度予算額) 厚労省:695百万円
文科省:216百万円



文科省・厚労省の緊密な連携のもと、医薬健栄研を中心に13機関*がー丸となって事業推進(*令和元年度実績)

出口戦略

本事業の成果である創薬ターゲットを企業導出するとともに、アルゴリズムや疾患データベース等の派生的成果物の利活用をはかる。

民間研究開発 投資誘発

創薬ターゲットの提供及びそれを見出す方法論を提供することで承認薬1剤当たり約3,000億円と言われる研究開発投資を誘発する。また、医療・ヘルスケア領域全般に展開できる成果を民間企業に提供し、医療情報システム市場(国内約6千億円)での研究開発投資を誘発する。

施策全体目標

- (データ収集) 特発性肺線維症 (IPF) 及び肺がん患者1500例の臨床情報及びマルチオミックス解析データを収集し、疾患統合データベースを構築する。
- (データ解析) 創薬ターゲット/患者層別化バイオマーカーの探索に必要な14種類のAI及び知識ベースを開発し、それらの性能を実証する。
- (結果解釈) 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。
- (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に、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。

令和元年度目標

- (データ収集) データ収集のさらなる加速化。IPF：200症例、肺がん：450症例の追加。(一部元施策で実施)
- (データ解析) AI解析技術開発体制の強化。
- (結果解釈) 統合プラットフォームの構築、疾患との関連予測システム、情報自動抽出及び推論技術開発のさらなる加速化。
- (オープンプラットフォーム) オープンプラットフォーム化の準備作業の開始。

事業名等	令和元年度目標	目標の達成状況
データ収集	<ul style="list-style-type: none"> • 診療情報が紐づいたオミックス情報の症例数 IPF : 200症例の追加。 肺がん : 450症例の追加。 	診療情報が紐づいたオミックス情報の症例数 ○ IPF : 216 症例の追加 (累計 516 症例) ○ 肺がん : 521 症例の追加 (累計 1,569 症例)
データ解析	<ul style="list-style-type: none"> • AI解析技術開発体制の強化。 	○ AI解析技術開発体制を強化するため、令和元年度の理研、国がん及び医薬健康栄研の 3機関体制 から令和2年度には理研での 参加研究者の増員 に加え、京大(医)、Karydo、九工大を加え 6機関体制 とした。
結果解釈	<ul style="list-style-type: none"> • 統合プラットフォームの構築 • 疾患との関連予測システム • 情報自動抽出及び推論技術開発のさらなる加速化 	AIによって疾患との関連が推測された分子から創薬ターゲット候補を見出すため、 ○ 7種類の公共・市販のデータベースをTargetMineに追加統合 した。 ○ 情報自動抽出システムについては、性能評価実験において良好な結果 が得られた。検索・推論基盤の構築に向けた準備を開始。 ○ 創薬ターゲット候補分子と疾患との関連を予測する手法を検証。 ○ オミックス解析の結果をより良く解釈するために、血中エクソソームの糖鎖解析の実施と肺のシングルセルオミックス解析の準備を行った。
オープンプラットフォーム	<ul style="list-style-type: none"> • オープンプラットフォーム化の準備作業の開始 	○ DBの制限公開及び成果物協調利用の義務化に関わる 包括協定 を本施策参加全機関と締結した。 ○ 企業・アカデミアの 利用意向調査 を実施した。

データ収集・・・AI解析のためのデータ収集

- IPF及び肺がんともに**目標を上回る診療情報が紐づいたオミックス情報を収集**した（IPF：目標200 実績216症例、肺がん：目標450 実績521症例）肺がんの症例数は、**同種のデータベースの中では世界最大規模**となった。
- 読影所見や診療記録等の文章3,000件以上を用いて医療表現アノテーション・コーパスと医療用語オントロジーを構築。
読影所見中の医療表現と、病名・症状の確実性を自動認識するシステムを構築し、**9割以上の高い認識精度を得た**。

オミックス項目	肺がん（累計数）		IPF（累計数） *一部元施策で実施	
	平成30年度	令和元年度	平成30年度	令和元年度
診療情報	1048例	1569例	300例	885例
全エクソーム解析	1048例	1569例	－	－
全ゲノム解析	30例	272例	－	56例
DNAメチル化アレイ	270例	430例	－	56例
ChIP-seq	－	150例	－	－
RNA-seq	840例	1494例	－	56例
プロテオーム	－	－	300例	602例

データ解析・・・疾患関連分子のAIによる探索

- AI解析技術開発体制を強化するため、令和元年度の理研、国がん及び医薬健康栄研の3機関体制から、令和2年度には理研での**参加研究者の増員**に加え、京大（医）、Karydo及び九工大を加えて**6機関体制**とした。
- 6機関で合計**8種の異なる方法**を用いたデータ解析手法を構築中。医薬健康栄研においては、**実データを用いてアルゴリズムの検証**を実施した。

結果解釈・・・疾患関連分子を手掛かりに創薬ターゲット候補を探索

- Data warehouseであるTargetMineに**7種類の公共・市販のデータベースを追加統合**した。
- 英語文献からの情報自動抽出システムについては、高精度に候補を挙げられるように手法を改良し、**性能評価実験において良好な結果が得られた**。また、検索・推論基盤の構築については、産総研で開発されたデータベースPoSSuMを例として基盤構築に向けた準備を開始した。
- 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情報に基づいて、**IPFや肺がんの創薬ターゲット候補の可能性を予測する手法の検証を行った（投稿準備中）**。
- オミックス解析の結果をより良く解釈するために、血中エクソソームの糖鎖解析の実施と肺がんのシングルセルオミックス解析の準備を行った。

オープンプラットフォーム

- データシェアリングを約した連携協定の下で、**他機関のAI研究者とも診療データ+オミックスの実データの利用が図られる**こととなった。
- **企業・アカデミアの利用意向についての一次調査を行った**。以下、概要。
 - 製薬・食品・ヘルスケア・診断分野・メディカルIT等の分野の民間企業を対象としたニーズ調査を実施した結果、総じて診療情報やオミックスデータを突合せたデータセットの**需要は高いにも関わらず、実際にそのようなデータセットを用いた研究を実施している例はごく一部**であった。その理由として、データセットの入手に関する障壁、法規制、情報管理・流出懸念、分析リソース（含むデータサイエンティスト）の不足、入手する情報の信憑性、費用対効果等が多数挙げられた。
 - 公的機関が構築したプラットフォームで公開される患者情報及びオミックスデータ等のデータベースや解析用AIを上記のようなコンセプトで公開提供した場合、企業が妥当とする取得・利用対価は数百万～2千万/年程度と幅広く捉えていた。

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年度及び令和元年度の2年間で9.5億円相当
 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：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）

令和元年度当初見込み	令和元年度実績
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）	
共同研究 人件費（4.5人年（約30百万円） / 21社） 資機材（計算機、ソフトウェア等）（約20百万円相当/2社） 共同研究 人件費（4人年（約60百万円） / 1社）	共同研究 人件費（4.8人年（約34百万円） / 22社） 資機材（計算機、ソフトウェア等）（約20百万円相当/2社） 共同研究 人件費（4人年（約60百万円） / 1社） 共同研究 遺伝子発現解析等 2件（6百万円） 共同研究 知識ネットワーク解析、データウェアハウス管理運営等 研究費（6.58百万円）、人件費（1.4人年（約14百万円））
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 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に関する共同研究（130百万円） 肺癌創薬因子を同定する新手法開発 企業研究者（2名20百万円） 自動アノテーションツールの開発 企業研究者（5名50百万円） 次世代ChIP-seq解析手法を用いた受託解析 企業研究者（1名10百万円） 共同研究（化合物の標的分子・薬効予測等）4件（18百万円）	共同研究 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に関する共同研究（130百万円） 肺癌創薬因子を同定する新手法開発 企業研究者（2名20百万円） 自動アノテーションツールの開発 企業研究者（5名50百万円） 次世代ChIP-seq解析手法を用いた受託解析 企業研究者（1名10百万円） 共同研究 化合物の標的分子・薬効予測等 5件（20百万円）、人件費（24百万円） 共同研究 ChIP-seqを用いたエピゲノムバイオマーカーの探索（5百万円） 共同研究 創薬ターゲット同定（6百万円）、人件費（約4百万円）

○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。

令和元年度当初見込み	令和元年度実績
<ul style="list-style-type: none"> 民間企業からの共同研究等に関する相談・調整等：10社 肺癌画像自動アノテーションシステムを企業導出 	<ul style="list-style-type: none"> 民間企業からの共同研究等に関する相談・調整等：12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） 肺癌画像自動アノテーションシステムを企業導出