

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）

「A I 技術領域」

令和3年度成果

令和4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:581百万円(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、患者の診療データと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）及び肺癌を対象に、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及び肺癌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。

事業の概要

- 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。（令和3年度291百万円）
- 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癌、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り企業と公的機関（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と肺癌を対象に、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1500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群とデータウェアハウスを開発する。医薬健栄研、理研、大学・国研など14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、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

- 民間からの貢献額：令和3年度実績として民間企業43社（製薬企業：17社、IT企業等：13社、化学・食品企業等：9社、その他：4社）と総額4.2億円相当の共同研究を行った。（内訳）：共同研究費（0.7億円）、人件費（3.5億円）
- 民間企業への導出（令和元年度）：AI搭載ITプラットフォーム（経済効果50億円；導入企業による試算）
（令和3年度）：医師がAI技術を開発できる研究基盤システム「AI開発支援プラットフォーム」の共同開発

資料2-1 「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ターゲット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

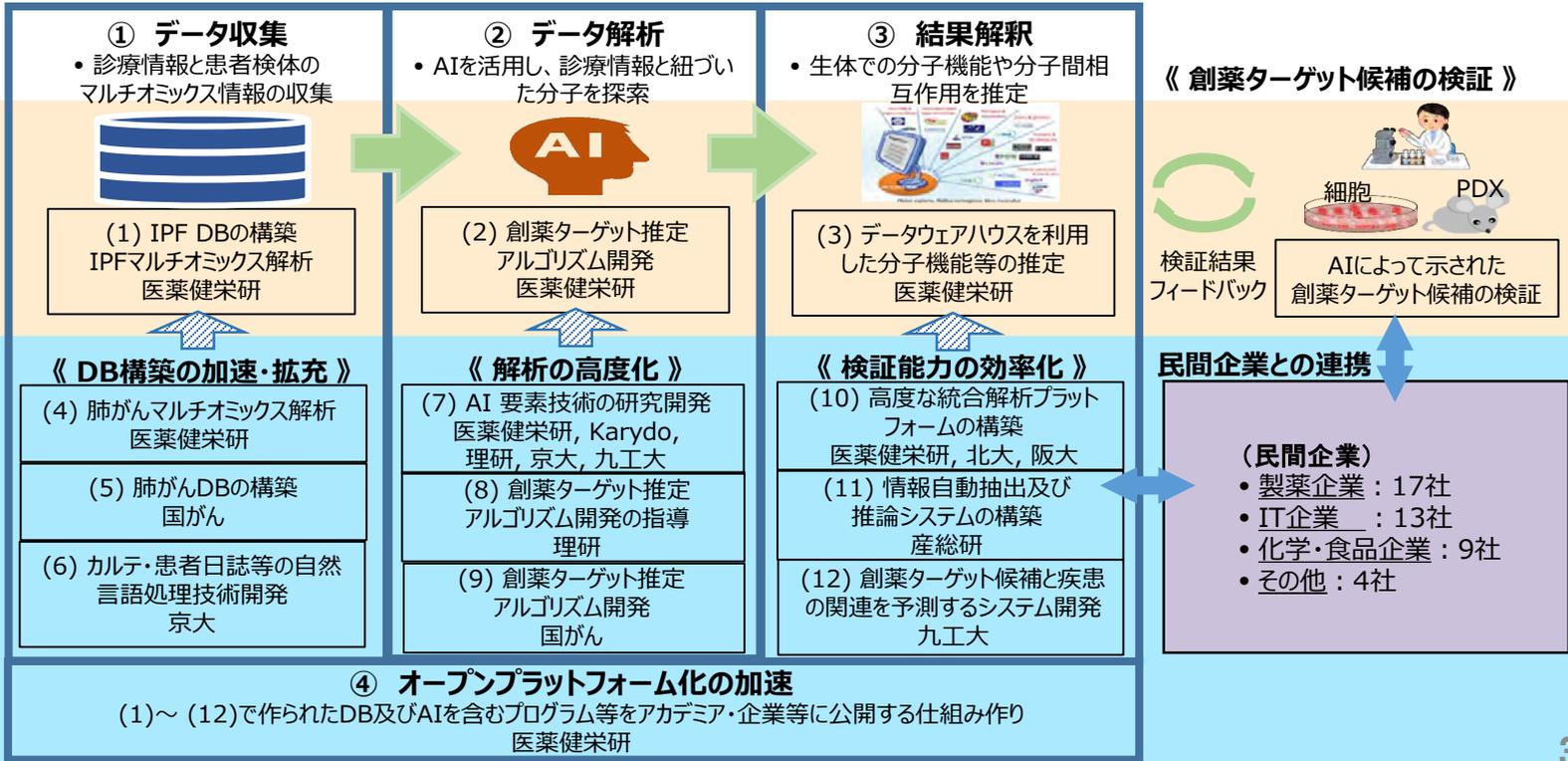
本施策では、① 診療情報及びマルチオミクスデータを収集し、それらを構造化した疾患統合データベースを構築する。② 疾患統合データベース及びAIを活用して患者を層別化し、診療情報と密接に関連した生体分子を探索する。③ AIにより探索された生体分子の分子機能や分子間相互作用を、既存知識を利用して推定することで、創薬ターゲット候補を決める。この技術を民間企業との連携を通じて社会実装させることで民間研究開発投資を促進する。

厚労省 元施策

「新薬創出を加速する人工知能の開発」
《令和3年度予算》
291百万円

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3年度予算》
厚労省：574百万円
文科省：7百万円

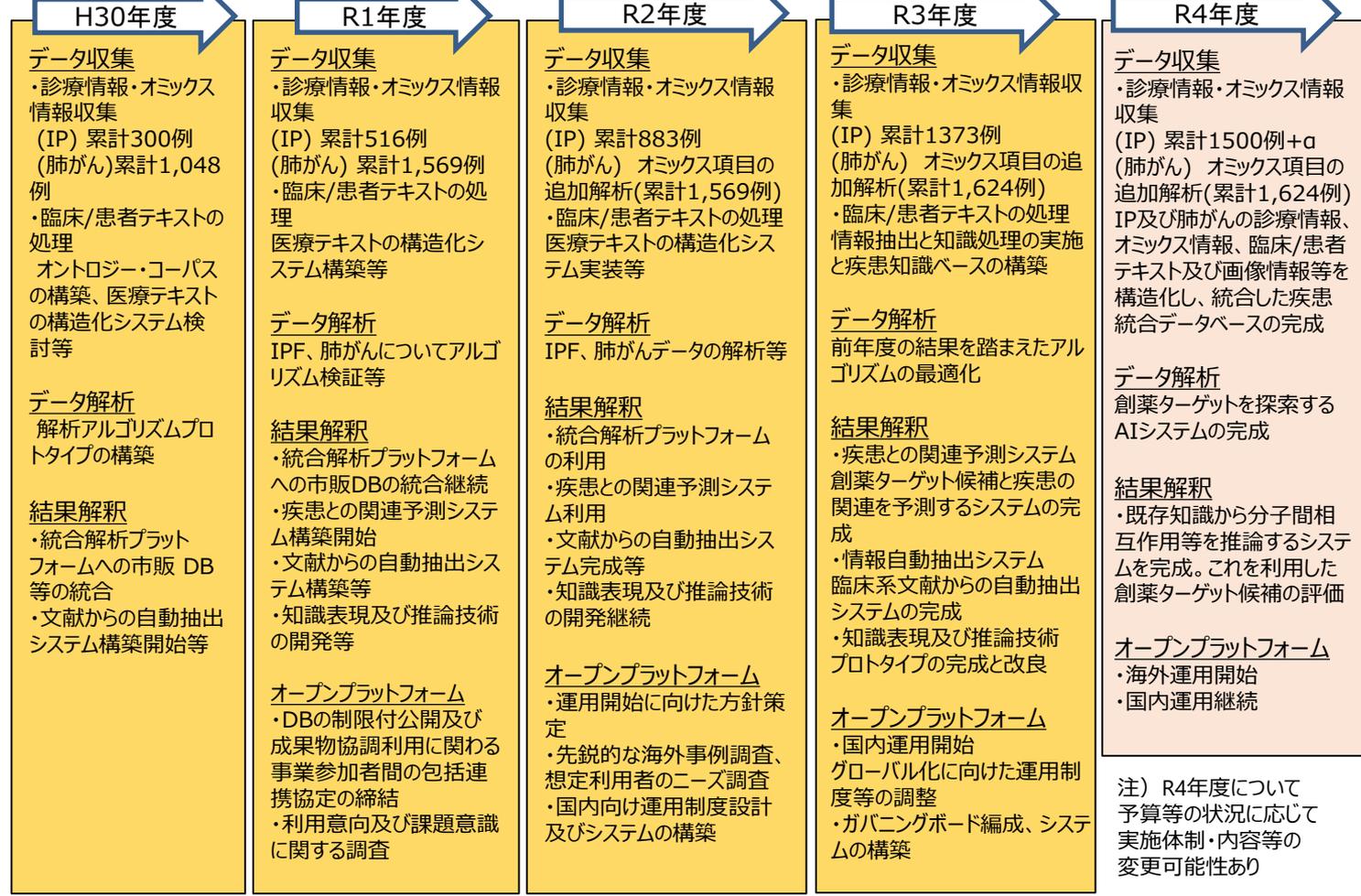


元施策

- データ収集 IP患者の診療情報、オミックス情報の収集（診療情報からのデータ抽出・構造化）
- データ解析 IPFにおける疾患・症状に特徴的な分子を推定するプログラムの開発
- 結果解釈 データウェアハウスの利用による結果解釈・仮説（創薬ターゲット候補）創出
- 仮説検証 仮説（創薬ターゲット候補）のwet実験による検証

アドオン施策

データベースの拡充／データ収集、解析技術の高度化・加速化
／オープンプラットフォームの構築準備



事業終了時の成果イメージ

- データ収集による成果
診療情報の収集・構造化のための処理技術
⇒ 症例報告自動作成システム(CDCS)
⇒ 画像アノテーションシステム
⇒ 医療テキスト(読影所見、病理所見、カルテ等)からの情報抽出・構造化システム
- データ解析による成果
患者の層別化、疾患や症状に特徴的な分子の抽出技術
⇒ 医療情報解析AI及びノウハウ
- 結果解釈による成果
文献からの知識の抽出/知識表現/生体分子の相互作用等の推論技術
⇒ 知識自動抽出システム
⇒ 新たな仮説の推論システム
- オープンプラットフォーム化による成果
医療情報を医療従事者以外に公開し利用してもらう新たなモデルケース(R3年度に国内向けの運用を開始)
⇒ データベース及びシステムの公開による研究開発の促進
⇒ 新たな事業モデルの提供

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するなど、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。

事業全体目標	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"> IPFデータベース構築：400症例の追加 肺がんデータベース構築：オミクスデータ欠損値補完 	<ul style="list-style-type: none"> IPF：累計1,373症例を達成見込み。新型コロナ感染拡大のために追加症例数が減少したが、神奈川コホートで目標数1000症例（累計）の同意を取得し、データベースに490症例を追加した（残りのオミクス解析は来年度実施）。 肺がん：目標症例数は令和元年度達成済み（現在1624症例）。令和2年度よりpan-negative肺がんを中心に欠損値を補填中。
<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）</p> <ul style="list-style-type: none"> 複数の異なるAIアルゴリズムを試行し、創薬ターゲット候補を追加提案 <p>③（結果解釈）</p> <ul style="list-style-type: none"> 創薬ターゲット候補の確からしさを検証する 2種類の情報自動抽出・推論用AI開発 	<ul style="list-style-type: none"> 本事業で収集したデータを共有し、各機関がそれぞれの技術でデータを解析中。 IPFでの関与が報告されていない分子/パスウェイを見出した（特許出願済み2件）。ヒト肺疾患部での発現を確認（免疫染色）したほか、家族性IPFの遺伝子変異導入マウスでの実験や患者でのメタボローム解析などを実施し、検証中。 Pan-negative肺がんの創薬ターゲットを探索し、関与が報告されていない重要分子1つを見出した。現在PDXマウスで検証中。 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について、症例テキストの時系列検索システムの開発を継続中。医療テキストを自動的に構造化・可視化するシステムを開発し、特許出願した。 医薬健栄研が保有するデータウェアハウス（TargetMine）に、これまでに市販データベース等を追加統合し、上記AIによる解析結果の解釈に活用。 英語文献からの情報自動抽出システムのプロトタイプ開発を完了し、オープンプラットフォームに実装した。推論技術の開発を継続中。
<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、研究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"> 先行事例等の調査結果に基づく制度策定 システム構築 	<ul style="list-style-type: none"> 昨年度の調査結果を基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。基本構想を基に運営規程案を作成したが、倫理手続きの遅れで、発行に至っていない。 今年度は、AI10個とこれまで収集した診療情報/オミクス情報を匿名加工情報にしたものをオープンプラットフォームに搭載する。 新たにIT関連企業1社と製薬企業2社と共同研究を開始。疾患情報を収集している他事業との連携（アカデミア5機関、民間企業7社）が合意され、覚書を締結した。LINC（アカデミア2機関、民間企業7社）との連携も開始。開発したAIを様々な疾患に適用すると共に、企業及びアカデミアの意見を反映させながらオープンプラットフォームの仕組みを改良していく予定。厚生労働省の全ゲノム解析等専門委員会において、本プラットフォームの活用を検討開始。 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。

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、490症例（目標400症例）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1,373症例。
- 肺がん症例数は世界最大規模（1,624症例）を達成済み。今年度はpan-negative肺がんを中心に欠損値補填としてChIP-seq解析（50例）、RNA-seq（75例）、全ゲノム解析（50例）の追加解析実施。

○ データ解析

- IPFについて、既存データベースや論文情報を利用してこれまでに開発してきた創薬ターゲット探索手法4種を統合することにより、より信頼度の高い創薬ターゲット提示システムへと改良した。さらに、新たに開発した予測手法によりIPFとの関連性が示唆される薬剤や遺伝子が追加して見出された。
- これまで効果的な治療方法が確立されていないPan-negative肺がんの創薬ターゲットを探索し、関与が報告されていない重要分子1つを見出した。今回同定された分子は、全ゲノムデータ・ChIP-seqデータ・RNA-seqデータを統合的に解析しない限り同定は困難であった（全ゲノム解析を中核とした多層化オミックスデータをマルチモーダルに解析する重要性を立証）。
- 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について、症例テキストの時系列検索システムの開発を継続中。医療テキストを自動的に構造化・可視化するシステムを開発し、特許出願した。事業成果を活用した国際コンペティションを開催し、事業成果のアピールに貢献。

○ 結果解釈

- 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。論文のIDを入力するだけで情報を自動的に抽出し、構造化データとして取得できる様になった。
- 従来の文献解析技術で必要となる学習データを必要とせず、欲しい情報を論文から直接得る新規技術を開発中。
- 情報自動抽出システム及び推論技術開発は継続中。COVID-19関連文献への技術適用も進め、ツール等の公開に向けてウェブサイトを作成。
- 患者層別化AIを用いたIPF臨床情報解析により見出されたIPFの特徴と紐づけられたタンパク質複数個について、肺がん合併IPF患者の肺線維化部位において実際に発現が亢進していることを確認（免疫染色）したほか、家族性IPFの遺伝子変異導入マウスでの実験や患者でのメタボローム解析などを実施し、検証中（創薬ターゲットについて特許出願2件）。Pan-negative肺がんで見出された重要分子についても、PDXマウスで検証中。
- 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した。

○ オープンプラットフォーム

- 昨年度にプロトタイプ構築が完了したオープンプラットフォームシステム上で簡便に利用できるようにするため、本事業成果のAIのうち9個をアプリ化しシステムに搭載した。既に搭載済みの患者層別化AIと合わせて合計10つのAIを搭載した。
- オープンプラットフォーム上での閲覧・解析を可能にするためには、要配慮個人情報である診療情報に匿名加工が必要となる。阪大より収集したIPF診療情報の匿名加工を完了し、プロテオームデータとともにオープンプラットフォームに掲載した。医療画像やテキストなどについても匿名加工準備中。
- これまでの共同研究に加え、IT関連企業1社と製薬企業2社と新たに共同研究を開始。また、疾患情報を収集する他事業とLINCとの連携を開始した。疾患情報を収集している他事業との連携（アカデミア5機関、民間企業7社）では、同事業で得られた様々な疾患データ（精神疾患、神経変性疾患、免疫炎症疾患など）に開発されたAIの適用を拡大していく。LINC（アカデミア2機関、民間企業7社）との連携においては、参加する製薬企業とIT企業にオープンプラットフォーム掲載されたAIを使用してもらい、企業及びアカデミアの意見を反映させながらオープンプラットフォームの仕組みを改良していく予定。
- 厚生労働省の全ゲノム解析等専門委員会において、本プラットフォームの活用を検討開始。
- 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。（令和二年度成果報告会、論文61報、学会発表77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3年度の4年間で18億円相当

- ① 平成30年度 5.2億円（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 4.3億円（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 4.5億円（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）
- ④ 令和3年度 4.2億円（内訳）共同研究等43社（製薬企業17社を含む）：共同研究費（0.7億円）、人件費（3.5億円）

当年度当初見込み	当年度実績
<p>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）</p> <ul style="list-style-type: none"> ① 共同研究 人件費（約29百万円） / 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 in silicoによる創薬研究等 5件 研究費（12.8百万円）、人件費（64百万円） ③ 共同研究 3件 研究費（27百万円）、人件費（5人年（36百万円） ④ 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） ⑤ 自然言語処理に関する共同研究 2件 研究費（2百万円）、人件費（20百万円） ⑥ インフォマティクスに関する共同研究 7件 研究費（28百万円）、人件費（52百万円） ⑦ がん関係の共同研究費 5件 研究費（10百万円）、人件費（90百万円） <p>-----</p> <p>計：共同研究等38社（製薬企業24社を含む）：共同研究費（0.8億円）、人件費（3.0億円）、資機材（約0.1億円相当）</p>	<p>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）</p> <ul style="list-style-type: none"> ① 創薬ターゲット探索に関する共同研究 人件費（40百万円、28社） ② in silico創薬に関する共同研究 3件 人件費（51.8百万円） ③ バイオマーカー関連の共同研究 2件 研究費（1.5百万円）、人件費（8百万円） ④ 創薬関連の共同研究 3件 研究費（30百万円）、人件費（70百万円） ⑤ 創薬関連の共同研究 1件 研究費（8.8百万円）、人件費（5.5百万円） ⑥ 自然言語処理に関する共同研究 3件 研究費（6.7百万円）、人件費（12百万円） ⑦ インフォマティクスに関する共同研究 5件 研究費（21百万円）、人件費（50百万円） ⑧ がん関係の共同研究費 3件 研究費（5百万円）、人件費（90百万円） <p>-----</p> <p>計：共同研究等43社（製薬企業17社を含む）：共同研究費（0.7億円）、人件費（3.5億円）</p>

○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。

当年度見込み	来年度見込み
<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：15社 （外資7社を含む15社の製薬企業等と共同研究についての話し合いを行い、新たに1社と共同研究契約を締結した。） • AI搭載ITプラットフォームを共同開発し、現在企業と製品化を目指し検証中。 	<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 （創薬ターゲット探索、薬効予測、読影所見の自動生成、AI技術を用いたがん医療プラットフォームの製品化及び副作用情報の自動抽出等の自然言語処理技術など）