

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」の全体像

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

- 民間からの貢献額：令和2年度実績（見込み）として民間企業44社（製薬企業：24社、IT企業等：6社、化学・食品企業等：4社、医療機器等機械メーカー：3社、ヘルスケアその他：7社）と総額4.5億円相当の共同研究を行った。（内訳）：共同研究費（1.0億円）、人件費（3.4億円）、資機材（約0.1億円）
- 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）

資料2 元施策の概要

PRISMによる加速・拡充

- 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。
- PRISMにより、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ネットワークラボのAI等基盤技術並びに産総研の知識抽出技術及び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
- ・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

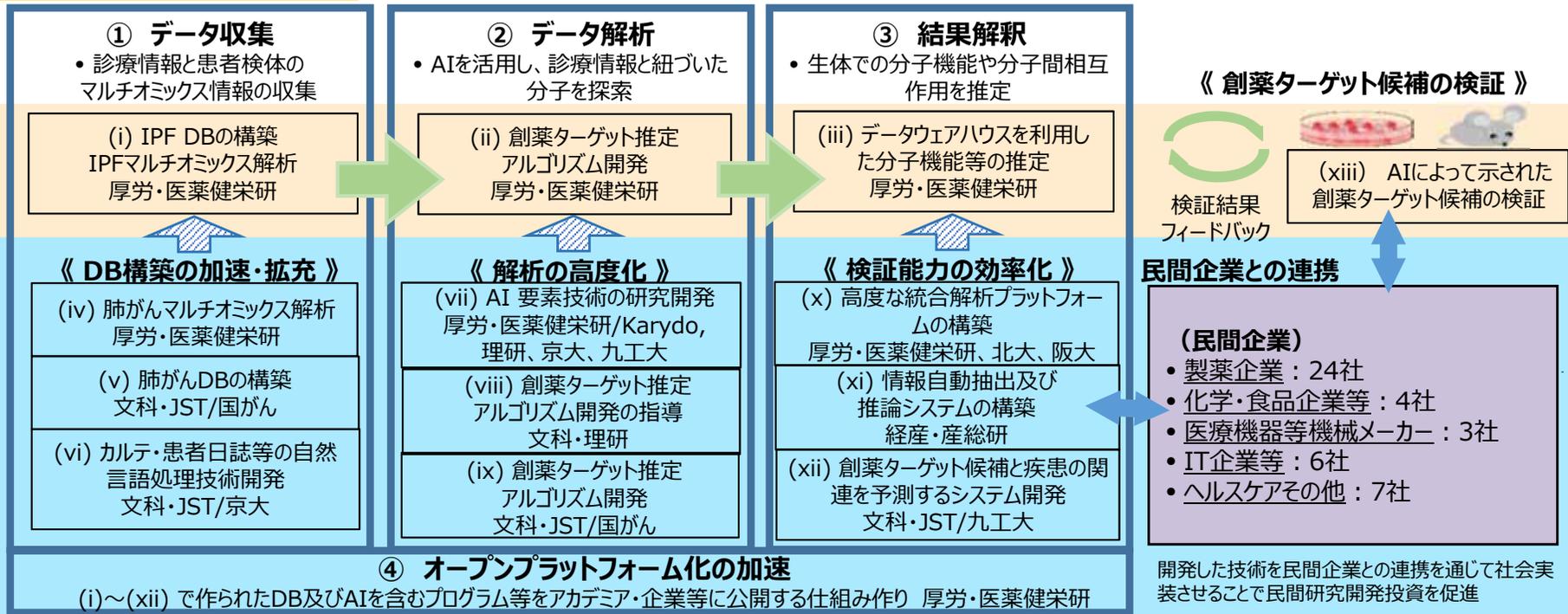
「創薬ターゲット探索プラットフォームの構築」の概要

厚労省 元施策

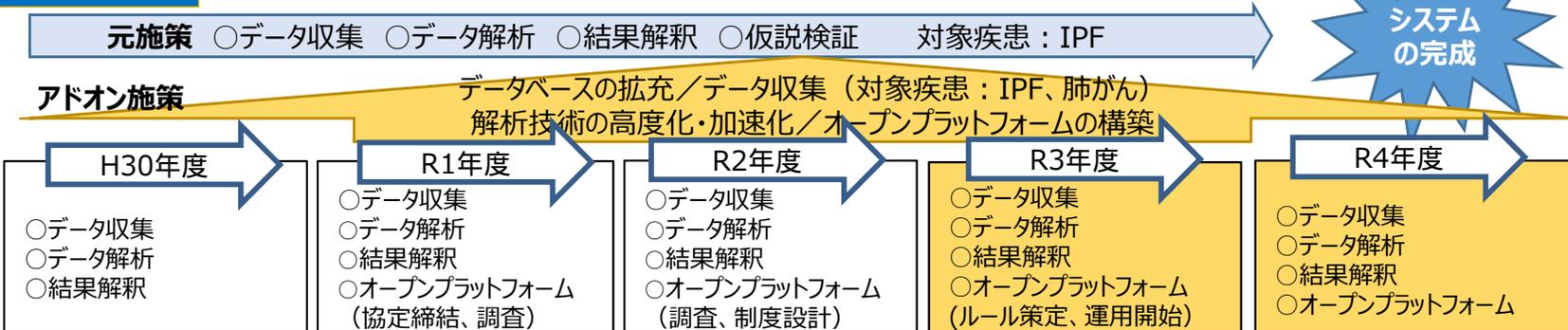
「新薬創出を加速する人工知能の開発」
《令和3年度》291百万円

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年度要求額》
厚労省：574百万円
文科省：7百万円



ロードマップ



事業終了時の成果イメージ
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

- ・情報の収集・構造化のための処理技術
- ・疾患統合データベース（IPF、肺がんともに 累計約1,500例）
- ・新たな仮説の推論システム
- ・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

資料3 「創薬ターゲット探索プラットフォームの構築」の概要①

(令和2年度 アドオン:683百万円/元施策:厚300百万円、文9,704百万円の内数、令和3年度 アドオン:581百万円/元施策:厚291百万円、文3,249百万円の内数)

背景・現状

◆ 背景

- 現在の創薬開発は、複数の機能分子や環境要因が関与する難病・稀少病等の複雑な疾患へとシフトしている。そのため、従来の遺伝子解析や動物実験頼みのアプローチでは新たな創薬ターゲットを見出すことが困難であり、「創薬ターゲットの枯渇」が産業界の深刻な問題となっている。
- 最近では、ゲノムから臨床情報までの疾患情報をAIで解析し、創薬ターゲットと患者層別化バイオマーカーを探索する新たなアプローチに「ゲームチェンジャー」として大きな期待が寄せられ、米国ではメガファーマや情報系ベンチャーによる熾烈な開発競争が繰り広げられている。
- 日本でも製薬企業等の意欲は高いが、企業だけで十分な臨床情報収集やデータ解析体制を整備することが困難なことから、国による先駆的取組に大きな期待が寄せられている。
- 厚生労働省では、平成29年4月より、医薬基盤・健康・栄養研究所（医薬健栄研）において、「創薬ターゲットの枯渇」問題を解決するために、創薬ターゲットがまだ見つかっていない特発性肺線維症（IPF）を対象疾患として、創薬ターゲットを探索する本格的なAI開発に着手した。
- 平成30年7月からは、内閣府による「官民研究開発投資拡大プログラム（PRISM）」の対象施策に採択され、肺がんを対象疾患に追加し、成果の社会実装に向けて5年間の研究開発目標を設定して文部科学省及び経済産業省が所管する国研等と連携を開始した。

◆ 現状

- （データ収集）新型コロナの影響により症例収集が半減したため、収集が容易な血液サンプルをオミックス解析の対象とすることにした。今年度中に380症例の解析を終了させる予定。
- （データ解析）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見出した。また、深層学習を含む機械学習を用いた複数の手法を用いて、IPFと関連する他の疾患やタンパク質等を予測し、意外な疾患との関連やタンパク質が見い出されてきた。AIからの出力結果の確からしさを検証する予定。
- （結果解釈）データ解析で見いだされてきた複数のタンパク質についてデータウェアハウスであるTargetMineを用いて疾患パスウェイ等の推定を行った。情報自動抽出システムのプロトタイプ及び抽出結果を用いた疾患ネットワークを構築している。推論技術開発については継続して実施中。
- （オープンプラットフォーム）国内外での先行事例等の調査を実施中。途中経過を確認し、逐次オープンプラットフォームの運営方針の策定に取り入れる予定。

研究開発目標

- ① 特発性肺線維症（IPF）及び肺がん患者1,500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。
- ② 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AI及び知識ベースを開発し、それらの性能を実証する。
- ③ 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。
- ④ 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。

戦略の位置付け等

- AI戦略2019（令和元年6月11日統合イノベーション戦略推進会議決定）
II 未来への基盤作り／II-2 研究開発体制の再構築／（1）研究環境整備
具体目標3 世界の研究者から選ばれる、本戦略に即した魅力的な研究開発の制度及びインフラの整備
- III 産業の基盤作り／III-1 社会実装／（1）健康・医療・介護
具体目標2 日本が強い医療分野におけるAI技術開発の推進と、医療へのAI活用による医療従事者の負担軽減
具体目標4 世界最先端の医療AI市場と医療AIハブの形成

資料3 「創薬ターゲット探索プラットフォームの構築」の概要②

(令和2年度 アドオン:683百万円/元施策:厚300百万円、文9,704百万円の内数、令和3年度 アドオン:581百万円/元施策:厚291百万円、文3,249百万円の内数)

出口戦略

- 本事業で創出される創薬ターゲット候補等の民間企業への導出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。具体的には、電子カルテ関係（医師の入力に係る労務時間を短縮できる電子カルテ入力支援システム、医療テキスト解析システムと疾患知識ベースを用いた症例報告自動作成システム等）や医用画像関連システム（画像アノテーションツール等）への応用を図る。
- 本事業で創出される疾患情報統合データベース、疾患知識ベース及びAIアルゴリズム等の「創薬ターゲット探索のための基盤」については、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。
- 本事業により創出される解析アルゴリズムは、生活習慣病の先制的介入要因の探索や認知症、精神疾患を始めとする慢性疾患の予後予測など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や機能統合を図ることによって統合ヘルスケア・オープンプラットフォームを構築し、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。
- 本枠組みを海外にも展開する。

PRISMで推進する理由

- （厚生労働省）
対象疾患にIPFと深く関連する肺がんを追加し2つに増やすとともに、医薬健栄研が保有するAI開発基盤に加えて理研、産総研などで培ってきた技術を相補的に活用することが可能となり、元施策事業で目指す事業目標達成の大幅な加速・高度化を期待できる。
- （文部科学省）
理研AIPセンターにて実施するAI等基盤技術の適用による世界初の患者層別化技術の開発に加え、JST-AIPネットワークラボにて実施していた研究課題であるAI技術を用いたマルチオミクス解析基盤技術や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の候補を予測する機械学習アルゴリズムなどの研究成果を、アドオンとして創薬ターゲットや患者層別化バイオマーカーの探索に向けた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、診断薬メーカー等）への技術移転等の加速が期待できる。

元施策の加速

- （対象疾患） 難病である特発性肺線維症（IPF）に加え、日本で多くの患者がおり、IPF患者で高頻度に併発する肺がんを加えることができた。
- （症例収集） 肺がんの統合データベースを構築するにあたり、これまで国がんにおいてバイオバンクやがん関連遺伝子のパネル検査等の目的で収集された肺がんの診療情報及びがん組織を利用し、短期効率的に肺がん疾患統合データベースを構築できた。
また、医療テキストからの情報の抽出にあたり、これまでJSTの事業で培ってきた知識に基づく頑健で高精度な構造的言語処理技術を利用できた。
- （データ解析） 理研及びJST事業で培ってきたAI技術を利用して、新規アルゴリズムを構築できた。
- （結果解釈） 産総研で培ってきた自然言語処理及び自動推論技術を利用して以下のことを実施できた。
 - ① 英語文献から知識を自動収集し知識ベースを構築。
 - ② 知識ベースから生体パスウェイ等を自動推論するシステムを構築。

資料4 令和2年度の成果

- 課題：細胞や動物実験を中心とした創薬ターゲットを探索する従来の研究スキームは、医薬品開発の成功率が長期に亘って低迷している要因となっていることから、製薬産業界では、患者の診療データとAIを活用してデータ駆動的に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。

目標

進捗

データ収集

- (最終目標) IPF及び肺がん各1,500例の診療・オミックスデータの収集
- (今年度目標) IPF：300例追加
肺がん：欠損値の補完

- 肺がん：目標の1,500例を達成（令和元年度）
今年度はpan-negative肺がん症例を中心に欠損値を補充
- IPF：令和元年度までの累積症例数516例に今年度380例を追加し、累積症例数890例（見込み）

データ解析

- (最終目標) 患者の層別化、疾患や症状に特徴的な分子の抽出技術の完成
- (今年度目標) IPF、肺がんデータの解析等

- 肺がん、IPFのオミックスデータからデータ駆動的に患者を層別化するAIを含む、10種類の創薬ターゲット探索AIの開発に成功。
- 複数個の肺がん及びIPF創薬ターゲット候補を捕捉。

結果解釈

- (最終目標) 文献からの知識の抽出/知識表現及び推論技術の完成
- (今年度目標) 文献からの自動抽出システム/知識表現及び推論技術の開発（継続）

- 医薬健栄研が保有するデータウェアハウスに3種類の市販データベースを追加統合し、上記AIによる解析結果の解釈に活用開始。
- 英語文献からの情報自動抽出システム及び推論技術開発を継続中。

オープンプラットフォーム

- (最終目標) 国内外・産学の研究者が本事業の成果を利活用できる仕組みを作る
- (今年度目標) ・運用開始に向けた方針策定
・国内向け運用制度設計及びシステムの構築

- 国内外での先行事例及び企業の参加意欲等の調査実施中。
- 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5 令和3年度の研究計画

- 体制 : JSTにおけるPRISM研究が終了するが、JSTでの成果を利用し、自然言語処理により構造化可能な医療テキスト情報の拡大及び実患者情報であるマルチオミクスデータ（全ゲノム、エピゲノム等）の解析を実施してもらうため、研究者は全員厚労省の枠組みの中で研究を継続する。
- 概要 : 令和2年度には、得られた特発性肺線維症（IPF）及び肺がんのデータを使って複数のAIにより創薬ターゲット候補を捕捉することができた。令和3年度は、マルチオミクスデータ（全ゲノム、エピゲノム等）解析のための解析法のブラッシュアップ、及び連携機関で作成したAIでの実患者情報の解析を進める。
- 予算 : 令和3年度予算要求 581百万円（厚労省：574百万円、文科省7百万円）
参考 令和2年度予算 683百万円（厚労省：513百万円、文科省170百万円）

目標

計画

データ収集

- (最終目標)** IPF及び肺がん各1,500例の診療・オミクスデータの収集
- (令和3年度目標)** IPF：400例追加（累計1,290例）
肺がん：欠損値の補完

肺がん：マルチオミクス
（データ収集・解析） pan-negative肺がん症例を中心にWGS及びChIP seq等の欠損値の補充を行う。AI開発を継続し、多層オミクスデータの網羅的解析により治療標的となる新規 driver (epi driver も含む) を探索する。

データ解析

- (最終目標)** 患者の層別化、疾患や症状に特徴的な分子の抽出技術の完成
- (令和3年度目標)** 複数のAIアルゴリズムの試行・最適化

IPF：マルチオミクス、連携機関でのデータ解析
（データ収集・解析） データ収集を継続する。AI開発を継続するほか、今年完成したAIの機能拡張を行い、マルチオミクスデータを使った解析を行う。また、連携機関による患者データの解析を進める。

結果解釈

- (最終目標)** 文献からの知識の抽出/知識表現及び推論技術の完成
- (令和3年度目標)** システムプロトタイプの完成

- データウェアハウスを継続して創薬ターゲット絞り込みに活用。
- 英語文献からの情報自動抽出システム及び推論技術開発の継続。

オープンプラットフォーム

- (最終目標)** 国内外・産学の研究者が本事業の成果を利活用できる仕組みを作る
- (令和3年度目標)** ・国内運用の開始
・グローバル化に向けた運用制度等の調整

- 先行事例等の調査結果に基づく国内運用制度の策定、並びに公開を見据えたシステム構築を完成し、国内運用を開始。
- グローバル化に向けた運用制度等の調整・検討を行う。

資料6 PRISM実施に伴う事業効果等

○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

- 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）
- ④ 令和3年度見込み（内訳）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）

令和2年度実績（見込み）	令和3年度見込み
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（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））	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人年（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）
創薬ターゲット推定アルゴリズムの開発（文科省） 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（29百万円）、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年度実績（見込み）	令和3年度見込み
• 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、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）