

令和3年5月19日

SIP AIホスピタル事業で構築した、新型コロナウイルス感染症例に関する 診療・検体解析情報等データベースの利用受付を開始します

❖ 概要

国立研究開発法人 医薬基盤・健康・栄養研究所（大阪府茨木市、理事長：米田 悦啓）では、内閣府が創設した戦略的イノベーション創造プログラム（SIP）第2期の「AI（人工知能）ホスピタルによる高度診断・治療システム」（プログラムディレクター：中村 祐輔）において、令和2年度、「新型コロナウイルス感染症（COVID-19）の治療薬・ワクチンの研究開発に資するデータ連携基盤の構築」プロジェクト（研究代表者：今井 由美子）として、COVID-19 症例のデータを連携して利用することが可能なデータ基盤の構築を進めて参りました。今般、診療情報 1,778 例（臨床検体解析情報が付随するもの 163 例、胸部 CT 画像が付随するもの 1,615 例からなる）のデータを、利活用形態に応じてデータベース化したクラウド型データベース利用システムを構築しました。このデータベースを広く研究に利活用いただけるよう、

- 当該データベースの概要をお知らせします。
- 当該データベースの利用手続きを定め、5月24日（月）よりデータ利用・共同研究の申請を受け付けます。

❖ データベースの概要、構成及び利用形式等

（データベースの概要）

データベースは、令和2年度に、国内23施設の医療機関の協力の下収集された、軽症～重症のCOVID-19入院患者における次の情報を対象としています。

- ① 急性期から回復期までの複数時点※の診療情報（年齢、性別、基礎疾患の有無、治療法、投与薬物、転帰等）に臨床検体（血液、気道液）の解析情報（ホストゲノム、免疫レパトア、抗体、サイトカイン、ウイルスゲノム等）が付随した、データの二次利用の同意がとれているもの（163例）
- ② 急性期から回復期までの複数時点※の診療情報（投与薬物等）に胸部CT画像情報が付随した、情報公開の下にオプトアウトで情報を収集したもの（1,615例）

※時点数は入院期間や観察期間によって異なります

図1. データベースの概要

データ構成	
└	患者IDフォルダ
└	└ 時点IDフォルダ
└	└ 診療情報ファイル (CSV)
└	└ ホストゲノムファイル (FASTQ)
└	└ ウイルスゲノムファイル (FASTQ)
└	└ 免疫レパトアファイル (FASTQ)
└	└ サイトカインファイル (CSV)
└	└ 抗体ファイル (CSV)
└	└ 画像情報フォルダ
└	└ DICOM画像.zip

データの種類		データ内容		匿名加工変換ルール等	
診療情報	A	基本情報	基本情報	年齢, 性別	年齢10歳ごとに表示, 0「男性」, 1「女性」
	B	診療情報 (匿名加工)	基礎疾患等	高血圧, 慢性肺疾患, 慢性腎疾患, 肝疾患, 糖尿病, 血液疾患, 悪性腫瘍等の有無	0「無」, 1「有」
			基本集計	入院日数, ICU滞在日数, 総酸素投与日数, 総人工呼吸日数, 総ECMO日数, 転帰等	変換なし
			重症度	重症度スコア, 酸素化係数	変換なし
			投薬等	アピガン, レムデシビル, トシリズマブ, ステロイド等の投薬の有無	0「無」, 1「有」
			治療法等	人工呼吸, ECMOの有無等	0「無」, 1「有」
			血液・生化学的検査	ヘマトクリット, 白血球数 (好中球数, リンパ球数), 血小板数, アルブミン, 総ビリルビン, AST(GOT), ALT (GPT), LDH, 尿素窒素, クレアチニン, Na, K, 血糖値, CRP	異常高値, 正常高値, 正常標準値, 正常低値, 異常低値等
画像情報	C	画像情報	胸部CT画像		
検体解析情報	D1	ホストゲノム情報	ホストゲノム	全ゲノムシーケンス	
	D2	ウイルス情報	ウイルスゲノム	SARS-CoV2全ゲノムシーケンス	
	D3	免疫情報	免疫レパトア	T細胞受容体 (TCR)、B細胞受容体 (BCR) のレパトアシーケンス	変換なし
			サイトカイン	CTACK, Eotaxin, Basic FGF, G-CSF, GM-CSF, GRO-a, HGF, IFN-a2, IFN-g, IL-1a, IL-1b, IL-1ra, IL-2, IL-2Ra, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12 (p70), IL-12 (p40), IL-13, IL-15, IL-16, IL-17, IL-18, IP-10, LIF, MCP-1, MCP-3, M-CSF, MIF, MIG, MIP-1a, MIP-1b, b-NGF, PDGF-BB, RANTES, SCF, SCGF-b, SDF-1a, TNF-a, TNF-b, TRAIL, VEGF	
			抗体	SARS-CoV2抗体 (N-IgG, S1-IgG, N-IgM)	

※ウイルスゲノムデータは GISAID (Global Initiative on Sharing All Influenza Data) に登録予定

(データのレベル分類・利用形式)

当研究所では上記の情報を利活用の形態によって次のようにレベル分類して、研究利用に提供します。

○レベル1：カタログ情報（ホームページ上に公開）

上記データベースに含まれるデータ内容の理解に資するよう、カタログ情報を作成し、当研究所ホームページ（<https://www.nibiohn.go.jp/COVID-database/>）にて5月24日（月）に公開致します。

○レベル2：匿名加工情報相当の診療・解析情報（申請に基づき研究者がデータ利用）

収集した情報から、匿名加工情報とすることが難しい画像、個人のゲノムデータ等を除き、診療・解析情報に対して、特定の個人を（再）識別することができない加工（匿名加工）に相当する加工を施して作成したデータベース。

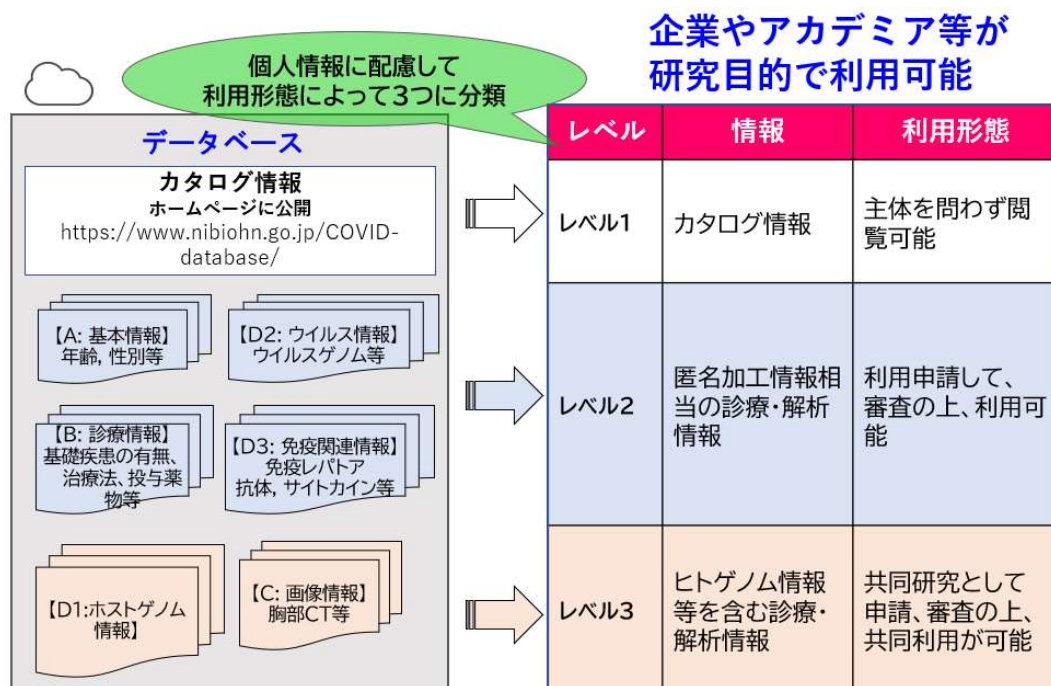
研究計画に基づき、当研究所に利用申請することで、審査の上、データ利用が可能となります。5月24日（月）より、利用申請を受け付けます。

○レベル3：ゲノム情報等含む診療・解析情報（申請に基づき、共同研究によりデータ利用）

収集した情報のうち、診療情報、ホストゲノム以外の解析情報等に加え、個人識別符号であるホストゲノム情報及び画像情報を含む情報により構成されるデータベース。それ自体が個人識別符号となるホストゲノム情報以外の情報については、特定の個人を識別することができる記述等の全部又は一部を削除（置換を含む）しています。研究目的によっては、個人のゲノム情報等完全に個人の識別性をなくしての実施が行えないものもあり、研究目的から当該データの利用が必要と認められる場合にのみ、当研究所との共同研究として共同利用を行います。

5月24日（月）より、共同研究申請を受け付けます。

図2 レベル1、2、3の構成とレベル2、3の利用スキーム



(期待されるデータ利活用)

上記の研究利用として、研究機関、医療機関、民間企業等において抗体薬等の治療薬やワクチンの研究開発、重症化の病態解析や重症化阻止の治療法に関する疫学研究や研究開発等の他、早期診断や重症化予測につながる画像解析 AI 技術の研究開発などに活用されることが期待されます。

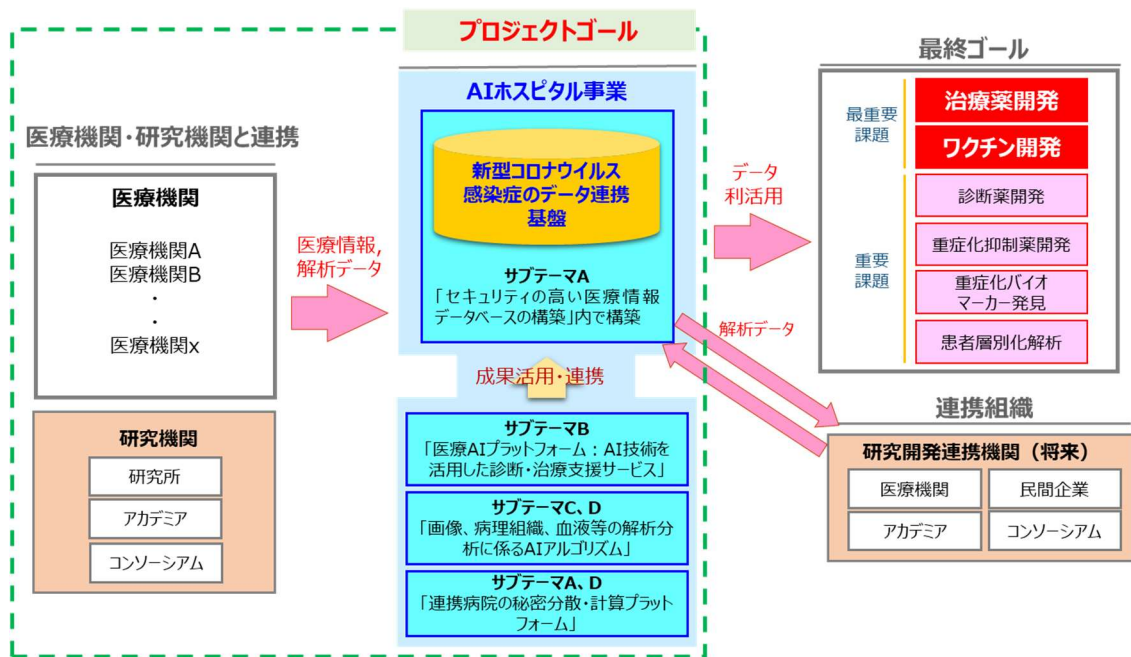
図3 利活用事例

	利用例 (他コロナ研究との連携を含む)
診療情報	<ul style="list-style-type: none"> ・ 病態の解明や治療に関する臨床的研究 ・ 病態予測、層別化システム開発の基盤情報 ・ ワクチンや抗ウイルス薬等の治験のプロトコール設定
ウイルス情報 (ウイルスゲノム)	<ul style="list-style-type: none"> ・ ウイルス変異に関するウイルス分子疫学研究 ・ 高病原性変異ウイルスに対するワクチンや抗ウイルス薬開発 ・ in silicoシミュレーションによる抗ウイルス薬開発
免疫情報 (レパトア・抗体・サイトカイン)	<ul style="list-style-type: none"> ・ 重症化のメカニズムや治療薬の効果に関する研究 ・ 重症化阻止免疫薬開発
ホストゲノム情報	<ul style="list-style-type: none"> ・ 重症化等に関するゲノム研究 ・ ゲノム情報による病態予測、層別化システム等の開発
画像情報 (胸部CT)	<ul style="list-style-type: none"> ・ COVID-19の画像診断に関する研究 ・ 画像診断AIプログラム研究 (教師データとして) ・ 画像による重症化予測システムの開発

❖ プロジェクトの概要

本事業は当研究所が、SIP第2期の「AI（人工知能）ホスピタルによる高度診断・治療システム」において、「新型コロナウイルス感染症（COVID-19）の治療薬・ワクチンの研究開発に資するデータ連携基盤の構築」プロジェクト（研究代表者：今井 由美子）として昨年度実施した研究事業（下図参照）に基づくものです。

「AI（人工知能）ホスピタルによる高度診断・治療システム」では、医療機器等やIoT機器を活用した、患者情報の網羅的収集、ビッグデータ化と共に、セキュリティの高い医療情報データベースの構築、AI技術を活用した診断・治療の支援システム（医療AIプラットフォーム）などの研究開発を進めております。



新型コロナウイルスの治療薬・ワクチンの開発に資するデータ連携基盤の模式図

本データベースは、令和2年度、検体の提供や解析データの第三者による利用にご理解をいただきご協力に同意いただいた患者様、本プロジェクトの趣旨に賛同いただいて、緊張したCOVID-19の医療現場において、検体や診療情報等の提供を行っていただいた医療従事者の皆様、コロナ禍の研究環境で、迅速に検体の解析やデータベースの構築を行っていただいた、下記協力研究機関を含む、関係した大勢の方々の献身的なご協力によって、構築され、公開が可能となったものです。また、今後、新興・再興感染症データバンク事業ナショナル・リポジトリ(REBIND)とも連携、協力していく予定です。

【代表研究機関】

- ・ 医薬基盤・健康・栄養研究所（研究の総括）

【協力研究機関】（五十音順）

- ・ 大阪大学医学部附属病院（検体・診療情報等の提供）
- ・ 大阪大学大学院医学系研究科 遺伝統計学（検体・画像の解析）
- ・ 岡山大学医学部附属病院（検体・診療情報等の提供）
- ・ がん研究会 がんプレジジョン医療研究センター（検体・画像の解析）
- ・ 札幌医科大学附属病院（検体・診療情報等の提供）

- ・ 昭和大学病院（検体・診療情報等の提供、検体・画像の解析）
- ・ 東京都医学総合研究所（検体・画像の解析）
- ・ 広島大学医学部附属病院（検体・診療情報等の提供）
- ・ 横須賀共済病院（検体・診療情報等の提供）

【協力機関】（五十音順）

- ・ NTT コミュニケーションズ株式会社（データベースの構築）
- ・ 株式会社情報通信総合研究所（データベースの構築）
- ・ 徳洲会インフォメーションシステム株式会社（検体・診療情報等の提供）

❖ 用語解説

○戦略的イノベーション創造プログラム(SIP)について

内閣府総合科学技術・イノベーション会議が、科学技術イノベーションを実現するために創設した国家プログラムです。平成30年度から第2期として12課題が採択されており、産学官連携の下、基礎研究から実用化・事業化、すなわち出口までを見据えて一貫通貫で研究開発を推進しています。

詳細は、<https://www8.cao.go.jp/cstp/gaiyo/sip/>をご参照ください。

○「AI（人工知能）ホスピタルによる高度診断・治療システム」について

SIP第2期の課題の一つで、中村 祐輔プログラムディレクター（公益財団法人がん研究会 がんプレジジョン医療研究センター所長）の指揮の下、AI、IoT、ビッグデータ技術を用いた「AIホスピタルシステム」を開発・構築・社会実装することにより、高度で先進的な医療サービスを提供するとともに、医療機関における効率化を図り、医師や看護師などの医療従事者の抜本的な負担の軽減を実現することを目指すプログラムです。

詳細は、<https://www.nibiohn.go.jp/sip/>をご参照ください。