資料1-1

「富岳」Society 5.0推進利用課題

「富岳」を機軸とした創薬DXプラットフォームの構築

2022年10月19日　第52回HPCI計画推進委員会

一般社団法人ライフインテリジェンスコンソーシアム
京都大学　大学院医学研究科
理化学研究所　計算科学研究センター
奥野恭史

一般社団法人ライフインテリジェンスコンソーシアム（LINC)

2016.11.16　日経新聞


日本経済新聞
11月16日
水曜日
創薬AIで50社連合
武田やNEC 新薬探し短縮
成功率高め国際競争力

Life Intelligence Consortium

2016年11月発足

30種の創薬AIプロトタイプの開発
内閣府オープンイノベーション大賞授賞

2021年4月
第２期開始

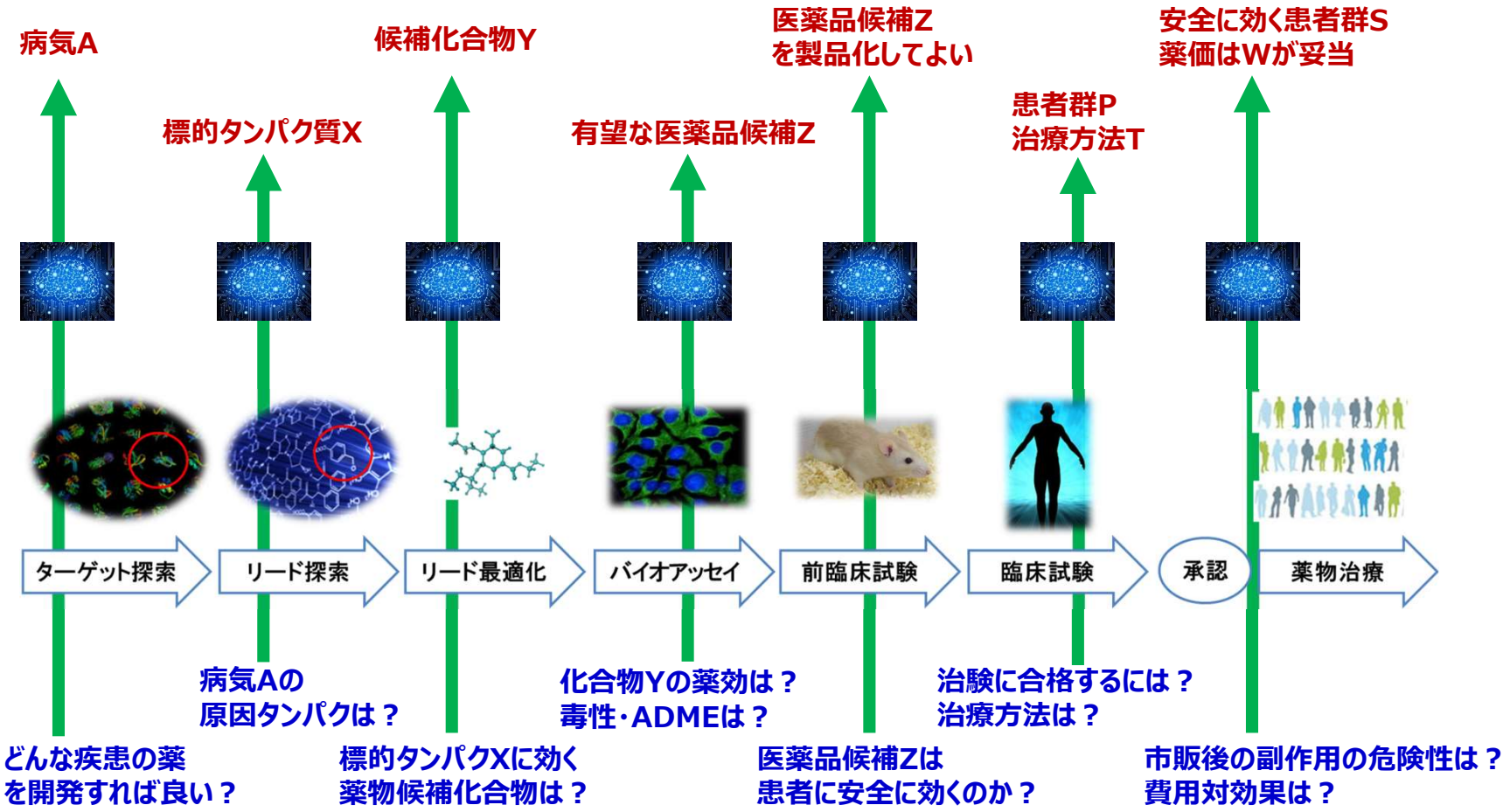
Society5.0の実現
創薬DXの推進

ライフインテリジェンスコンソーシアム（LINC）
京大・理研・医薬健栄研等、ライフ系企業、IT系企業等
約125企業・団体が参画

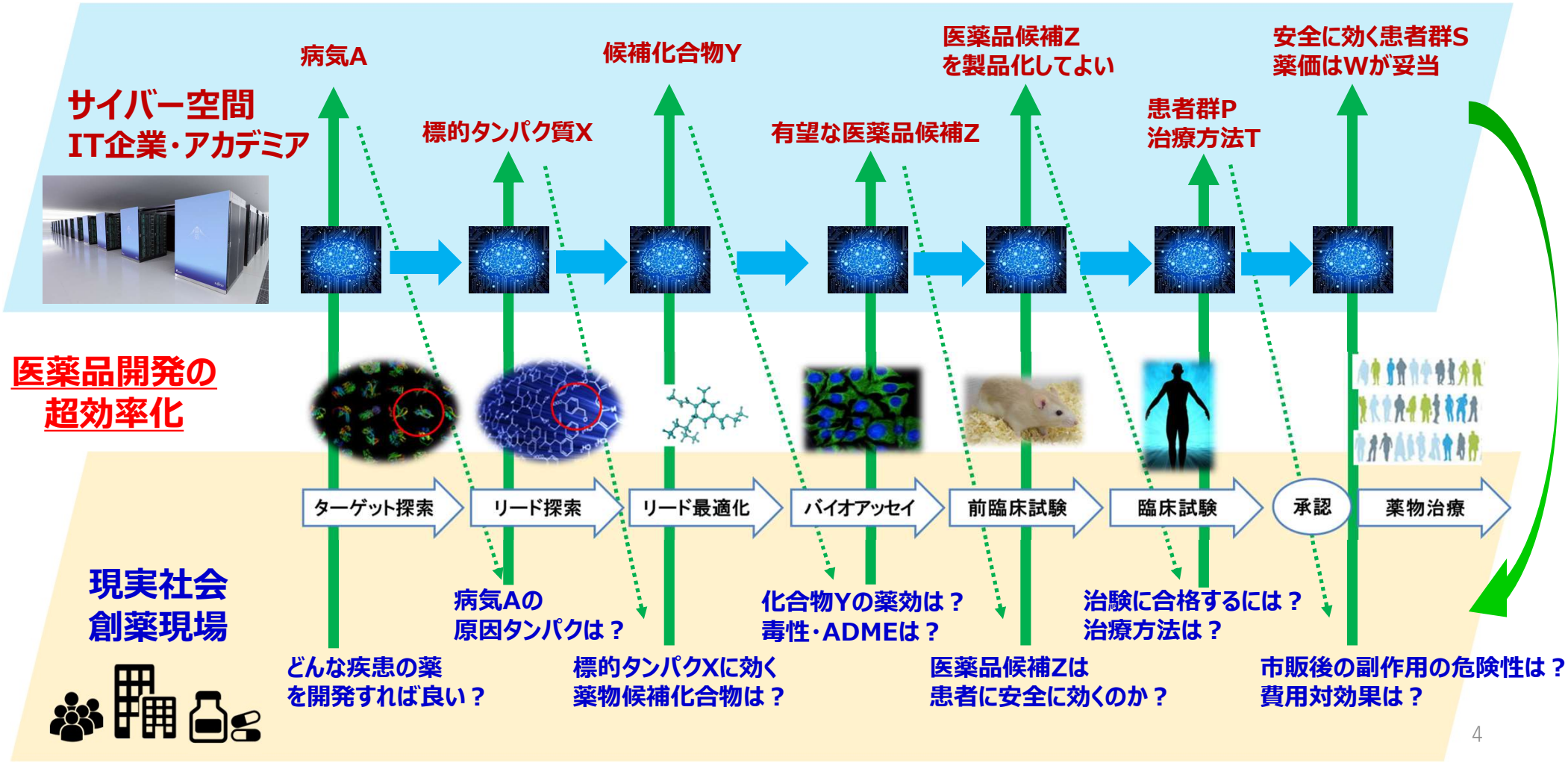
一般社団法人
ライフインテリジェンスコンソーシアム
LINC

LINCでは創薬プロセスにそった約30種のAI技術を開発

開発期間10年以上
開発費1000億以上



LINCが目指す創薬DXプラットフォーム



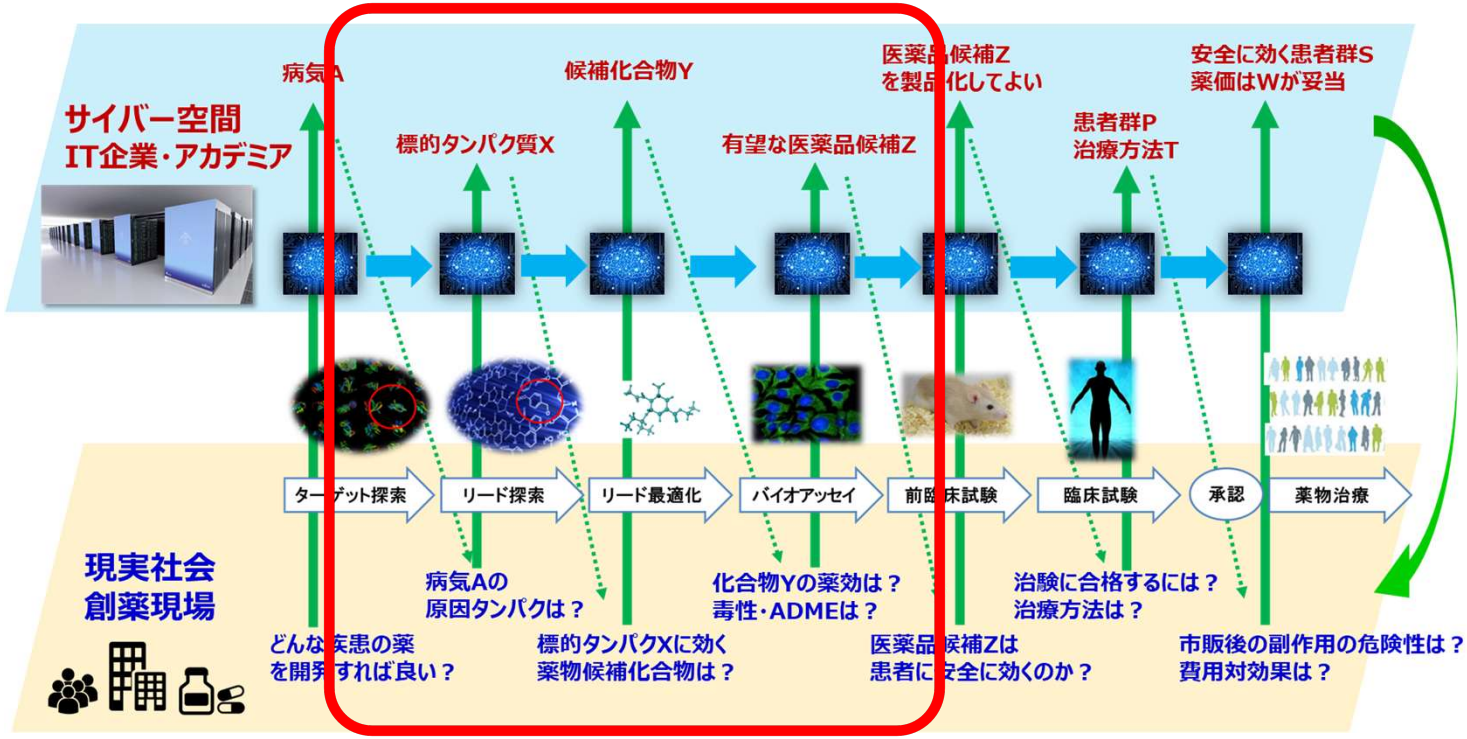
本課題で開発する創薬DXプラットフォーム

HPC/AI 創薬プラットフォーム

LINC、理化学研究所、京都大学、医薬基盤・健康・栄養研究所で開発済/中のAI・HPC技術を「富岳」に実装し、HPC/AI駆動型創薬プラットフォーム（HPC/AI創薬PF）を構築し、サービス化を目指した試験研究を行う。

創薬データベース

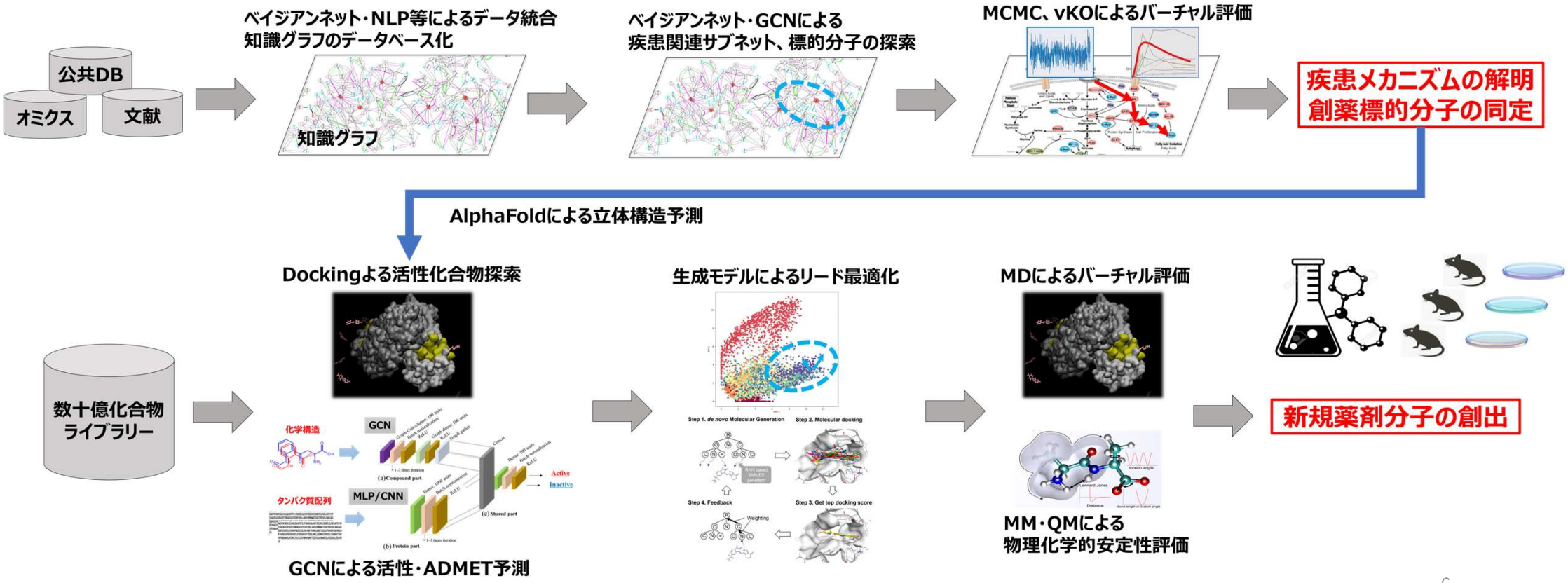
製薬企業等のライフ系企業・アカデミアにおいて必要な創薬知識データベース（創薬DB）を構築する。



HPC/AI 創薬プラットフォーム

- 「創薬ターゲット探索」 ⇒ 「リード化合物創出」に至るHPC/AIフローを構築
- 「富岳」を中心に、HPC/AIフローの自動化を図ることで、創薬の超効率化を実現

創薬ターゲット探索： 疾患名・患者サンプルデータ等を入力して、疾患メカニズムや標的タンパクを推定するHPC/AIフロー

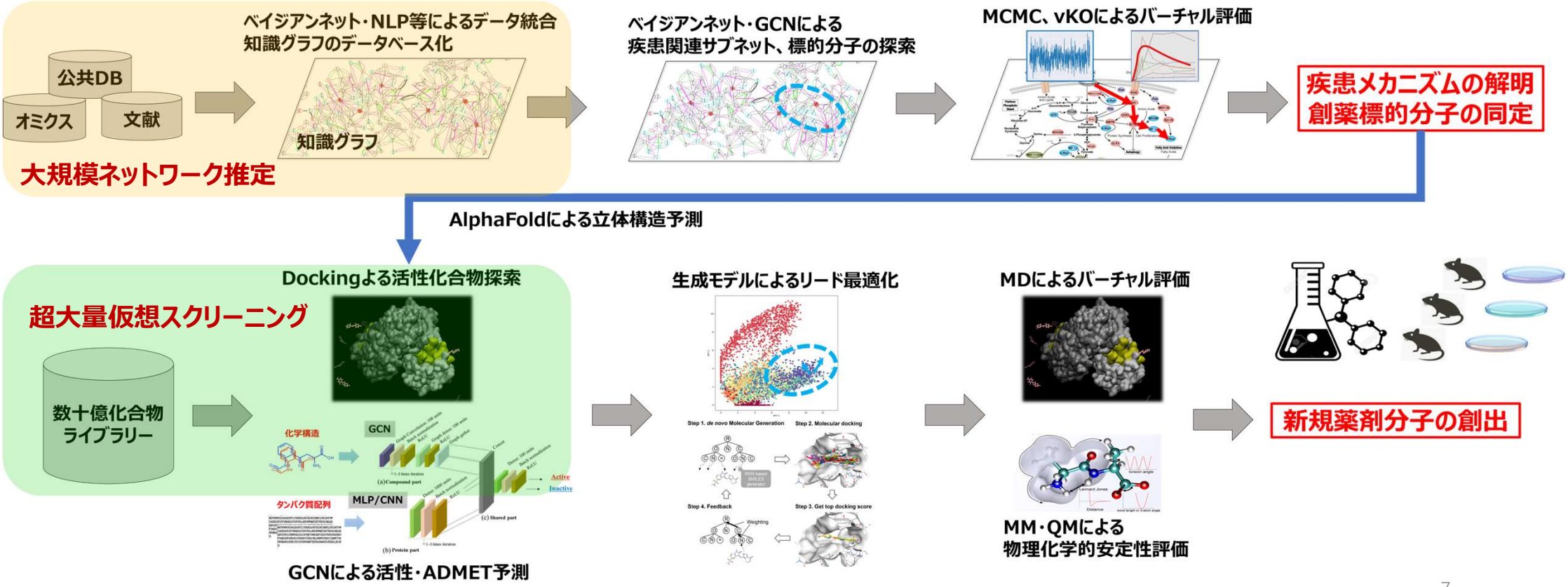


リード化合物創出： 標的タンパク質名を入力して、リード化合物を推定するHPC/AIフロー

創薬データベース

「超大量仮想スクリーニング」と「大規模ネットワーク推定」を行いデーターベース化

創薬ターゲット探索：　疾患名・患者サンプルデータ等を入力して、疾患メカニズムや標的タンパクを推定するHPC/AIフロー

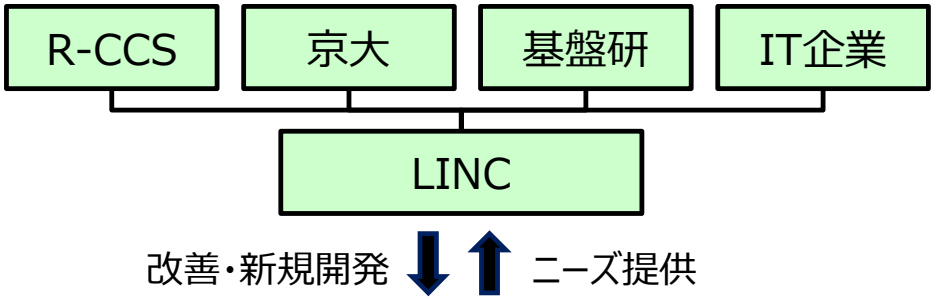


リード化合物創出：　標的タンパク質名を入力して、リード化合物を推定するHPC/AIフロー

実施体制

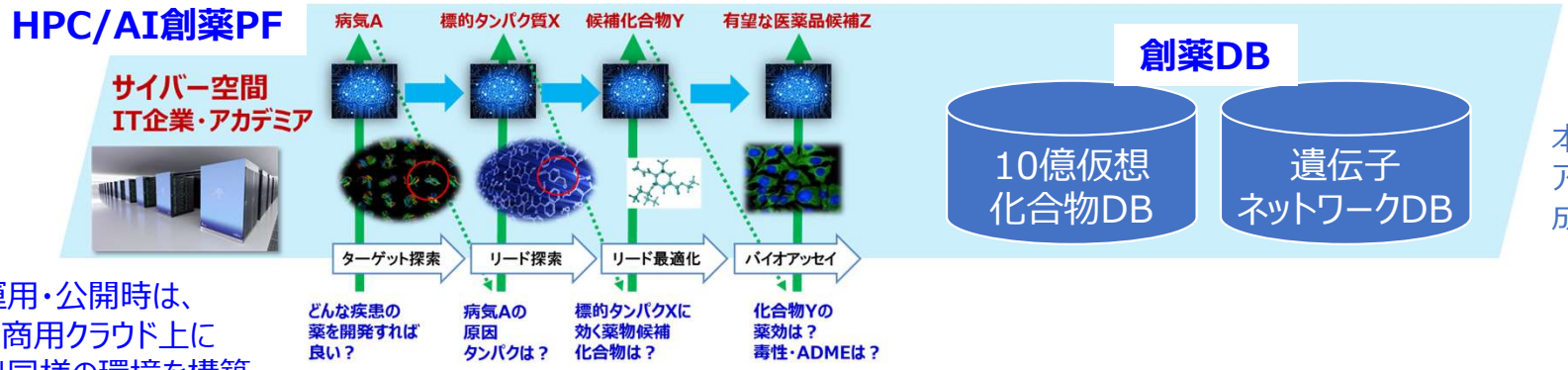
	研究開発	評価	事業化検討
HPC/AI創薬プラットフォーム 理研 井阪悠太	理研R-CCS、京大 LINC IT企業	LINC全企業 （AMED/BINDS事業での評価も実施）	LINC事務局 理研R-CCS S5推進拠点
創薬DB：超大量仮想スクリーニング 理研 大田雅照	理研R-CCS、京大、 LINC IT企業、製薬企業・ライフ系企業	LINC WG04参画企業 （AMED/DAIIA事業での連携も検討）	
創薬DB：大規模ネットワーク推定 弘前大学 玉田嘉紀	京大、弘前大、医薬基盤研	LINC WG02参画企業 （内閣府PRISM事業での評価も実施）	

社会実装イメージ



LINC（＋R-CCS、京大、基盤研、IT企業）が
LINC等の事業体からのフィードバックを受けて、
既存AIを改善、新規AIを開発

改善・新規開発 ⇩⇧ ニーズ提供



本S5課題利用期間中は、
アカデミアも企業も無償利用
成果は公開

本格運用・公開時は、
民間の商用クラウド上に
「富岳」同様の環境を構築

計算結果 ⇩⇧ プラットフォームの利用申請

LINC等の事業体

LINC等の事業体が、製薬企業に
PF/DBをサービスとして提供

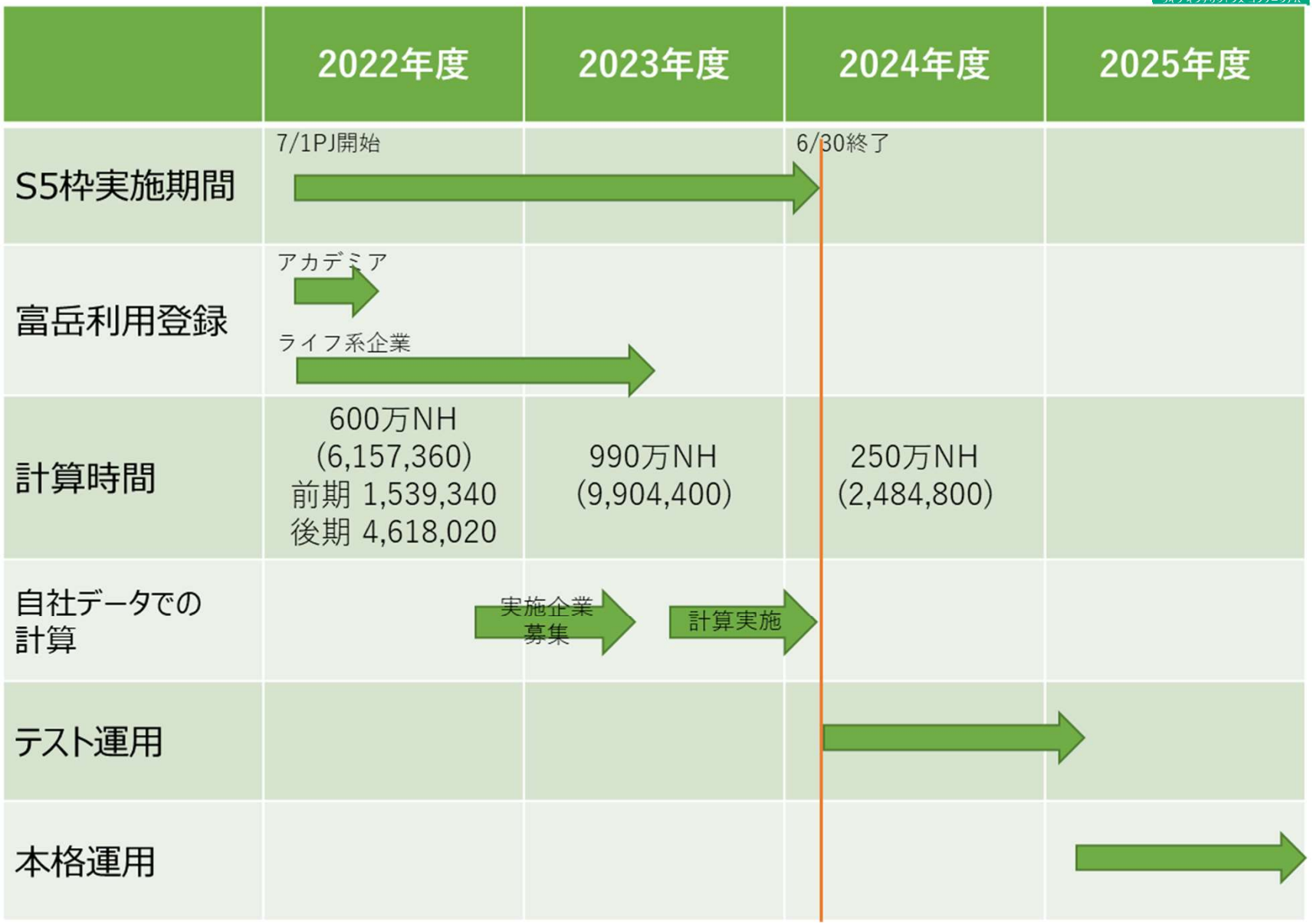
計算結果 ⇩⇧ サービス利用申請

製薬企業

スケジュール

実施状況
7月：
- **「富岳」利用登録**
- **参画企業との契約**

8月から：
- **計算開始**



本事業の参加規則を策定

スーパーコンピュータ『富岳』を機軸とした創薬 DX プラットフォームの構築事業
参加規則

（趣旨）

第１条　スーパーコンピュータ『富岳』を機軸とした創薬 DX プラットフォームの構築事業参加規則（以下「本規則」という）は、一般社団法人ライフインテリジェンスコンソーシアム（以下「LINC」という）が、一般社団法人ライフインテリジェンスコンソーシアム定款（以下「LINC 定款」という）第４条に基づき、事業として実施する「富岳」Society5.0 推進利用課題「スーパーコンピュータ『富岳』を機軸とした創薬 DX プラットフォームの構築」（以下「本事業」という）について、LINC に入会した者（以下総称して又は個別に「会員」という）が本事業に参加するにあたって遵守し、且つ、誠実に対応しなければならない事項を定める。

（用語の定義）

第２条　本規則において、次の各項に掲げる用語の定義は、それぞれ各項の定めるところによる。

２　本規則において「知的財産権」とは、特許権、特許を受ける権利、実用新案権、実用新案登録を受ける権利、意匠権、意匠登録を受ける権利、回路配置利用権、回路配置利用権の設定の登録を受ける権利、育成者権、種苗法第３条に規定する品種登録を受ける地位、商標権、商標登録出願により生じた権利、著作権（著作権法第２１条から第２８条までに規定するすべての権利を含む）及び日本以外の国又は地域における前記各権利及び地位に相当する権利及び地位をいう。

３　本規則において「営業秘密」とは、不正競争防止法第２条第６項に該当する情報をいう。

４　本規則において「営業秘密等」とは、営業秘密若しくは本条第２項の権利の対象とな

- **本事業期間における企業参加は、評価検証のためとし、成果は公開**
- **ただし、自社の非公開データでの評価検証できるスペシャルメンバーを設定**

（メンバーシップと役割）

第５条　本事業に参加する者（以下「事業参加者」とする）は、LINC 内に設けられた本事業を実施するためのプロジェクトに所属していなければならない。また、参加区分及び役割は次のとおりとする。

参加区分	役割
A）アカデミアメンバー	LINC のアカデミア会員であり、本事業における創薬 DX プラットフォームのプロトタイプの開発及び他の事業参加者の検証及び評価を受け改善等（以下「本開発」という）を実施する
B）ユーザーメンバー	LINC の正会員であり、本事業における創薬 DX プラットフォームのプロトタイプをアカデミアメンバーと協力し、テストユーザーとして使用、検証、又は評価（以下「本検証」という）を実施する

２　ユーザーメンバーが、自社の非公開情報を用いて本検証を行うことを希望する場合、別途 LINC 等と契約を交わすものとする（以下別途契約を交わしたユーザーメンバーを「スペシャルメンバー」という）。

進捗：HPC/AI 創薬プラットフォーム

- 理研・理事長裁量経費のサポートにより開発
- 富岳を中心に、「創薬ターゲット探索」 ⇒ 「リード化合物創出」に至る自動フローを構築中

Confidential

リード化合物創出部
6 modules and 37 core applications

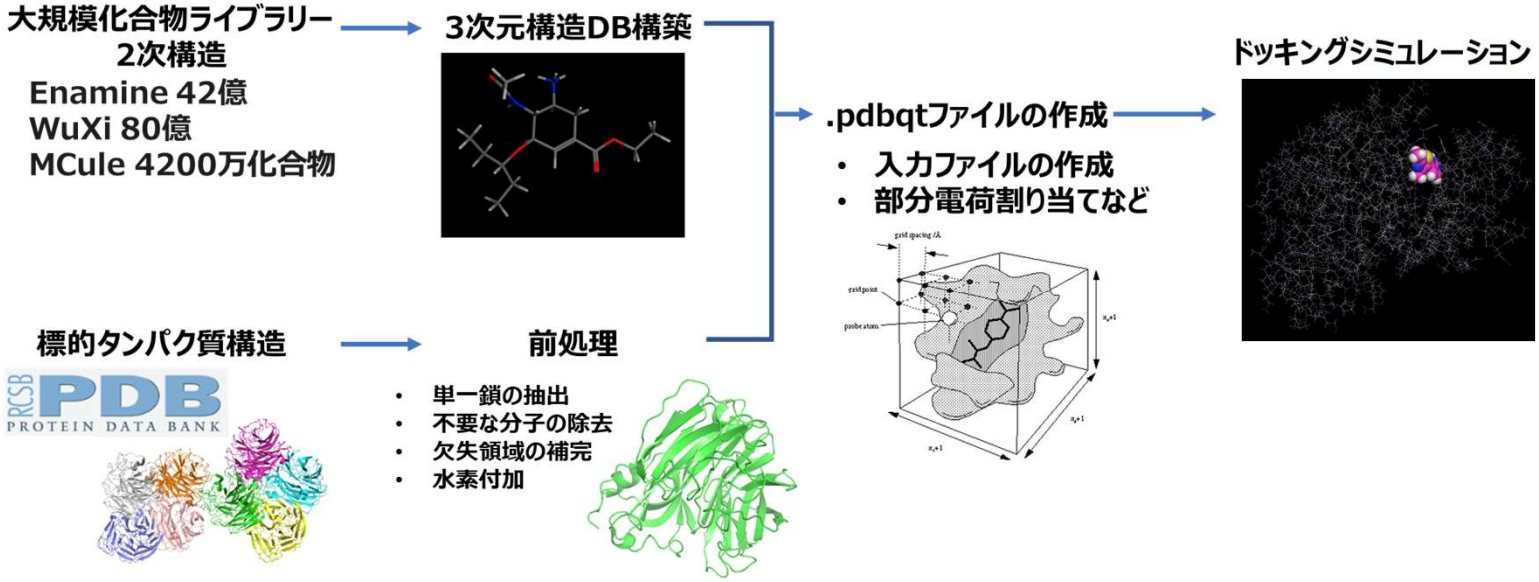
Confidential

ターゲット探索部
14 modules and 20 core applications

進捗：創薬データベース 「超大量仮想スクリーニング」

目標：「富岳」により海外メガファーマの100倍規模の数10億化合物バーチャルスクリーニングを実現

大規模化合物ライブラリー（約250億化合物）の3次元構造ファイルDBを構築中



Dockingフローの高速化と計算時間の検討

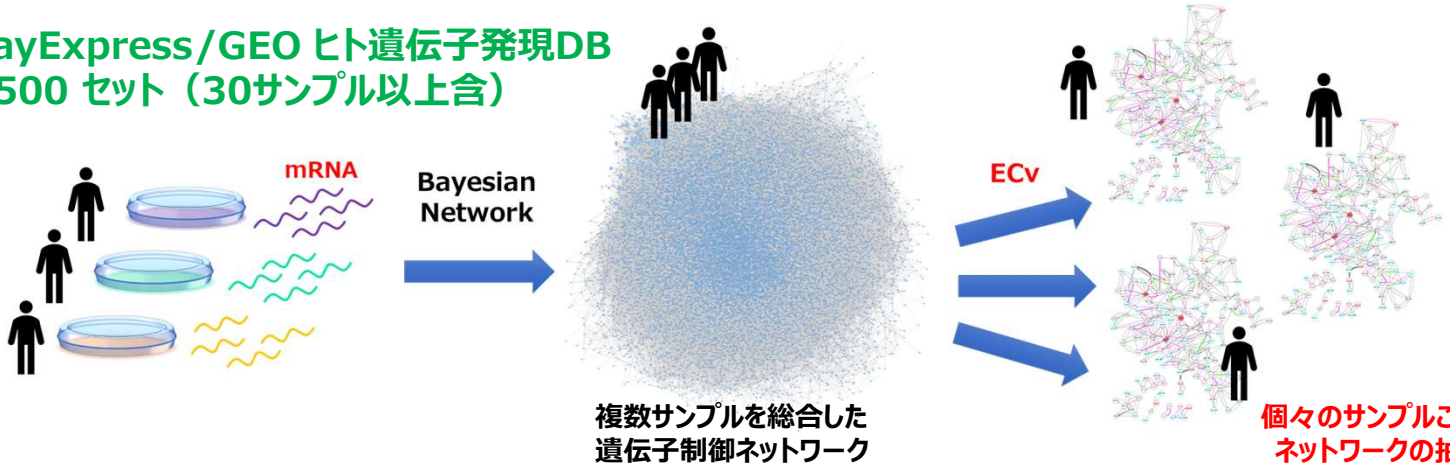
化合物数	計算時間	費用
100万個	6job投入 1h	約21万円
1000万個	6job投入 9h	約210万円
1億個	30job投入 18h	約2100万円
10億個	100job投入 55h	約2億1000万円

現場で汎用的に利用される費用感は200万円程度。そのため、本研究では、250億化合物のプロダクトラン結果を通じて、1000万個スタートのスクリーニングのスクリーニングアルゴリズムの構築を目指す。

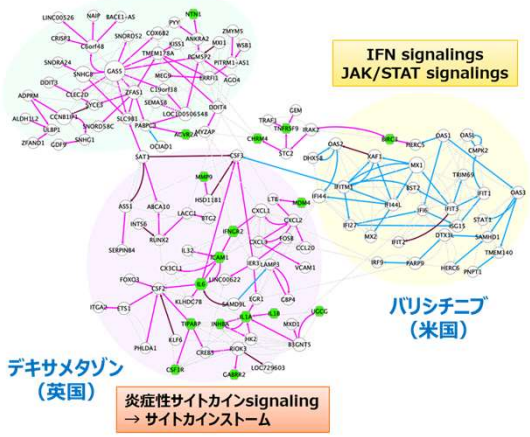
進捗：創薬データベース 「大規模ネットワーク推定」

背景：遺伝子発現データのDBは多数存在するが、個々のサンプルの遺伝子ネットワークDBは存在しない（我々が初めて個別ネットワーク推定に成功）
目標：「富岳」による世界初のネットワークDBを実現

ArrayExpress/GEO ヒト遺伝子発現DB
16,500 セット（30サンプル以上含）



新型コロナの重症化ネットワークの推定と創薬標的分子の推定に成功



- **入力ファイルの準備**
- **前期で144セットの計算を完了（1セット（約100サンプル）のネットワーク化： 256ノード並列で1時間）**

	RNAseq (mRNA)	Microarray (1 color, mRNA)
All studies (human_*_differential_experiments_study_info.txt)	316	1092
All groups (human_*_differential_experiments_group_info.txt)	1169	4875
All samples (human_*_differential_experiments_sample_info.txt)	9953	37746
Studies with disease annotation	241	729
Samples with disease annotations	8282	29689
Unique disease names	152	436
Studies with compound annotation	67	270
Samples with compound annotations	1246	5591
Unique compound names	91	358

参考

計算資源量

R4年度
（2022年度7月～）

項目	NH	備考
HPC/AI創薬PF構築作業　1/2(a)*3/4	37,500	
超大量仮想スクリーニング		1万化合物8NHで算出
検証ドッキング(eMolecule)（b)	1,440,000	
検証ドッキング(DNA encode)　の一部(c)	800,000	
化合物生成　(d)	167,360	
大規模遺伝子ネットワーク推定　(f)	3,712,500	200NH/network*16500/network*3回/2*3/4
合計	**6,157,360**	

R5年度（2023年度)

項目	NH	備考
HPC/AI創薬PF構築作業　1/2(a)	50,000	
超大量仮想スクリーニング		1万化合物8NHで算出
検証ドッキング(DNA encode)*3/4-30万(c)	4,904,400	
大規模遺伝子ネットワーク推定　1/2/(f)	4,950,000	200NH/network*16500/network*3回/2
合計	**9,904,400**	

計算資源量

R6年度
（2024年度 ～６月）

項目	NH	備考
HPC/AI創薬PF構築作業　1/2(a)*1/4	12,500	
超大量仮想スクリーニング		1万化合物8NHで算出
検証ドッキング(DNA encode)*3/4-50万(c)	1，234,800	
大規模遺伝子ネットワーク推定　(f)/2*1/4	1,237,500	200NH/network*16500/network*3回/2*1/4
合計	**2,484,800**	

計算資源量の詳細内訳

項目	NH	備考
超大量仮想スクリーニング		
(b) eMolecule 900万化合物×2異性体*50回	1,440,000	
(c)検証ドッキング(DNA encode)	6,939,200	
DEL（5セット） 1億化合物×2異性体	800,000	
Enamine Real 21億化合物×2異性体	3,360,000	
WuXi 17億化合物×2異性体	2,720,000	
Mcule purchasable 3700万×2異性体	59,200	
(d)化合物生成	167,360	2000万化合物で385NH
DNA encoded ライブラリー（1億化合物）	19,250	385NH×1億/2000万×2異性体×5ライブラリ
Enamine Real 21億	80,850	385NH×5×21億×2異性体
WuXi 17億	65,450	385NH×5×17億×2異性体
eMolecule 900万	385	385NH×5×１億×2異性体
Mcule purchasable 3700万	1,425	385NH×5×0.37×2異性体
(f)大規模遺伝子ネットワーク推定	9,900,000	200NH/network*16500*3回

実測値として以下のデータに基づいて算出
ドッキング：1万化合物8NH, 1億化合物80,000NH、化合物生成：385NH　2000万化合物：大規模遺伝子ネットワーク推定：1 ネットワークあたり 124.78 NH≒200NH

社会実装に向けた計画
（課題開始から3年後までに社会実装を目指すこと）

HPC/AI創薬PF及び創薬DBを利活用するため、民間企業の観点から課題を抽出し解決策を検討することにより、商用で提供されているクラウドサービスやSaaSなどを意識した商用サービス化を視野に入れた体制をLINCを中心に構築する。具体的には、産学コンソーシアムならではの強みを活かし、データや成果、さらに知財の考え方など（下記に検討項目）を、創薬PFをベースに検討し課題解決策を見いだし、業界標準となるようなの考え方につなげ、継続的な運用ができる体制の構築を目指す。

事業化に向けた検討項目：

- 企業が単に利用するだけでなくPF/DB構築に貢献できる仕組み作り（共有財産として利用できるようなデータや知財等の企業からの提供）
- 創薬に関わる全ての企業がPF/DBを必要なときに必要な機能を使える仕組み
- PF上や利用にあたってのデータや成果の共有・公開方法、知財の考え方

年次計画：

2022年度：創薬PF/DB実装・運用における課題・ニーズの抽出（製薬中心）
2023年度：課題解決策の具体化とLINC内コンセンサス
2024年度：テスト運用（テスト終了後、本格運用・公開）
※ 本S5課題利用期間中は、アカデミアも産業目的も無償で公開

本格運用・公開の考え方：

- 民間の商用クラウド上に「富岳」同様の環境を構築し、本格運用・公開
- HPC/AI創薬PF及び創薬DBを、アカデミアは無償利用、産業目的は有償利用（富岳利用料は富岳利用規定に準ず）