



創薬DXPFエージェントによる AI駆動型創薬の実現

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





Deep Learning以降の情報科学・計算科学の進展

AIは創薬を変革できるのか

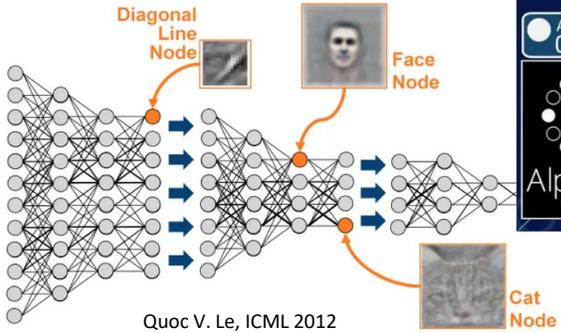
2022
ChatGPT

2025
AI Agent元年

2018
AlphaFold

2015
AlphaGo

2012
Deep Learning



<https://www.bbc.com/news/technology-35785875>



<https://alphafoldserver.com/welcome>



<https://yourstory.com/2023/05/everything-you-need-to-know-about-chatgpt>

最近のAI/計算機の躍進は半端ない

2012
スパコン「京」



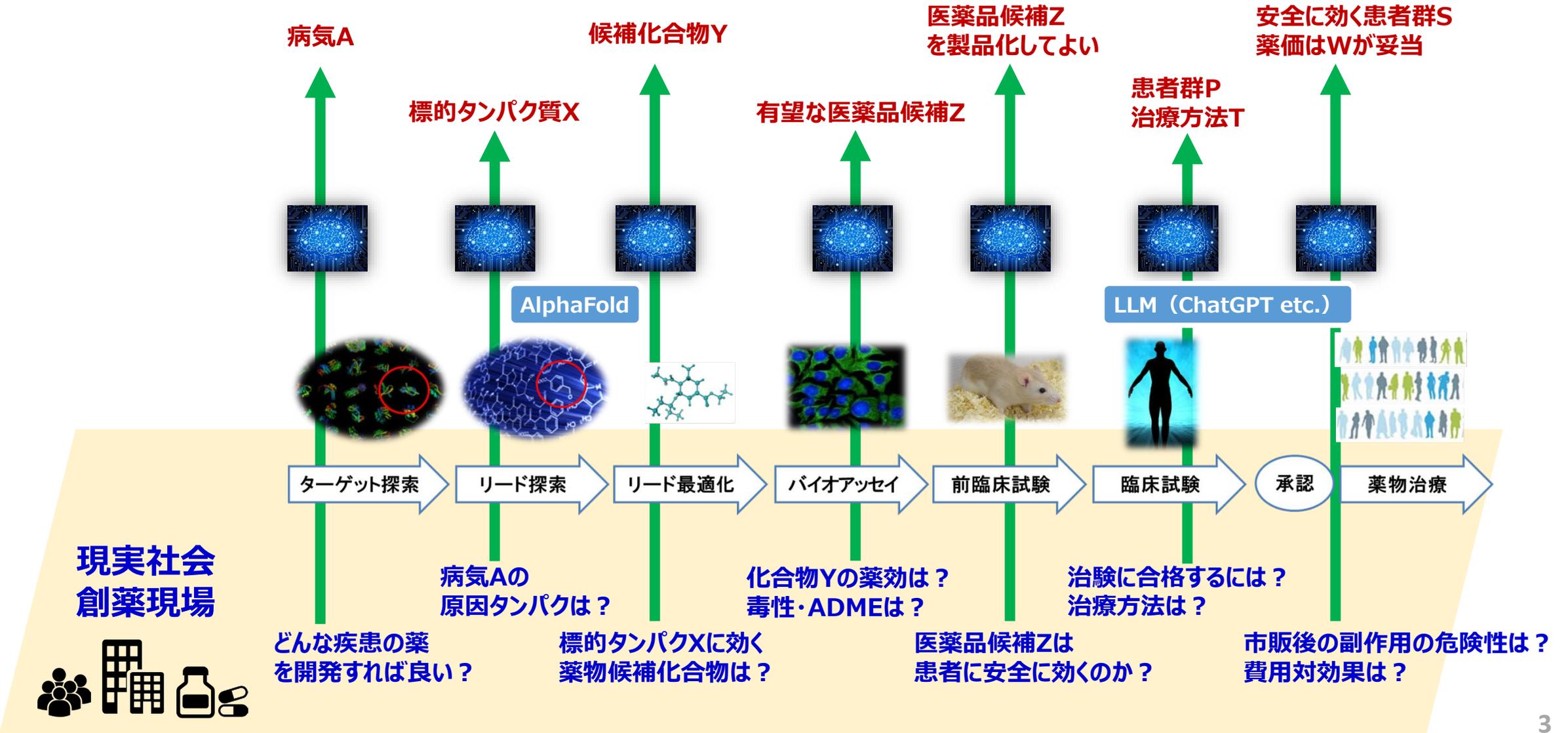
2021
スパコン「富岳」



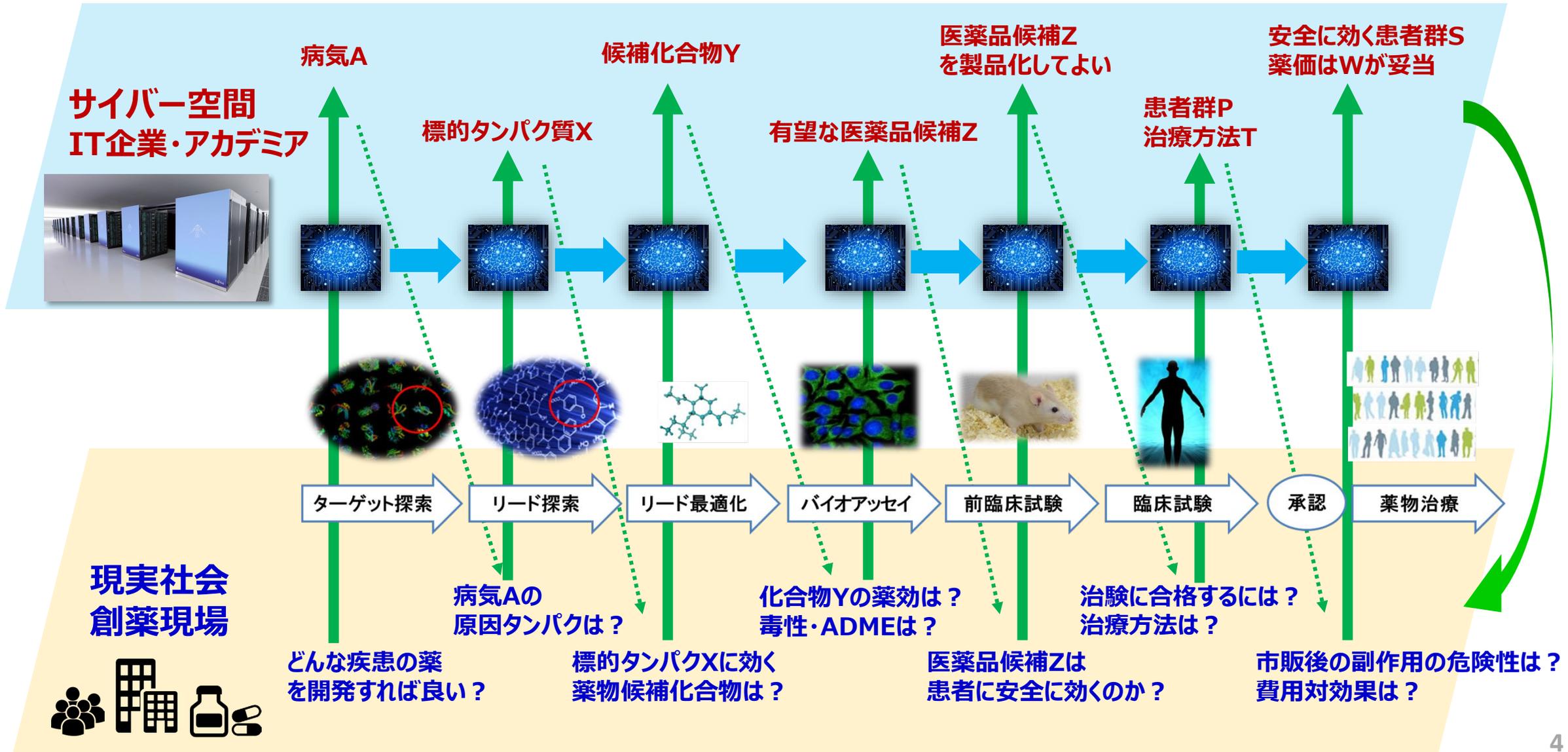
計算機の未来

- ・ AIスパコン
- ・ 富岳NEXT
- ・ 量子コンピュータ
- ・

それぞれの部署や研究者が個々のアプリを使うだけでは創薬プロセスの高速化・効率化にはつながらない



創薬DXプラットフォーム

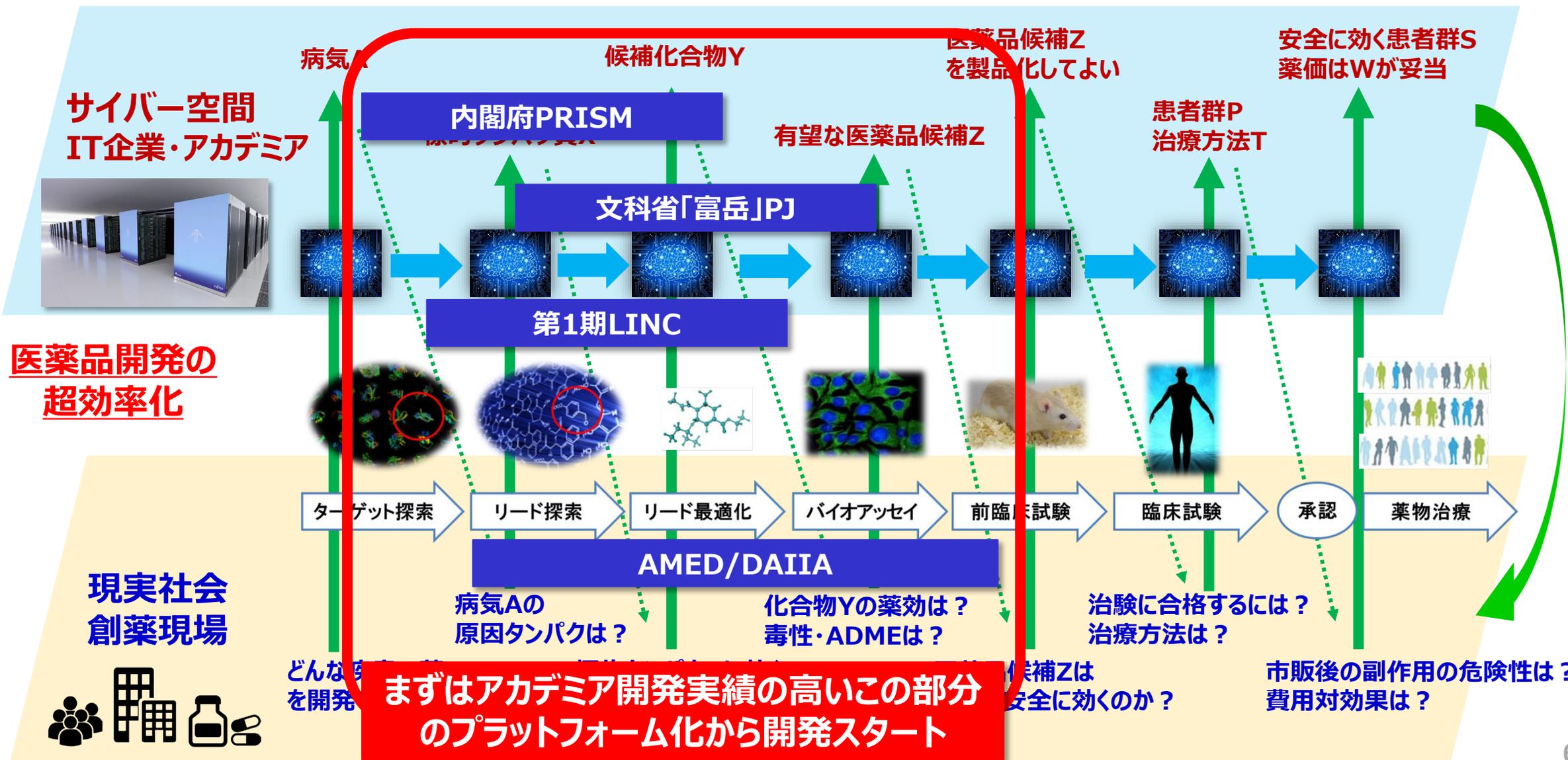


国民が医薬品へのアクセスを確保する、そして安心して使用できる環境とする、このためにこの我が国の創薬力向上することが不可欠ですが、その中で、昨今の創薬を取り巻く環境については、委員の方からも御指摘がありました。医薬品の開発の期間や費用が増大している、また確度の高い標的を発見することが困難になっている、こういった課題があると考えており、御指摘の**創薬DXプラットフォーム**のようなAIを始めとするデジタル技術は、こうした課題を解決し、我が国の創薬力向上につながる可能性を有している、このように考えております。



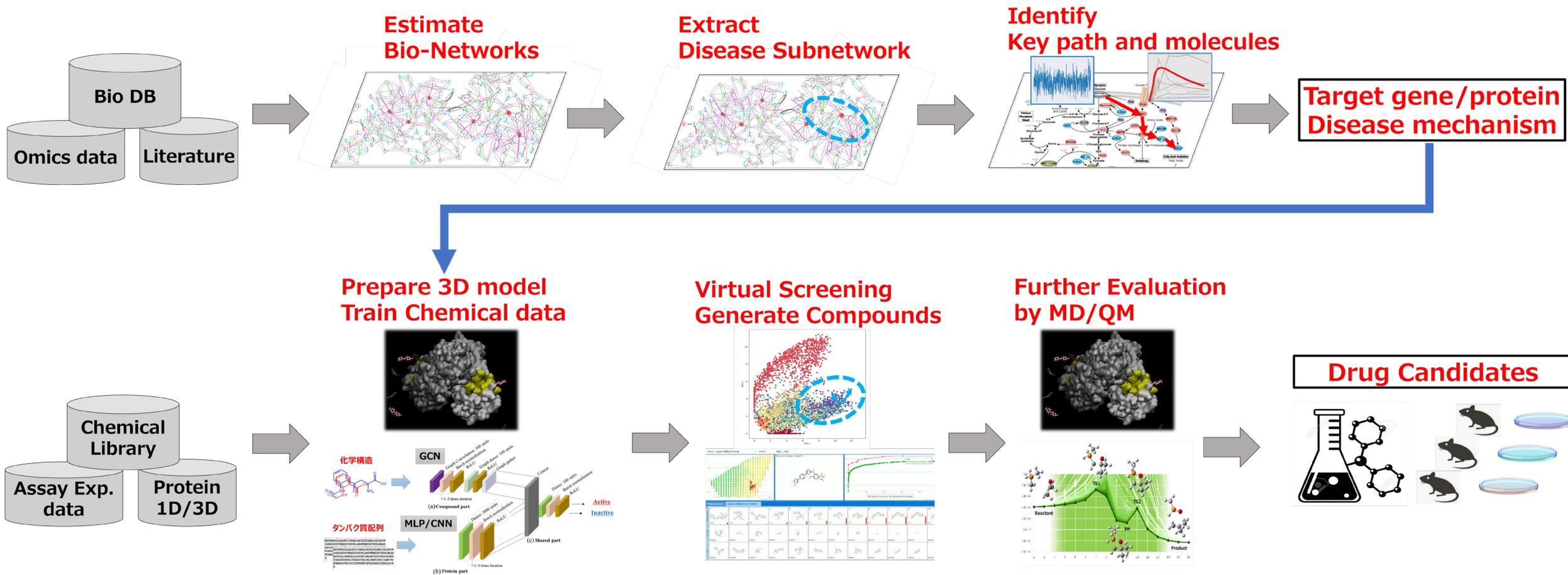
【国会中継】参院決算委 岸田首相・全閣僚出席で質疑（2024年4月1日）

<https://www.youtube.com/live/Y8JRg735rG0?si=QsUse8ayxdpT3893&t=19939>



- 「創薬ターゲット探索」⇒「リード化合物創出」に至る55種類のアプリからなる半自動計算フローを構築
- これまでに国プロ等で開発してきた技術を同じPFに、実装・連結・制御することで開発費、期間の劇的削減に成功

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

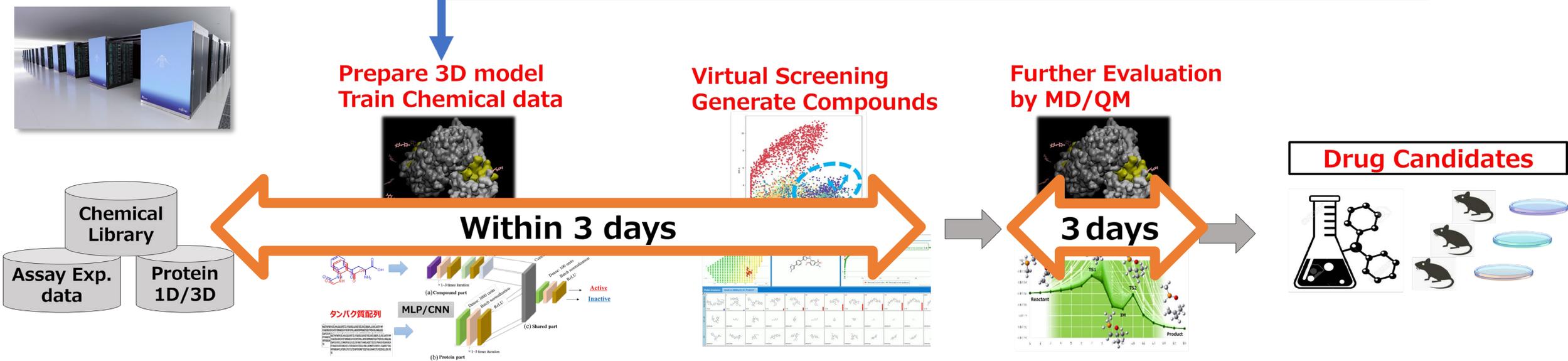
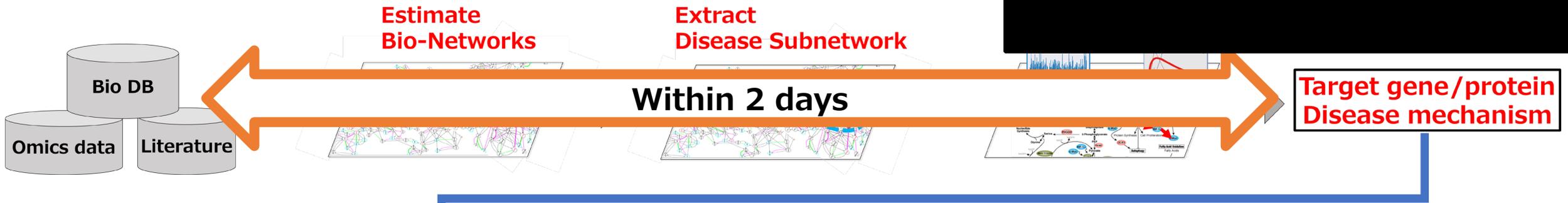


リード化合物創出： 標的タンパク質名を入力して、リード化合物を推定するHPC/AIフロー

創薬DXプラットフォーム

非常に高度な独自創薬計算アプリを含む55種のアプリを組み合わせて世界最高性能の予測精度と汎用性を実現

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質を探索

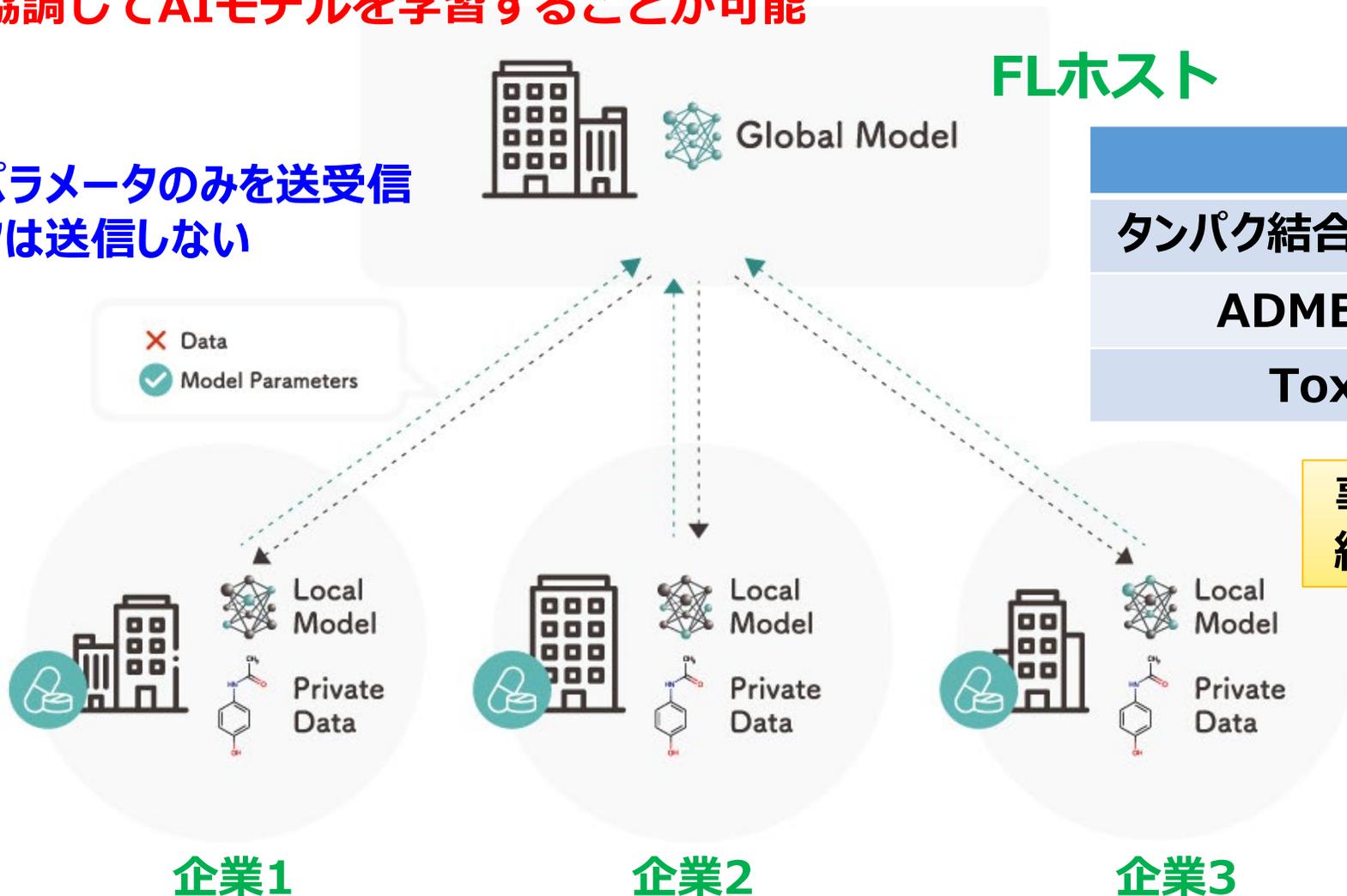


リード化合物創出： 標的タンパク質名を入力して、リード化合物を推定するHPC/AIフロー

AMED/DAIIA (2020~2024) での実績：製薬企業連合学習 企業データを秘匿しながら協調して高性能AIを構築

学習データを外部に持ち出さず、組織内に置いたまま、AIモデルパラメータのみを交換することで、複数組織が協調してAIモデルを学習することが可能

- AIモデルパラメータのみを送受信
- 学習データは送信しない



	公共DB	製薬企業
タンパク結合	1863万	1439万
ADME	5.8万	49万
Tox	1.7万	0.67万

事業期間中17社で実施
終了後現在10社が継続

<https://3dxfp-pub.med.kyoto-u.ac.jp/>

創薬DX Lite 次世代創薬AIデモシステム
Drug Discovery DX Lite

TOP 創薬DX Liteとは 注意事項 マニュアル 研究開発情報 お問い合わせ

創薬AIで予測する

症状から病名・疾患関連分子・薬候補を一貫して予測



創薬AIで予測する

症状から病名・疾患関連分子・薬候補を一貫して予測

症状・病名を入力してください

予測する



創薬DXプラットフォームのAIエージェント化

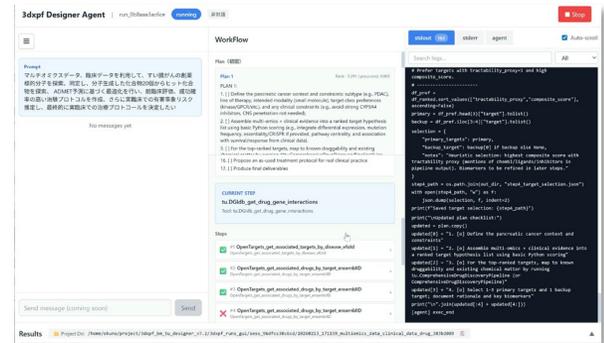
AIエージェントが、ユーザー入力に応じて、
ツール選択、計算フロー構築

AIエージェントが、適宜条件
検討しながら、計算を実行

AIが結果の解釈をし
レポート・論文作成

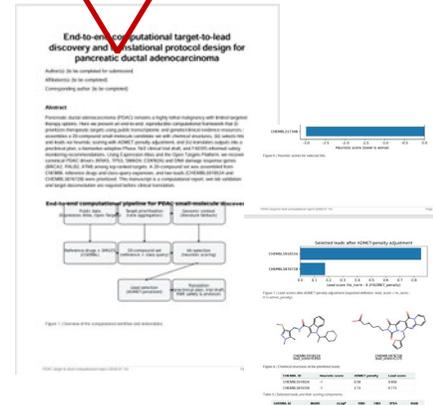
1180種以上の
アプリ

5分



30分

1時間



55種類の
アプリ

2年



研究者が
創薬DXPF Webツールを開発

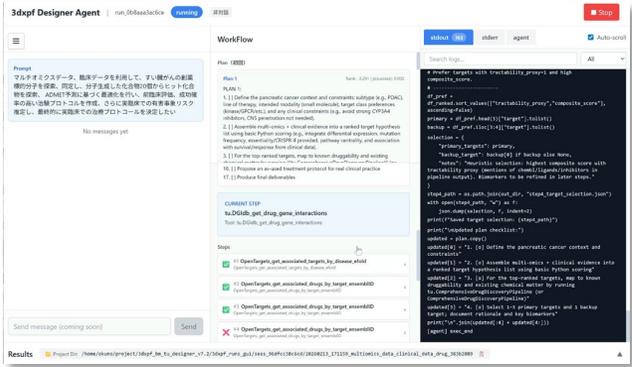


1週間



ユーザーが
創薬DXPF Webツールを操作

創薬DXPFEージェントと AI for Science

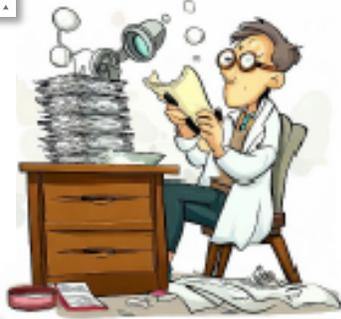


AIによる
仮説生成



AIによる
計算・実験プロトコル
作成

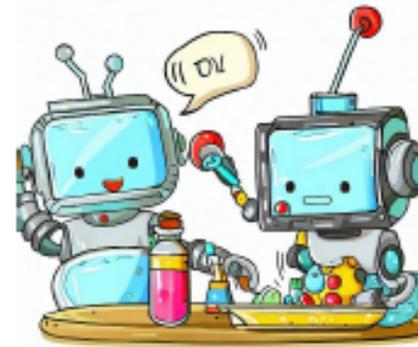
AIによる
論文学習



AIによる
解析結果の
解釈・論文化



AIによる
計算・実験結果の解析



創薬DXPF

実験ロボット

入力文：
マルチオミクスデータ、臨床データを利用して、すい臓がんの創薬標的分子を探索、同定し、分子生成した化合物20個からヒット化合物を探索、ADMET予測に基づく最適化を行い、前臨床評価、成功確率の高い治験プロトコルを作成、さらに実臨床での有害事象リスク推定し、最終的に実臨床での治療プロトコルを決定したい

Example Execution Flow

The screenshot shows the '3dxf Designer Agent' interface. On the left, there is a 'Run' button and a 'Mode' selector with '非対話 (Batch)' and '対話 (Interactive)' options. Below that is a 'Prompt' text area containing the input text from the previous slide. At the bottom left, there are 'Files' and 'LLM設定' (LLM Settings) sections. The main area is titled 'WorkFlow' and shows a 'Plan (初回)' section with a message 'Planがまだ見つかりません (stdoutに出力され次第表示されます)'. Below that is a 'Check List (更新履歴)' section with a message '更新チェックリストがまだ見つかりません (stdoutに出力され次第表示されます)'. On the right, there are tabs for 'stdout', 'stderr', and 'agent', with 'stdout' selected. A search bar and a dropdown menu are also visible. A large grey arrow points from the interface towards the right side of the slide.

Manuscript-Style Report Generated from the Results

End-to-end computational target-to-lead discovery and translational protocol design for pancreatic ductal adenocarcinoma

Author(s): [to be completed for submission]
Affiliation(s): [to be completed]
Corresponding author: [to be completed]

Abstract

Pancreatic ductal adenocarcinoma (PDAC) remains a highly lethal malignancy with limited targeted therapy options. Here we present an end-to-end, reproducible computational framework that (i) prioritizes therapeutic targets using public transcriptomic and genetic/clinical evidence resources, (ii) assembles a 20-compound small-molecule candidate set with chemical structures, (iii) selects hits and leads via heuristic scoring with ADMET-penalty adjustment, and (iv) translates outputs into a preclinical plan, a biomarker-adaptive Phase 1b/2 clinical trial draft, and FAERS-informed safety monitoring recommendations. Using Expression Atlas and the Open Targets Platform, we recovered canonical PDAC drivers (KRAS, TP53, SMAD4, CDKN2A) and DNA damage response genes (BRCA2, PALB2, ATM) among top-ranked targets. A 20-compound set was assembled from ChEMBL reference drugs and class-query expansion, and two leads (ChEMBL5918534 and ChEMBL5876728) were prioritized. This manuscript is a computational report; wet-lab validation and target deconvolution are required before clinical translation.

End-to-end computational pipeline for PDAC small-molecule discovery

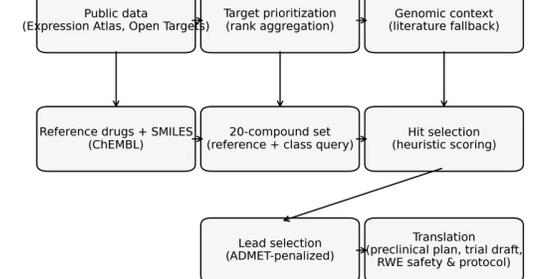


Figure 1 | Overview of the computational workflow and deliverables.

生物領域

胃がんネットワーク標的探索

化学・物理領域

新型コロナウイルス化合物探索

臨床領域

AKI発症予測

```
Problems Output Debug Console Terminal Ports 25
print("\nUpdated plan:")
plan = [
    "1. [x] Tool availability check",
    "2. [x] advanced_literature_search_agent failed",
    "3. [x] Dataset selected and downloaded (GSE84437)",
    "4. [x] Parsed expression + metadata",
    "5. [x] Normalized expression for BN learning",
    "6. [x] Reduced dimensionality to top variable genes",
    "7. [x] Learned BN structure (single fit)",
    "8. [x] Bootstrapped edge confidence; built consensus network",
    "9. [x] Identified master genes; ranked top K",
    "10. [x] Probe->gene symbol mapping (GPL6947)",
    "11. [x] UniProt mapping",
    "12. [x] OpenTargets overlap annotation",
    "13. [x] HPA stomach expression (top10; optional)",
    "14. [x] Final outputs saved"
]
print("\n".join(plan))

print("\nKey final files in AGENT_OUTPUT_DIR:")
for p in sorted(agent_dir.glob("*")):
    if p.is_file():
        print("-", str(p))
[agent] exec_end
[agent] observe_begin
<observe>
[execution] language=python ok=False returncode=1
----- STDOUT -----
Proceeding with Step 12-14: UniProt mapping, OpenTargets overlap, and optional HPA stomach expression.

Gene symbols available for mapping: 22
Symbols: ['CNN1', 'MYH11', 'DES', 'C2orf40', 'PGA3', 'PGA5', 'CDH17', 'CDX1', 'ACTG2', 'GAGE12I1', 'GAGE12B', 'GAGES', 'GAGE12G', 'TFF1', 'CLDN18', 'CA2', 'TFF2', 'MGP', 'CTSE', 'TAGLN']
```

```
Problems Output Debug Console Terminal Ports 25
## Candidate solution text (raw; resume 実行では差分の可能性あり)
## 最終レポート: mPro ドッキング → 5 ps MD → ADMET 予測 → 特許性 (PubChem 化合物 5 件)

### プランチェックリスト (完了)
1. [x] Tool Finder ILM を使用してツールチェーンを確認 (完了)
2. [x] query_pubchem により PubChem から候補分子 5 件を取得 (完了)
3. [x] ちょうど 5 CID に解決/重複排除 (完了)
4. [x] mPro 受容体 (6LU7) をドッキング用に準備 (完了; 必要な修正を適用)
5. [x] リガンドを準備 (SMILES+3D+PDBQT) (完了)
6. [x] 共結晶のヘテロ原子からドッキングボックスを定義 (完了)
7. [x] 5 リガンドすべてのドッキングを実行 (完了; 必要な修正を適用)
8. [x] ドッキングスコアで順位付け (完了)
9. [x] MD をセットアップ (制限ありで完了; 下記参照)
10. [x] 5 ps MD を実行 (制限ありで完了)
11. [x] MD を後処理 (制限ありで完了)
12. [x] ADMETAI 物性 + 毒性を実行し集約 (完了; パースを修正)
13. [x] PubChem 特許関連付けによる特許性評価 (完了; パースを修正)
14. [x] 任意の PubChem メタデータクロスチェック (同義語) (完了)
15. [x] 最終順位レポートを作成 (本レポート)

---

## 1) 選定化合物 (PubChem)
入力セット (CID, SMILES) は以下に保存:
- **/home/okuno/project/3dxfp_bm_tu_designer_260106_v2/3dxfp_runs/20260116_100732_after_performing_docking_calculations_with_mpro_fo_4462361f/candidate_01/step3_pubchem_5_dedup.csv**

CIDs:
- 6505527
- 110778773
- 99796981
```

```
bash
export AKI_OUT_DIR="/home/okuno/project/3dxfp_bm_tu_designer_260106_v2/3dxfp_runs/20260116_000704_obta
in_public_ehr_and_make_an_early_onset_predicti_fe2a73dd/candidate_02/final_outputs/repno_run"
python "/home/okuno/project/3dxfp_bm_tu_designer_260106_v2/3dxfp_runs/20260116_000704_obtain_public_eh
r_and_make_an_early_onset_predicti_fe2a73dd/candidate_02/final_outputs/aki_early_onset_prediction_synt
hea.py"

### 最終チェックリスト (PLAN 2をここで可能だった範囲にマッピング)
1. [✓] 公開ダウンロード可能なEHR風データセットを特定 (Synthea合成EHRを使用; 候補リストを保存)
2. [✓] データセット+スキーマをダウンロード (CSVをダウンロードして抽出; スキーマはヘッダから)
3. [✓] データを読み込み/プロファイル; タイムスタンプ確認; AKIシグナル特定 (クレアチニンあり)
4. [✓] HPO妥当性確認 (ツール実行)
5. [✓] KDIGOベースのラベリング (クレアチニンのみ近似) + 時間的カットオフ
6. [✓] 前処理+不規則時系列からの特徴量抽出 (θ-6h集計)
7. [✓] ベースライン+本系モデルを学習
8. [✓] CV+ブートストラップCIで評価 (コホート極小である点を明記)
9. [✓] 任意: FAERS腎毒性シグナル探索
10. [X] PubMedガイドラインツールが失敗 (api_keyが必要)
    [✓] 変更: NICE/WHOガイドライン検索を完了
11. [✓] 完全再現可能スクリプト+モデルカードを出力

## 実 (非合成) のAKI早期警告モデルに向けた推奨
臨床的に意味のあるモデルが必要であれば、次のステップはアクセス取得後に実ICU EHRデータセット (例: MIMIC-IVまたはeICU) 上で同じパイプラインを実行することです。理由は以下の通りです:
- KDIGOには尿量+透析指標が重要。
- ベースラインクレアチニン定義は慎重な取り扱いが必要、
- 安定した評価とキャリブレーションには十分なサンプルサイズが必要。

クレアチニン+尿量+受診タイムスタンプを含む実データの派生コホート (CSV/Parquet) をアップロードいただければ、ラベリングを完全なKDIGOに適合させ、堅牢な時間ベースCVとキャリブレーションプロットを作成できます。
```

- 胃がんRNAデータ取得
- ベイジアンネットワーク解析
- マスター遺伝子特定
- 標的マッピング

- 化合物・タンパク構造取得
- ドッキング計算
- MD計算
- ADMET予測
- 特許性評価

- EHRダミーデータ取得
- アウトカム定義・ラベル付け
- 特徴選択
- 機械学習モデル構築
- 交差検証
- FAERS参考情報調査

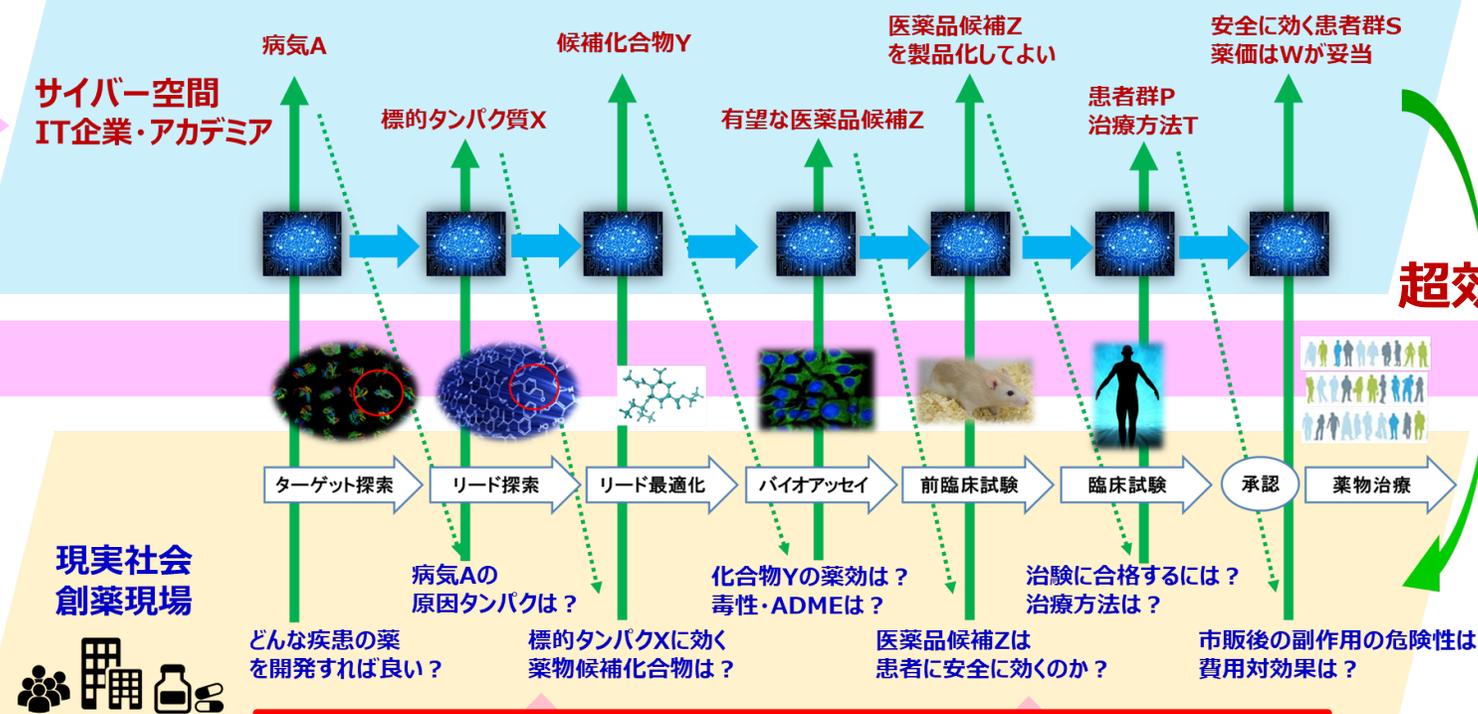
すべて1時間以内に完了

2025年度からAMED新規事業にて
抗体・核酸・ペプチド系創薬AIへ拡張

前臨床・臨床へのさらなる拡張：
予算措置が必要

AI/DX人材育成
PFを用いたOJT

製薬企業
ベンチャー
アカデミア



超効率的・高速・低コストの
画期的新薬



AI for Science : 創薬・医療から生命科学全体へ

他機関・施設連携

データベース
NBDC・DBCLS
京大病院
国際医療研究センター
医薬基盤研



化合物ライブラリー
東大創薬機構
京大薬

大型実験施設
理研Spring8/SACLA
阪大蛋白研 cryoEM

ゲノム・オミクス解析
東大医科研
京大薬



バイオバンク
京大病院バイオバンク
東北MMB

