

CBI学会2021年大会 企業セッション SS-13

# 中外製薬の創薬研究におけるAWS活用事例紹介

中外製薬株式会社

創薬化学研究部 荒川 晶彦

創薬基盤研究部 角崎 太郎, 西藤 ゆかり

2021年10月26日

# Agenda

01

中外製薬のDXに関する取り組み

02

創薬計算化学での活用事例紹介

03

Bioinformaticsでの活用事例紹介

# 中外製薬の概要

## ◆ がん・バイオに強みを持つ、研究開発型製薬企業

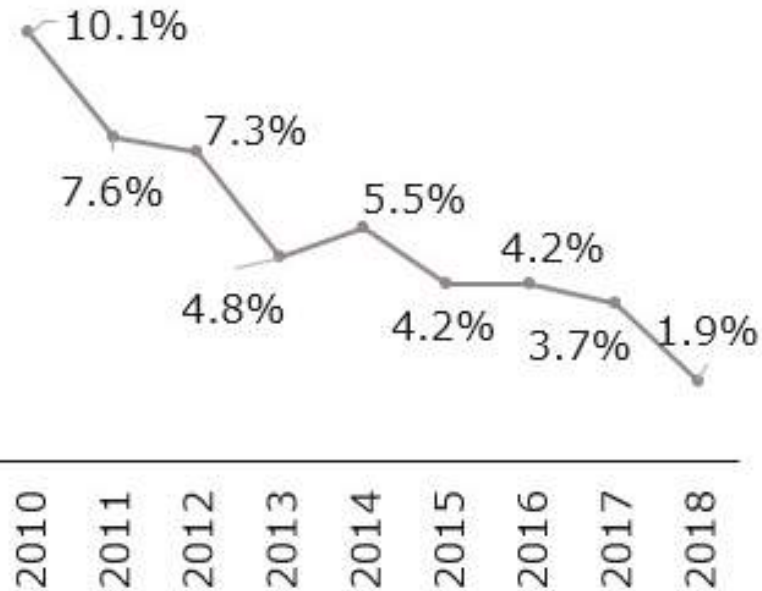
- 医療用医薬品メーカーとして日本トップクラス
  - 売上収益：7,869億円、営業利益：3,012億円
    - がん領域売上シェア 第1位 【2020年度決算ベース】
  - 時価総額：7兆2,334億円（第1位） 【2021年8月末時点】
- 独自の創薬技術力
  - 国産初の抗体医薬を創製。抗体・中分子等で世界最先端の技術力
- ユニークなビジネスモデル
  - 戦略パートナーであるロシュ社が株式59.89%を保有
  - 独立した上場企業として自主的経営を実行



# 新薬創出におけるAI活用への期待

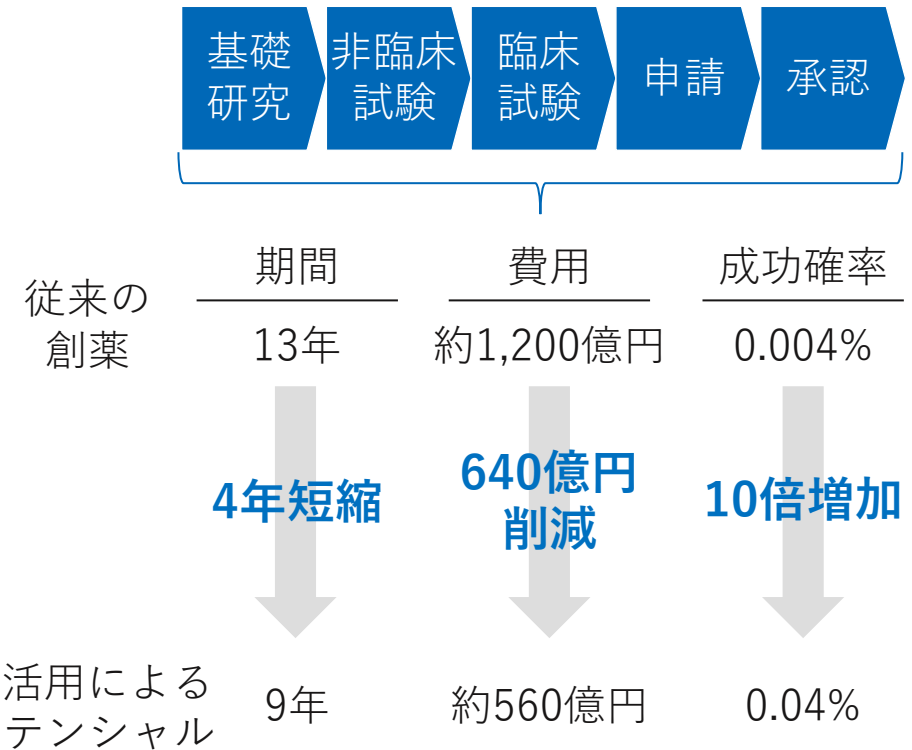
- ◆ 医薬品のR&D生産性は年々下落しており、これまでできなかったことができるようになる”デジタル”を活用した創薬に取り組む必要性が高まっている

メガファーマのR&D生産性 (IRR) \*1



2010年以降下落しており、直近では1.9%と最低値を記録している

創薬におけるAI活用のポテンシャル\*2



\*1: Deloitte "Unlocking R&D productivity Measuring the return from pharmaceutical innovation 2018" (グローバル大手12社を対象)

\*2: 「第3回 保健医療分野におけるAI活用推進懇談会(平成29年3月) 奥野構成員提出資料」 (厚生労働省) を加工して作成

<https://www.mhlw.go.jp/file/05-Shingikai-10601000-Daijinkanboukouseikagakuka-Kouseikagakuka/0000154209.pdf> (2021/3/12アクセス)

デジタル技術によって中外製薬のビジネスを革新し、  
社会を変えるヘルスケアソリューションを提供する  
トピックイノベーターになる

## “ビジネスを革新する”とは

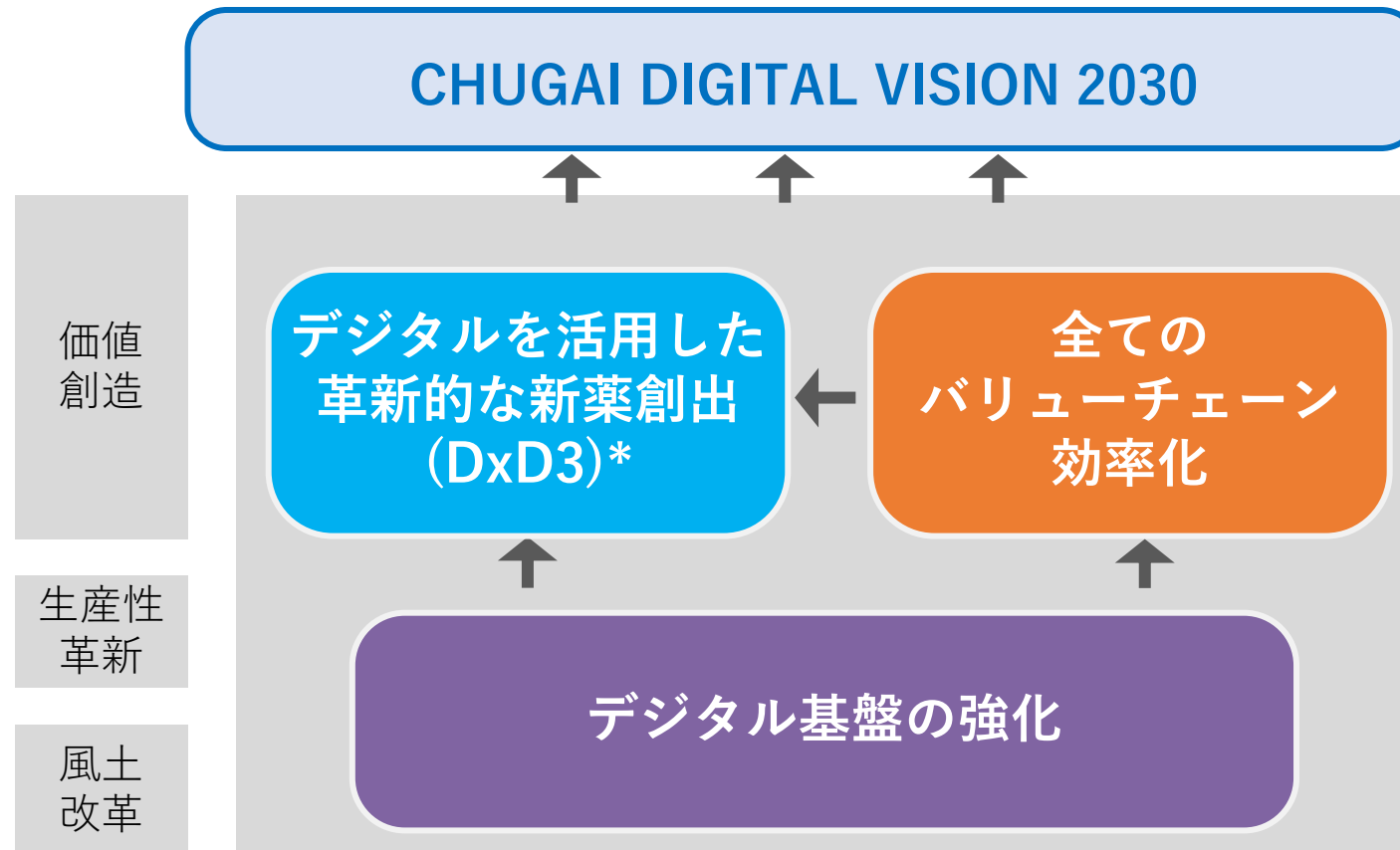
- ◆ デジタルを活用した**革新的医薬品**の継続的な提供を実現する
- ◆ 全ての**バリューチェーンの大幅な効率化**を実現する
- ◆ **革新的なサービスが提供**できる体制を構築する
- ◆ 中外製薬の**社員の意識、組織・風土を変える**

## “社会を変える”とは

- ◆ 個々人に寄り添った**最適な個別化医療の提供**を実現する
- ◆ 超早期診断・予防・治癒の実現による**ライフタイムを通じた高いQOLを実現**する
- ◆ 人口減少、少子高齢化の社会でも**sustainableな社会保障制度を実現**する

# 中外製薬の3つの基本戦略

- ◆ デジタル基盤を強化し、全てのバリューチェーンを改革することで圧倒的に生産性・効率性を向上
- ◆ デジタルを活用することにより革新的な新薬を創出し、社会を変えるヘルスケアソリューションを提供する。

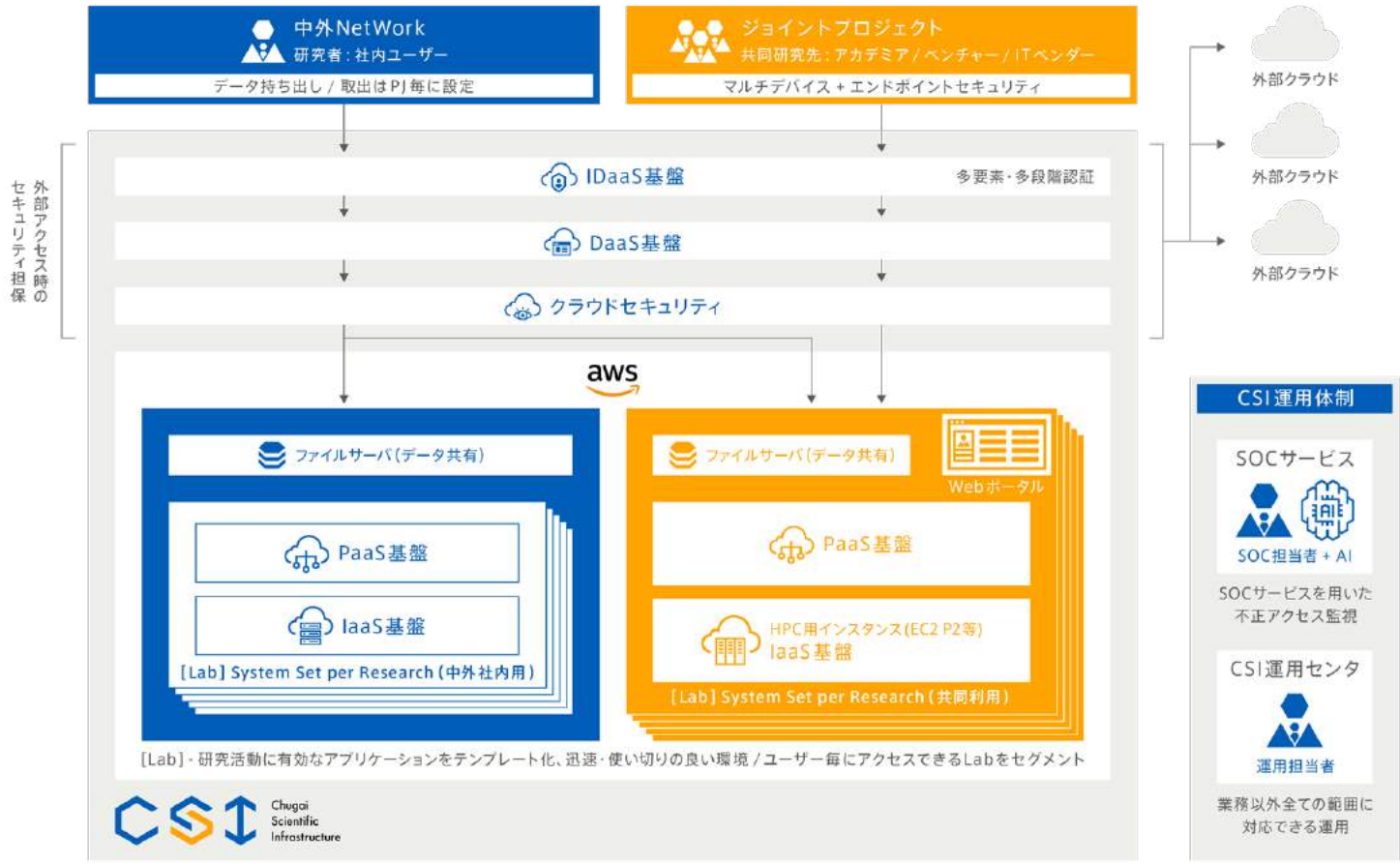


# 中外製薬DXの2030の絵姿



# CSI(Chugai Scientific Infrastructure)の構築

- ◆ 全社データ利活用の推進を目的とした、大容量のデータを安全に利用、移動、保管するためのIT基盤



- 社内データの部門横断的な活用の促進
- アカデミアや病院との共同研究プロジェクトを迅速に推進する研究環境の提供
- 作業の共通化・自動化による環境構築コスト削減(1/10)・期間の短縮
- ゲノムデータ等の高いセキュリティが要求されるデータの安全な取扱
- 情報流出、外部攻撃等の安全性リスク低減

# Agenda

01

中外製薬のDXに関する取り組み

02

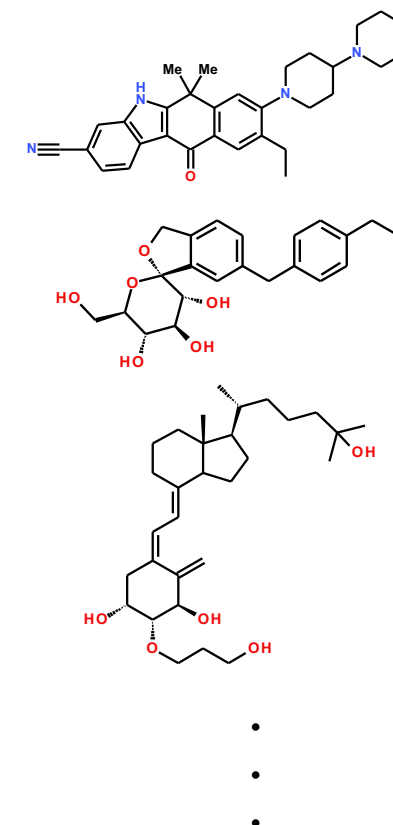
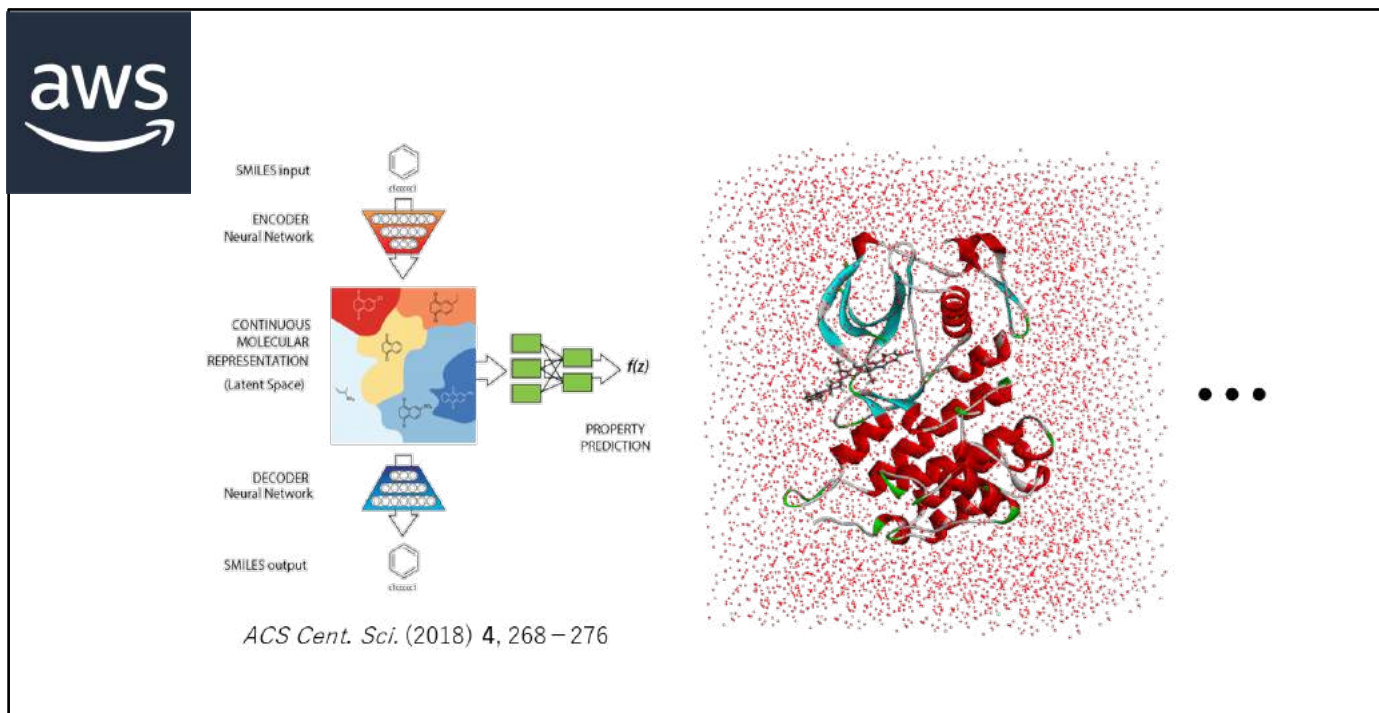
創薬計算化学での活用事例紹介

03

Bioinformaticsでの活用事例紹介

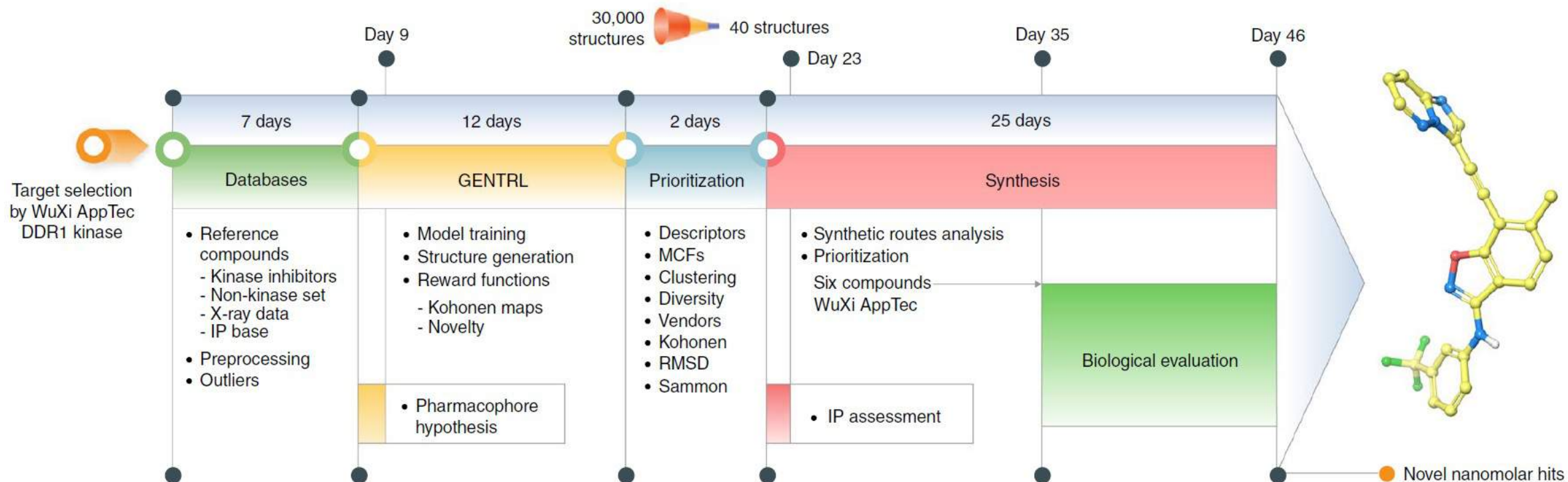
# Overview

近年創薬への適用が普及してきたAIや分子シミュレーションを大規模に実行して有望化合物を提案したい。その計算機資源としてAWSを活用している。



# 創薬計算化学の動向 1 / 2

AIを活用した新規Hit/Lead創出に関する事例が報告されている。

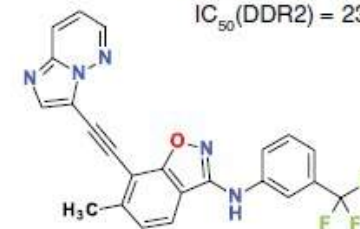


● Novel nanomolar hits

$IC_{50}(\text{DDR1}) = 10 \text{ nM}$

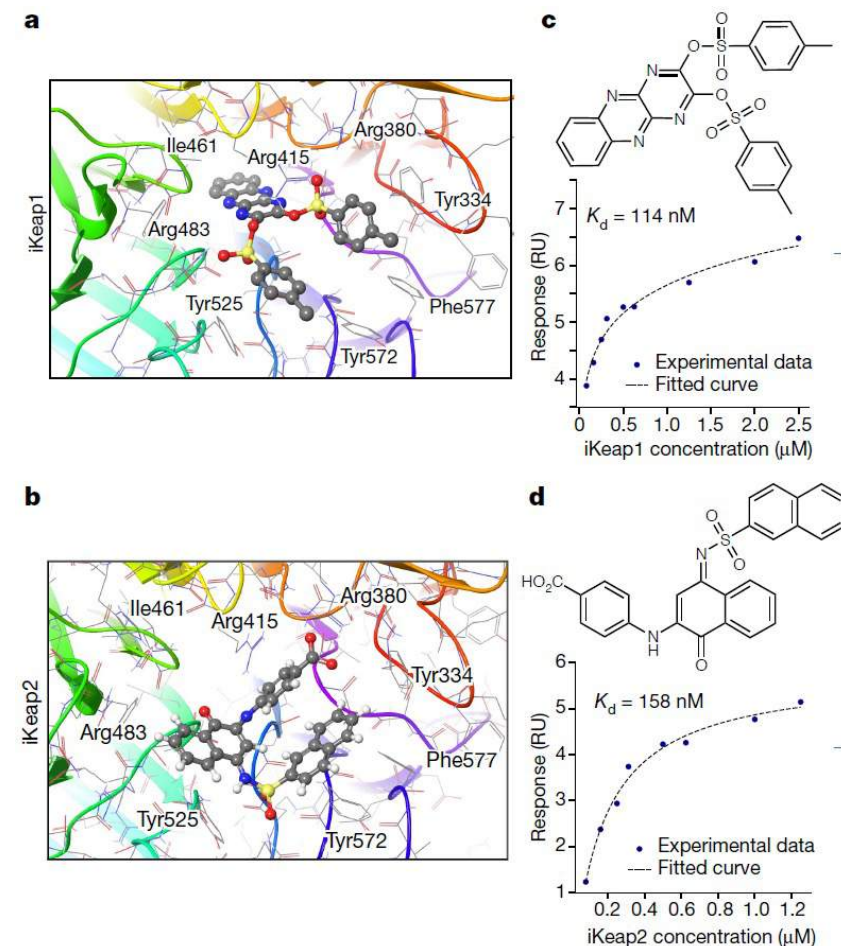
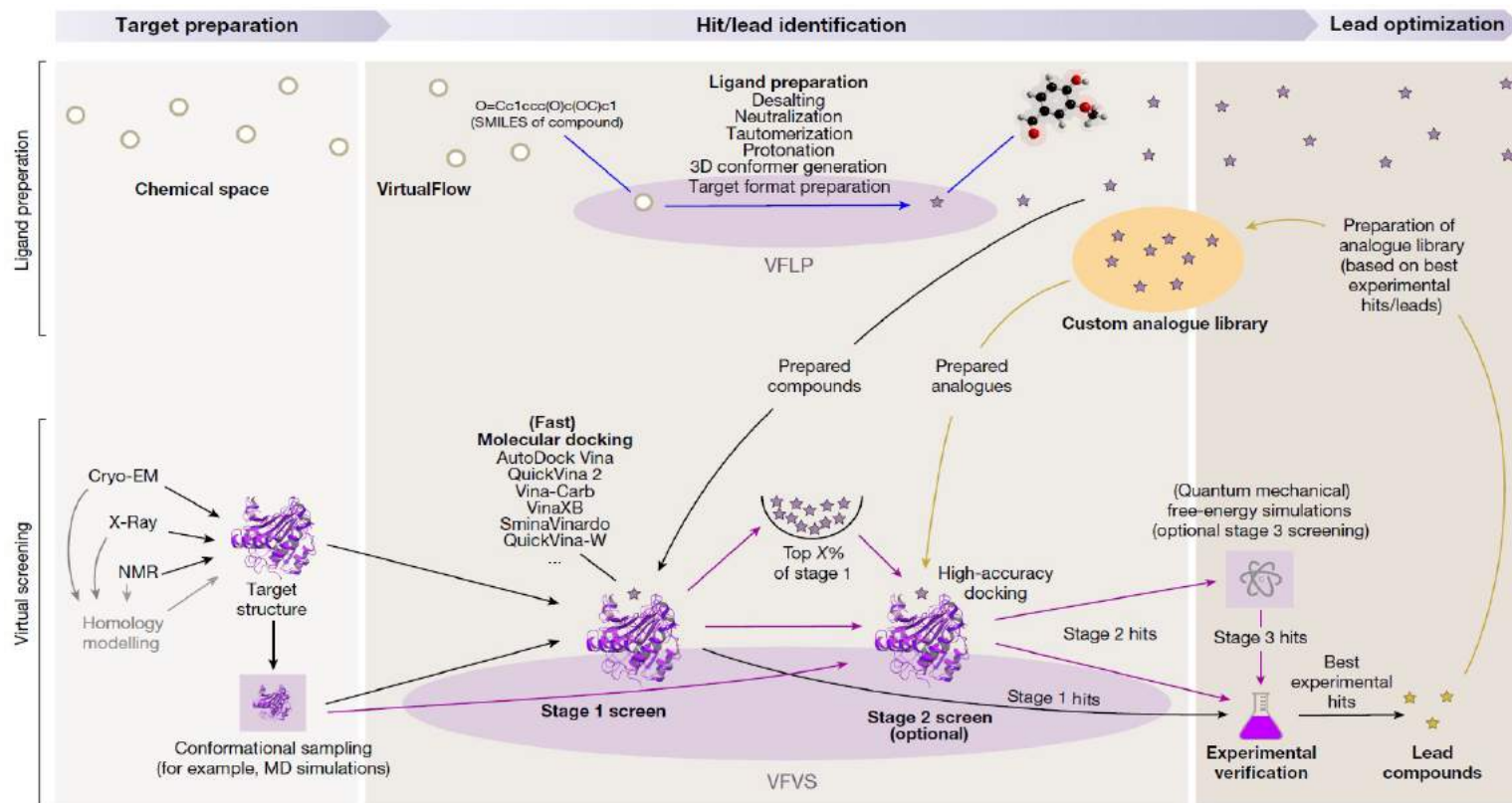
$IC_{50}(\text{DDR2}) = 234 \text{ nM}$

*Nature Biotechnology* (2019) **37**, 1038–1040



## 創薬計算化学の動向 2 / 2

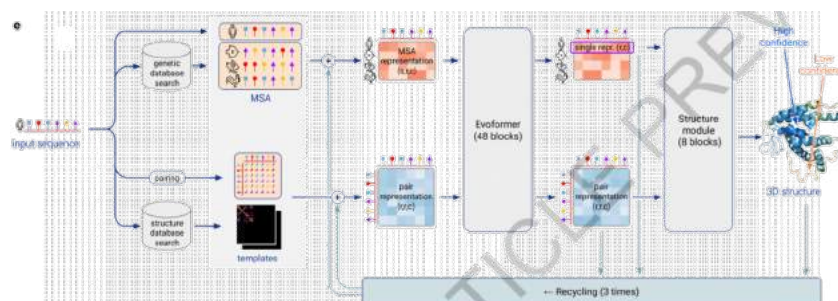
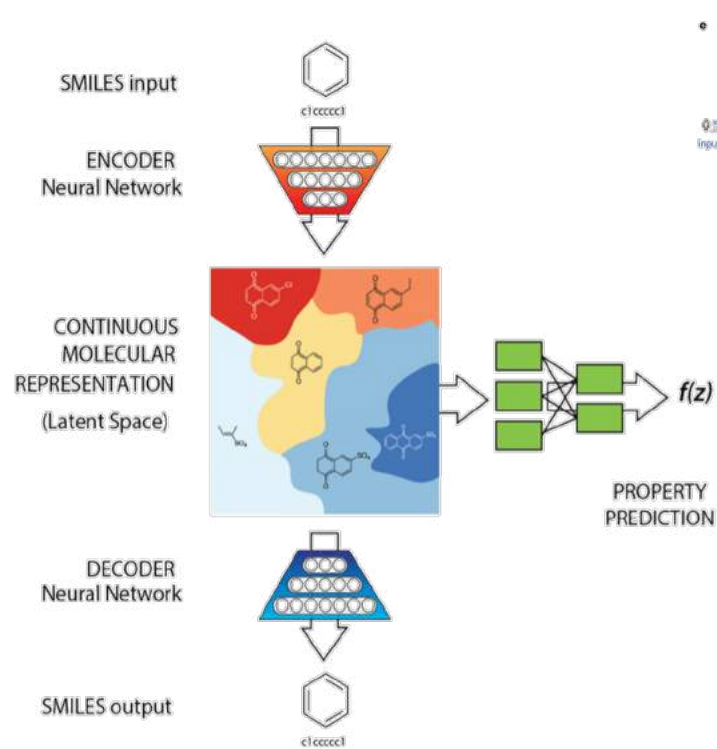
分子シミュレーションからのHit/Lead同定も報告されている。



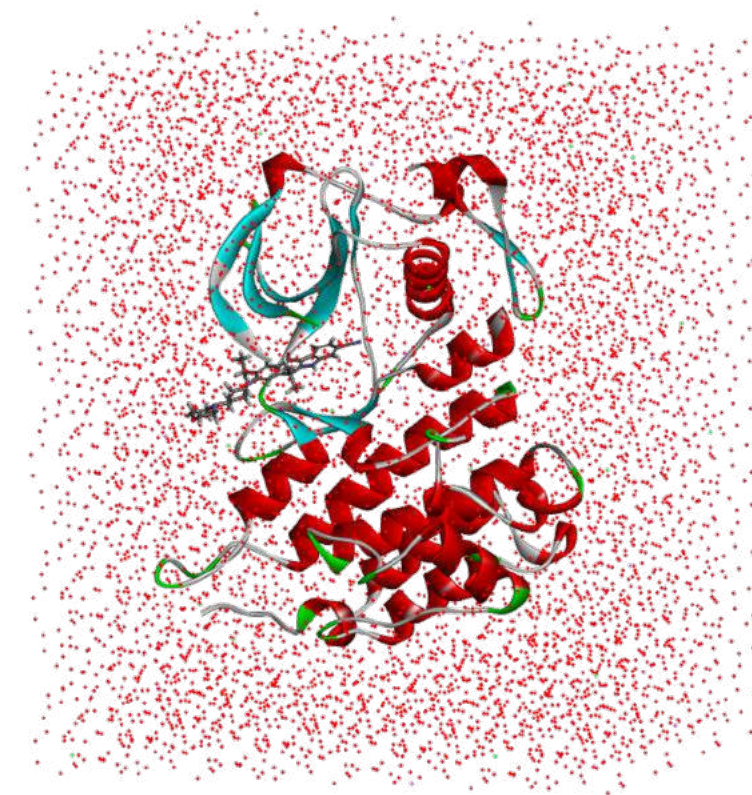
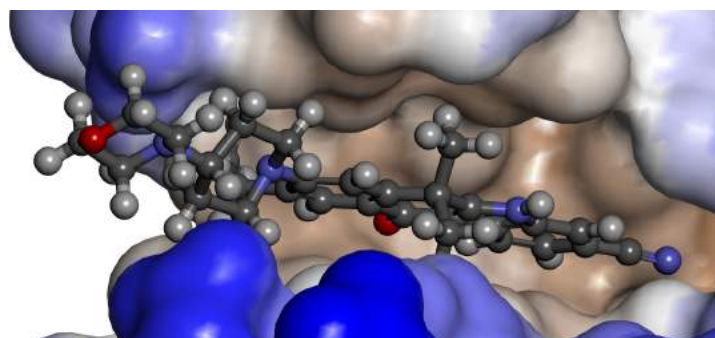
*Nature* (2020) **580**, 663–668.

# 社内で計算化学を活用した創薬を実施するためには？

AIや分子シミュレーション関連の高負荷計算を大規模に実行できる環境が必要である。



*Nature* (2021) 596, 583–589

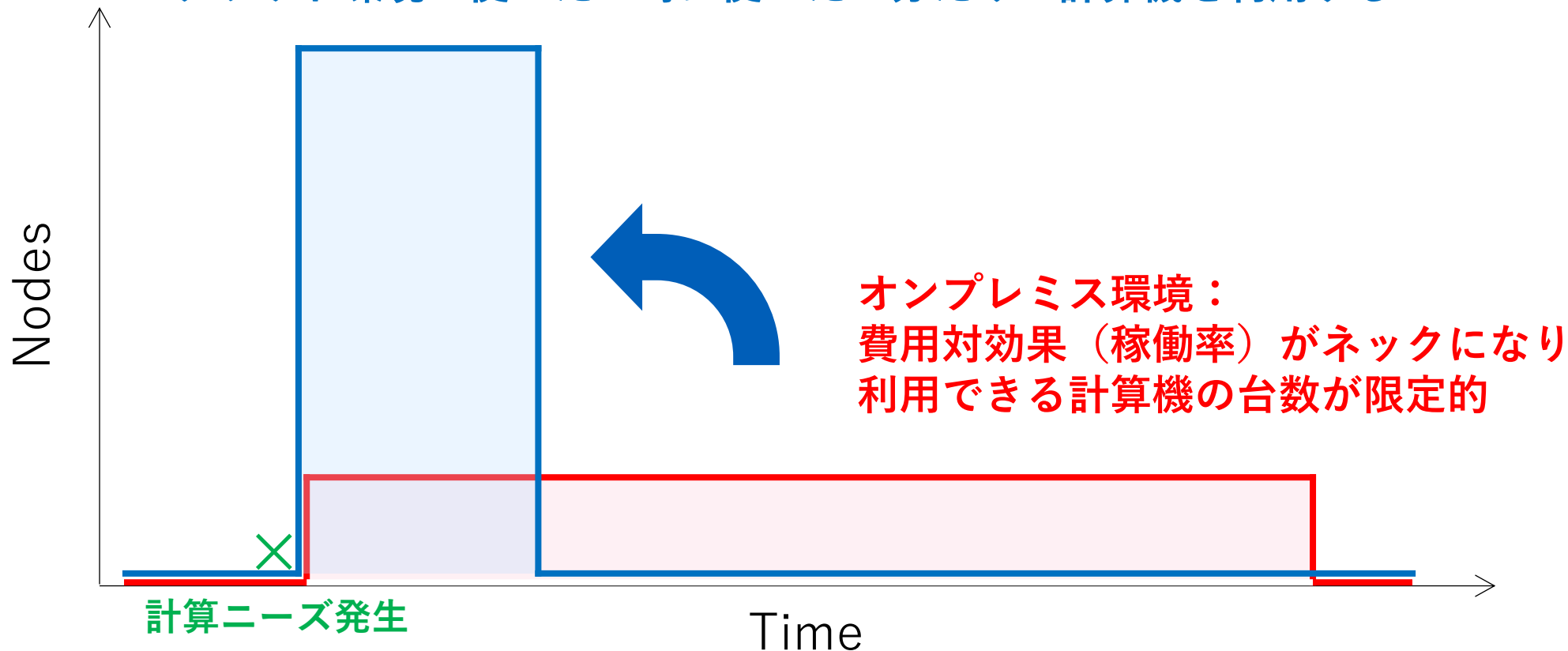


*ACS Cent. Sci.* (2018) 4, 268 – 276

# クラウドを利用する理由

突発的な計算ニーズに対して大規模計算を短時間で実行したい。

クラウド環境：使いたい時に使いたい分だけの計算機を利用する



# AWSを採用した理由

社内外で十分な利用実績があるAWSが第一選択肢になった。

## 社内要因

- ✓ 社内の既存システムとの連携
- ✓ AWS担当者との定期的な意見交換  
→ 構想段階からの相談  
ベンダーを交えた打ち合わせ

## 外部要因

- ✓ 豊富な外部の情報ソース
- ✓ 豊富なベンダー候補
- ✓ 計算機リソース・サービスの種類が潤沢
- ✓ 創薬関連ソフトウェアの稼働実績

# 創薬計算化学用クラウド環境の構築に向けて

低コストでスケーラブルな計算環境を実現したい。

## 要件

- ✓ 必要な種類・数の計算機の確保
- ✓ 利用コストの抑制
- ✓ オートスケーリング
- ✓ 既存環境と同じ計算環境

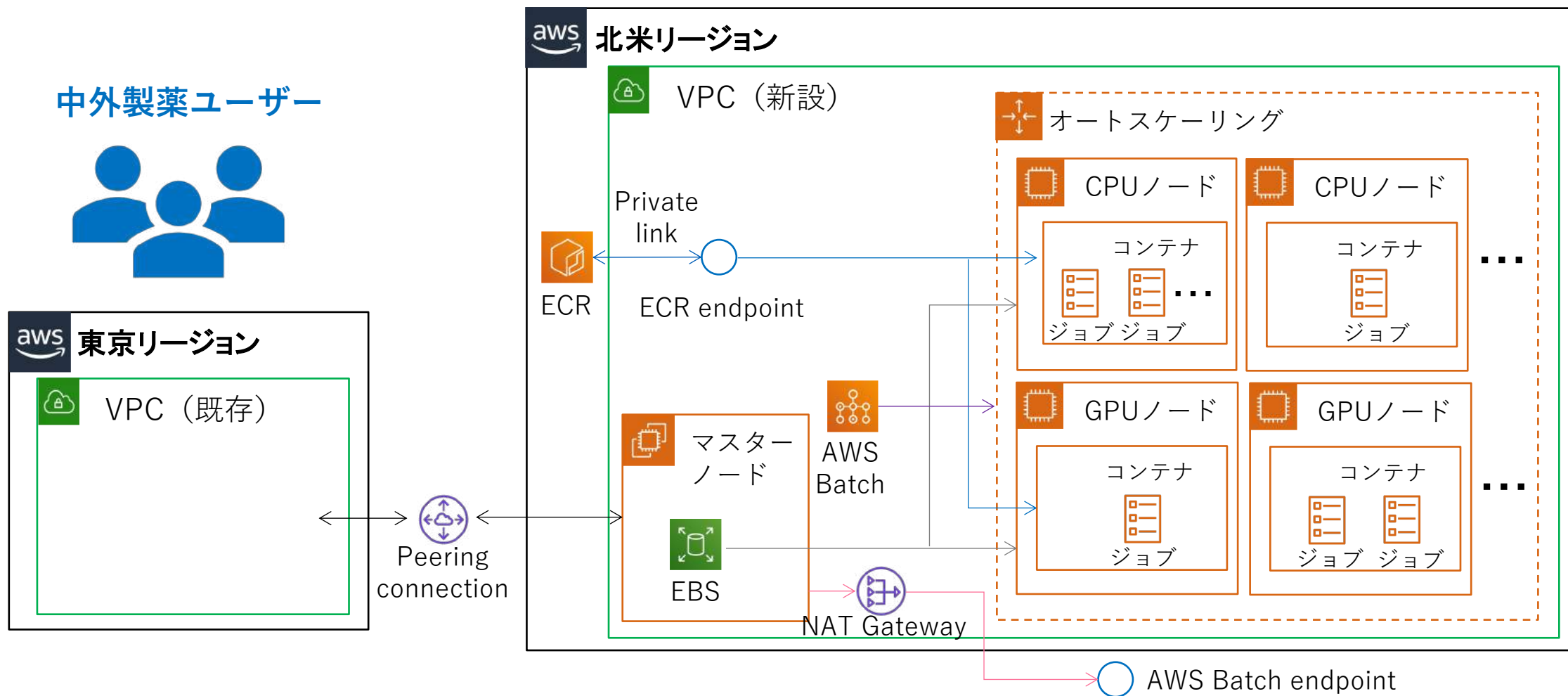


## 解決策

- 北米リージョンに環境構築
- スポットインスタンスを積極利用
- AWS Batchを採用
- コンテナ型仮想化

# 創薬計算化学用クラウド環境

AWS Batchを利用した計算化学用クラウド環境を構築した。



# 今回構築したクラウド環境の使用感について

いくつかの懸念点が想定されたものの実際に利用してみて大きな問題に直面することはなかった。

## 利用前に想定された懸念点

- ✓ 北米リージョンとのデータ通信の遅延
- ✓ インスタンス稼働までの待機時間
- ✓ ジョブ管理の動作・使いやすさ
- ✓ 1ノードに多数のジョブを投入した際の性能
- ・
- ・
- ・

## 利用してみた印象

- 計算時間と比較して律速にならない
- 同時投入時に最大15分の待機時間あり
- 期待通り
- メモリ超過等なければ問題なし
- ・
- ・
- ・

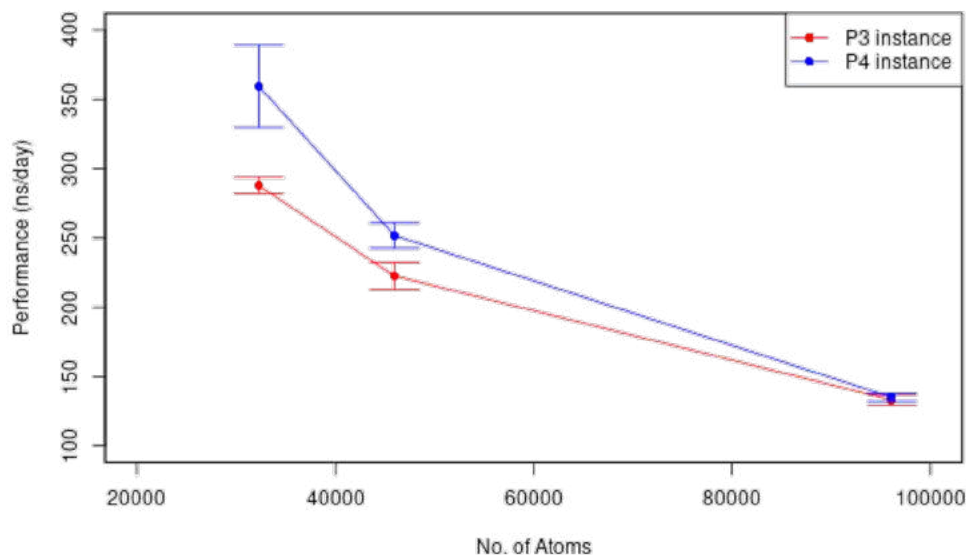
# 今後の展望

計算機性能の向上は現状の計算化学の限界を打破する足がかりになると考えている。  
最適な計算機環境を構築して創薬に貢献していきたい。

## 最新の計算機にアクセスできる環境

e.g. P3インスタンス → P4インスタンス  
Tesla V100                      Tesla A100

MDシミュレーションでの性能比較



## 社内に導入できないような計算機の活用

e.g. 量子コンピュータ

### 量子デバイス (QPU)

- ゲートベース量子コンピュータ
  - Rigetti : 超伝導量子ビット (The Rigetti Aspen-9)
  - 32量子ビット、結合は部分的
  - IonQ : イオントラップ量子ビット (IonQ linear trap)
  - 11量子ビット、全結合が特徴
- 量子アニーリング
  - D-Wave : 超伝導量子ビット
  - DW\_2000Q\_6 (2048量子ビット)
  - DW\_Advantage (5760量子ビット)

デバイスの開発・構築実行は非常に高度な技術が必要とする上、運用も高コスト。クラウドベースのオンデマンドに適している。

© 2021 Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.



<https://aws.amazon.com/braket/hardware-providers/>



AWS様提供資料

# Agenda

01

中外製薬のDXに関する取り組み

02

創薬計算化学での活用事例紹介

03

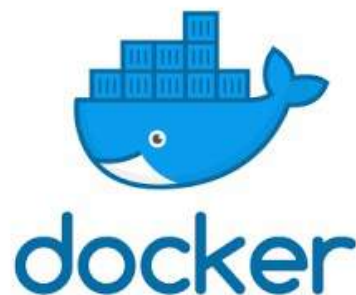
**Bioinformaticsでの活用事例紹介**

# Overview

- クラウドベースのNGS解析プラットフォームを構築中.



- 本システムを支える要素技術は, 以下の3つ.

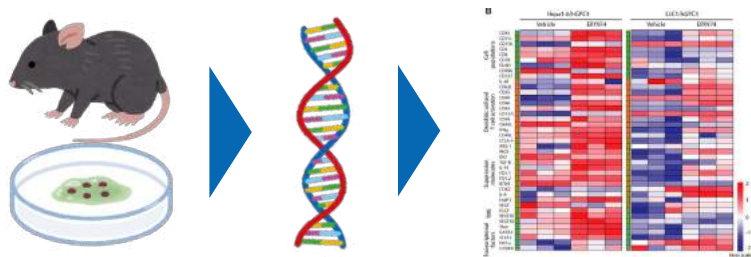


本プラットフォームで実験・解析・解釈をシームレスにつなげる.

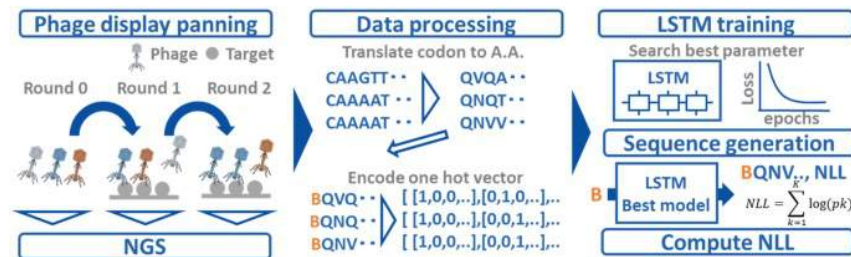
# 弊社におけるNGSデータの活用事例

- NGSデータはオミックス解析から、抗体配列解析まで広く使われる。

## ○ 遺伝子発現解析・ゲノム解析



## ○ 抗体配列最適化



- これまで、弊社では**オンプレミス**のクラスター計算機にパイプラインを構築し、解析を行っていた。

### ○ 短所

- 維持管理に労力がかかる。
- 大量のデータの解析に対応できない。
- 利用開始までに購入期間が必要。



NGS解析環境のAWSへの移行に現在取り組んでいる。



# NGS解析の課題解決に採用した要素技術

- NGS解析の課題解決に採用した要素技術が3つある。

## ○ コンテナ仮想化



環境管理が便利に



NGS Toolのコンテナ化の活動

## ○ クラウド



計算資源増  
運用コスト低減



様々なサービスが用意されている

## ○ Workflow言語



ツール管理が容易に



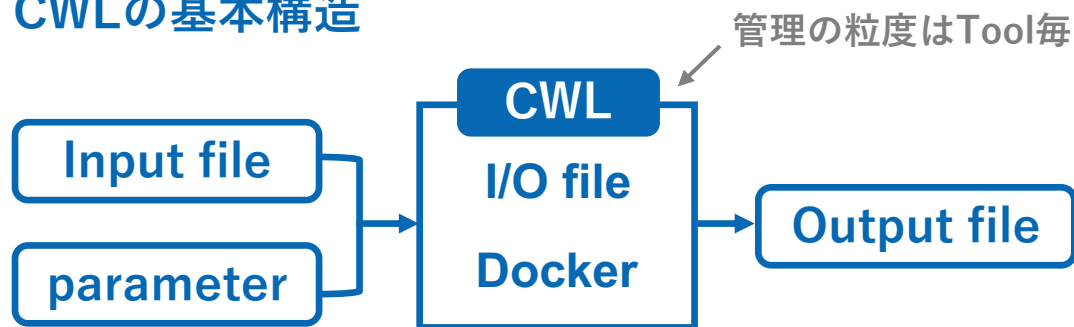
Workflow言語実行システム

コンテナ仮想化・クラウドは一般的、Workflow言語は馴染みがない？

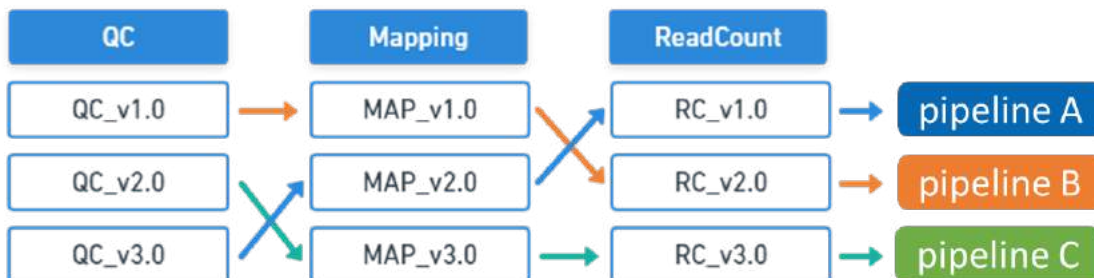
# CWL : Common Workflow Language

- 弊社ではWorkflow言語のひとつである, CWLを採用した。

## ○ CWLの基本構造



## ○ CWLのパイプライン管理



## ○ CWLのコード例

```

1 #!/usr/bin/env cwl-runner
2
3 cwlVersion: v1.0
4 class: CommandLineTool
5 baseCommand: [ samtools, view, -Sb ]
6
7 hints:
8   - class: DockerRequirement
9     dockerPull: biocontainers/samtools:v1.9-4-deb_cv1
10
11 inputs:
12   sam:
13     type: File
14     inputBinding:
15       position: 99
16     out_name:
17
18 nthreads:
19
20
21
22
23
24
25
26
27
28 outputs:
29   bam:
30     type: File
31     outputBinding:
32       glob: $(inputs.out_name)
33
34 stdout: "stdout_sam2bam.txt"
  
```

▼実行コマンド

▲コンテナの指定

◀ Input の定義

◀ Output の定義

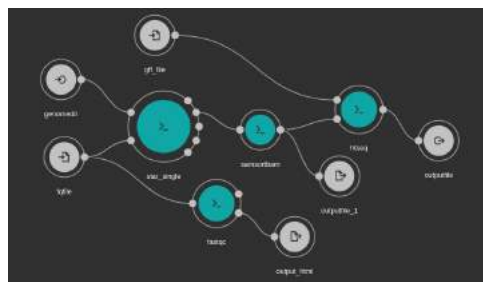
言語仕様がシンプル、yml形式

I/Oと環境を定義し、パイプライン開発・管理が簡便に行える。

# 技術選定について

- なぜ **CWL** なの？
  - コミュニティが活動的で情報も多く、NGS系ツールが充実している。
  - ML型のため構文解析が容易であり、周辺ツールが開発されやすい

## ○ 周辺ツールが多い



GUIでパイプラインを構築

## ○ 活動的なコミュニティ

**Pitagora Workflows in CWL**

DOI: 10.5281/zenodo.2563024

This repository is to share the Common Workflow Language of Pitagora Network, as known as Galaxy Community Japan.

**Tools**

- Get Data
  - download-sra
  - fastq-dump



**Michael R. Crusoe**

@mrcrusoe Co-founder & SCComm-05, Project Lead, learn more at @commonwlorg

#CWL #Workflows

📍 Techstars in Berlin

📧 @mrcrusoe

📅 2018年11月からTwitterを利用しています

630 フォロワー • 1,738 コレクター

<https://github.com/pitagora-galaxy/cwl>  
<https://github.com/common-workflow-library>

## ○ 外部機関と解析パイプラインの共有



- なぜ **Cromwell** なの？
  - AWSにサービスとして展開済み、すぐに動かせる。
  - 最低限の機能が実装されている。



**Cromwell on AWS**

A Collaboration with Broad Institute to Accelerate Genomics Research

Contact Public Sector Sales



弊社は **CWL(Workflow言語)** と **Cromwell(実行エンジン)** を選択。

# NGS解析プラットフォーム

- AWS上にNGS解析プラットフォームを構築中, 本格的な運用は年明けを予定.

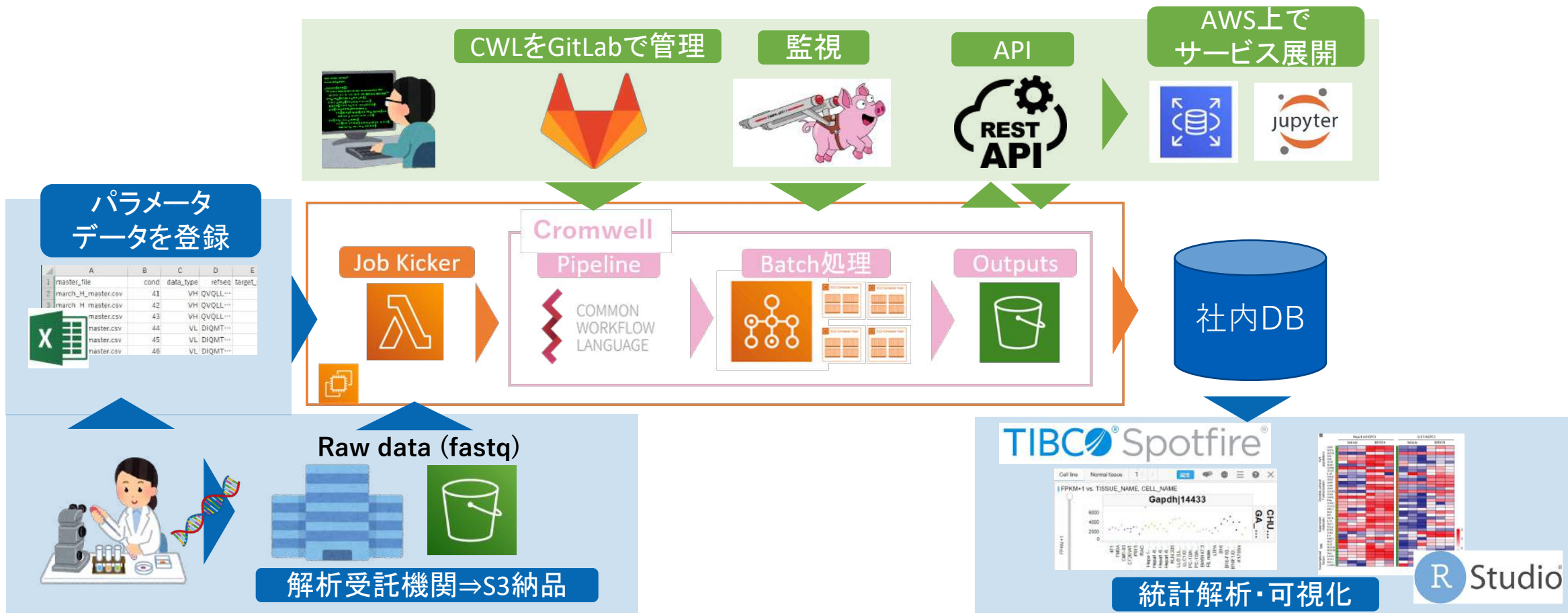
## ○ AWS上に構築した解析プラットフォームの概略図



- ① 実行パラメータの登録を検知し, Cromwellに実行ジョブを投げる.
- ② CromwellにCWLのパイプラインを登録しておく.
- ③ 実行ジョブは, AWS Batchでスケーリングされる.
- ④ 出力結果はCromwell上でID管理され, S3に保存される.
- ⑤ S3を参照して, データ解析を行える.

本プラットフォームで実験・解析・解釈がシームレスにつながる。

# 遺伝子発現データにおける運用例



※Cromwell サーバーの構築・運用は「三菱スペースソフトウェア社」様に依頼

データの入手, 解析の実行, 結果の取得までAWS上で実施可能なシステムを現在構築中.

# まとめ

- 本システムで達成すること
  - 解析パイプラインの再現性・冪等性・拡張性の向上.
  - データ解析の敷居を下げ, 仮説検証のサイクルを高速にまわす.
  - 統一的なデータ蓄積により, データ資産を作る.
- クラウド( AWS ) に構築するメリット
  - スケーラブルな計算リソースとストレージ.
  - 解析結果がクラウド上にあり, DB構築等の新規サービスの展開が容易になる.
  - サーバー自体の保守/運用の軽減.
- CWLのメリット
  - パイプラインの管理・共有・デプロイが容易になる.
  - 環境とI/Oを固めて再現性のある解析が実現できる.
  - 他のWLに比べて, チームでの開発に向いている.

クラウドで研究活動のDX化を促進していきます！

# 中外製薬研究本部の主なデジタル人材募集状況（キャリア採用）

募集要綱など詳細は右記HPを参照：<https://www.chugai-pharm.co.jp/recruit/career/index.html>

問い合わせ先：中外製薬株式会社 研究業務推進部 採用担当 E-mail：[rcrecruit@chugai-pharm.co.jp](mailto:rcrecruit@chugai-pharm.co.jp)

職種名	仕事の内容
AIを活用した化学創薬技術開発の研究者	世界最高水準の創薬の実現のために、AIなどの最新の情報科学技術とCADD技術を融合させた他社にはない低・中分子創薬技術の開発に取り組んでいただきます。
創薬化学研究におけるCADD研究者	世界最高水準の創薬の実現および独自の低・中分子創薬技術の確立のために、以下の課題解決に取り組んでいただきます。 CADD技術を用いた低・中分子医薬品創製プロジェクトの推進 社内外の技術・データを活用した新規CADD技術の開発
医薬品研究のデータサイエンティスト	創薬研究および臨床開発のプロジェクトにおいて、以下の研究を推進する。 <ul style="list-style-type: none"> <li>抗体、低中分子創薬研究に必要な機械学習アルゴリズムの開発</li> <li>臨床ゲノム情報やリアルワールドデータの解析</li> <li>各種解析に必要なデータベース設計および構築、システム設計</li> <li>各種オミックスデータの統合解析 など</li> </ul>
医薬品研究のデータエンジニア	創薬研究プロジェクトを加速させる様々なデータベースやシステムの企画・構築・運用や、オミックスデータ等の各種データ解析等の業務 <ul style="list-style-type: none"> <li>抗体や低中分子の創薬研究における各種データベース</li> <li>ゲノム情報やリアルワールドデータ、オミックスデータの統合解析</li> <li>上記に関するデータ解析基盤の企画・構築・運用 など</li> </ul>
ゲノミクス・プロテオミクスのデータ解析スペシャリスト	創薬研究および臨床開発のプロジェクトにおいて、以下の研究を推進する。 <ul style="list-style-type: none"> <li>各種オミックスデータ（ゲノミクス、プロテオミクス）の解析</li> <li>LC-MS/MSおよびNGSデータ解析ツールの改良・開発</li> <li>各種オミックスデータを起点とした創薬仮説構築、機序解明、研究計画立案 など</li> </ul>

創造で、想像を超える。