



# エーザイのオミクス解析におけるAmazon Omics を用いたアジャイルなプロトタイピング事例

エーザイ株式会社

DHBL ヒューマンバイオロジーインテグレーションファウンデーション  
ヒューマンバイオロジーデータエコシステム  
高橋 健太郎

2023.6.29

*hvc*  
human health care



- 自己紹介
- モチベーション
- Amazon Omicsを用いたアジャイルなプロトタイピング事例
- まとめ

# Company Profile & Corporate Philosophy



設立: 1941年 (昭和16年) 12月6日  
本社: 東京都文京区小石川 4-6-10  
代表者: 代表執行役 CEO 内藤 晴夫  
売上高: 7,562 億円 (2021年度)  
従業員数: 連結: 11,322人, 個別: 3,034人 (2022年3月末現在)  
事業内容: 医薬品の研究開発、製造、販売および輸出入

*hke*

*human health care*

患者様と生活者の皆様の喜怒哀楽を考え

そのベネフィット向上を第一義とし

世界のヘルスケアの多様なニーズを充足する

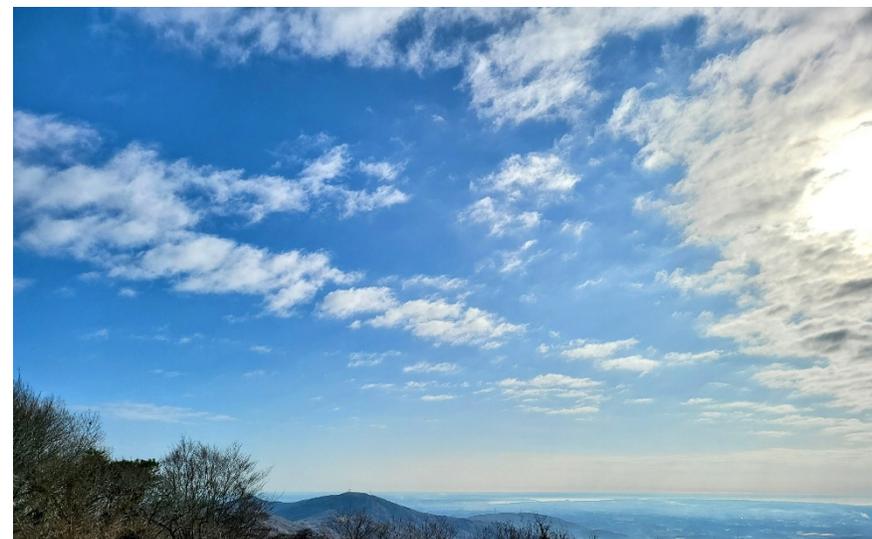


筑波研究所



小石川ナレッジセンター

- 名前： 高橋 健太郎 ニックネーム: takaken
- 所属組織： エーザイ株式会社  
DHBL ヒューマンバイオロジーインテグレーションファウンデーション  
ヒューマンバイオロジーデータエコシステム部長
  - 2008 エーザイ入社
  - 2008 コアテクノロジー研究所 (筑波研究所)
  - 2010 バイオマーカー&パーソナライズドメディスン機能ユニット (筑波研究所)
  - 2016 hhcデータクリエーションセンター (東京)
  - 2022- DHBLヒューマンバイオロジーインテグレーションファウンデーション (筑波研究所)
- 業務
  - データサイエンティスト  
ゲノム解析、プロテオーム解析、バイオマーカー解析、がん・認知症研究に従事
  - Ruby, R, Python, Java, AWS ... Emacs
- 趣味：
  - スキー



- 背景

現在、エーザイではRNA-Seq等のNGS 1次解析は、日本では社内開発システム(Rera)を、USでは商用のSaaSであるDNANexusを利用している。

理想的には、同一の環境を利用するのが望ましいが、現在様々な理由からそのようになっていない

- 目的

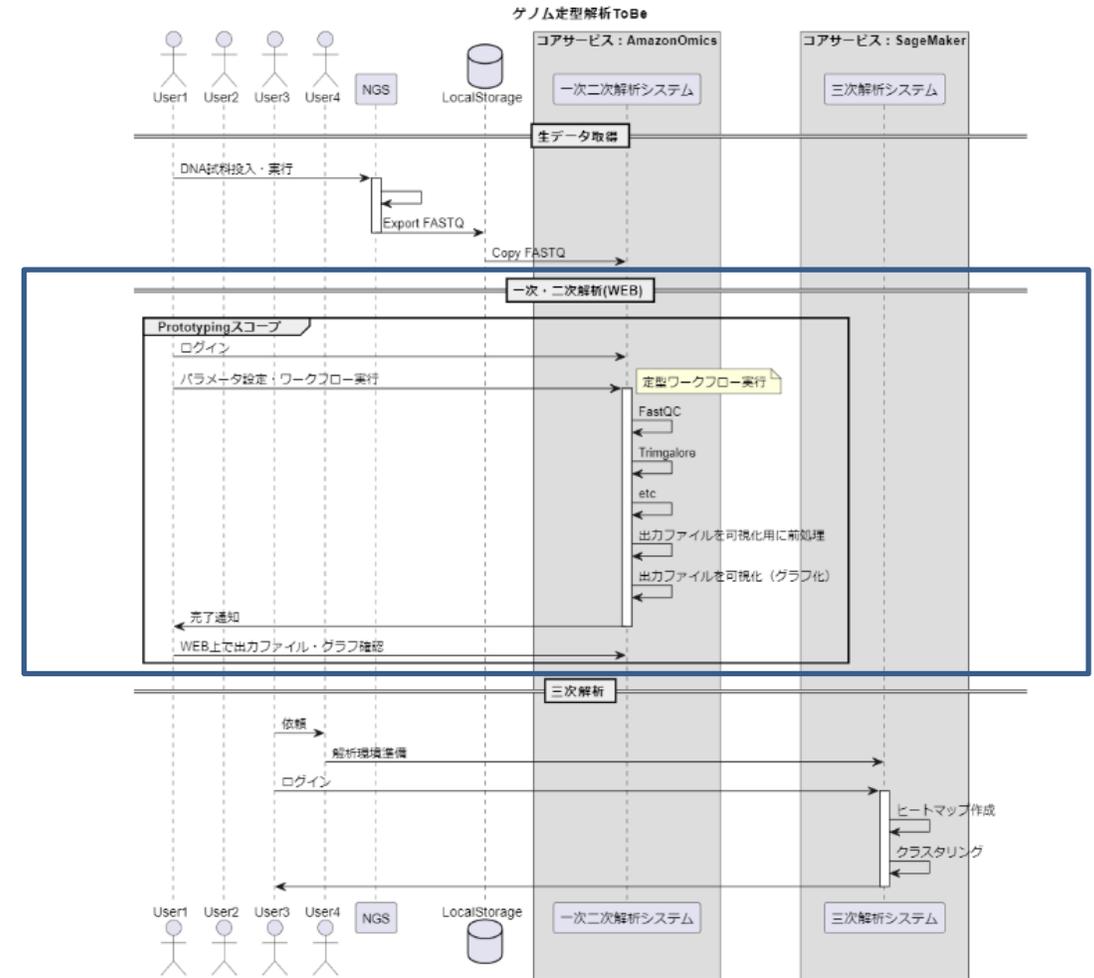
今回、Amazon Omics\*のPrototypeを通じて、社内システムのリプレースの実現性の検討及び、DNANexus等商用の解析Platformとの違いを知るとともに、Amazon Omicsに代わり、自前で開発が必要なものは何かを明確にし、今後のエーザイでのNGS解析環境構築に向けての検討材料の取得を目的とする

→AWSとのプロトタイピングへ

\* AWS フルマネージドなライフサイエンス向けの解析・保存・分析向けの専用サービス

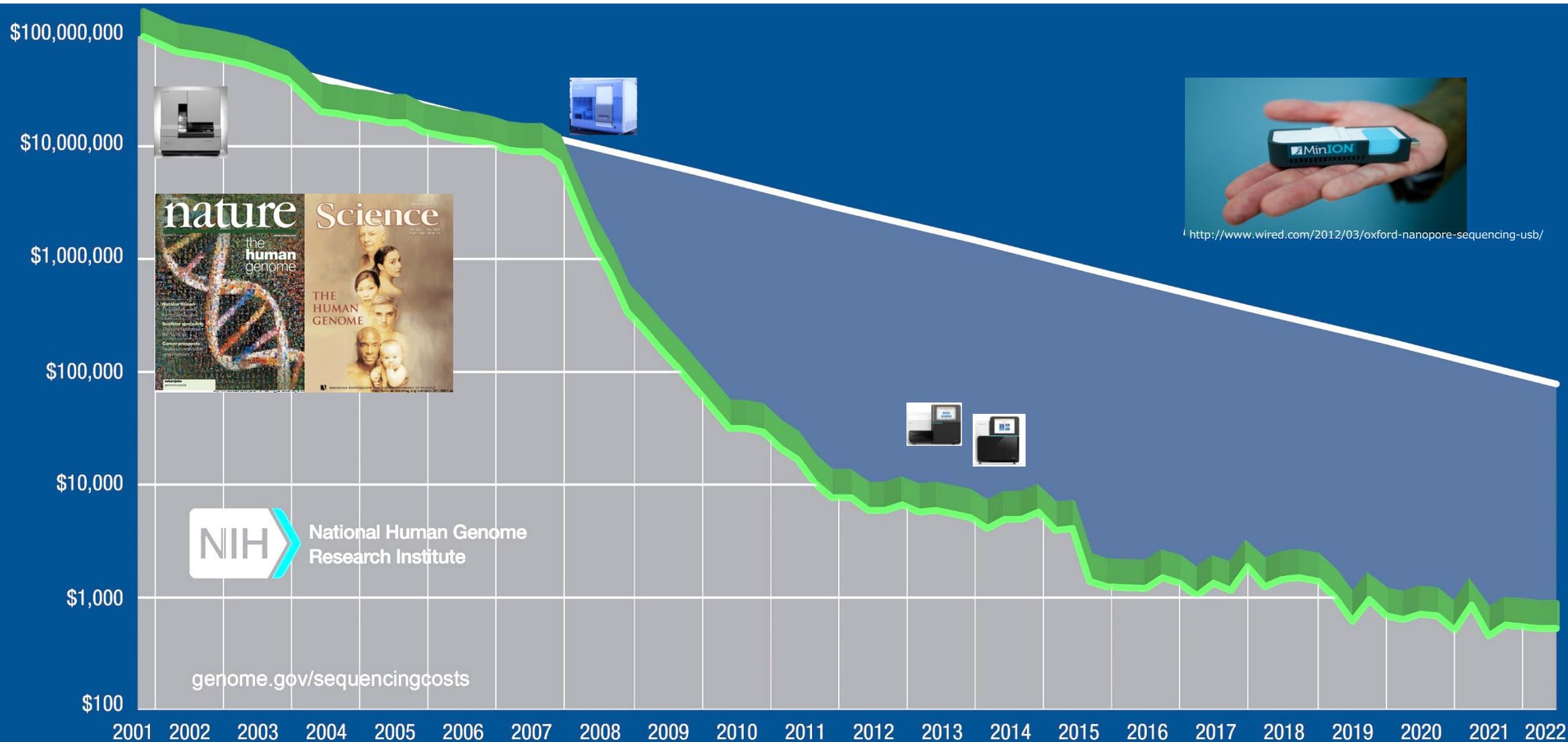
# プロトタイピング概要

- 体制
  - AWS: 4名  
アーキテクト1名、開発1名、全体1名、営業1名
  - エーザイ: 4名  
データプラットフォーム2名、全体2名
- 期間
  - 2か月
    - ToBeをAWS側とエーザイで打ち合わせ後、複数回アジャイルにAWSにて開発を行いFBをイテレーティブに実施
- 対象
  - Amazon Omics を用いたBulk RNA-seq解析  
1次解析から可視化
- アウトプット
  - Infrastructure as Code向けのインフラ構築用デプロイコード (CDK/Cloud Formation)
  - プロトタイプアプリケーションコード類



NGS定型解析ToBe

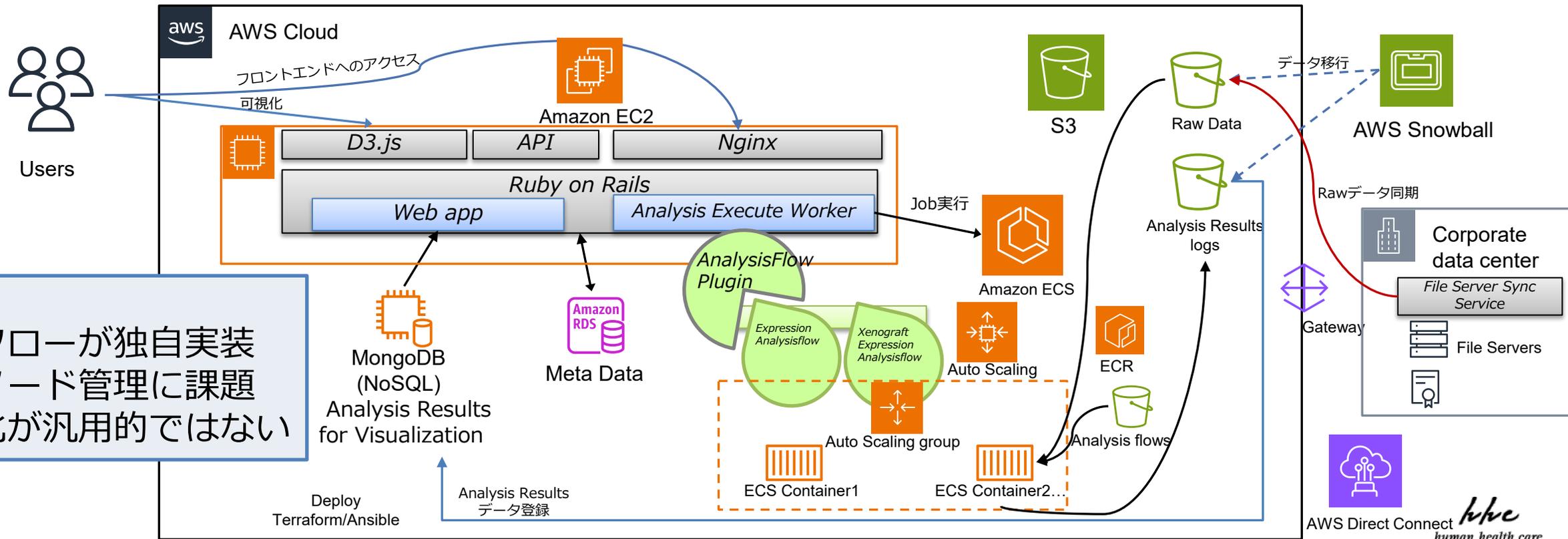
# ヒトゲノムのコスト



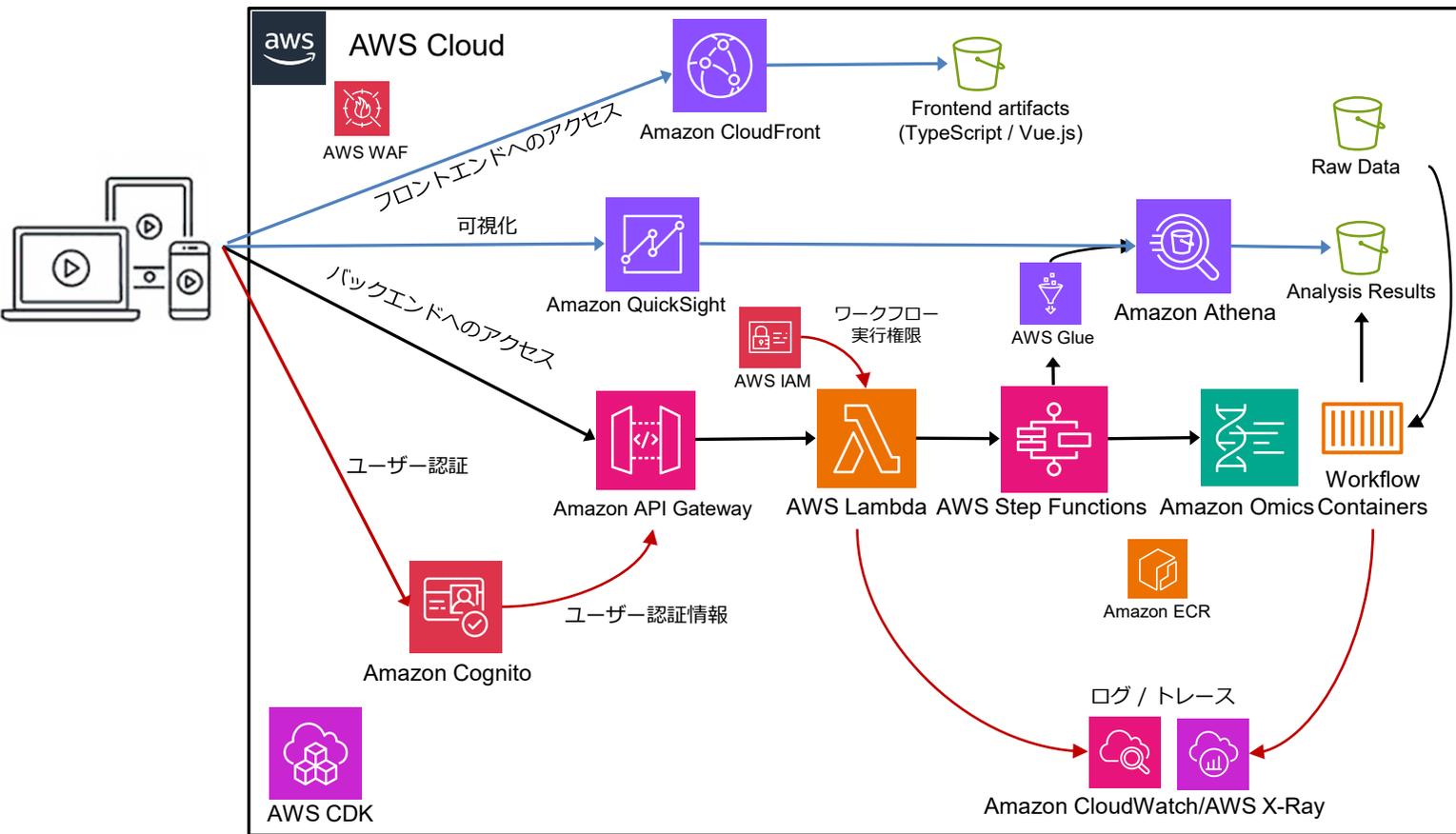
# これまでのシステム NGS用社内開発解析システム(Rera, RNA-seq向け)



- 2017年：
  - 旧オンプレミスのNFSシステムからSnowballを用いて50 TBの既存Raw/解析データをS3に転送、容量無制限のS3へ移行
  - ネットワーク回線としてDirect Connectを利用
    - ECSを利用したImmutable Infrastructureによる解析 (m4.16xlarge→c6in.32xlarge)
    - Auto scaling による計算ノード管理 (カスタムプログラムでAWS APIで対応)
- 2023年：現在、約16,000サンプル、S3: 325TB (50TB増/年)



# プロトタイピング全体アーキテクチャー



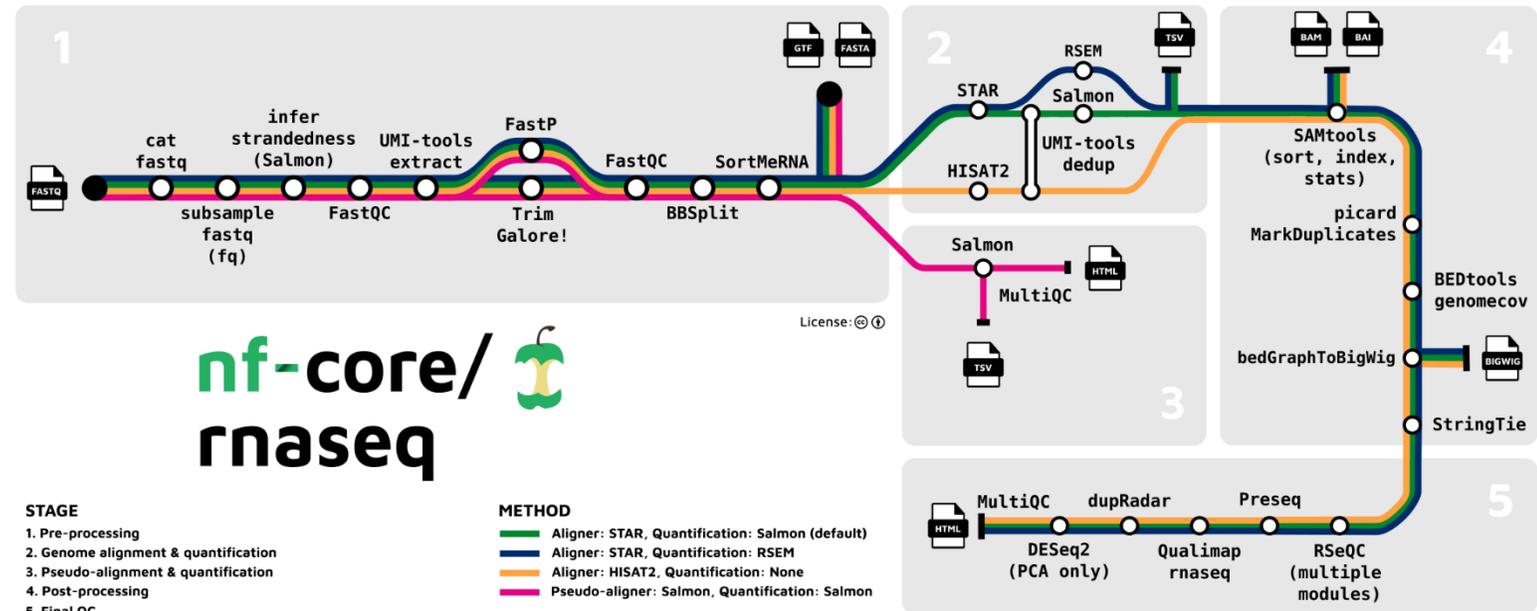
サービス名	システムでの利用用途
Cognito	Prototypingシステムのユーザ、グループ管理
Cloud Front	Prototypingシステムのendpoint (WEB URL)
API Gateway	Lambda関数へのRest API
Lambda	各種AWSサービスの実行関数
<b>Amazon Omics</b>	<b>フルマネージドなライフサイエンス向けの解析・保存・分析向けの専用サービス</b> PipelineによるRNA-Seqの解析機能
Quick Sight	解析結果の可視化
S3	リフェレンスゲノム及び、解析サンプルファイル、解析結果の格納場所、また、PrototypingシステムのJavaScriptファイルの配置場所
Step Function	サーバレスオーケストレーションサービスでAmazon Omicsで利用する各種ソフトウェアイメージ等の登録
IAM	ユーザ管理
Cloud Watch/X-Ray	各サービスのログの収集と閲覧、検索
CDK/Cloud Formation	各サービス等へのdeploy

## 特徴

汎用的なNextflowベースの解析フロー  
 Amazon Omics 自動的に計算ノード・リソース管理・ログ出力  
 サーバレスなAthena/QuickSightを用いた可視化+サービス化

nf-core  
<https://nf-co.re/>

- Nextflow (DSL)で定義されたバイオインフォマティクス向けPipelineコミュニティ



<https://nf-co.re/rnaseq>

最小限の変更でAmazon Omics (WDL及びNextflowに対応)で実行可能

# プロトタイピング画面サンプル



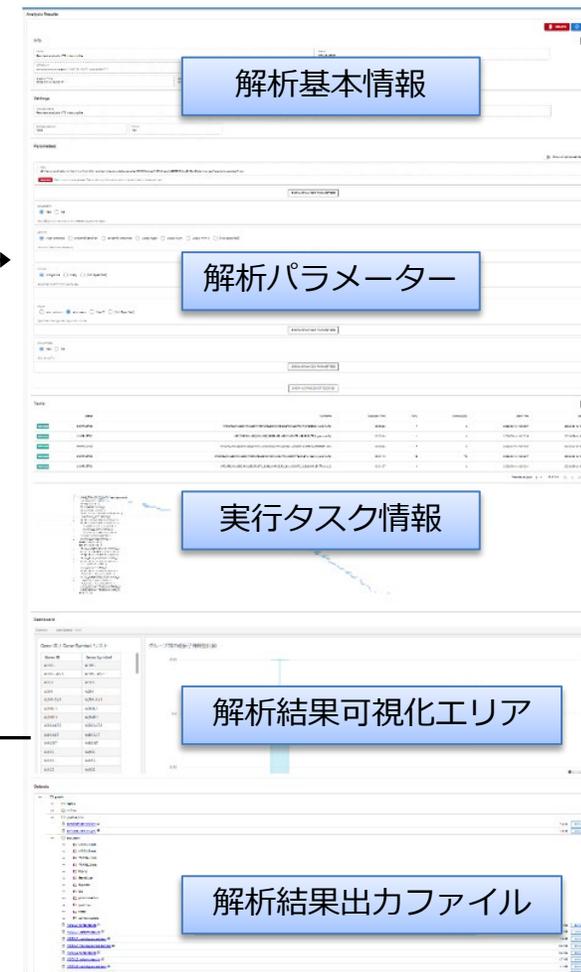
ログイン画面

解析一覧画面

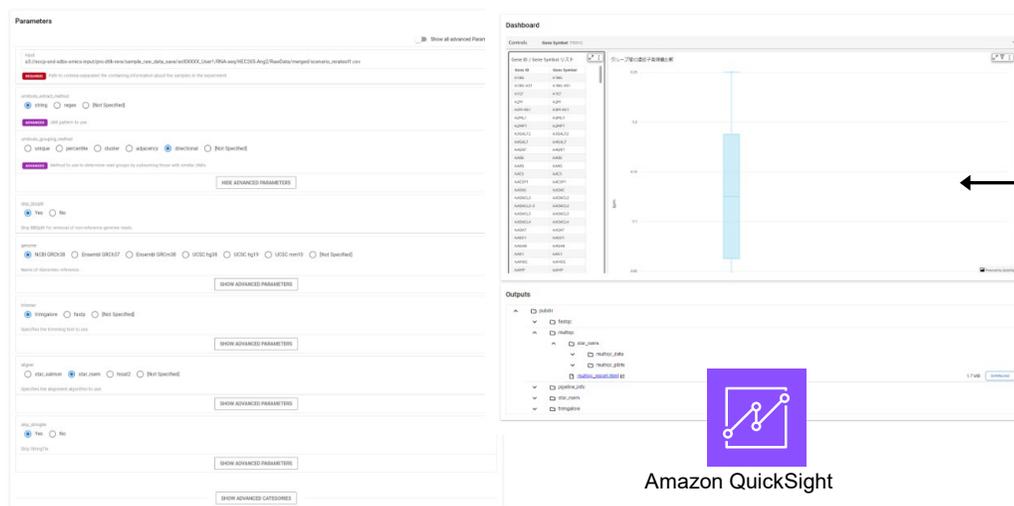
解析詳細画面



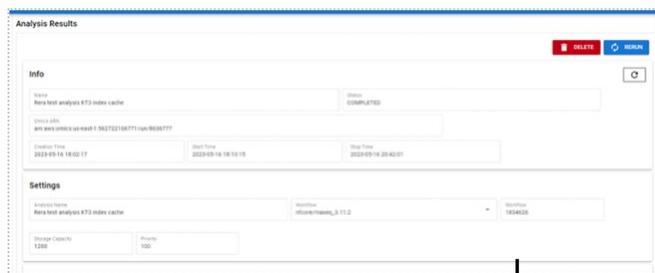
Status	Name	Execution Time	Creation Time	Start Time	Stop Time
FAILED	Rera test analysis GSE104704 test1	00:14:23	2023-05-18 09:54:00	2023-05-18 10:05:21	2023-05-18 10:19:41
COMPLETED	Rera test analysis KT3 index cache	02:28:45	2023-05-16 18:02:17	2023-05-16 18:13:15	2023-05-16 20:42:01
COMPLETED	Rera test analysis KT2	02:55:17	2023-05-16 17:47:19	2023-05-16 17:58:18	2023-05-16 20:53:36
COMPLETED	Rera test analysis KT1	02:56:28	2023-05-16 17:37:50	2023-05-16 17:47:49	2023-05-16 20:44:15
COMPLETED	test_kt1	00:13:28	2023-05-16 09:32:49	2023-05-16 09:44:14	2023-05-16 09:57:43



解析可視化画面



新規解析画面



解析実行



Amazon Omics

Amazon QuickSight

解析結果出力ファイル

# サービス比較Pros/Cons概要



## Amazon Omicsの使いどころ

- ・ nextflow/WDL等公開(nf-core)や利用している解析フローをそのまま利用できる
- ・ 上記のworkflowの自動的な分散処理によるコスト/リソース最適化が可能
- ・ AWSの各種サービスとの連携が簡便に可能

比較項目	Amazon Omics/ Prototyping	DNAnexus	内製システム (Rera)
ワークフローを簡単に定義できる (nf-coreの定義を流用できる)	◎	△	×
ワークフローを簡単にカスタマイズ できる	○	◎	×
データや結果に対して細かいアクセ ス制御ができる	△	○	◎
GDPR等法規制への十分な対応がな されている	×	◎	×
国内での解析処理が可能（サーバを 国内に設置）	×*	?	◎
サンプルファイルやリファレンス データを簡単にアップロードできる	○	○	◎
バックアップなどの体制が整ってい る	○	○	△
解析リソース（サーバの性能）を変更 できる	◎	○	△
解析処理のログをユーザ自身が確認 できる	◎	◎	△
二次解析（可視化や統計解析）がGUI 上で行える	△	?	○
AWSサービスとの連携	◎	×	△

\* 2023/6現在 USリージョンでのみ利用可能

# プロトタイピングを行ったことにより得られたもの

- AWSとの緊密なコミュニケーション
  - 迅速なプロトタイピングが可能
  - 最新のAWSの情報をタイムリーに利活用が可能 (Amazon Omics 発表後すぐに検討)
- 最新のIaC (Infrastructure as Code) / サーバレスを中心にした開発スタイルやアーキテクチャーを素早く社内に取り込むことができる
  - 今後の社内プロジェクトのテンプレートとして検討
- アジャイルな開発スタイル
  - 以前の社内開発システムが1年半程度で開発したものが、約2か月で最新のアーキテクチャーベースになりえることを認識

# Data driven Innovation

## 仮説検証から仮説生成へ

従来

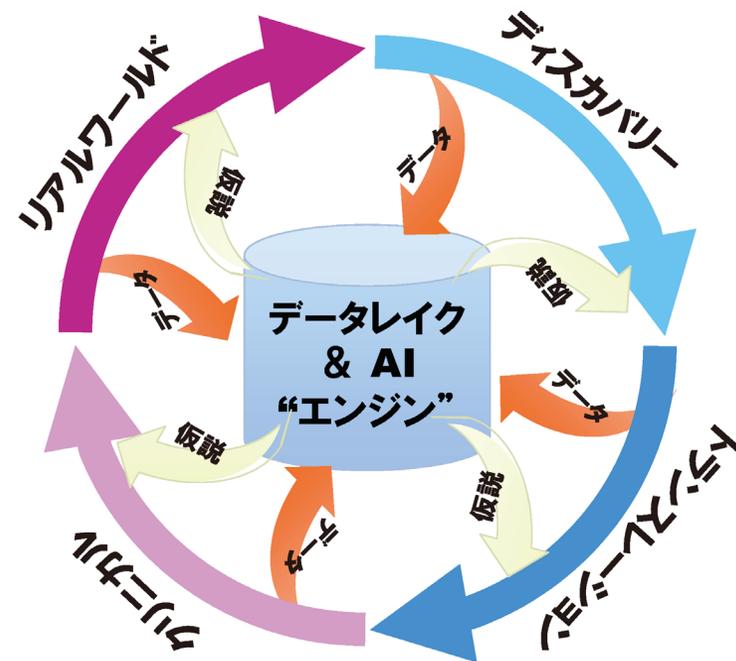
これから

### データによる仮説検証

- 論文情報から研究者が創薬仮説を立て、それを確認するためにデータを取る
- 得意とする反応系や過去の成功体験を元に化合物を合成し薬理データを取る
- 自らが取ったデータを最も信頼性の高いデータと位置づけ意思決定を行う
- 自ら立てた仮説に合わないデータが出た時には実験が間違っていると考え

### データからの仮説生成

- リアルワールドデータを解析し、今まで知られていなかったエビデンスを得てリサーチクエスチョンとし新たな試験を計画
- 臨床サンプルのゲノム解析とマルチオミクス解析からターゲット仮説を生成させ、システムティックに検証実験を行う
- コンピューター上で化学構造をバーチャルに発生させ、その中から改善が予測される化合物のみを合成し活性測定





- AWS
  - 岡田さん
  - 嶺さん
  - 益子さん
  - 水木さん
- エーザイ
  - 三浦さん
  - 酒寄さん
  - 赤田さん

参考: AWS ビルダーズフレッシュ記事リンク

<https://aws.amazon.com/jp/builders-flash/202307/eisai-omics-prototyping/>



薬が必要なくなる未来を、  
めざす製薬会社でありたい。

<https://eisai-recruit.jp/>

[博士採用Entry ▶](#)

[キャリア採用Entry ▶](#)