

**AWS RE:INVENT**

**re:Cap**





# re:Invent 2023 で発表されたゲノミクス領域における最新事例および関連サービスのご紹介

鳥羽 祐輔

ソリューションアーキテクト

アマゾンウェブサービスジャパン合同会社

# 自己紹介

## 鳥羽 祐輔

アマゾンウェブサービスジャパン  
ソリューションアーキテクト

製薬企業を中心にクラウド活用に関する技術的な  
ご支援を提供しています。

好きな AWS サービス



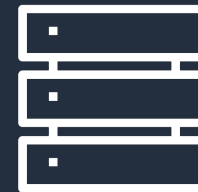
# ゲノミクス領域におけるよくある課題



シーケンサ



ゲノムデータ  
ストレージ



解析環境

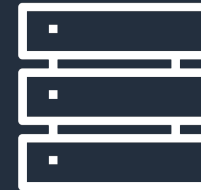
# ゲノミクス領域におけるよくある課題



シーケンサ



ゲノムデータ  
ストレージ



解析環境

増加し続けるゲノムデータに  
オンプレミスのストレージの  
拡張が間に合わない

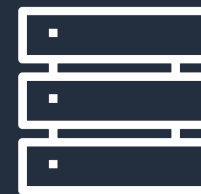
# ゲノミクス領域におけるよくある課題



シーケンサ



ゲノムデータ  
ストレージ



解析環境

増加し続けるゲノムデータに  
オンプレミスのストレージの  
拡張が間に合わない

オミクス解析処理実行環境の  
設定管理が複雑で研究者のニーズに  
スピーディーに対応できない

# アジェンダ

## 1. ゲノムデータのクラウドへの移行

- Caris Life Sciences 様: 25 PB のゲノムデータをクラウドへ移行

## 2. AWS 上でのゲノム解析処理の実行

- AWS HealthOmics とは
- Amgen 様: AWS HealthOmics の活用によるイノベーション



# アジェンダ

## 1. ゲノムデータのクラウドへの移行

- Caris Life Sciences 様: 25 PB のゲノムデータをクラウドへ移行

## 2. AWS 上でのゲノム解析処理の実行

- AWS HealthOmics とは
- Amgen 様: AWS HealthOmics の活用によるイノベーション





# Caris Life Sciences様: 25 PB のゲノムデータをクラウド移行

NTA213

## 0 to 25 PB in one year

<b>Mason Hensley</b> Chief Technology Officer Caris Life Sciences	<b>Joe Howe</b> Principal, Enterprise Sales Amazon Web Services	<b>David Cathey</b> Senior Solutions Architect Amazon Web Services
---	---	--

 © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



<https://www.youtube.com/watch?v=IPEsJ5UtwS8>

[https://d1.awsstatic.com/events/Summits/reinvent2023/NTA213\\_0-to-25-PB-in-one-year.pdf](https://d1.awsstatic.com/events/Summits/reinvent2023/NTA213_0-to-25-PB-in-one-year.pdf)



# Caris Life Sciences様: 29 PB のゲノムデータをクラウド移行

NTA213

## 0 to 25 PB in one year

Mason Hensley  
Chief Technology Officer  
Caris Life Sciences

Amazon S3 snapshot – Nov 28, 2023

# 29.3 PB

aws © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



<https://www.youtube.com/watch?v=IPEsJ5UtwS8>

[https://d1.awsstatic.com/events/Summits/reinvent2023/NTA213\\_0-to-25-PB-in-one-year.pdf](https://d1.awsstatic.com/events/Summits/reinvent2023/NTA213_0-to-25-PB-in-one-year.pdf)



# ゲノムデータの定常的な増加により課題が顕在化


- 1日に数百TBのゲノムデータがシーケンスされることも
  - 新しいシーケンスデータの保管やデータサイエンティストが作業を行うためのオンプレミス環境のストレージ容量が限られている
  - データ退避先のテープライブラリも限界を迎えつつあった
    - 32のドライブ/約6000のテープを管理しながらのデータ管理は複雑に
- データサイエンティストがデータにアクセスするのに時間がかかる
  - 必要なコホートデータを取得するのに数週間かかる事態



# AWS のソリューションの採用

-  Amazon S3
  - 適切なストレージクラスを使うことで迅速なアクセスと費用対効果を両立

-  AWS DataSync
  - ログ記録、データ検証を含むマネージドな転送機能を提供

-  Amazon Athena
  - S3 インベントリへのクエリにより様々なデータファイルの場所を把握

-  AWS Control Tower
  - 適切な人が適切な操作を行えるようランディングゾーンを提供

# プロジェクトを通して実現したこと

## • データ連携によるイノベーションの実現

- 社内: 適切な人々が適切な時に適切なデータにアクセスできるようにして、研究開発の能力を向上させ、患者の命を救う
- 社外: 主要な製薬会社や学術機関と協力し最高水準の治療法を開発

## • コスト削減

- オンプレミスのデータセンターや研究所のインフラへの支出の削減

## • スケーラビリティ

- ITリソースが患者サンプルの受け入れや患者の命を救う活動においてボトルネックにならないようにスケールさせることができる

# アジェンダ

## 1. ゲノムデータのクラウドへの移行

- Caris Life Sciences 様: 25 PB のゲノムデータをクラウドへ移行

## 2. AWS 上でのゲノム解析処理の実行

- AWS HealthOmics とは
- Amgen 様: AWS HealthOmics の活用によるイノベーション





# Amgen様: AWS HealthOmics 活用によるイノベーション

AIM215

## Omics innovation with AWS HealthOmics: Amgen's path to faster results

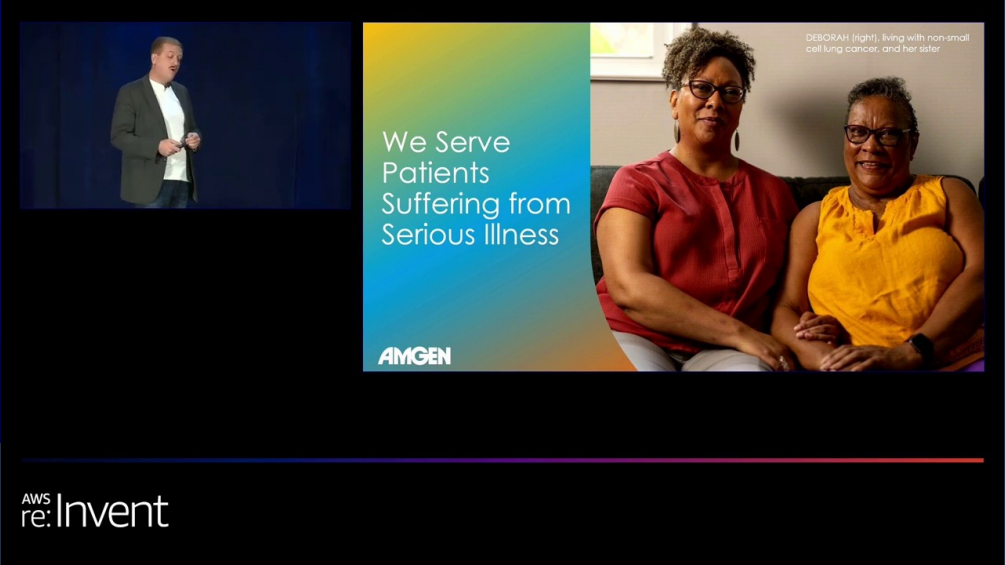
<b>Ariella Sasson</b> (she/her) Principal Solutions Architect AWS	<b>Ittai Eres</b> (he/him) Senior Scientist   CRADI Computational Biology & Bioinformatics Technologies Research & Development Amgen
--	--

 © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

## Agenda

- 01 Year in review
- 02 AWS HealthOmics: deeper dive
- 03 Innovation with AWS HealthOmics: Amgen's path to faster results

 © 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



DEBORAH (right), living with non-small cell lung cancer, and her sister

We Serve Patients Suffering from Serious Illness

AMGEN

AWS re:Invent

<https://www.youtube.com/watch?v=Cwo83c4xvPQ>

[https://d1.awsstatic.com/events/Summits/reinvent2023/AIM215\\_Omics-innovation-with-AWS-HealthOmics-Amgens-path-to-faster-results.pdf](https://d1.awsstatic.com/events/Summits/reinvent2023/AIM215_Omics-innovation-with-AWS-HealthOmics-Amgens-path-to-faster-results.pdf)



# AWS HealthOmics

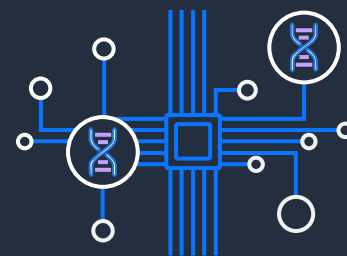
プロダクションレディなオミクス解析環境をフルマネージドで迅速に提供



マルチオミクスと  
マルチモーダル分析



集団ゲノム解析レベル  
の規模に対応



フルマネージドな  
バイオインフォマティクス  
計算環境



組み込みのセキュリティ、  
プライバシー、  
コンプライアンス  
(HIPAA 適格)

- 過去10年間にAWSがGenomics England, Stanford, Philips, AstraZeneca, Illumina, DNA nexus などのお客様と取り組んだゲノミクス関係の活動の知見に基づいて設計されている
- 一般利用開始。米国東部（バージニア北部）、欧州（アイルランド）、欧州（ロンドン）、欧州（フランクフルト）、アジアパシフィック（シンガポール）にて



# AWS HealthOmics 全体像

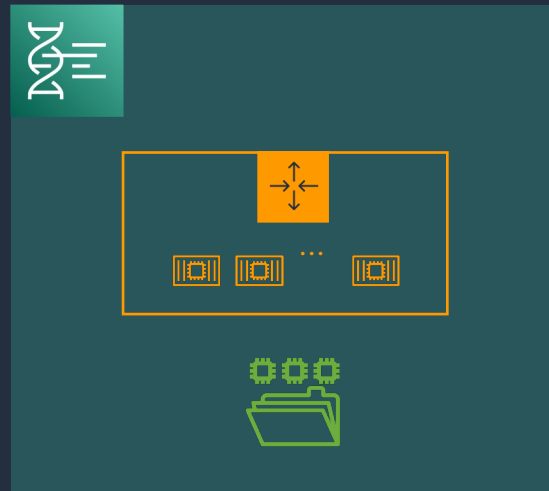
3つの主要コンポーネントから構成

## HealthOmics Storage



オミクスデータの保存

## HealthOmics Workflow



オミクスデータの2次解析

## HealthOmics Analytics



オミクスデータの3次解析と  
マルチモーダル解析

# HealthOmics ワークフローに事前定義ワークフローが追加

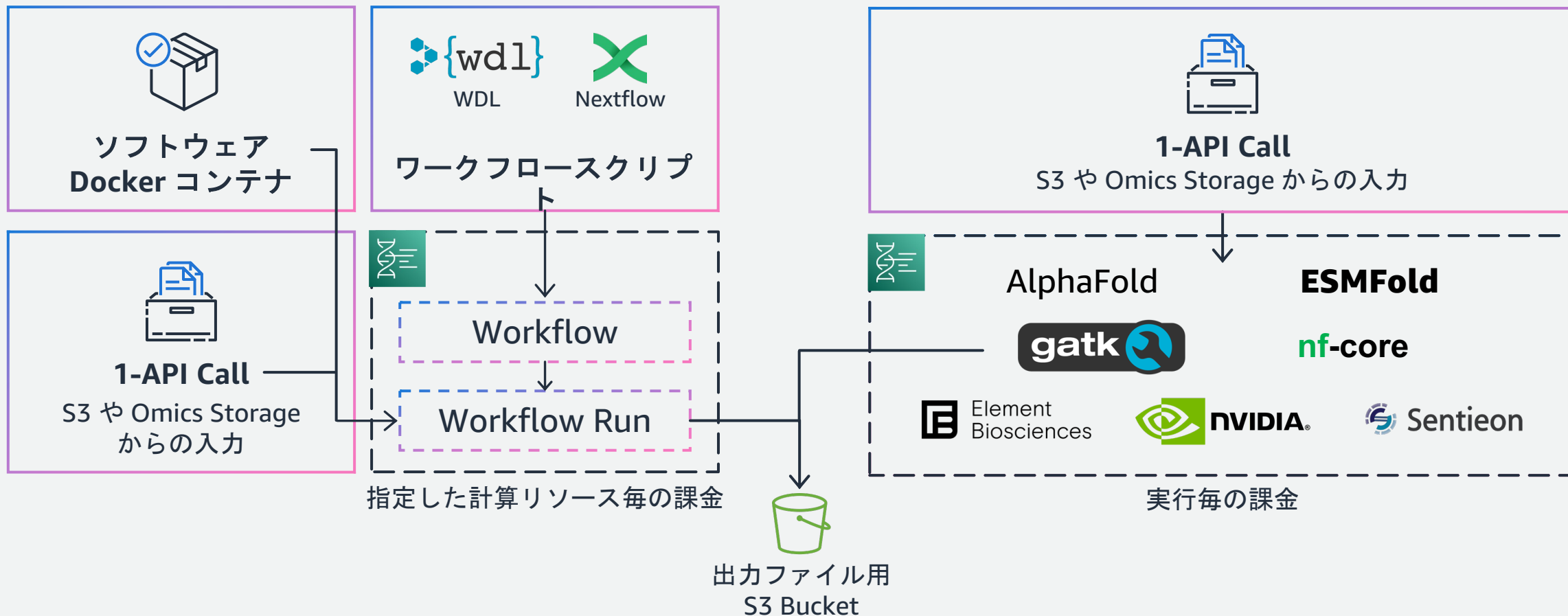
プライベートワークフロー

Bring your own

New!

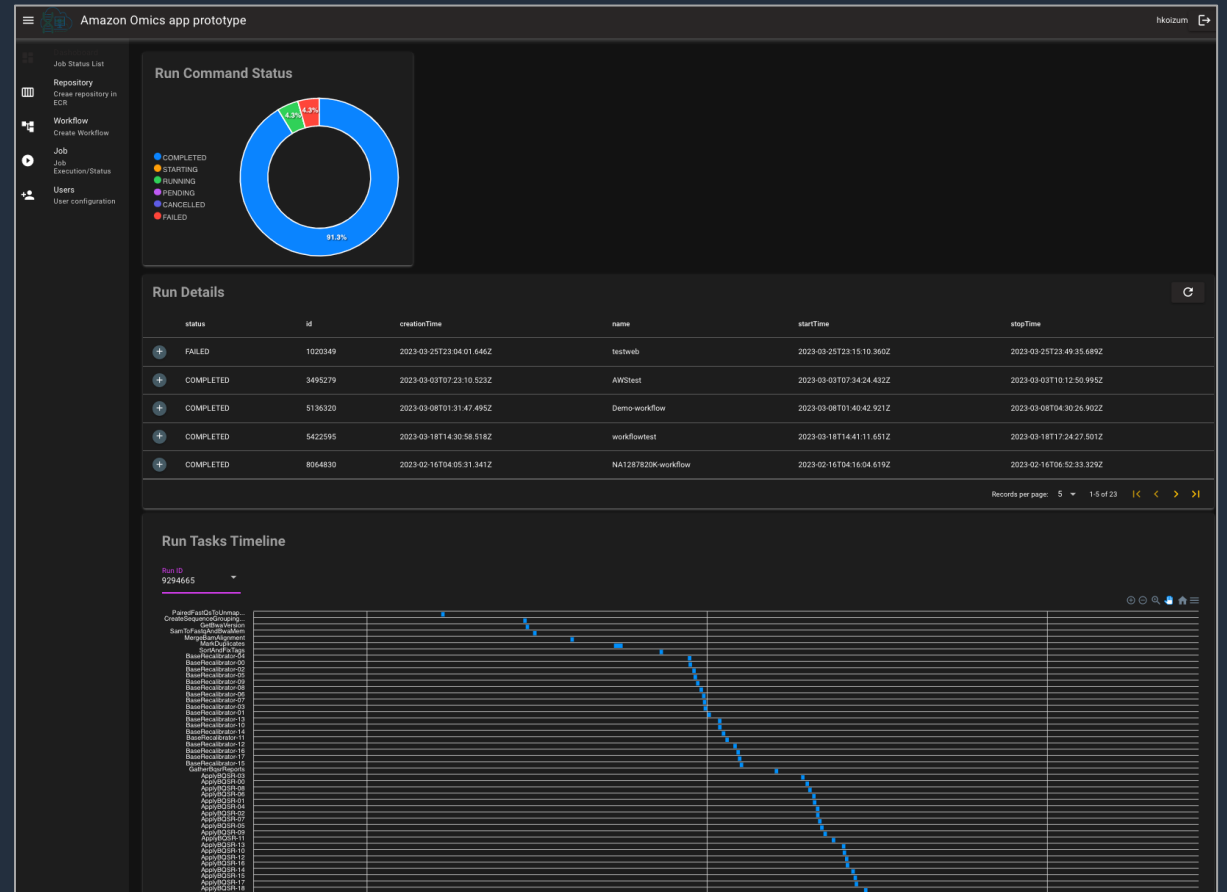
Ready2Run ワークフロー

Hosted



# AWS HealthOmics Webapp の公開

- バイオインフォマティクスに関わるバイオインフォマティシャンやリサーチャーのためのウェブアプリ
- AWS HealthOmics ワークフローをウェブアプリから操作可能
- 実行中のワークフローのステータスをチャートとテーブルで視覚化
- aws-samplesにオープンソースとして提供  
GitHub - aws-samples/amazon-omics-webapp-integrations  
<https://github.com/aws-samples/amazon-omics-webapp-integrations>



# Amgen 様の従来の計算環境における課題

## Computational biology at Amgen

### Amgen's first AWS solution for omics on AWS was powerful, but...

Required up-front times from IS (instead of scientists) to set up pipelines

Difficult for scientists to troubleshoot and track tasks

Individuals created bespoke siloed compute environments for their own solutions

AWS HealthOmics help address these problems & more

## 従来ソリューションの課題

- パイプライン利用のために情報システム部門との連携が必要
- 研究者がトラブルシューティングや各タスクを追跡することが困難
- 各利用者は使いたいツールに応じて計算環境を独自カスタマイズすることに

# 差別化に繋がる業務に注力するために HealthOmics を採用

**Why does Amgen use AWS HealthOmics?**

WHY: WE'RE FOCUSED ON DISCOVERY NOT I.T.

**AWS HealthOmics workflows:**

- Stores and manages our workflows in one place with the same authorization model
- Supports what we need in Nextflow
- Futureproof: Gives us flexibility to use other languages (e.g., WDL) if we need to later
- Gives us scale and stability our previous solution didn't

**No need to re-invent the wheel**

## HealthOmics ワークフローのポイント

- 1つの場所で全てのワークフローを保存・管理することができる
- Nextflow をサポート、かつ別のワークフロー言語 (WDL等) に切り替えられる柔軟性
- 従来できなかったようなスケーラビリティと安定性が手に入る

# 体感した AWS HealthOmics 活用のメリット

- 研究者がセルフサービスでパイプラインを実行できるように
- 最大 40-60% の TCO コスト削減を実現
- 従来と比べ同等もしくははより速い実行時間
- ジョブ実行データの透明性: トラブルシューティングが容易、他システムとの連携も可能に

## Advantages of AWS HealthOmics

Scientists are **empowered** to perform self-service

Shared pipelines lead to **robustness, reproducibility, reduced cycle times**

TCO **cost savings up to 40–60%** (varies w/ pipeline & inputs)

Compute **times** are as **fast or faster** than prior solution

Debugging is **transparent**; run metadata retained in CloudWatch

Enables quick troubleshooting, easy real-time monitoring of run progress, connection with other systems (LIMS)

Run-level **visibility into billing**—large implications for company of Amgen's scale & billing complexity



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



For scientists interested in **broadening accessibility, usage, and visibility** of omics pipelines, AWS HealthOmics represents a uniquely **appealing and fit-for-purpose solution**



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# Nextflow nf-core の活用

nf-core: コミュニティの努力により、ベストプラクティスに基づいた様々なデータタイプのバイオインフォマティクス分析パイプラインを公開

“**nf-coreをお勧めします**。素晴らしいコミュニティリソースです。あなたの労力を軽減し、行っていることがそのデータを見る有効な方法であるという自信を高めてくれます。”

“私が Nextflow について知っていることは全て AWS HealthOmics 活用の過程で学んだものです。興味はあるものの Nextflow 開発者ではない方がいたら、**学ぶことはそれほど難しくなく**ことをお伝えします。”

The screenshot displays the 'Pipelines' page on the nf-core website. The page title is 'Pipelines' and the subtitle is 'Browse the 94 pipelines that are currently available as part of nf-core.' The page features a search bar and filters for 'Released' (55), 'Under development' (27), and 'Archived' (12). The pipelines are sorted by relevance and displayed in a grid. Each pipeline card includes a title, a brief description, a list of associated tags, and a version number with a release date.

Pipeline Name	Stars	Version	Release Date
ampliseq	141	2.8.0	released 1 day ago
taxprofiler	76	1.1.3	released 6 days ago
scrnaseq	112	2.5.0	released 8 days ago
rnaseq	719	3.14.0	released 10 days ago
rnasplice	19	-	-
methyelseq	121	-	-

# nf-core パイプライン活用の現状とこれから

## Specific use cases

### Business use case



Analysis of Cut & Run data across high throughput samples with varied antibodies enabled wet lab scientists to decide which antibodies were most specific to target, and hence best to use in future standard workflow

### Data type

- ATAC-seq
- Cut & Run
- Hi-C
- RNA-seq
- scRNA-seq

Implemented nf-core pipelines already in-use by Amgen scientists

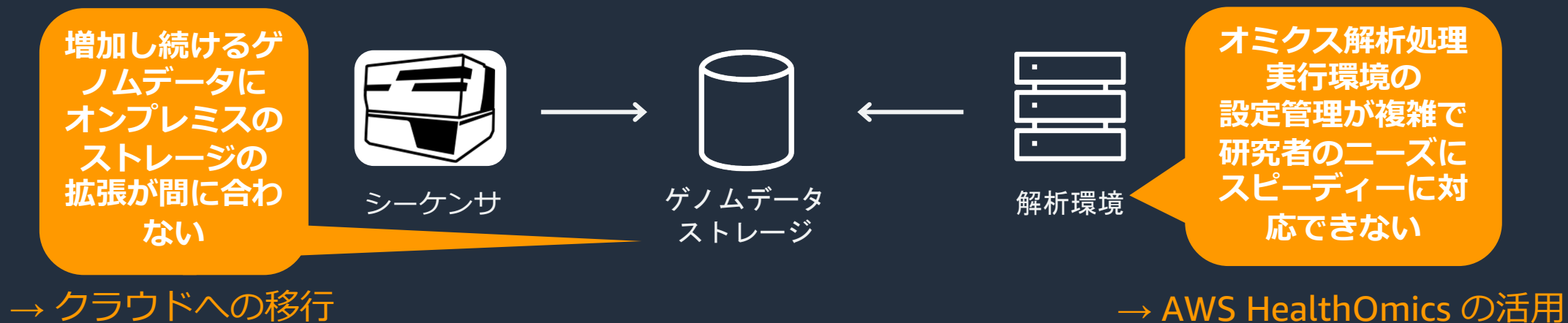
With plans to migrate ~30 other nf-core + custom pipelines to AWS HealthOmics

- 既に5つのnf-coreパイプラインをAWS HealthOmics上に実装
- 約1週間かかる解析作業が1日程度に短縮
- nf-coreの他のパイプラインや独自のカスタムパイプラインの30以上を移行することを予定

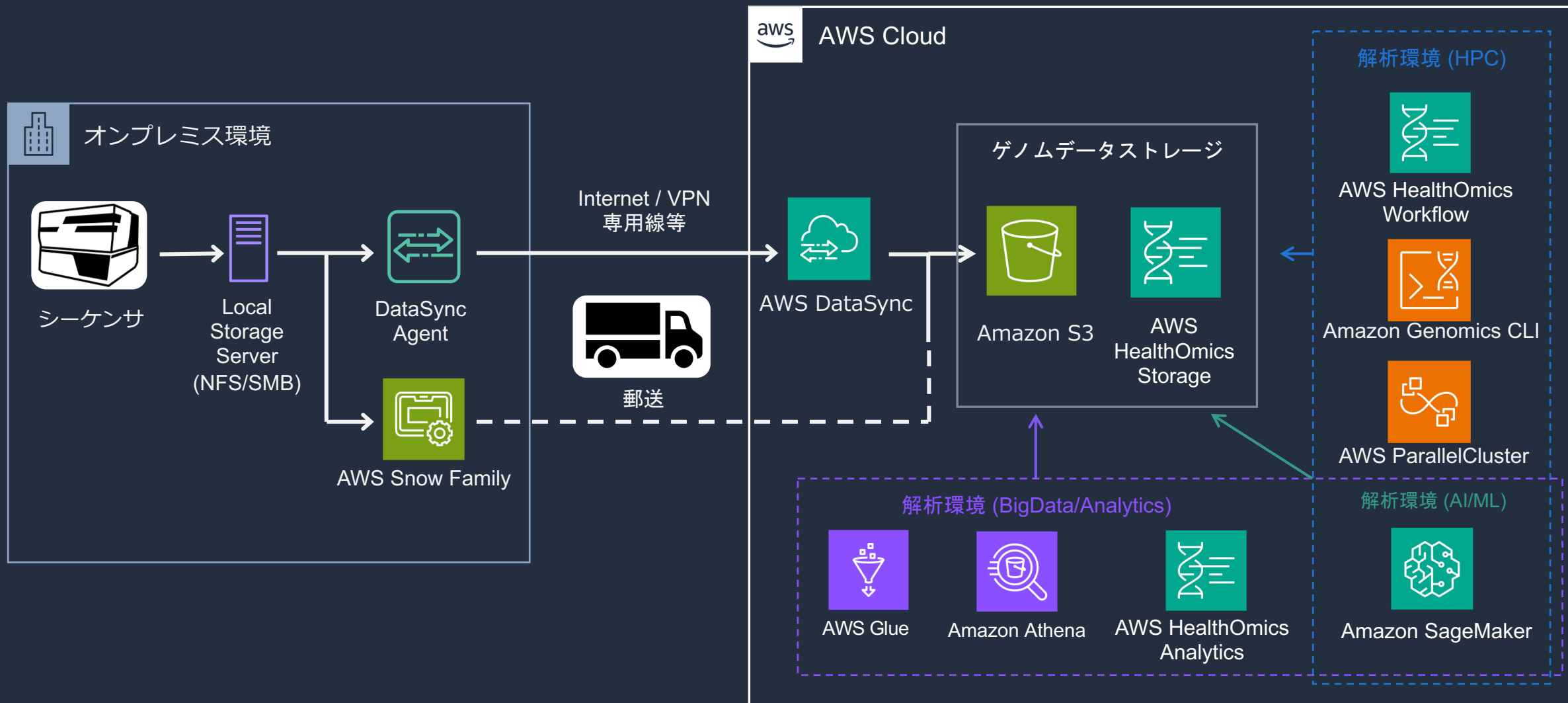


# 本セッションのまとめ

- ゲノミクス領域において AWS を活用することで、ITリソースの調整ではなく、**ビジネスの差別化に繋がる業務にフォーカス**できる
- Caris Life Sciences 様: 増え続けるテープドライブとディスク容量の調整業務からの開放
- Amgen様: スケーラブルな解析計算環境を研究者がセルフサービスで利用できるように



# ゲノミクス領域の多様なニーズをサポートするAWSソリューション



# AWS HealthOmics 関連資料

Amazon Omics のご紹介 – ゲノムデータおよび生物学的データを大規模に保存、クエリ、分析するための専用サービス

<https://aws.amazon.com/jp/blogs/news/introducing-amazon-omics-a-purpose-built-service-to-store-query-and-analyze-genomic-and-biological-data-at-scale/>

Part 1: Amazon Omics のご紹介 – セキュアかつ大規模に生物配列データから洞察を得る

<https://aws.amazon.com/jp/blogs/news/part-1-introducing-amazon-omics-from-sequencedata-to-insights-securely-and-at-scale/>

Part 2: Amazon Omics を使用したゲノミクスデータの保管と分析のエンドツーエンド自動化

<https://aws.amazon.com/jp/blogs/news/automated-end-to-end-genomics-data-storage-and-analysis-using-amazon-omics/>

AWS HealthOmics - End to End ワークショップ

<https://catalog.workshops.aws/amazon-omics-end-to-end/ja-JP>



**Thank you!**