

製薬業界におけるゲノム解析・解釈のための イルミナインフォマティクス

古川 貴久 (Takahisa Furukawa)

イルミナ株式会社 大規模ゲノムプロジェクトチーム

インフォマティクス ストラテジックアカウントマネージャー

Agenda

SECTIONS

イルミナについて

ゲノミクスについて

AWS環境におけるイルミナの
インフォマティクスソリューション

製薬業界における活用事例

イルミナについて

illumina®



OUR MISSION

To improve human health
by unlocking the power of the genome

(ゲノムのパワーを解き放ち、人々の健康をより豊かにする)

<主な事業内容>

遺伝子変異や遺伝子機能の解析のためのライフサイエンスツールや統合システムの開発、製造、販売

illumina®



イルミナ会社概要

イルミナ株式会社



2004年1月
設立



100名 +
従業員



グレッチェン・シャンテール・
ウェイトマン
代表取締役



1,000台以上
シーケンサー
導入済



東京オフィス
大阪オフィス



NextSeq Dx
2018年クラスI医療
機器として届出

お客様：

国立研究センター、大学、病院、
企業（製薬会社、検査会社、食品企業など）

コラボレーター：

大塚製薬株式会社、シスメックス株式会社、
国立がん研究センターなど

illumina, Inc.



1998年4月
設立



約9,100名
従業員



サンディエゴ
カルフォルニア州
本社



18,000台以上
シーケンサー
導入済

約 4,970億円

2021 売上

約 4.4兆円

時価総額

受賞歴



最もスマートな企業 第3位

2016年 MIT テクノロジーレビュー



最もイノベーティブな会社 バイオテック部門 第10位

2017年 Fast Company

イルミナの各国のオフィス



United States

- San Diego (Headquarters)
- Foster City
- Hayward
- Baltimore
- Madison

Netherlands

- Eindhoven

Russia

- Moscow

Turkey

- Istanbul

Brazil

- São Paulo

United Arab Emirates

- Dubai

United Kingdom

- Cambridge

China

- Beijing
- Shanghai
- Guangzhou
- Hangzhou
- Taipei City

Belgium

- Mechelen

France

- Évry
- Rennes

Japan

- Tokyo
- Osaka

Germany

- Berlin

Singapore

Italy

- Milan

Australia

- Melbourne

Israel

- Tel Aviv

South Korea

- Seoul

ゲノミクスについて



ゲノミクスの応用分野



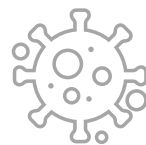
がん



治療法の選択と
モニタリング



遺伝病



希少疾患の診断の確立



NIPT



出生前診断の
リスクの最小化



集団ゲノム



医療システムの
質と効率の向上のための
データ蓄積と活用

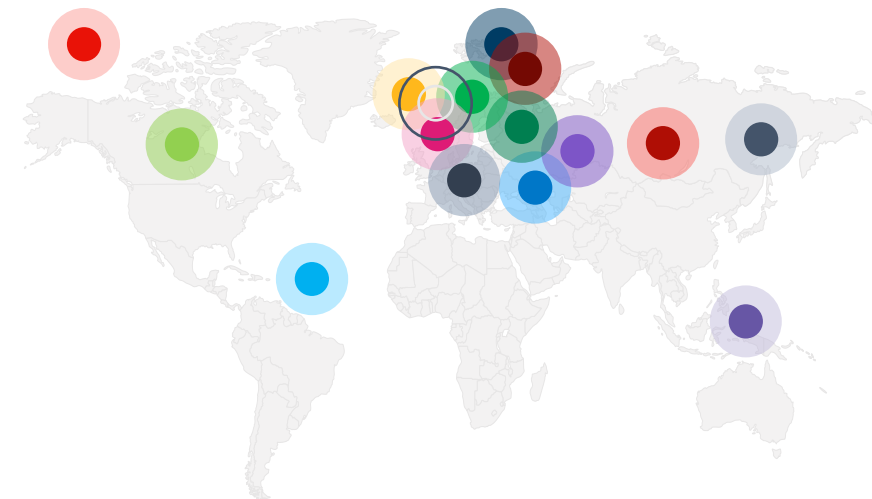
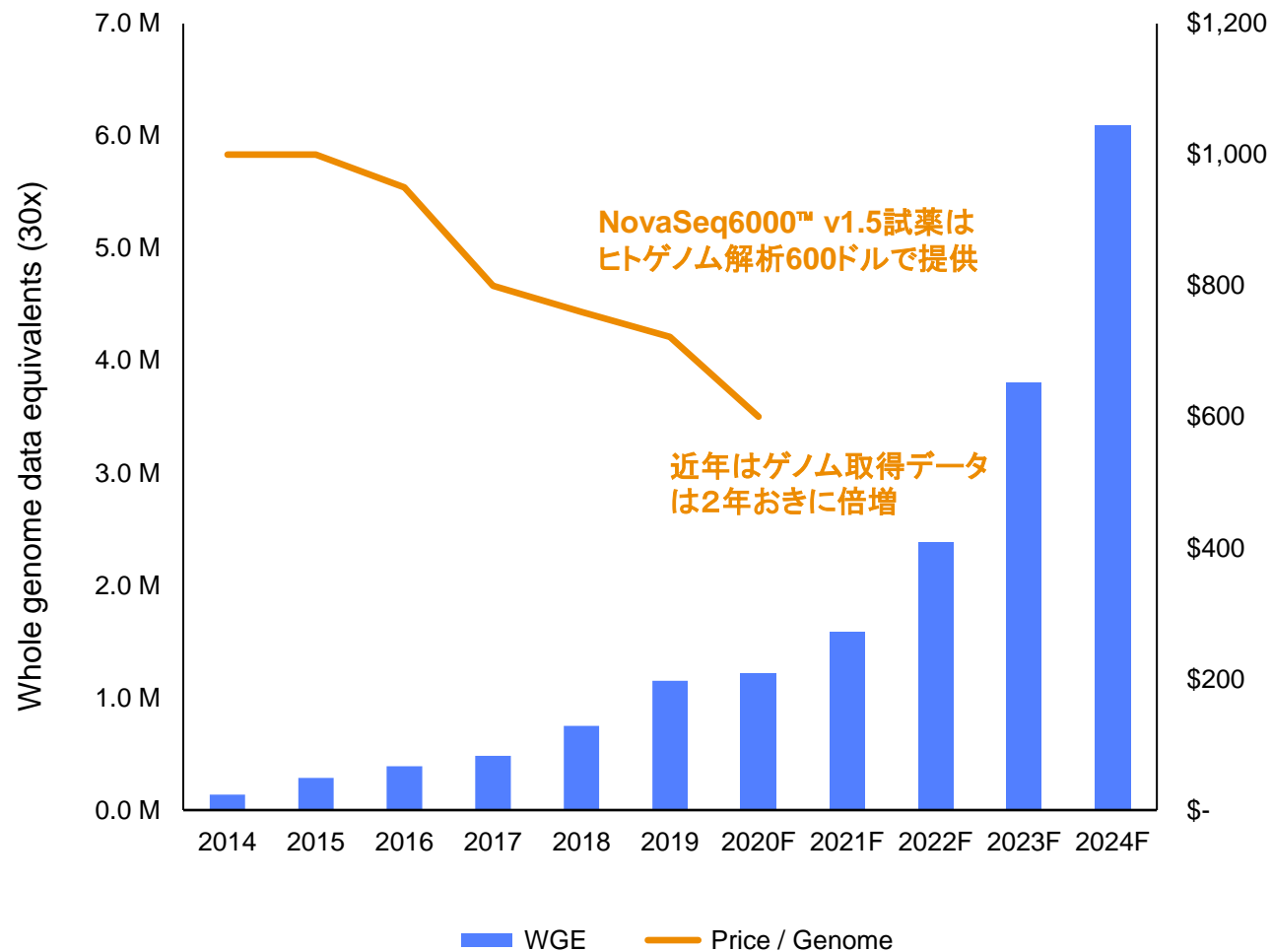


COVID-19 調査観察



今回のパンデミックに対処し、
将来のパンデミックを防止

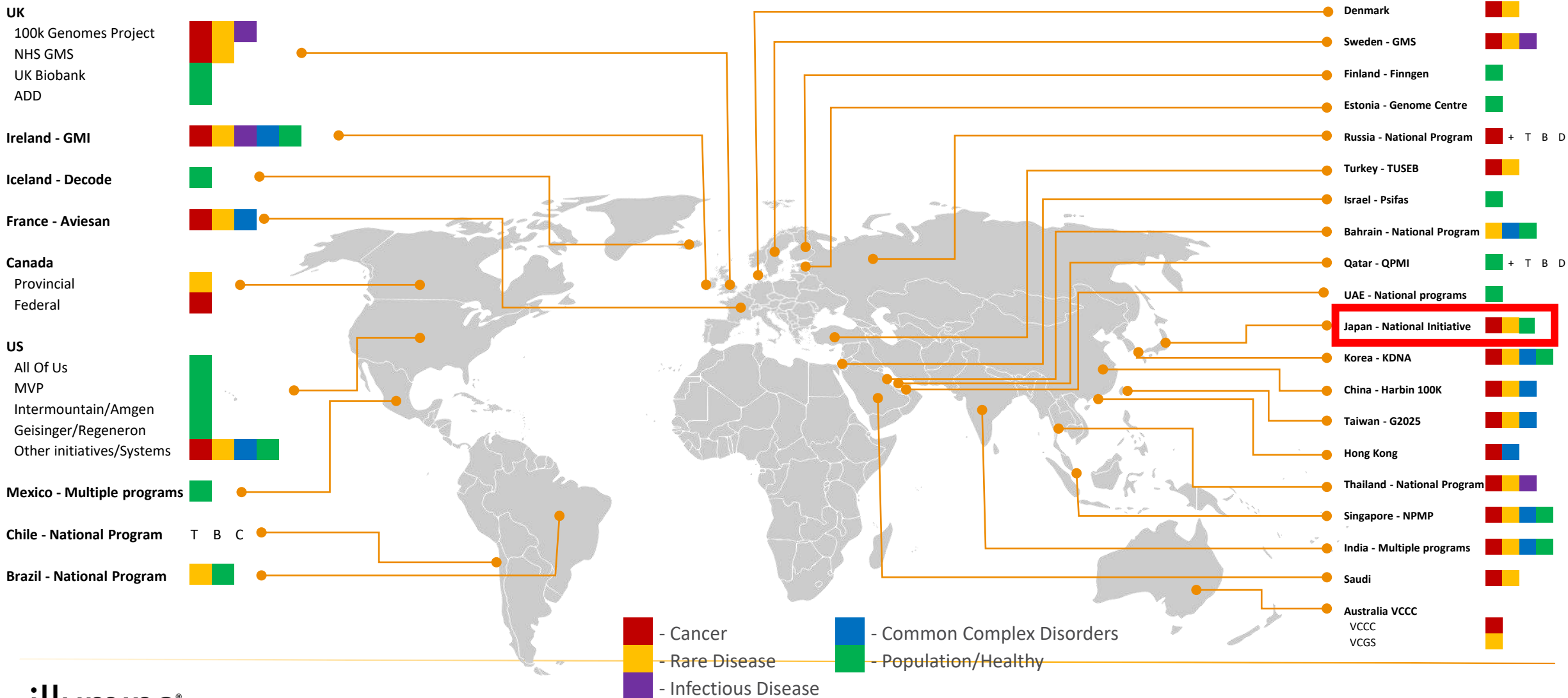
ゲノムデータ生成加速の継続



10M+
サンプル

50+
大規模ゲノムプロジェクト

50か国を超える多くの国々が大規模なプログラムを準備



急速に増加するデータ量がもたらす課題

急速に増加するデータ量は、強力なインサイトを引き出すためのユニークなデータの機会の創出とともにデータ解析・管理の課題も

>18,000

世界で利用される
イルミナシーケンサーの数



50倍

2010-2020年のバイオテクノロジーにおけるデータの成長

50-70 GB

30xのヒトゲノム解析占有するおおよそのストレージスペース

~150Pb

2019年にイルミナシーケンサーで生成されたデータ量



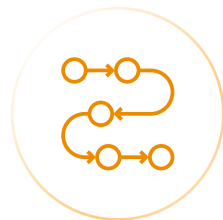
大規模ゲノムプロジェクトによりさらにその成長は加速

ゲノムデータ処理や管理のための課題



インフラの拡張

増大する大規模なデータに対応するためには、コストとリソースが必要



新しいワークフローの構築と展開、管理

多くのシステムや技術を駆使する必要があり、非効率になりやすい



セキュリティ、プライバシー、コンプライアンスの維持

多額の投資と定期的な監査チェックが必要となる



限られたインフォマティクスリソースの活用の最大化

グループ間で共有されてしまい、科学的な焦点と新しい洞察の推進を阻害

ENABLING TOMORROW'S DISCOVERIES

AWS環境におけるイルミナの インフォマティクスソリューション

illumina®



Illumina[®] Connected Analytics (ICA)

インフォマティクス活用と科学的発見を
促進するセキュアなゲノクスデータプ
ラットフォーム

- ① 解析パイプラインの構築とカスタマイズ
- ② ワークフローの大規模な実行
- ③ データおよび結果の探索と共有



ICAの世界での活用とAWS採用の理由



9

世界でのICAの配置数

AWS

クラウド環境

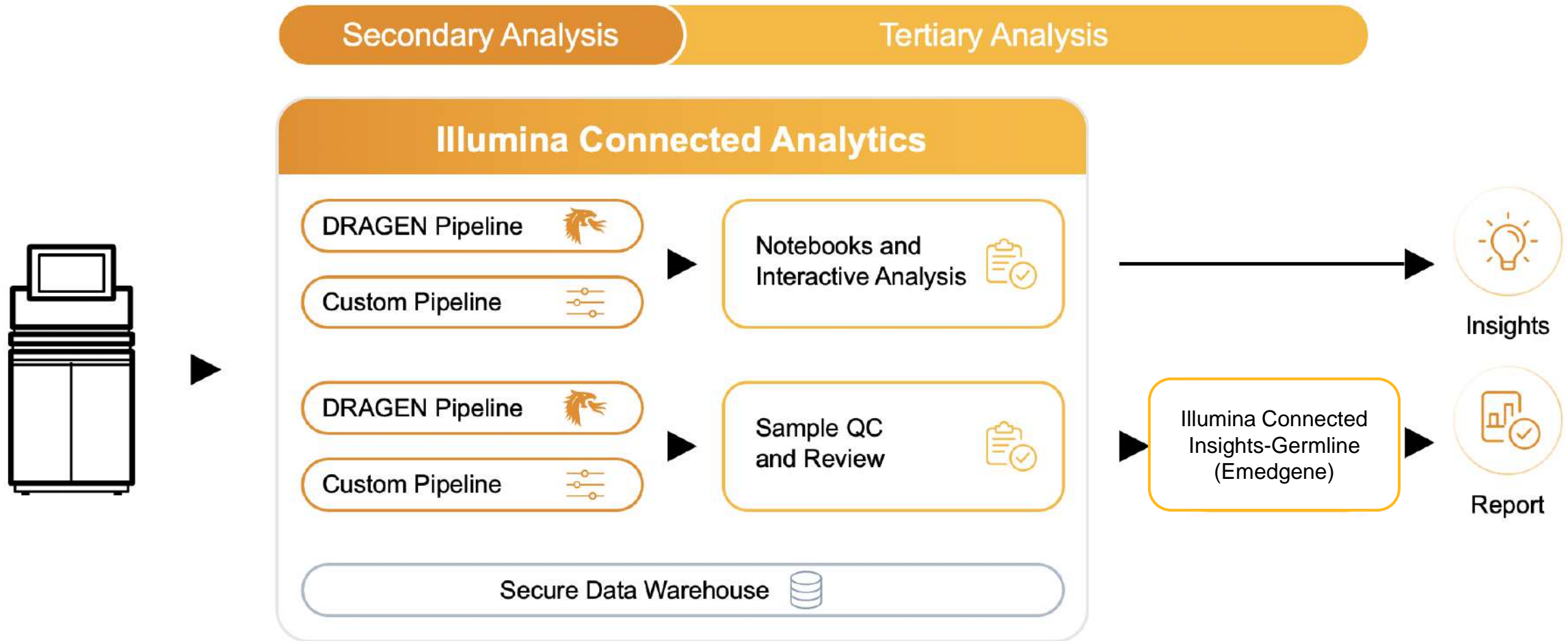
< 採用理由 >

- ・グローバルなフレームワーク
- ・2013年以来10年以上の活用の実績と経験
- ・EC2やS3 Glacierの利用によるコスト減
- ・ISO、HIPAA、GDPRなどのセキュリティ標準とコンプライアンス認定をサポート

● Available today

● Planned – 2023 (Subject to change)

ICAでのデータ処理の流れ



ICAによるインフォマティクスの効率化 主な特徴

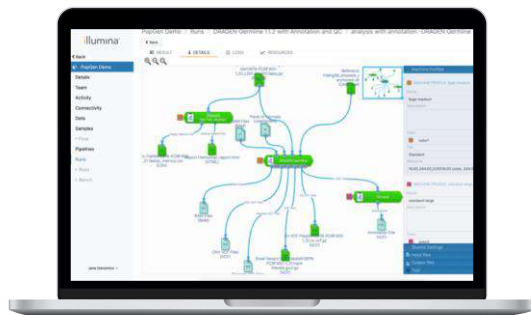
生産的なワークフロー



FPGAを用いるDRAGEN解析パイプラインの活用



CWLやNextflowなどのツールでワークフローを管理



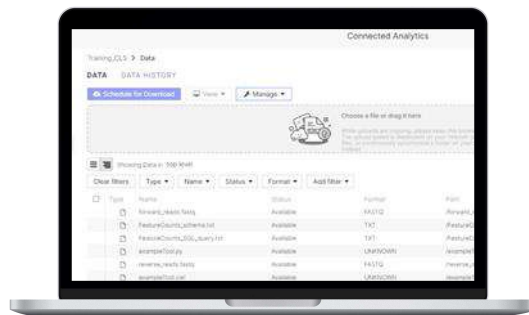
データの管理



安全なワークスペースに整理、コンプライアンスに則ってグローバルに共有可能



AWS S3にデータを保持しながら、ICA機能の利用が可能



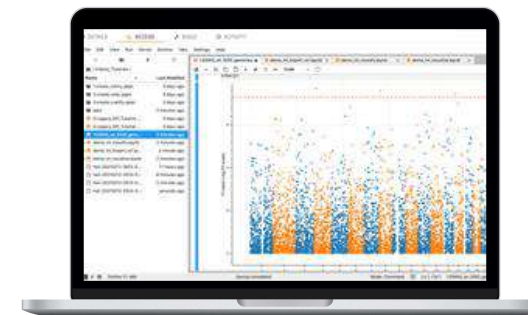
データの分析



Jupyterノートブックを含む柔軟なコンピューティング環境でのデータ解釈



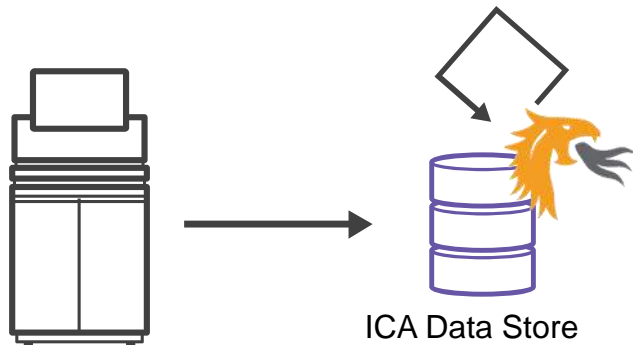
スケーラブルなデータウェアハウスでサンプルデータと母集団データを集計、クエリ、可視化





ICA上でDRAGEN™パイプラインへのアクセスが可能

DRAGENは、業界をリードする正確性で、
包括的かつ効率的な二次解析ソリューションを提供



Key Pipelines & Capabilities

- Germline
- Somatic
- Enrichment
- Joint Calling
- RNA-Seq
- Single-cell RNA
- Methylation
- Map & Align
- CNV, SV calling
- CYP2D6, HLA
- TMB, MSI

正確性

SNP、INDEL、CNVなどを高い感度と特異性で検出可能

包括的

全ゲノム、トランスクリプトーム、パネル、シングルセルデータ、メチル化データなどを1つのプラットフォームで解析可能

効率的

スピード: 30x全ゲノムあたり17分で解析可能

圧縮: FASTQのサイズをgz形式から最大1/5に圧縮

DRAGEN全ゲノム解析は集団ゲノム解析に最適

> 英国バイオバンク (UKBB)

UKBB は分析に DRAGEN バージョン 3.7.8 を使用します。

[UKBBについてもっと知る](#)

[コマンド](#)

> All of Us リサーチプログラム

All of Us Research Program では、DRAGEN バージョン 3.4.12 を使用しています。このプログラムは最近、DRAGEN バージョン 3.7.8 を検証し、必要な変更管理の承認を取得中です。

[All of Us リサーチプログラムの詳細](#)

[コマンド](#)

> PRECISE (シンガポール)

PRECISE (シンガポール) プログラムでは、DRAGEN バージョン 3.7.6 を全ゲノム サンプルの最初のパイロットとして使用し、現在は SG100K プロジェクトの一部として DRAGEN バージョン 3.7.8 を使用します。

[PRECISE \(シンガポール\) の詳細](#)

[コマンド](#)

<https://developer.illumina.com/dragen/dragen-popgen>

All of Us Research Program

Input

Download the input files

[Sample 1](#) [Sample 2](#) [Coverage Region 1 BED file](#)
[Coverage Region 2 BED file](#) [Cross Contamination VCF file](#) [Hash Table](#)

Final Command for DRAGEN v3.7.8 Production

```
dragen ¥  
-r <hg38-ref-dir/ ¥  
--fastq-list <path-to>/fastq_list.csv ¥  
--fastq-list-sample-id <id> ¥  
--output-directory <out-dir> ¥  
--output-file-prefix <prefix> ¥  
--enable-variant-caller true ¥  
--vc-emit-ref-confidence GVCF ¥  
--vc-enable-vcf-output true ¥  
--enable-duplicate-marking true ¥  
--enable-map-align true ¥  
--enable-map-align-output true ¥  
--output-format CRAM ¥  
--vc-hard-filter "DRAGENHardQUAL:all:QUAL<5.0;LowDepth:all:DP<=1" ¥  
--vc-frd-max-effective-depth=40 ¥  
--vc-enable-joint-detection true ¥  
--qc-cross-cont-vcf <path-to>/SNP_NCBI_GRCh38.vcf ¥  
--qc-coverage-region-1 <path-to>/wgs_coverage_regions.hg38_minus_N_interval_list.bed  
¥  
--qc-coverage-reports-1 cov_report ¥  
--qc-coverage-region-2 <path-to>/acmg59_all_ofus_19dec2019.GRC38.wGenes.NEW.bed ¥  
--qc-coverage-reports-2 cov_report ¥  
--qc-coverage-ignore-overlaps true  
--qc-coverage-count-soft-clipped-bases true  
--read-trimmers polyg  
--soft-read-trimmers none
```

[Copy to clipboard](#)

Output

Compare your VCF files to HG002 sample output data [here](#)

ICAのクラウド環境でDRAGEN™パイプラインを利用するメリット



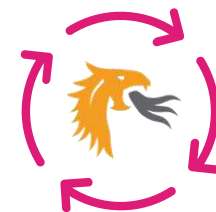
容易なスケールアップ、高速化

DRAGENパイプラインの並列実行による
解析スループットの最大化



費用対効果の向上

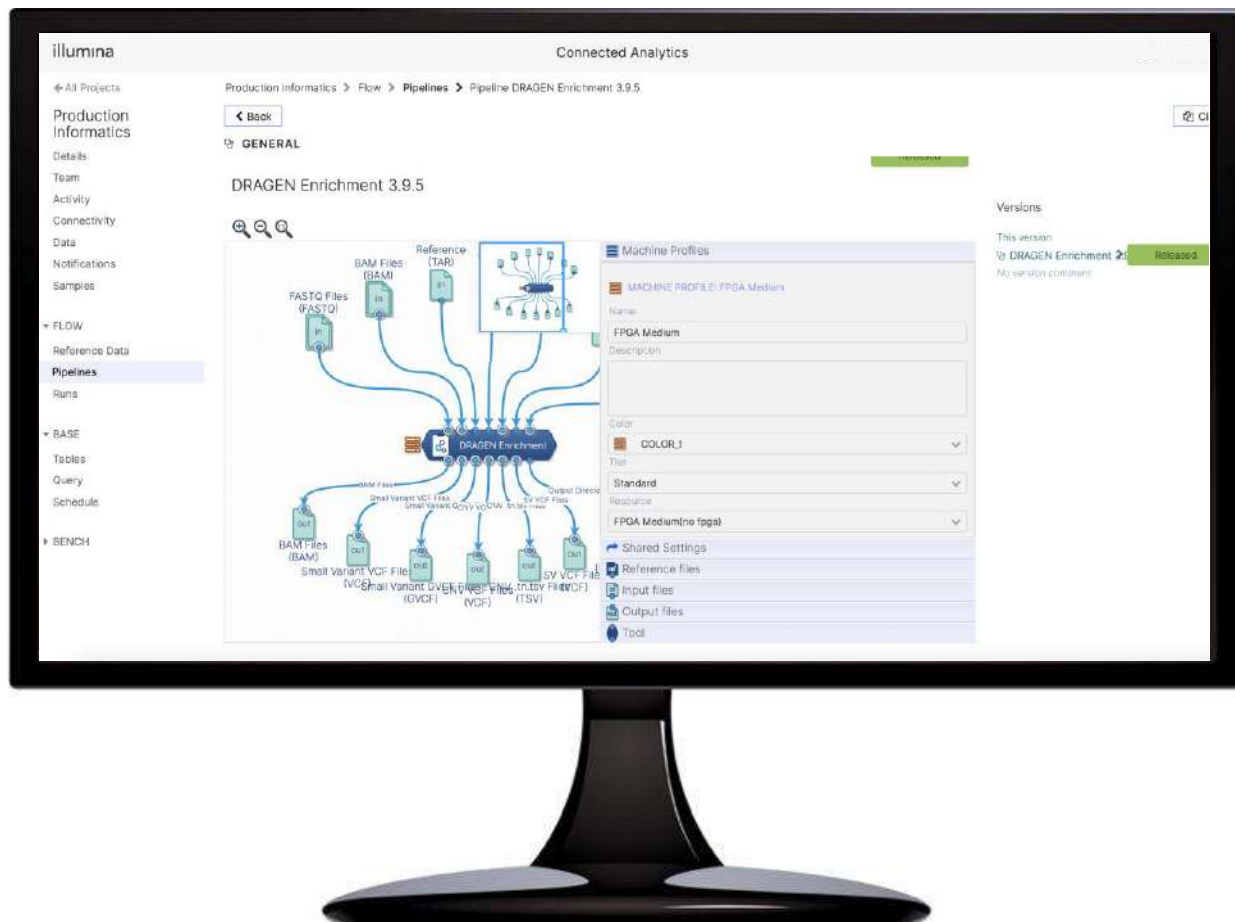
ローカル環境やAMIを利用する際に必要な
DRAGEN解析のライセンス購入必要なし
(ICAライセンス費用と計算・管理費用は別途必要)



簡単な導入

ハードウェアのメンテナンスが不要
ソフトウェアのアップグレードが不要
ブラウザでDRAGENの機能をすぐに利用可能

分析パイプラインの構築、展開、バージョン管理



DRAGEN

解析にFPGAを用いるDRAGENパイプラインが利用可能
既存パイプラインをお客様のカスタムパイプラインの一部に組み込むことも可能



nextflow



パイプラインのインポート、構築、カスタマイズ

CWL、Nextflow、Docker、CLI、API など業界標準のツールを使用して、カスタムパイプラインの構築や実行可能



パイプラインの作成やバージョン管理

検証済みバージョンのロック、自動化の設定(CLIやAPIを使用)、計算リソースの制御したワークフローの運用などが可能

セキュアでコンプライアンスに準拠したデータ環境



AWS環境の利用により、安全なグローバルフレームワークを容易に実現



ISO27001および13485認証取得の情報セキュリティおよび品質管理システム



全世界のインスタンスでGDPRプライバシー規制の原則に準拠



規制環境下でのHIPAAコンプライアンスをサポート

*Health Information Portability and Accountability Act; only applicable in the US.

データの維持・管理

お持ちのAWSのS3バケットをICAで接続することでデータ管理とコスト管理を実現



お持ちのS3上のデータに対するICA解析ワークフローの実行

ICAが管理するツールと計算リソースで外部管理データを処理



双方向のデータフロー

データを分析ツールに読み込み、結果をバケットに書き戻すことが可能



ストレージコストをS3で直接管理

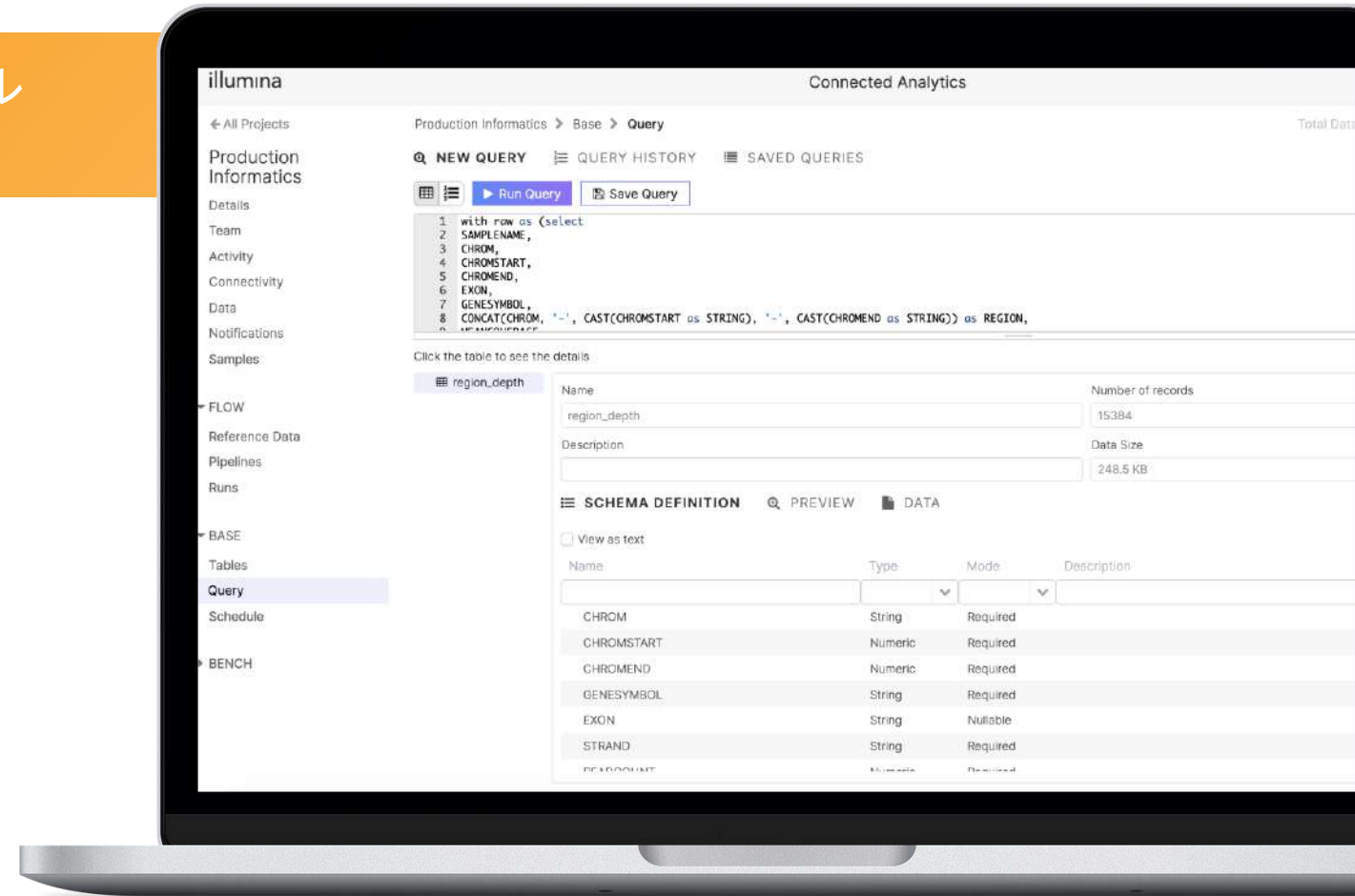
AWSアカウントに保存されたデータを完全に制御し、コストを管理することが可能



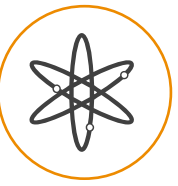
ICA Base: 大規模スケーラブルなデータウェアハウス

データウェアハウス内の複雑なマルチモーダルデータセットの集計やクエリが可能

- **カスタムデータテーブルの作成が可能**
- NGS、アレイ、プロテオミクス、RWE/RWD*などのマルチモーダルデータを集計可能 (Ex. csv, tsv, VCF, etc)
- **簡単なデータ取り込み**
- 解析結果の自動取り込み可能
- カスタムデータセットやパブリックデータセットを組み込んだのクエリを強力にサポート
- **Association Studiesのためのクエリ**
- GWASやPheWASのクエリをオンデマンドで実行可能
- SQLクエリを保存して共有し、将来に利用することが可能



*Real World Evidence, Real World Data



ICA Bench: データサイエンスの共同ワークスペース

新しいインサイトを生み出すセキュアでコラボレーティブなノートブック環境を提供

GWAS analysis

Builder for Python and R - 2.0.0

Demo of GWAS analysis on subset of 1000 genomes data using HAIL

Medium 5 GB

Stopping

デジタルノートブック(JupyterLab notebook)によるインタラクティブなデータ処理

- 社内外のあらゆるデータ、パイプライン、データベースを取り込み可能

様々なデータ処理に対応

- 標準的なゲノミクスおよびディープラーニングライブラリをサポート
- RとPythonがインストールされた標準ワークスペースを利用可能
独自のワークスペースを設計することも可能

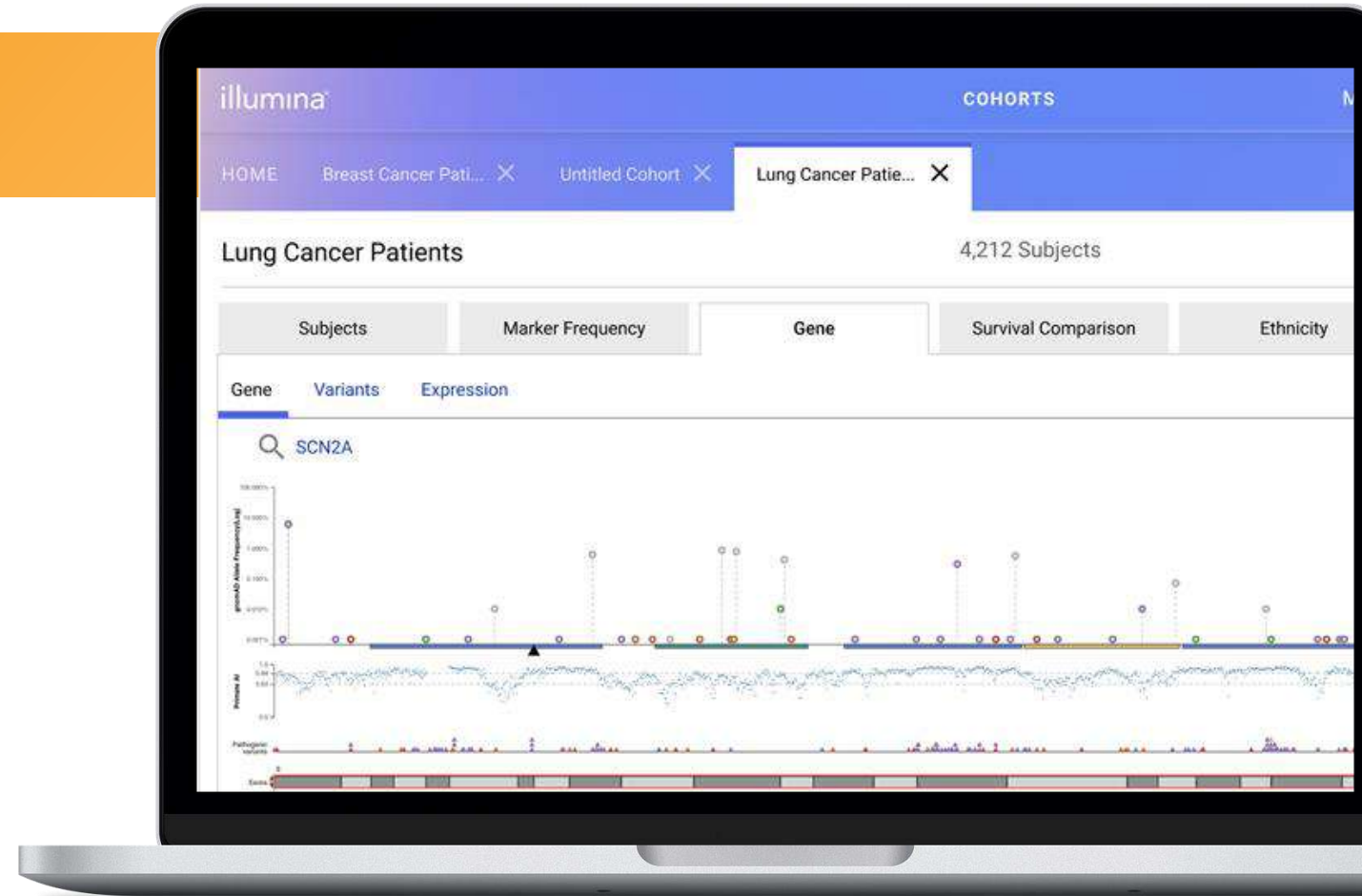
ICA Cohorts

集約されたゲノムデータを解析して
遺伝子型と表現型の関係を理解

個々のサンプルおよびコホート全体のアップロード
UI、CLI、またはカスタマイズされたパイプラインの一部を利用

既存もしくはアップロードしたコホートデータをブラウジング
数値、ブーリアン、階層的な属性を使用してサンプルのサブセットを選択

遺伝子ごとのバリエーション表示
表現型や他の属性との潜在的な関連性を発見する



ENABLING TOMORROW'S DISCOVERIES

活用事例

illumina®

4

ゲノム医療による癌患者への貢献



早期発見



包括的な評価



高感度モニタリング



治療方法の特定



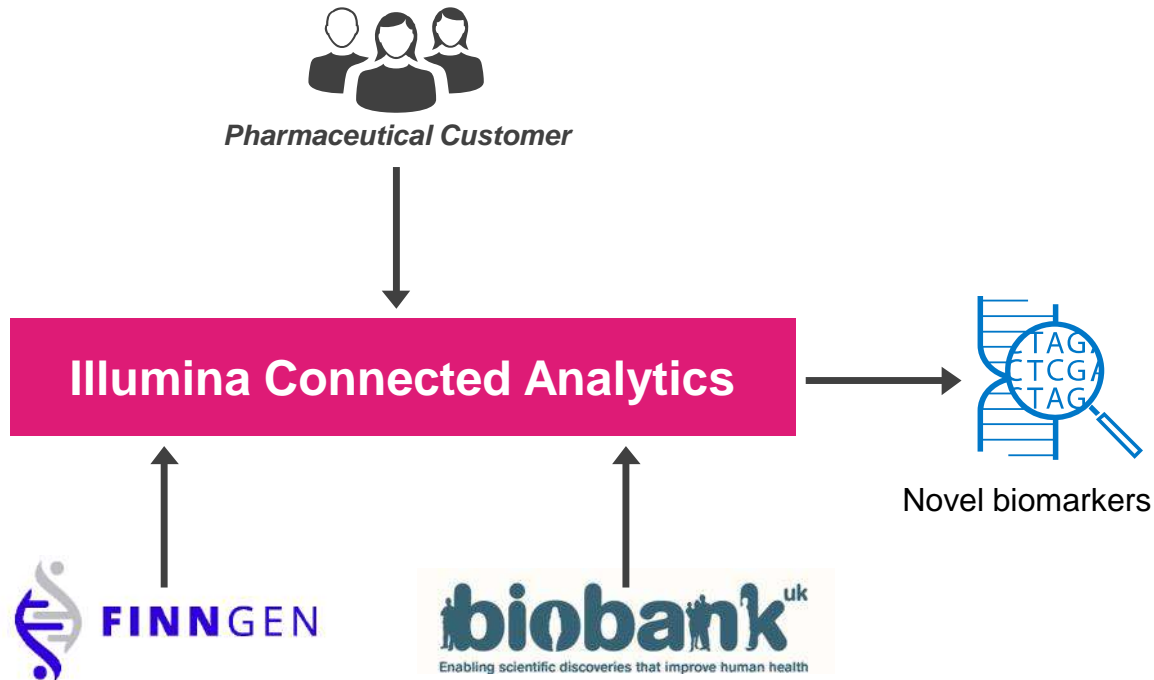
パートナー例





ICAのバイオマーカーとターゲット探索での活用事例

製薬企業で、公共データソースから80万人の被験者と1万人の表現型を組み合わせ、1兆4千億行のデータを探索し、医薬品開発および投資の意思決定に活用



目標

- 集団規模のゲノムデータセットを用いて、糖尿病やその他の疾患領域における新規バイオマーカーを探索

ワークフロー

- 80万人の被験者と1万人の関連する表現型の組み合わせ
- 20Mバリエーションに対するGWASとPheWASの実施
- データウェアハウス内の1.4T行を調査

今後の取り組み

- データ、ノートブック、AI、その他のデータサイエンスツールセットを活用し、医薬品開発をさらに促進
- 投資判断のための臨床試験結果の予測にゲノムを活用するワークフローを開発

イルミナによるゲノミクスに基づく創薬への取り組み

>\$123B

創薬における年間
研究開発費

>\$2B

創薬におけるゲノミクスの
市場規模

>2×

ゲノミクスによる成功率の向上

原因が複雑な疾患ではまだまだ治療薬
の開発が求められています

創薬の大きな機会と貢献の可能性

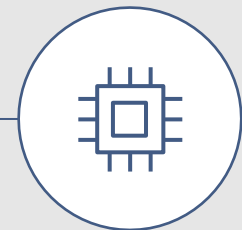
illumina®



NovaSeq™ 6000



DRAGEN™ ICA



AIやMLに基づくゲノム解釈

イルミナの創薬開発プラットフォーム

製薬企業様

AMGEN

REGENERON AstraZeneca

大規模コホートプロジェクト



NEW PARTNER

研究用のみで、診断目的には使用できません。



ICAを活用したデータ解析や共有で新たな発見を推進

医療システム／国家プログラム
データ所有者

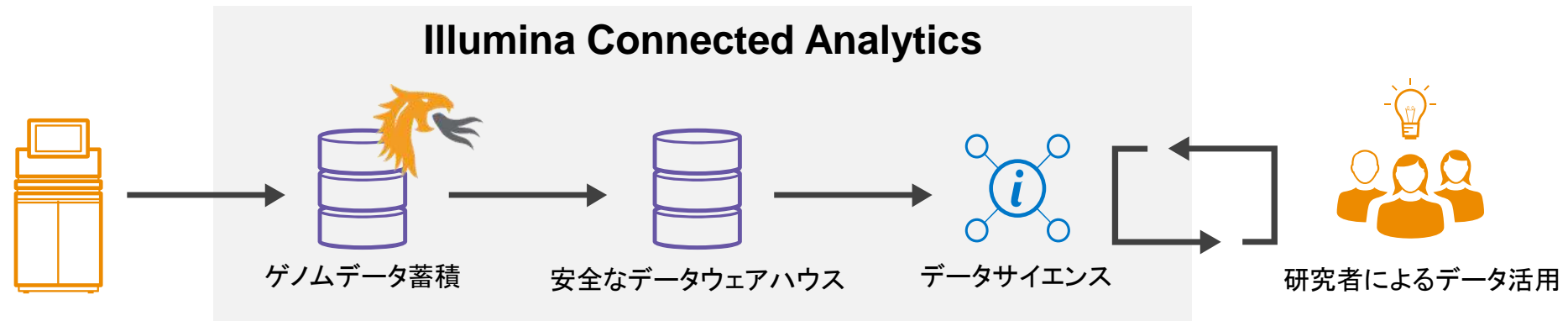
GOAL

- 5年以内に15K+のWGSを配列する
- 産業界のパートナーを通じてデータセットを収益化
- 追加プロジェクトの資金獲得

製薬会社・バイオテクノロジー企業・アカデミア
研究者、データサイエンティスト

GOAL

- 発見目的のためのデータマイニング
- 結果予測、リスク層別化による価値創出



ご清聴ありがとうございました

お問い合わせ先 古川貴久 (Takahisa Furukawa)

tfurukawa@illumina.com