

Illumina Connected Analytics

大規模インフォマティクス ワークフローを生み出す

- CWL (Common Workflow Language)やNextflowなどのツールを用いたワークフローのインポート、構築、編集
- 安全なワークスペースでデータを整理し、コンプライアンスに遵守した形でグローバルに共有
- JupyterLab Notebookを含む柔軟なコンピューティング環境でデータを解釈

illumina®

はじめに

次世代シーケンサー(NGS)テクノロジーの進歩により、ライフサイエンス研究や臨床研究の実施速度は劇的に変化してきました。シーケンスの高速化とコスト低下により、データ生成能力はデータから生物学的洞察と臨床的洞察を抽出する能力をはるかに上回ると考えられます。安全なデータ管理、インフラストラクチャーの拡大、新しいインフォマティクスワークフローの構築と展開という課題を解決するには、柔軟かつ包括的なプラットフォームが必要です。Illumina Connected Analytics (ICA)では、ユーザーが規模を拡大してデータプライバシー、セキュリティおよびコンプライアンスを維持しながら、柔軟な解析パイプラインを構築、バージョンアップ、展開することができます。

ICAは、インフォマティクスの運用と科学的洞察を推し進める、セキュアなゲノムデータプラットフォームです。(図1、表1)。ICAにより以下のことが可能になります:

- 解析パイプラインの構築とカスタマイズ
- 大規模なプロダクションワークフローの実行
- データと結果の探索と共有

効率化されたワークフロー

ICAはイルミナのシーケンスシステムを用いてNGS研究を実施しているラボにとって、中心的なコンポーネントです。クラウドコンピューティングによるリソースの順応性を生かし、複雑なシングルセルプロジェクトの何万もの細胞の不定期なスクリーニングから、集団規模の全ゲノムシーケンスまで、同一のアーキテクチャーを用いて、あらゆる規模でオペレーションをサポートします。ユーザーはシームレスに手持ちの機器をICAと統合できます。

ICA内部では、指定したワークフローに応じて、すぐに使えるDRAGEN™パイプラインまたはカスタムパイプラインでデータを自動的に解析できます。幅広い解析オプションは、品質管理からデータアグリゲーションや高度なデータサイエンスツールにまで及び、迅速かつ拡張性のあるデータ処理を実現します。ICAは、豊富なRESTfulアプリケーションプログラムインターフェース(API)およびコマンドラインインターフェース(CLI)ツールを備え、拡張性のあるプラットフォームを提供します。これらのAPIは、データのライフサイクル全体で転送、アクセス、使用されるときにワークフロー効率を最大にし、Global Alliance for Genomics and Health(ゲノミクスと健康のための世界連合、GA4GH)に遵守したAPIも含んでいます。¹

表1: ICAの概要

	特長	利点
セキュリティおよびプライバシー	コンプライアンス	ローカル、地域、世界的な規制基準、HIPAAコンプライアンスおよびGDPR原則、ISO 27001認証に遵守
	セキュリティ管理	厳格なデータ分離、「転送中」(TLS 1.2) および「保管中」(AES 256)のデータ暗号化を維持
	監査証跡	誰がどのデータにいつアクセスしたかを追跡する活動記録を維持
	シングルサインオン(SSO)(オプション)	アクセスを管理するために施設承認情報を活用
リソース	要望に応じた計算リソース	パイプラインエンジン中の計算リソースのみの支払いで、費用を削減
	要望に応じて規模を拡大	現在要求されるレベルのクラウドストレージと計算ニーズに応じた規模拡大
	プラットフォームおよびユーザーダッシュボード	ニーズを効率的に理解、管理、予測するためにリソース要求を視覚的に表示
管理	プロジェクトおよびユーザーの管理	細やかなプライバシーに対するユーザーアクセスとアクティビティの管理
	データ共有	大規模なグローバルな連携のためのデータサイロの橋渡し
	データアーカイブ	低コストのストレージ階層に使用しないデータを保存することでコスト削減
使いやすさ	シーケンスシステムの統合	イルミナシーケンスシステムからシームレスにデータを転送
	ビジュアルパイプラインビルダー	コードを書かずにパイプラインを作成
	ツールおよびパイプライン	追加設定なしのパイプラインを活用し、カスタムツールをインポート
	API および CLI	お客様の要望に合ったツールで、プラットフォームとプログラマティックにインタラクト
	「Bring your own bucket」	プライベートに管理されたクラウドアカウント内に保存されたデータにアクセス
高度なツール	データ視覚化	動的な視覚化プロットおよびインタラクティブなウェブアプリを作成し、R および Pythonパッケージを用いてデータを表示
	Docker、Nextflow および CWL に対応	Common Workflow Language でパイプラインを書き出し、クラウドで簡単に解析を起動
	RESTful、GA4GH- 準拠 API	ツールとデータへのプログラムアクセス、およびその他のソフトウェア環境との相互運用が可能
	JupyterLab との統合	高度なデータ解析の実施。R および Python を用いた AI/ML モデルの構築と訓練
	データアグリゲーションおよびクエリー	SQL を用いて集団レベルのデータクエリーを実行



図1: ICAはデータの管理と解析のための基盤を形成

リードからデータへの変換

ICAはデータ二次解析用のさまざまなオプションがあり、リードから結果取得までのワークフローを効率化します。ICAは、既存のパイプラインや製品を使用し、カスタマイズされたパイプラインを構成するための柔軟性を備えているため、実質的にあらゆるインフォーマティクスアプリケーションに対応できます。

すぐに使えるオプション

ICAは、シーケンスデータの迅速かつ正確な二次解析を実施するDRAGEN Bio-IT Platform²へのアクセスをはじめとした、パワフルですぐに使えるデータ処理ツールおよびパイプラインを提供します(図2)。

パイプラインのカスタム化

バイオインフォーマティクスはDockerイメージレポジトリから既存のツールをインポートしたり、Nextflow、CWLおよびグラフィカルパイプラインエディターを使って新しいパイプラインを構築したり編集したりできます。ラボオペレーターやその他のサイエンティストは、直感的にデザインされたユーザーインターフェースを使って簡単にパイプラインを起動することが可能です。

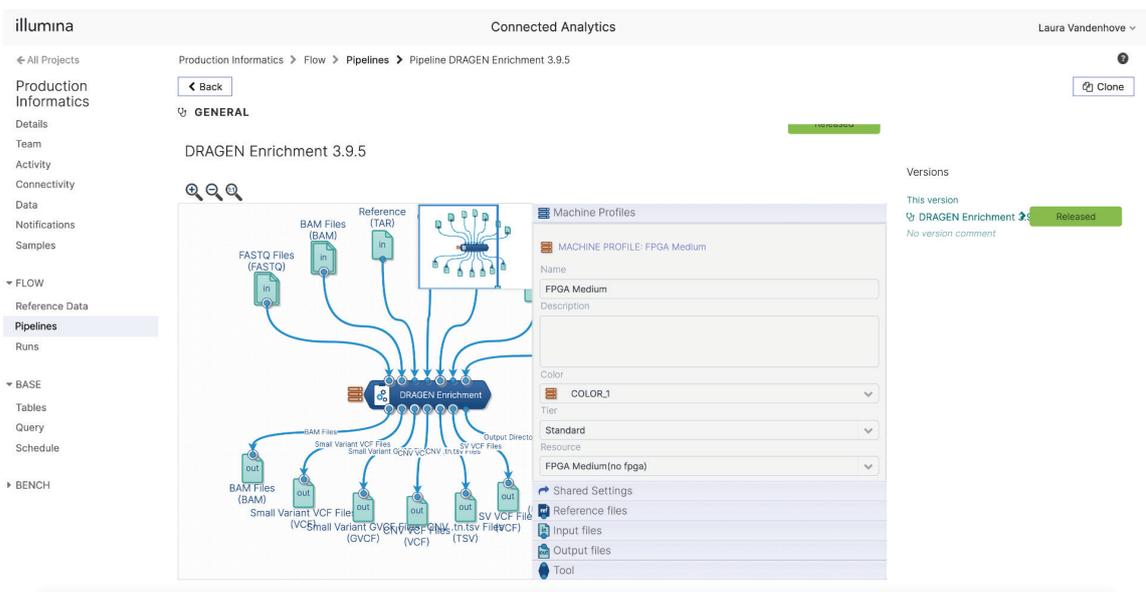


図2: ICA内のDRAGENパイプライン: ICA内のすぐに使えるDRAGENパイプラインは、リードからレポート作成まで迅速かつ正確な二次解析を実現します。

データ管理と制御

データの増加に伴い、個々のデータセットの価値を増大させるために、科学界の中でのデータ共有、再利用、統合に対応するためのインフラストラクチャーに対するニーズが高まっている時代が到来しています。このニーズに対処するために、ICAはデータ管理におけるベストプラクティスの導入が実現するようにデザインされた、いくつかの性能を搭載しています。

アクセス制限

きめ細かいアクセス制限により、管理者が許可を設定し、既存の施設承認の利点を生かしてアクセスを制限することが可能です。監査ログはイベントおよび変更の記録として役立ち、プラットフォームを使用しながら、そのプラットフォームにアクセスしたときの各ユーザーのその行動を記録し、コンプライアンスと説明責任の実施を実現します。

オープンフォーマット

ICAはデータに依存しないプラットフォームとして設計されています。分子データ、臨床データ、表現型データおよび画像などの非構造化データを含む多種類のデータ解析に対応します。

コラボレーション

ICAはコンプライアンスを維持する方法で地理的境界線全体にわたるコラボレーションを後押しします。データおよびツールを、データ完全性とプライバシーを維持した形で瞬時に送信し、他のユーザーと共有できます。さらに、解析と共有のために、外部クラウドソースにホストされたデータツールと解析ツールをICAにインポートできます。

データアグリゲーションおよびクエリー

ICAは複雑なアグリゲーションステップと統合ステップを自動化し、数百万個のサンプルのデータを網羅する機能的なナレッジマネジメントシステムを構築します(図3)。実質的にあらゆる種類のデータ、ジェノタイプ、表現型、メタデータ、アノテーション、その他の関連する情報を取得することができます。ユーザーは自身のデータモデルを定義し、自身のクエリーを書き出し、必要なデータセット間の関連を探索できます。ICAに集積されたデータは、新規バイオマーカーの探索、患者集団の層別化、経時的なアッセイ性能のモニタリングなどに利用できる豊富な情報を含んでいます。

The screenshot displays the Illumina Connected Analytics web interface. The top navigation bar includes the logo, the product name 'Connected Analytics', and the user name 'Laura Vandenhove'. The main content area is titled 'Production Informatics > Base > Query'. It features a sidebar with navigation options like 'All Projects', 'Production Informatics', 'Details', 'Team', 'Activity', 'Connectivity', 'Data', 'Notifications', 'Samples', 'FLOW', 'Reference Data', 'Pipelines', 'Runs', 'BASE', 'Tables', 'Query', 'Schedule', and 'BENCH'. The central panel shows a 'NEW QUERY' editor with a SQL query: `1 with row as (select`
`2 SAMPLENAME,`
`3 CHROM,`
`4 CHROMSTART,`
`5 CHROMEND,`
`6 EXON,`
`7 GENESYMBOL,`
`8 CONCAT(CHROM, '-', CAST(CHROMSTART as STRING), '-', CAST(CHROMEND as STRING)) as REGION,`
`9 FROM ...)`. Below the query editor, there's a table with columns 'Name' and 'Number of records'. The 'region_depth' table is highlighted, showing 15384 records and a data size of 248.5 KB. At the bottom, a 'SCHEMA DEFINITION' table is visible, listing fields like CHROM, CHROMSTART, CHROMEND, GENESYMBOL, EXON, STRAND, and REFERENCE with their respective types and modes.

図3: ICAはデータアグリゲーション、検索および継続的な学習が可能: ユーザーはデータセット間の関連を探索し、ユーザー主導型の課題に応えることができます。

洞察を押し進めるためのセキュアなノートブック環境

進行中の無数のデータ探索には、アルゴリズムを開発しカスタマイズする能力が必要不可欠です。データサイエンティストは、定評のある JupyterLab Notebook (PythonおよびR) を活用した、インタラクティブなプログラミングモジュールにより、シームレスかつセキュアな環境で集積されたデータを解析できるようになります(図4)。

メソッドおよびアルゴリズムの開発段階では、サンドボックス環境でパイプラインを開発、変更できます。その環境では、ニーズに応じて機械学習モデルを迅速に構築、検証、反復することができます。ユーザーは TensorFlow³ または scikit-learn⁴ などの幅広い標準的なライブラリーにアクセスし、自身のカスタムライブラリーにそれらのライブラリーを簡単に持ち込むことができます。生産段階に移行する準備ができたときに、ICAによってノートブックをツールに変換できます。この後、これらのツールはICAツールのレポジトリで使用できるようになり、生産パイプラインに組み込まれることになります。

コアとなるセキュリティおよびコンプライアンス

研究、臨床治療およびヒトの診断を目的としたゲノムデータを扱う際は、セキュリティは最も重要です。ICAは次のようなさまざまなデジタル対策および管理手段を使用し、最も厳しいデータセキュリティ要件をも満たします:

- シーケンス機器からアップロードされたデータを AES 256 標準を使って暗号化し、トランスファーレイヤーセキュリティ (TLS) で保護
- ICA内のデータはアマゾンウェブサービス (AWS) 上でホストされ、AWS Well-Architected のベストプラクティスを用いてさまざまな業界認証セキュリティ基準への遵守を維持⁵
- 認証サービスは SAML 2.0 でサポートされ、施設ユーザーおよびパスワードを管理 (オプション)
- データ起源のトレーサビリティをサポートする監査証をレポート

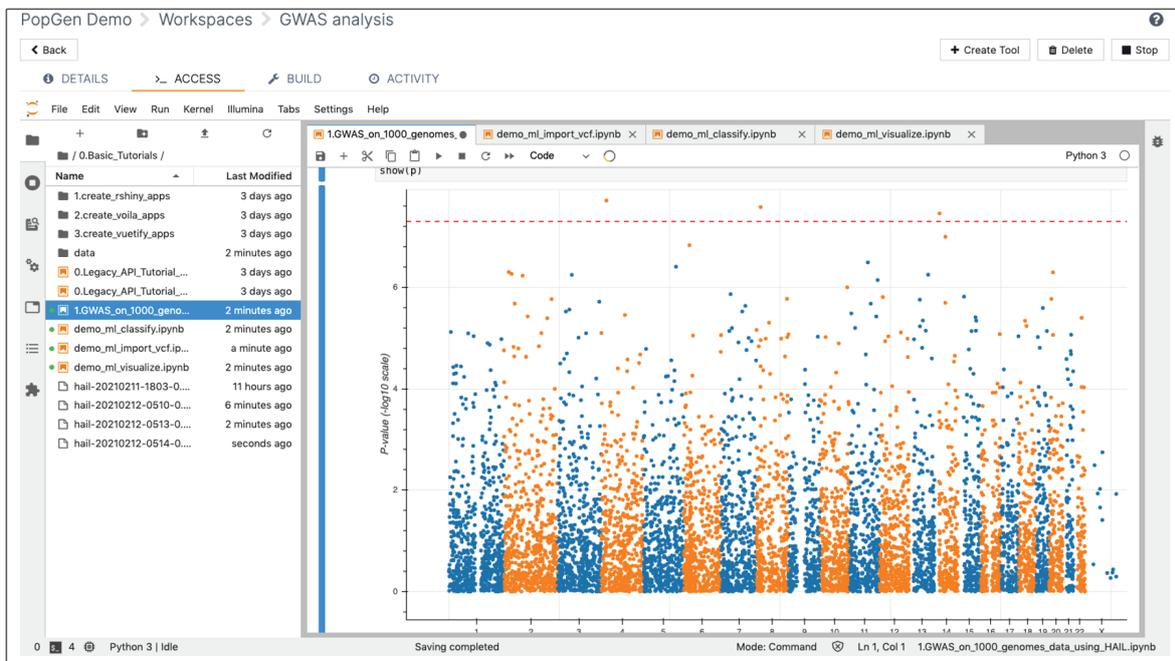


図4: インタラクティブ解析と視覚化: ICAは多次元データの視覚的探索のためにJupyter Notebookの使用に対応します。

ICAは、規制された環境でオペレーションを行う、次のような厳しい環境を遵守する必要があるカスタマーもサポートします：

- EU一般データ保護規則(GDPR)⁶および医療保険の携行性と責任に関する法律(HIPAA)などのデータ保護に関する現行の法律⁷
- 国際標準化機構(ISO)27001情報セキュリティマネジメントシステム⁸
- ローカルの規制およびコンプライアンス要件に対応するための保証付きのデータレジデンシー

製品情報

製品名	カタログ番号
ICA Professional Annual Subscription	20044876
ICA Enterprise Annual Subscription	20038994
ICA Enterprise Compliance Add-on	20066830
ICA Training and Onboarding	20049422

詳細はこちら

jp.illumina.com/ConnectedAnalyticsをご覧ください。

参考文献

1. Enabling responsible genomic data sharing for the benefit of human health. Global Alliance for Genomics & Health website. www.ga4gh.org. Accessed October 22, 2020.
2. Illumina DRAGEN Bio-IT Platform | Variant calling & secondary genomic analysis. Illumina website. jp.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html. Accessed October 22, 2020.
3. TensorFlow. TensorFlow website. tensorflow.org. Accessed January 11, 2021.
4. scikit-learn: machine learning in Python. scikit-learn website. scikit-learn.org/stable/. Accessed January 11, 2021.
5. Cloud Security—Amazon Web Services (AWS). Amazon website. aws.amazon.com/security. Accessed October 22, 2020.
6. General Data Protection Regulation (GDPR) Compliance Guidelines. GDPR website. gdpr.eu. Accessed January 11, 2021.
7. US Department of Health & Human Services. Health Information Privacy. HHS website. hhs.gov/hipaa/index.html. Accessed January 11, 2021.
8. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. ISO website. iso.org/isoiec-27001-information-security.html. Accessed January 11, 2021.
9. iCredits for Data Storage and Analysis | Illumina Analytics. Illumina website. jp.illumina.com/products/by-type/informatics-products/credits.html. Accessed October 22, 2020.

イルミナ株式会社

〒108-0014 東京都港区芝 5-36-7 三田ベルジュビル 22 階
Tel (03) 4578-2800 Fax (03) 4578-2810
jp.illumina.com

 www.facebook.com/illuminakk

販売店

本製品の使用目的は研究に限定されます。診断での使用はできません。販売条件：jp.illumina.com/tc

© 2022 Illumina, Inc. All rights reserved.

すべての商標および登録商標は、Illumina, Inc. または各所有者に帰属します。
商標および登録商標の詳細は jp.illumina.com/company/legal.html をご覧ください。
予告なしに仕様および希望販売価格を変更する場合があります。

illumina[®]