

データシート

DRAGENTM 二次解析

次世代シーケンスデータを使用した、
精確で包括的かつ
効率的なバリエーションコール

illumina[®]

はじめに

次世代シーケンサー (NGS) を介してゲノムの力を解き放つことは生物医学研究とプレジジョンメディシンの進歩に必要不可欠です。NGSから最大の遺伝学的洞察を得るために、研究者には精確かつ効率的に生のシーケンスデータを意味のある結果に転換することができるデータ解析ツールが必要です。さらに、NGSを利用するために施設や組織が必要とするのは、幅広いユーザーに対応し、経済的負担が少なく、採用までの技術的障壁が低い、簡単に使えるソリューションです。

イルミナのDRAGEN (Dynamic Read Analysis for GENomics) 二次解析は、全ゲノム、エクソーム、トランスクリプトームおよびメチローム研究などの幅広いアプリケーションに対してNGSデータ解析に関連する重要な課題に取り組むために開発されました。DRAGEN二次解析ソフトウェアは、NGSデータを処理し、洞察を促すための三次解析を可能にするアプリケーションスイートです。使用可能なツールは非常に高い精度で、包括的かつ効率的なソリューションから成り、あらゆる規模や分野のラボがゲノムデータを用いて、より多く解析することが可能です。

精確な結果

DRAGEN二次解析は非常に精確な結果を生成します。DRAGEN二次解析v3.7は、2020年のPrecision FDA Truth Challenge V2 (PrecisionFDA V2) において、イルミナシーケンスデータを用いて、すべてのベンチマーク領域とマッピングが困難な領域で最も精確な結果を勝ち取りました。^{1,2} その後のリリースでは、機械学習 (ML) やDRAGENマルチゲノム (グラフ) テクノロジーなどの分野での進歩により、新たなスタンダードとなる精度を設定し続けています。DRAGEN二次解析v4.3の最新バージョンは、すべてのベンチマーク領域で99.89%のF1スコア (適合率と再現率を組み合わせた指標) という前例のないスモールバリアントコール精度を提供します (図1)。これは、Human Pangenome Reference Consortium (HPRC) データからの256のハプロタイプを含む128のサンプルに基づいて構築された第3世代DRAGENマルチゲノム (グラフ) リファレンスによって可能になり、より大きな遺伝的多様性を捉えることができます。また、アリル頻度がわずか3%のモザイクバリアントを検出できる新しく統合されたモザイクコーラーも、精度の向上に貢献しています。

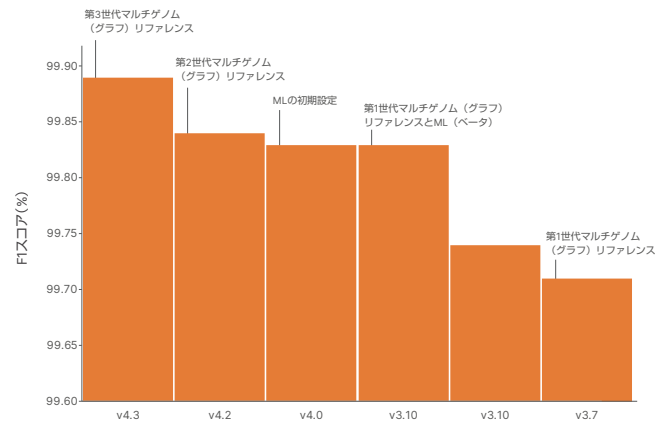


図1: DRAGEN二次解析の精度: Y軸のF1スコア (%) は真陽性と真陰性の結果の合計値の割合を算出したものです。^{3,4}

包括的な解析

DRAGEN二次解析は、ゲノムを包括的にカバーし、幅広いアプリケーションに対応するため、NGS解析を実施するラボの多様なニーズを満たします。DRAGENパイプラインは、全ゲノムシーケンス (WGS)、全エクソームシーケンス、濃縮パネル、シングルセルRNA-Seq、シングルセルATAC-Seq、バルクRNA-Seqおよびメチル化解析など、さまざまな種類の実験に対応します (表1)。DRAGENソフトウェアの幅広い機能を部分的に再現するには、30以上のオープンソースツールが必要となります。^{3,4}

DRAGEN二次解析では、生殖系列解析向けに、ExpansionHunterおよびSMN、GBA、CYP2B6、CYP2D6およびHLAなどのターゲットコーラーといったバリエーションコーラーが含まれています。DRAGEN v4.3では、MRJDという新しい特殊なコーラーも導入され、例えば、PMS2、SMN1、SMN2、STRC、NEB、TTN、およびIKBKGなどのセグメント重複領域における困難な遺伝子のカバレッジを得ることも可能になります。これらのツールを使用すると、1塩基変異、挿入と欠失 (Indel)、リピート伸長、大きなゲノム領域の構造変異など、広範囲の遺伝的変異を解析できます。また、DRAGENマルチゲノム (グラフ) リファレンスによってマッピング品質が向上するため、バリエーションコール精度の向上につながり、配列の複雑性のためにアクセスが困難なゲノムの領域を解消します。これにより、医学的に関連する可能性のある遺伝子のカバレッジが増加し、マッピング困難な領域で1塩基変異、小さなIndel、コピー数変異および構造変異のコールが可能になります。

表1: DRAGEN二次解析は幅広い二次解析アプリケーションに対応^a

アプリケーション	オンサイトサーバー	イルミナシーケンシングシステム内蔵型		イルミナクラウドプラットフォーム	
	DRAGENサーバー	NovaSeq Xシリーズ	NextSeq 1000、 NextSeq 2000 システム	BaseSpace Sequence Hub	Illumina Connected Analytics
BCL変換	✓	✓	✓	✓	カスタムのみ
DRAGEN ORA圧縮	✓	✓	✓		カスタムのみ
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
全ゲノム	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞
濃縮 (エクソームを含む)	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞	生殖系列 + 体細胞
DRAGEN Amplicon	✓		DNAのみ	✓	✓
RNA	✓	✓	✓	✓	✓
シングルセルRNA	✓		✓	✓	✓
NanoString GeoMx NGS			✓	✓	
メチル化	✓	✓		✓	✓
メタゲノム	✓ ^b			✓	
RNA pathogen detection				✓	
COVID	COVIDSeq.COVID Lineage		COVIDSeq (クラウドのみ)	COVIDSeq.COVID Lineage	
TruSight Oncology 500 ポートフォリオ	✓			✓ ^c	✓
scATAC-Seq	✓			✓	✓
インビュテーション	✓			✓	✓
PGx Star Allele Caller	✓	✓	✓	✓	✓
Illumina Complete Long Reads				✓	✓
RPIPおよびUPIP用の DRAGEN二次解析	✓			✓	ベータ

a. コアDRAGENソフトウェアのバージョンはプラットフォームによって異なります。詳細については、地域の担当者にお問い合わせください。

b. Kmer分類子によって実現されるメタゲノムアプリケーション。さらに多くのツールが近日中に利用可能。

c. Illumina Connected Analyticsのサブスクリプションが必要。

効率的な解析

DRAGENソフトウェアは、NGSデータセット処理の効率を最適化するために必要なデータ解析速度をラボに提供するように設計されています。DRAGEN二次解析は、ハードウェアアクセラレーションとフィールドプログラマブルゲートアレイ (FPGA) アーキテクチャーを使用して、高速のターンアラウンドタイムを達成します。DRAGEN解析アルゴリズムは極めて効率が高く、ゲノムデータ解析に対する2つの世界最速記録を生み出しています。^{5,6} 実運用では、オンサイトのDRAGEN二次解析は、40 ×カバレッジ相当の全ゲノムNGSデータをすべてのコーラーを用いて約35分で処理できるのに対し、限られた数のバリエーションタイプをコールする一般的に使用されるオープンソースメソッドでは処理に8時間以上かかります。⁷

大きなNGSデータファイルの保存、管理、共有に対応しやすくするために、DRAGEN Original Read Archive (ORA) テクノロジーは、従来のfastq.gz形式のFASTQファイルを最大5倍の可逆圧縮で提供します。DRAGEN ORAの可逆圧縮はFASTQファイルの詳細を維持し、非常に高速であるため、50~70 GBのFASTQファイルを約8分

* DRAGENサーバーv4上のHG001-HG007スタンダードに基づいたイルミナ社内データに基づく。DRAGEN v4.3で利用可能なMRJDやVNTRなどの新しい特殊なコーラーはなし。

で圧縮でき、一般的な研究対象研究生物種に幅広く対応します。また、DRAGEN二次解析は多彩なパイプラインを搭載しており、パイプラインのさまざまな段階で、入力データファイルを受け入れ、出力ファイルを作成することができます (図2)。

FPGAおよびハードウェアアクセラレーション

高度に設定可能なFPGAにより、ベースコール (BCL) ファイル変換、マッピング、アライメント、ソーティング、重複マーキング、およびハプロタイプバリエーションコーリングなどのゲノム解析アルゴリズムに対して、超効率的なハードウェアアクセラレーションによる実行が可能になります。FPGAの柔軟性により、イルミナは広範なDRAGENアプリケーションパイプラインスイートを開発し、可能な限りの最高精度、包括性、効率をもたらすための頻繁なアップデートと追加が可能になります。

カスタムリファレンス

DRAGEN二次解析ではカスタムのヒト、非ヒト、または非標準リファレンスをユーザーが作成することができます。作成したリファレンスは、カスタムリファレンスファイルに対応するすべてのDRAGENアプリケーションに入力として使用できます。ほとんどのDRAGENパイプラインにはhg19、hg38 (HLAありまたはなし)、GRCh37、

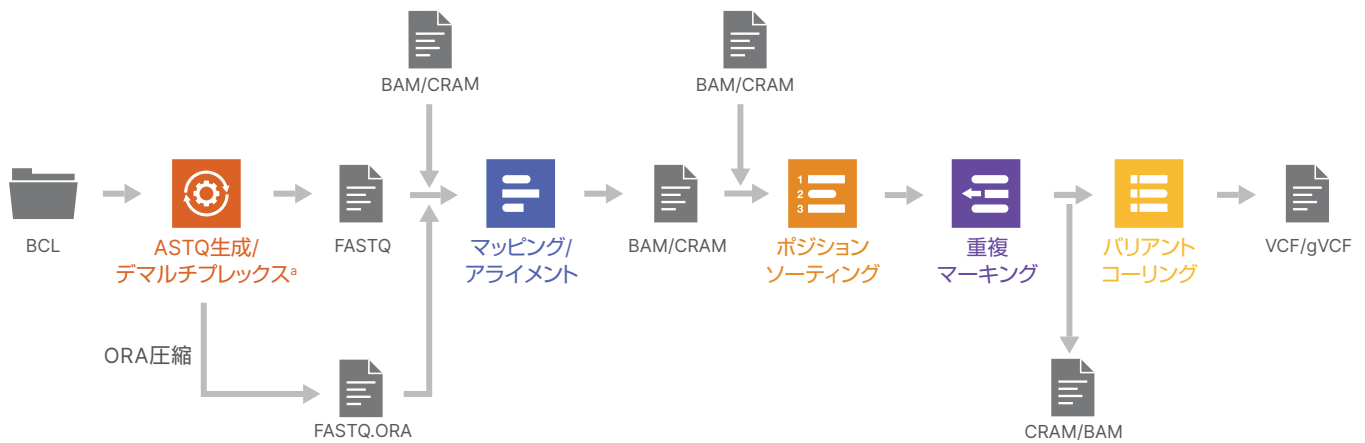


図2: DRAGEN二次解析パイプラインの柔軟性: 各DRAGENパイプラインは正確かつ効率的な解析に対応するために特異的なステップが含まれます。一例として示したDRAGENの全ゲノム生殖系列パイプラインでは、さまざまな入力ファイルを受け入れるための柔軟性があり、幅広い種類の出力を生成することから、ユーザーは実験をカスタマイズし、希望するファイル形式を生成することが可能です。

a. BCL変換はスタンドアロンツールとしても利用できます。

CHM13v2およびhs37d5に対するビルトインサポートが搭載されています。DRAGENソフトウェアを使用すると、多様な集団と特定の集団の両方に対してグラフ標準のマルチゲノムリファレンス機能を拡張できます。

拡張性

DRAGEN二次解析は、低コストと短いターンアラウンドタイムを維持しながら、必要に応じてオペレーションを拡大することが可能です。DRAGENソフトウェアにより、次のいくつかの方法で研究の可能性を容易に広げることができます。

- 1. NovaSeq™ XシリーズとNextSeq™ 1000およびNextSeq 2000システムの最新情報**：内蔵型のDRAGENでは、1回のランでフローセルごとに複数のアプリケーション（最大1つのBCL変換と選択した他の3つのパイプラインを含む4つのアプリケーション）を同時に実施できます。
- 2. 処理能力をバースト**：サンプル数の増加により作業負荷が増加する時には、Illumina Connected Analytics上のDRAGEN二次解析やBaseSpace Sequence Hub上のDRAGENアプリを使ってクラウド上で追加の容量を利用することができます（図3）。
- 3. オペレーションの拡大**：1台のDRAGENインスタンスで、幅広いDRAGENパイプラインとサポートされているサンプルタイプのランを実行できます。DRAGENソフトウェアの包括性と効率性により、ユーザーはターンアラウンドタイムや結果のクオリティを損なうことなくオペレーションを拡大できます。
- 4. ゲノムへの移行**：DRAGENの事前構築されたパイプラインにより、ターゲットパネルからエクソーム、ゲノムへ容易に移行できるようになります。
- 5. 大規模集団ゲノミクスに対応する実行力**：DRAGEN二次解析は、大規模コホート解析用に簡便なワークフローを提供しており、複数のパイプラインを併用することで、高い精度で遺伝的バリエーションをコールします。DRAGEN gVCF Genotyperは数千から数百万のゲノムバリエントコール形式（gVCF）ファイルの集約が可能であり、既存のバッチを再処理することなく新しいバッチを取り込みます。ORA圧縮によりストレージコストが節約されます。
- 6. ディープシーケンスアプリケーション**：DRAGEN二次解析は、ゲノムの場合は300倍以上、エクソームの場合は1,000倍以上の平均カバレッジで、高効率の高深度シーケンス解析をサポートします。ディープシーケンス機能は、腫瘍学研究や希少遺伝性疾患の研究などのアプリケーションに役立ちます。

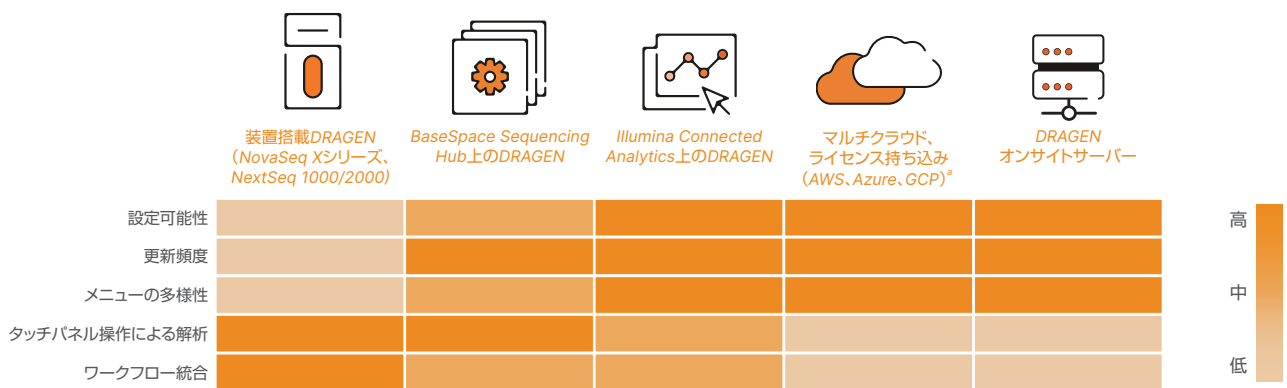


図3：あらゆるラボのNGS解析ニーズに合うように設計された機能を備えたDRAGEN/パイプラインアクセスオプション

a. Amazon Web Service (AWS)、Azure、またはGoogle Cloud Platform (GCP、早期アクセス) へのアクセスに関する情報については、イルミナの担当者にお問い合わせください。

マルチプラットフォームの アクセシビリティ

DRAGENパイプラインは、オンサイト、装置上またはクラウドソリューションで利用でき、ニーズに最も合うソリューションを選択できます (図3)。

DRAGENオンサイトサーバー

DRAGENオンサイトサーバーには、NGSデータを収集し保管するためのローカルストレージソリューションが別途必要です。生のシーケンスデータがシーケンス装置からローカルネットワーク接続を経由してローカルストレージに転送された後、DRAGENサーバーに移され、選択されたワークフローを実行します。解析後、ソフトウェアはローカルストレージがある場所に生成した出力ファイルを書き出します。DRAGENオンサイトサーバーは以下の特徴があります。

- コマンドラインインターフェースを通じてDRAGEN機能の柔軟な構成をサポート
- 最大30の従来のコンピューティングインスタンスを置き換え
- 約35分で40 xカバレッジの全ヒトゲノムのNGSデータを処理

DRAGENを内蔵したNovaSeq™ Xシリーズ

NovaSeq Xシリーズには内蔵型のDRAGEN二次解析が搭載されており、NovaSeq Xシリーズによって生成される膨大な量のデータに対応するように設計された、精確で自動化され、効率化された解析を提供します。内蔵型のDRAGENソフトウェアスイートは、BCL変換、生殖系列、体細胞、濃縮、RNAおよびメチル化に対応するNGSアプリケーション (表1) を使用して二次解析とORA圧縮を提供します。内蔵型DRAGENは以下の特長があります。

- 複数の二次解析パイプラインを並行して実行
- 1回のランでフローセルあたり同時に最大4つのアプリケーションを実施
- 最大5倍の可逆データ圧縮とストレージコストの節約
- 解析コストを節約でき、5年間でNovaSeq Xシステムの購入価格を超える可能性がある

DRAGEN内蔵NextSeq 1000および NextSeq 2000システム

NextSeq 1000およびNextSeq 2000システムは、迅速かつ高精度二次解析のためにDRAGENソフトウェアを内蔵しています。本ソフトウェアは、使いやすいグラフィカルインターフェースを通してアクセスするため、エキスパートユーザーと非エキスパートユーザーが必要な解析を実施し、迅速に結果を生成することができます。内蔵型のDRAGENソフトウェアは、一般的なさまざまなNGSアプリケーションに対応するために選り抜かれたパイプラインを取り揃えており (表1)、高品質なバリエーションに対応するために受賞歴のあるMLとマルチゲノム (グラフ) リファレンス解析を搭載しています。内蔵型DRAGENは以下の特徴があります。

- 内蔵型DRAGEN二次解析によりベンチトップシーケンスシステムの中で最高の精度を提供
- DRAGENインフォマティクスパイプラインへのアクセスを提供
- わずか2時間で結果を生成することが可能
- 直感的なパイプラインアルゴリズムを使用して、外部のインフォマティクス専門家への依存を縮小

BaseSpace Sequence Hub

BaseSpace Sequence Hub上で利用できるクラウドベースのDRAGENパッケージは、精確かつ効率的な解析と、安全なエコシステムおよび汎用性のあるスケーラビリティとを組み合わせています。BaseSpace Sequence Hub上のDRAGENソフトウェアにより、あらゆる規模と分野のラボが二次解析をボタン操作で実施できます。BaseSpace Sequence Hubはイルミナ装置と接続可能です。暗号化されたデータは装置からBaseSpace Sequence Hubに送られるため、キューレーションされた一連のアプリケーションで簡単にデータを管理し解析できます。アマゾンウェブサービス (AWS) 上のBaseSpace Sequence Hubは以下の特長があります。

- ボタン操作による使いやすいDRAGEN解析を提供
- エキスパートユーザーと非エキスパートユーザーによる効率的なオペレーションのために直感的なグラフィカルユーザーインターフェースを使用
- 追加のインフラストラクチャーの設備投資なしでパワフルなコンピューターリソースにアクセス

illumina Connected Analytics

illumina Connected Analyticsはクラウドベースの包括的なバイオインフォマティクスプラットフォームであり、セキュアかつ拡張性と柔軟性のある環境で、大量のマルチオミックデータを管理、解析、解釈できるようになります。illumina Connected Analyticsにアクセスすると、パッケージ化されたパイプラインとして、またはカスタムパイプラインに組み込むための個別のツールとして、DRAGEN二次解析スイートを利用できます。

まとめ

DRAGENはNGSデータを精確、包括的かつ効率的に解析するパワフルな二次解析ソフトウェアツールです。複数のDRAGENソフトウェア導入オプションから、プロジェクトの種類とスケールに最適なソリューションを選択できます。さらに、ユーザーはパフォーマンスとワークフローのニーズに合わせてさまざまな導入オプションを組み合わせることができます。NGSテクノロジーの継続的な進歩に合わせて、DRAGEN二次解析へのタイムリーなアップデートを提供することで、現在のパイプラインの最高の性能の可能性を確保しつつ、最新のアプリケーションに対応する新たなパイプラインが追加されます。

詳細はこちら

[DRAGEN二次解析](#)

[DRAGEN二次解析サポートページ](#)

[お問い合わせ](#)

参考文献

1. The food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. <https://precision.fda.gov/challenges/10/results>. Accessed March 14, 2022.
2. Catreux S, Jain V, Murray L, et al. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. Illumina website. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Published January 12, 2022 Accessed March 14, 2022.
3. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. Illumina website. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Published November 9, 2020. Accessed March 14, 2022.
4. Internal data on file. Illumina, Inc., 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Accessed March 14, 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Published February 12, 2018. Accessed March 14, 2022.
7. Betschart RO, Thiéry A, Aguilera-Garcia D, et al. Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment. *Sci Rep.* 2022;12(1):21502. Published 2022 Dec 13. doi:10.1038/s41598-022-26181-3
8. Gross A, Maciuca S, Cox A, et al. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. Illumina website. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Published May 24, 2021. Accessed March 14, 2022.

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

© 2024 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。

