

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研究と医学の進歩に必要不可欠です。NGSから遺伝学的洞察を最大に得るために、研究者には正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、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データの解析と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プラットフォームは、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、二次解析ソフトウェアスイートです。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 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対象としたAll Benchmark RegionsおよびDifficult to Map Regionsで最も精確な結果を勝ち取りました。^{1,2} グラフゲノムおよびイルミナ機械学習(ML)の革新により、DRAGEN 4.0ソフトウェアはAll Benchmark Regionsにおいてあらゆるシーケンステクノロジーの中で、非常に優秀なデータ精度である、F1スコア(精度とコール率を組み合わせた基準)99.83%を実証しました(図1)。^{1,2} また、DRAGEN 4.0 + Graph(初期設定でML有効)は、主要組織適合複合体(MHC)領域において、PrecisionFDA V2に提出されたすべての結果と比較して最高のF1スコアを示し、最も精確にコールしました。

包括的な解析

DRAGEN二次解析は、1台でさまざまな種類の実験を包括的にカバーし、幅広いNGSアプリケーションを実施するラボのご要望に応えます。DRAGENパイプラインは、全ゲノムシーケンス(WGS)、濃縮パネル、シングルセルRNA-Seq、シングルセルATAC-Seq、シングルセルマルチ

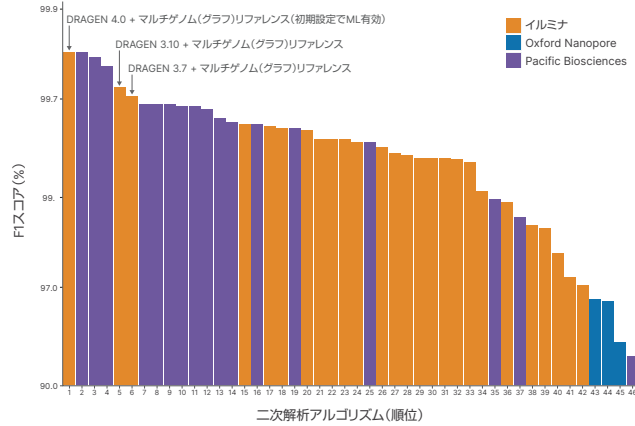


図1: All Benchmark RegionsのデータセットにおけるDRAGEN 4.0 + Graph (初期設定でML有効)とPrecisionFDA Truth Challenge v2に提出された結果の精度比較: DRAGEN 4.0 + Graph (初期設定でML有効)は、Pacific Bioscienceのシーケンスデータ上のGoogle DeepVariantと並び、非常に優れた精度を示しています。DRAGEN 3.10 +Graphでは、Graphおよびリファレンス/ALTコンティグの処理方法が向上したため、DRAGEN 3.7 + Graphよりも改善が示されています。Y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}

オミックス、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ています(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}

リピート伸長、構造多型(SV)、コピー数バリエーション(CNV)、ExpansionHunterおよびSMN、GBA、CYP2B6、CYP2D6、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おり、DRAGENソフトウェアは幅広くゲノムをカバーできます。また、DRAGEN Multigenome(Graph)リファレンスによって実際上イルミナリードが伸長し、複雑性の低い領域に到達できるため、繰り返し配列によって評価が困難なゲノム領域を解消します。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改善され、マッピング困難な領域で1塩基、コピー数および構造多型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します

アプリケーション	オンプレミスサーバー	内蔵型のイルミナシーケンスシステム		イルミナのクラウドプラットフォーム上	
	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	✓	✓	✓	✓	利用可能(カスタムのみ)
DRAGEN ORA圧縮	✓	✓	✓		利用可能(カスタムのみ)
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
全ゲノム	生殖細胞 + 体細胞	生殖細胞のみ 体細胞は近日中に利用可能	生殖細胞のみ	生殖細胞 + 体細胞	生殖細胞 + 体細胞
エンリッチメント (エクソームを含む)	生殖細胞 + 体細胞	生殖細胞 + 体細胞	生殖細胞 + 体細胞	生殖細胞 + 体細胞	生殖細胞 + 体細胞
DNAアンプリコン	✓		✓	✓	✓
RNA	✓	✓	✓	✓	✓
シングルセルRNA	✓		✓	✓	✓
発現差		✓	✓	✓	
NanoString GeoMx NGS			✓	✓	
RNAアンプリコン	✓			✓	近日中に利用可能
メチル化	✓	近日中に利用可能		✓	✓
メタゲノム解析				✓	
RNA pathogen detection				✓	
COVID	COVIDSeq。COVID Lineage		COVIDSeq(クラウドのみ)	COVIDSeq。COVID Lineage	
TruSight Oncology 500	cfDNAは利用可能、 固形腫瘍は近日中に 利用可能			3.10で有効	✓
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				ベータ

効率的な解析

DRAGENソフトウェアは、NGSデータセットを最も活用するために必要なデータ解析スピードとファイルオプションを提供できるよう、特別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} 実際のアプリケーションでは、DRAGEN二次解析は、34 × カバレッジ相当の全ゲノムNGSデータを、従来のCPUベースのシステムでは15時間以上で処理するのに対して、約30分で処理することができます。⁷

大きなNGSデータファイルのストレージ要求に対応するために、DRAGEN Original Read Archive (ORA) テクノロジーはFASTQファイルを1/5に可逆圧縮します。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分で圧縮します。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 Reference Builderで、ハッシュテーブルと呼ばれるヒト、非ヒト、または非標準リファレンスを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。BaseSpace™ Sequence Hub上のDRAGEN Reference BuilderアプリケーションはFASTAファイルが必要です。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 Graph Toolkitにより、グラフィカルなリファレンス性能をより多様なヒトグラフィカルなリファレンスに拡張できます。

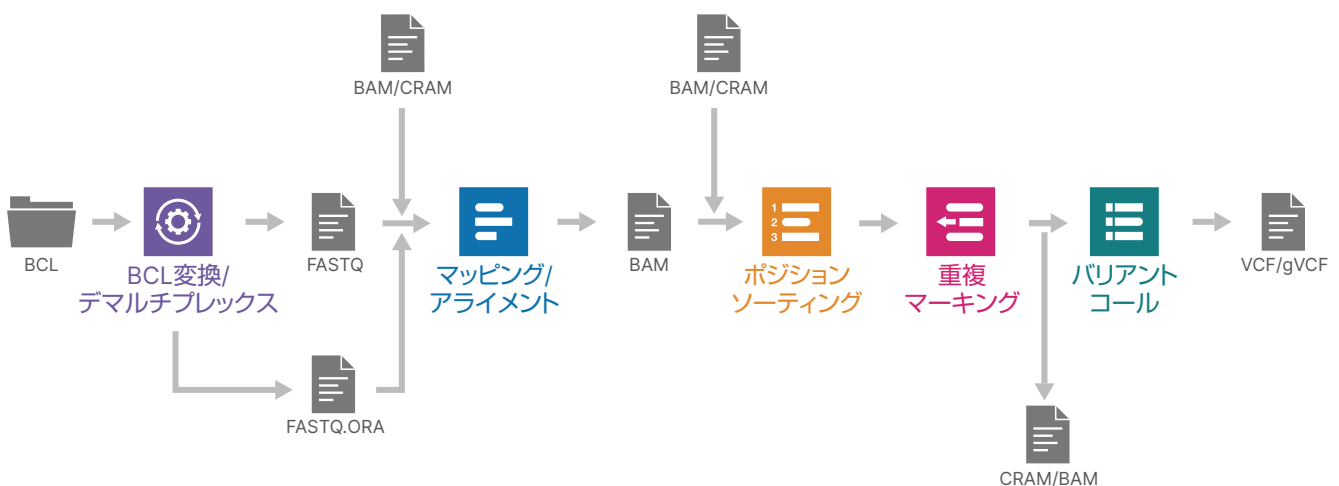


図2: DRAGENパイプラインの柔軟性: 各DRAGENパイプラインは正確かつ効率的な解析に対応するために特異的なステップが含まれます。DRAGENパイプラインはさまざまな入力ファイルを受け入れるための柔軟性があり、幅広い種類の出力を生成することから、ユーザーは実験をカスタマイズし、希望するファイル形式を生成することが可能です。

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シリーズの調和**: 内蔵型のDRAGENは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、異なるDRAGENソフトウェアを並行して利用することで、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、すべての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ソフトウェアの精度、包括性、効率により、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. **エクソームからゲノムへ**: 全エクソームシーケンス(WES)からWGSへの拡大により、生成したデータが大幅に増加します。DRAGENソフトウェアなら、大規模な投資やハードウェアのインフラストラクチャーまたはクラウドベースのソリューションを追加することなく、エクソームからゲノムまで解析できます。

5. **非常に大量のデータセット**: 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ます。複数のパイプラインを併用することで、コホートサンプリングから高い精度で小さなバリエーションおよび大きなバリエーションをコールします。DRAGENソフトウェアは数千から数百万のゲノムバリエーションコール形式(gVCF)ファイルの集約とジェノタイピングが可能であり、既存のバッチを再処理することなく新しいバッチを集約します。DRAGEN Joint Genotyping Pipelineは複数のゲノムから同時にバリエーションをコールし、精度を損なうことのない高速解析で大規模なコホートに拡張します。⁸ 例えば、**1000人ゲノムプロジェクト**データのDRAGEN二次解析では、多様なサンプルの大規模かつ正確なバリエーションコールが実現し、カバーレッジデータが不均一な領域、または想定から逸脱している領域を同定できました。

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)。

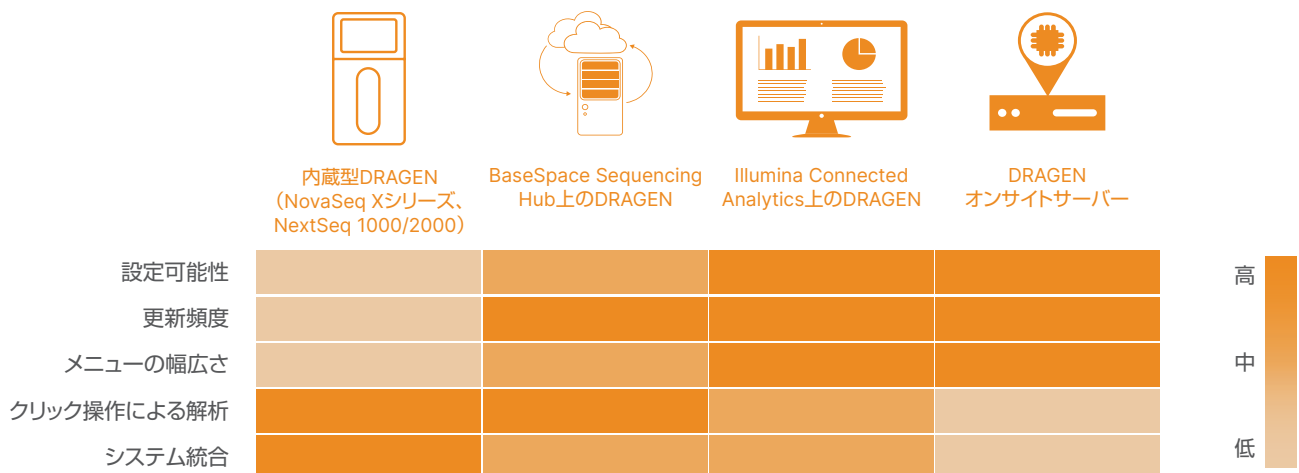


図3: あらゆるラボのNGS解析ニーズに合うように設計された機能を備えるDRAGENパイプラインアクセスオプション。

DRAGENオンサイトサーバー

オンサイトのDRAGEN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:

- さまざまなレベルのコマンドラインインターフェースに対応
- 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30分で34 × 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シリーズは、これまでで最も強力なDRAGENソフトウェアを内蔵しており、精度で自動化された包括的な二次解析を提供します。内蔵型のDRAGENソフトウェアスイートは、BCL変換、Germline、RNA、Enrichmentに対応するNGSアプリケーションを使用してバリエーションコールとORA圧縮を提供します(表1)。内蔵型DRAGEN:

- 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サポートされているアプリケーションによる解析
- 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内蔵型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。内蔵型のDRAGENソフトウェアパッケージは、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び抜かれたパイプラインを取り揃えており(表1)、使いやすいインターフェースにより、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DRAGEN:

- 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はイルミナ装置から直接アクセスできます。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:

- DRAGEN解析用のボタン操作による使いやすいソリューションを提供。
- 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のDRAGEN二次解析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、セキュアで拡張性と柔軟性のある環境で大量のマルチオミックスデータを管理、解析、解釈できるようになります。Illumina Connected Analytics:

- パッケージ化されたパイプラインまたは、カスタムパイプライン用の個別のツールとして、すべてのDRAGENソフトウェアへのアクセスを提供
- 高度に自動化されたワークフローと最適化されたハイスループット研究用のカスタムソリューションに対応
- 保証されたデータレジデンシー、シングルサインオンアクセス、監査ログ、Health Insurance Portability and Accountability Act (HIPAA)コンプライアンスおよび欧州連合General Data Protection Regulation (GDPR)に対応したアクセスコントロールを用いた高度なセキュリティ環境を提供

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。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、アプリケーションが使用可能になるにつれて、新しいパイプラインが継続して追加されます。

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. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Accessed March 14, 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. jp.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Accessed March 14, 2022.
3. Illumina. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. jp.illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. 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. bioitworld.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. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015;7:100. doi: 10.1186/s13073-015-0221-8.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. jp.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. 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

© 2023 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。

