

データシート

DRAGEN™ Bio-IT Platform

NGSデータのための精確で包括的
かつ効率的な二次解析

- 精確な結果
- 包括的なアプリケーション
- 効率的な解析

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、研究者には精確かつ効率的に生のシーケンスデータの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) Bio-IT Platform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 Bio-IT Platformは、非常に精確な結果を生成します。2020年のPrecision FDA Truth Challenge V2(PrecisionFDA V2)では、イルミナシーケンスデータを対象としたAll Benchmark領域およびDifficult to Map領域カテゴリーで、DRAGEN v3.7が最高精度を達成しました。^{1,2} DRAGEN v3.10を用いたGraph GenomesおよびIllumina Machine Learningの革新により、すべてのシーケンステクノロジーに対するデータ精度の新しいスタンダードが確立され、All Benchmark領域でF1スコア99.83%を達成しました(図1)。^{1,2} また、DRAGEN 3.10、Graph、およびIllumina Machine Learningは、主要組織適合性複合体(MHC)領域においてPrecisionFDA V2から提出されたすべてのコールと比較して、最も精確なコールである最高のF1スコアも獲得しました。

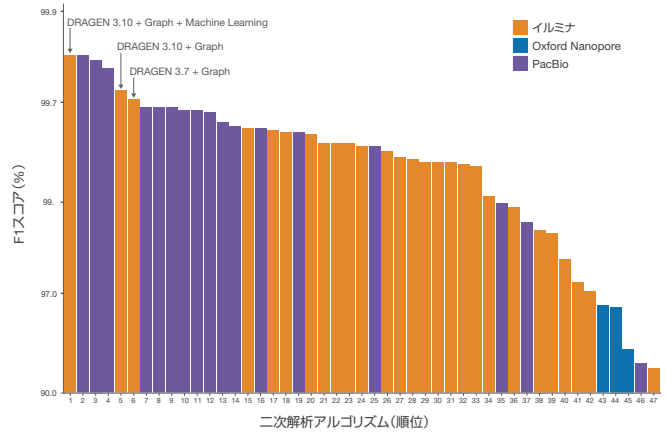


図1: All Benchmark領域でのデータセットにおけるDRAGEN 3.10、GraphおよびMachine LearningとPrecisionFDA Truth Challenge v2に提出された結果の精度の比較: DRAGEN 3.10 + Graph + Machine Learning解析メソッドはPacBioのリードスコアと同率で1位となりました。DRAGEN 3.10 + Graphでは、Graphおよびリファレンス/ALTコンティグの処理方法が向上したため、DRAGEN 3.7 + Graphよりも改善が示されています。Y軸のF1スコア (%) は合計結果の割合として真陽性および真陰性の結果を計算したものです。

包括的なカバレッジ

DRAGENソフトウェアはバリエーションコーラースイート(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、CYP2D6、HLAなどのターゲットコーラー)を搭載し、幅広いゲノムカバレッジを提供します。また、DRAGEN Graphと並行コンピューターアーキテクチャーによって効率的にイルミナリードを伸長し、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。

効率的な解析

DRAGEN Bio-IT Platformは、データ解析スピードと必要なファイルオプションを提供するために特別に設計されており、NGSデータセットから最大の利益を享受できます。DRAGENソフトウェアは、ハードウェア

* HLA:ヒト白血球抗原

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解析アルゴリズムの効率はゲノムデータ解析の2つの世界速度記録をもたらしました。^{3,4} 実際のアプリケーションにおいて、従来のCPUを用いたシステムでは8時間以上かかる34xカパレッジに相当する全ゲノムのNGSデータを、オンサイトのDRAGEN Bio-IT Platformは約36分で処理できます。

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にロスレス圧縮します。DRAGEN ORAのロスレス圧縮はFASTQファイルの詳細を維持し、非常に高速であるため、50 GB~70 GBのFASTQファイルを約8分で圧縮します。

FPGAおよびハードウェアアクセラレーション

高度な構成が可能なFPGAはDRAGEN Bio-IT Platformの効率に必要不可欠です。FPGAプログラミングは、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ルなどのゲノム解析アルゴリズムをハードウェアアクセラレーションによって極めて効率的に実行することが可能です。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継続的なアップデートと追加が可能になります。

複数のアプリケーションに対応

DRAGEN Bio-IT Platformは幅広いNGSアプリケーションを実施するラボのニーズを満たします。機能には、入力データファイルを受け入れ、パイプラインのさまざまな段階で出力ファイルを作成できる汎用性のあるパイプラインを含みます(図2)。DRAGEN二次解析パイプラインは、全ゲノムシーケンス、エンリッチメントパネル、シングルセルRNA、バルクRNAおよびメチル化解析などの多岐にわたる種類の実験をサポートしています(表1)。

カスタムリファレンス

DRAGEN Reference Builderは、ハッシュテーブルとも呼ばれる非ヒトまたは非標準リファレンスを生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アプリケーションの入力として使用できます。BaseSpace Sequence HubのDRAGEN Reference BuilderアプリケーションにはFASTAファイルが必要です。ほとんどのDRAGEN/パイプラインは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を搭載しています。

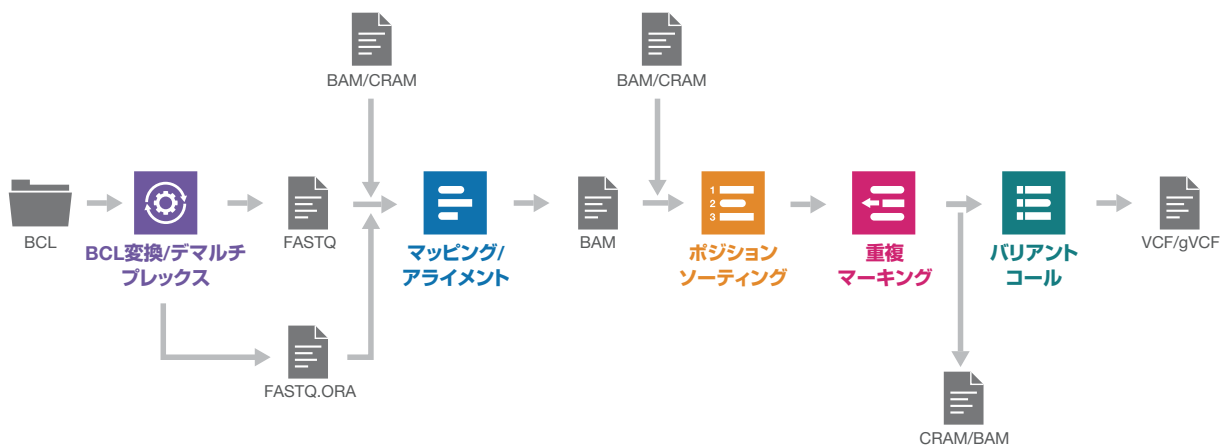


図2: DRAGENパイプラインの柔軟性: 各DRAGENパイプラインにはそれぞれ特別なステップがあり、正確かつ効率的な解析をサポートします。DRAGENパイプラインはさまざまな入力ファイルを受け入れるための柔軟性があり、幅広い種類の出力を生成することから、ユーザーは実験をカスタマイズし、希望するファイル形式を生成することが可能です。

表1: DRAGEN Bio-IT Platformは多数の広範な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	DRAGENオンサイト サーバー	NextSeq 1000 システム NextSeq 2000 システム	BaseSpace Sequence Hub	Illumina Connected Analytics		Multi-Cloud ライセンス持ち込み
				事前設定	持ち込み	
BCL変換	✓	✓	✓		✓	推奨されない
DRAGEN ORA圧縮	✓	✓			✓	✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
全ゲノム	生殖細胞 + 体細胞	生殖細胞のみ	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
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 (CNVなし)	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	✓		✓	✓	✓	✓
メタゲノム解析			✓			
RNA pathogen			✓			
COVID	COVIDSeq, COVID Lineage	COVIDSeqのみクラウド	COVIDSeq, COVID Lineage	COVIDSeq, COVID Lineage	COVIDSeq, COVID Lineage	
TruSight Oncology 500 ctDNA	ctDNAは利用可能、 固形腫瘍は 近日中に利用可能			ctDNAは利用可能、 固形腫瘍は 近日中に利用可能		

拡張性

DRAGEN Bio-IT Platform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™ 6000システムとの調和**：1台のDRAGENサーバーで、S4フローセルを用いたNovaSeq 6000システムのランを2時間未満でマルチプレックスできます。
- 2. 処理能力をバースト**：サンプル量の増加に伴って作業負担が増加している間、DRAGENプラットフォームの並行アクセスオプションを使うことで更なる処理能力を獲得できます (図3)。

- 3. オペレーションの拡大**：1台のDRAGENプラットフォームで、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ツールの精度、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)から全ゲノムシーケンス(WGS)へ移行することにより、生成される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。

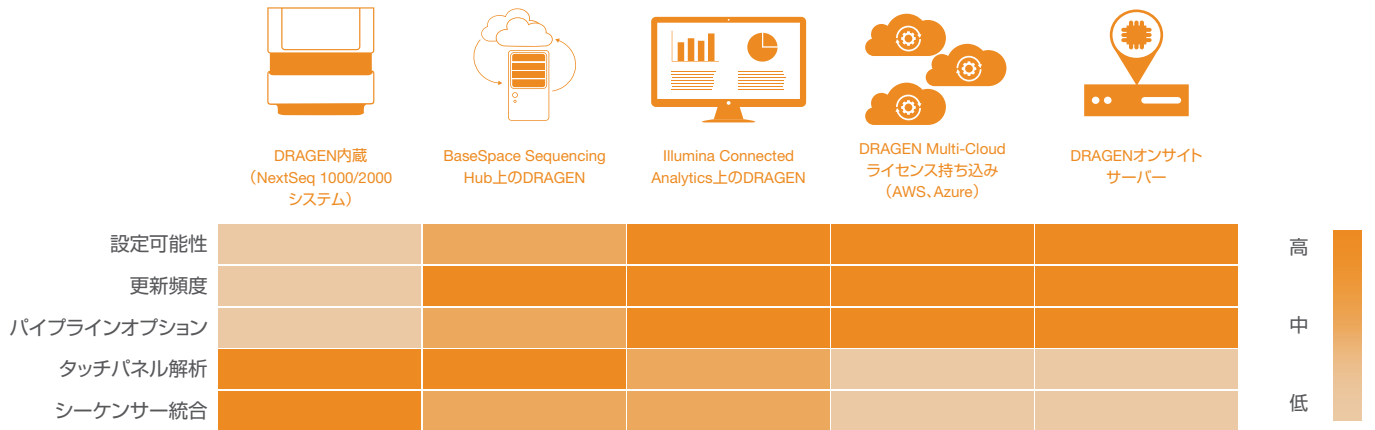


図3: DRAGEN Bio-IT/パイプラインは、あらゆるラボのNGS解析のニーズに合うようにデザインされた性能があるオプションにアクセスします。

5. 非常に大量のデータセット: DRAGENプラットフォームは、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。⁵ 例えば、1000人ゲノムプロジェクトデータにDRAGENプラットフォームを使うことにより、多様なサンプルの大規模かつ高精度なバリエーションコールが実現し、カバ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 Bio-ITパイプラインスイートは、オンサイト、装置上、またはクラウドソリューションでの利用を介してアクセスできるため、ニーズに最も合うソリューションを選択できます。

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では以下が可能になります。

- 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
- 25分未満で30×カバレッジのヒトゲノム全体のNGSデータを処理

NextSeq™ 1000およびNextSeq 2000システム内蔵DRAGEN Bio-IT Platform

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。内蔵されたDRAGENソフトウェアスイートは、ユーザーフレンドリーなインターフェースで幅広い一般的なNGSアプリケーション(表1)に対応するために選ばれたパイプラインを提供しており、エキスパートユーザーでも非エキスパートユーザーでも必要な解析を素早く実施し、結果を生成できます。内蔵型DRAGEN Bio-IT Platformでは以下が可能になります。

- 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で使用可能なクラウドベースのDRAGENスイートは、精確かつ効率的な解析とセキュアなエコシステムおよび多様な拡張性とを兼ね備えています。BaseSpace Sequence HubのDRAGEN Bio-ITソフトウェアはあらゆる規模と分野のラボがタッチパネルから二次解析を実施できます。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では以下が可能になります。

- DRAGEN解析を、プッシュボタンによる簡単で使いやすいソリューションとして提供
- 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 Bio-IT Platformはクラウドベースの包括的なデータ管理と解析を行う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データを管理、解析、解釈できるようになります。Illumina Connected Analyticsでは以下が可能になります。

- DRAGEN Bio-ITスイートにアクセスし、プレパッケージ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使用可能
- 最適化したハイスループット研究用に、高度に自動化されたワークフローとカスタムソリューションに対応
- 保証済みの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、およびHealth Insurance Portability and Accountability Act (医療保険の携行性と責任に関する法律)コンプライアンスおよび欧州連合General Data Protection Regulation(一般データ保護規則)の指針に対応したアクセスコントロールによって非常にセキュアな環境を提供

DRAGEN Multi-Cloudのライセンス持ち込み(BYOL)

既存のクラウドベースのサービスのあるラボは、仮想イメージとしてDRAGEN Bio-IT Platformを導入することでAWSまたはMicrosoft Azure上でDRAGEN解析を実行できます。DRAGEN Multi-Cloudのライセンス持ち込みは以下が可能になります。

- DRAGENオンサイトサーバーとして同一ソフトウェアとコマンドラインオプションを使用(TruSight™ Oncology 500パイプラインを除く)
- インフォマティクスに精通したお客様がDRAGENパイプラインを既存のクラウド環境とワークフローに統合することを検討している場合に最適

まとめ

DRAGEN Bio-IT Platformはパワフルなソフトウェアツールであり、NGSデータの精確、包括的かつ効率的な二次解析を提供します。複数のDRAGENプラットフォームオプションから、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テクノロジーが継続的に進歩に合わせて、DRAGEN Bio-IT Platform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

イルミナのDRAGEN Bio-IT Platform jp.illumina.com/dragen

DRAGENサポートページ jp.support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html

お問い合わせ jp.illumina.com/company/contact-us.html

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. 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.
4. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Accessed March 14, 2022.
5. 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

© 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。

Pub. No. M-GL-00680 v2.0-JPN 07JUN2022

illumina®