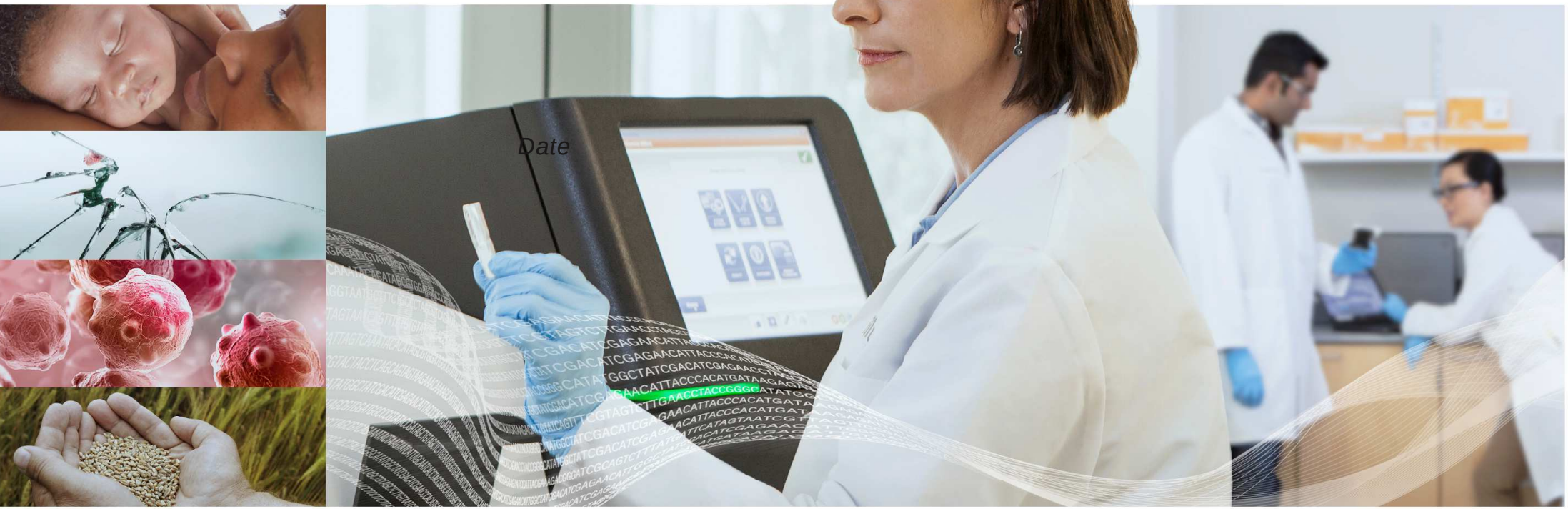
製品ミニ紹介: **Illumina DRAGEN™**

どらげん？どらじぇん？ - やってきた NGS 高速解析の竜

ver. 2019年6月

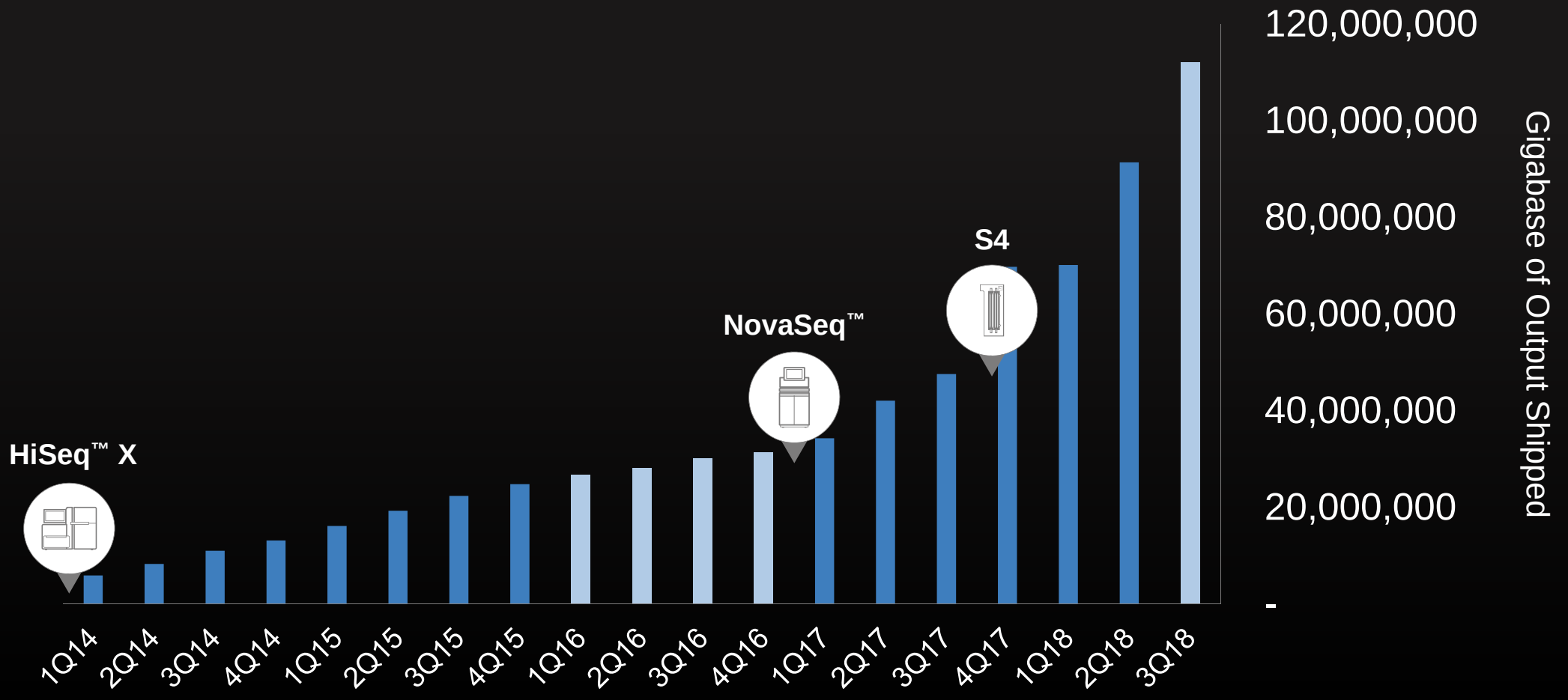
イルミナ株式会社
シニア バイオインフォマティクス スペシャリスト
癸生川 絵里(Eri Kibukawa)



お問合せ:contactJPN@illumina.com

For Research Use Only. Not for use in diagnostic procedures

ハイスループットシーケンスデータの増大



For Research Use Only. Not for use in diagnostic procedures.

そこで FPGA ソリューション

FPGA - Field Programmable Gate Array

ハードウェアアクセラレーションのテクノロジーの中でも
高度で先進的なFPGAによる高速化，をゲノミクスの世界で既に実現.
導入先からも厚い信頼を得ている.

https://www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html

illumina

無敵の速さ. 大量でも速い.

Dragenは ハードウェアによるNGS二次解析の高速化を実現した製品です.

01 **高速性**
ヒト全ゲノムを25分で解析，業界最速レベルのスピード

02 **精確性**
SNPとINDELを高い感度，特異度で検出

03 **経済性とフレキシビリティ**
状況に応じてオンサイトサーバとクラウドを選択，
併用できる柔軟性

For Research Use Only. Not for use in diagnostic procedures.

illumina®

ゲノミクス イングランド
（Genomics England , 英） https://www.genomicsengland.co.uk/

Population Sequencing

背景
- Conducting the 100,000 Genomes Project
- Sequencing genomes from NHS subjects, with a focus on rare disease and common cancers

課題
- Re-process 5,000 WGS from individuals with rare disease using latest reference genome (GRCh38)

ソリューション
Process (align and call variants) samples with latest reference using:
- DRAGEN Germline Pipeline
- DRAGEN Joint Genotyping Pipeline

"Genomics England has found working with Illumina and formerly Edico, in the development and implementation of DRAGEN for our realignment and variant calling programme, very collaborative and constructive. DRAGEN is the first platform to utilise FPGA cards to optimise alignment, which has accelerated our project significantly, leading to more benefits... Genomics England is finalising work to reprocess all 100,000 genomes using DRAGEN and rolling it out prospectively for our production pipeline."

-Dr. Liz Edwards, Head of Bioinformatics Partnerships

ハドソンアルファ・バイオテクノロジー研究所
（Hudson Alpha Institute for Biotechnology，米）

High-volume genomics services laboratory

背景

- Large genomic services lab in Alabama, providing sequencing services to research, academic, and commercial customers globally
- Scaling up sequencing operations

課題

- With HiSeq™ X Ten, HAIB's output tripled, creating a need to analyze the data rapidly to avoid bottlenecks

ソリューション

DRAGEN on-premise enabled HAIB to phase out local analysis servers and speed up sample analysis to increase sample throughput and eliminate bioinformatics bottlenecks

使用パイプライン

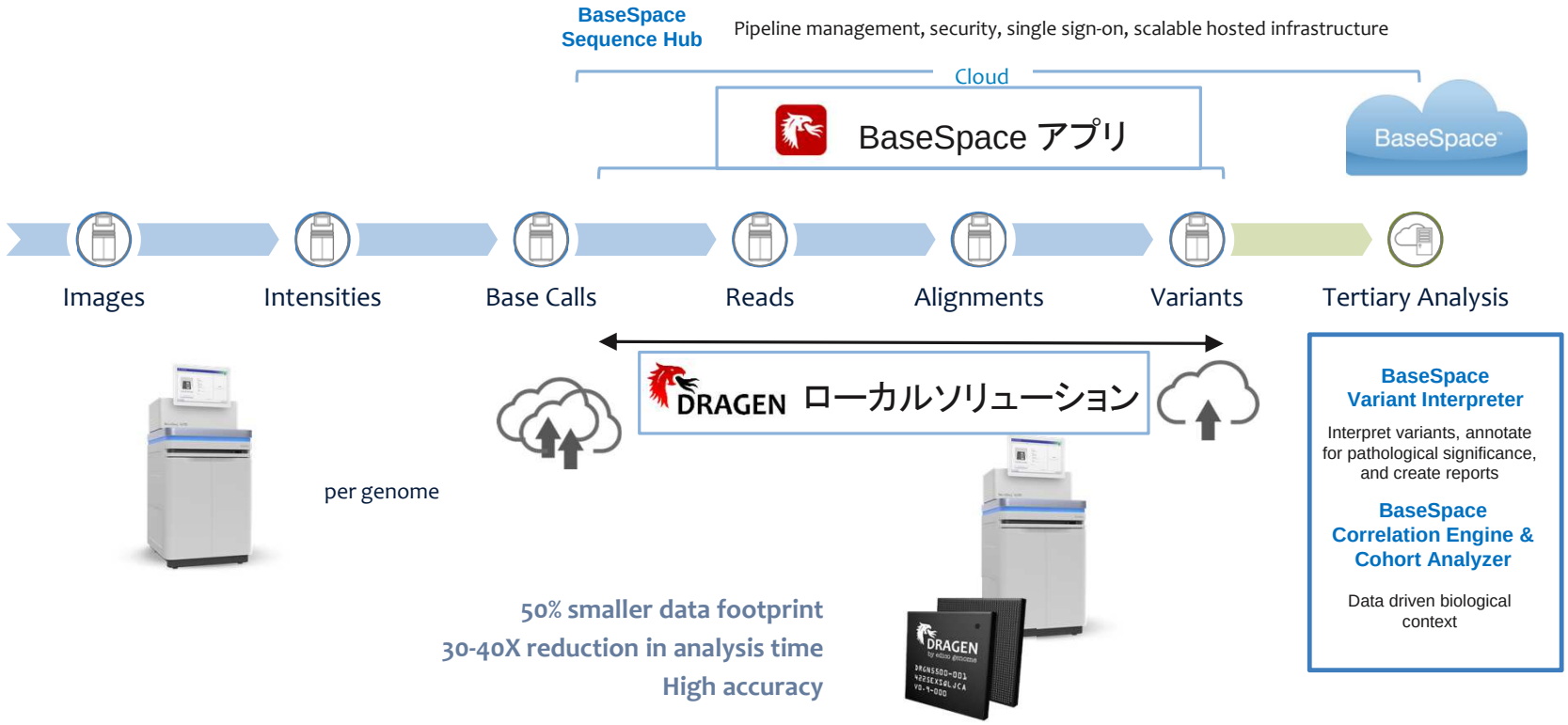
- DRAGEN Germline Pipeline
- DRAGEN Joint Genotyping Pipeline
- DRAGEN CNV Pipeline
- DRAGEN Somatic Pipeline
- DRAGEN RNA Pipeline
- DRAGEN VLRD Pipeline

~20,800サンプル/年/台

DRAGEN 1台で年間あたりに解析可能なサンプル数（x30, WGS）

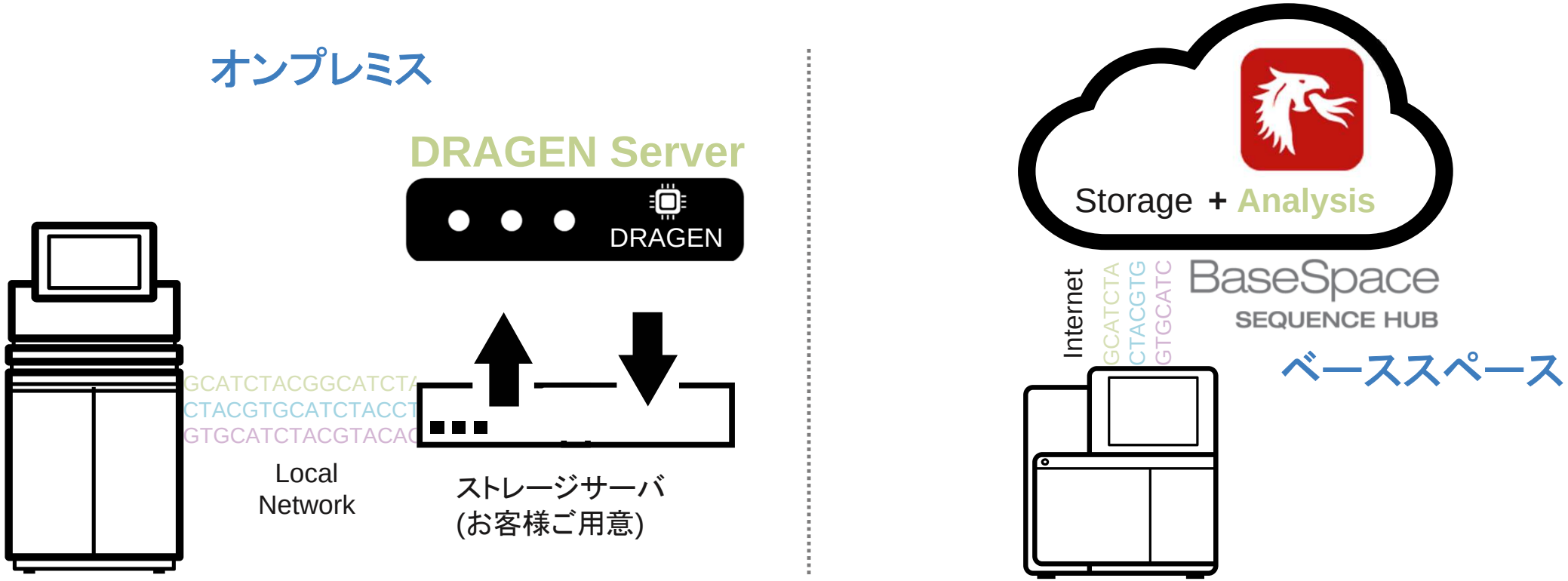
illumina®

どの工程を置き換えるもの？



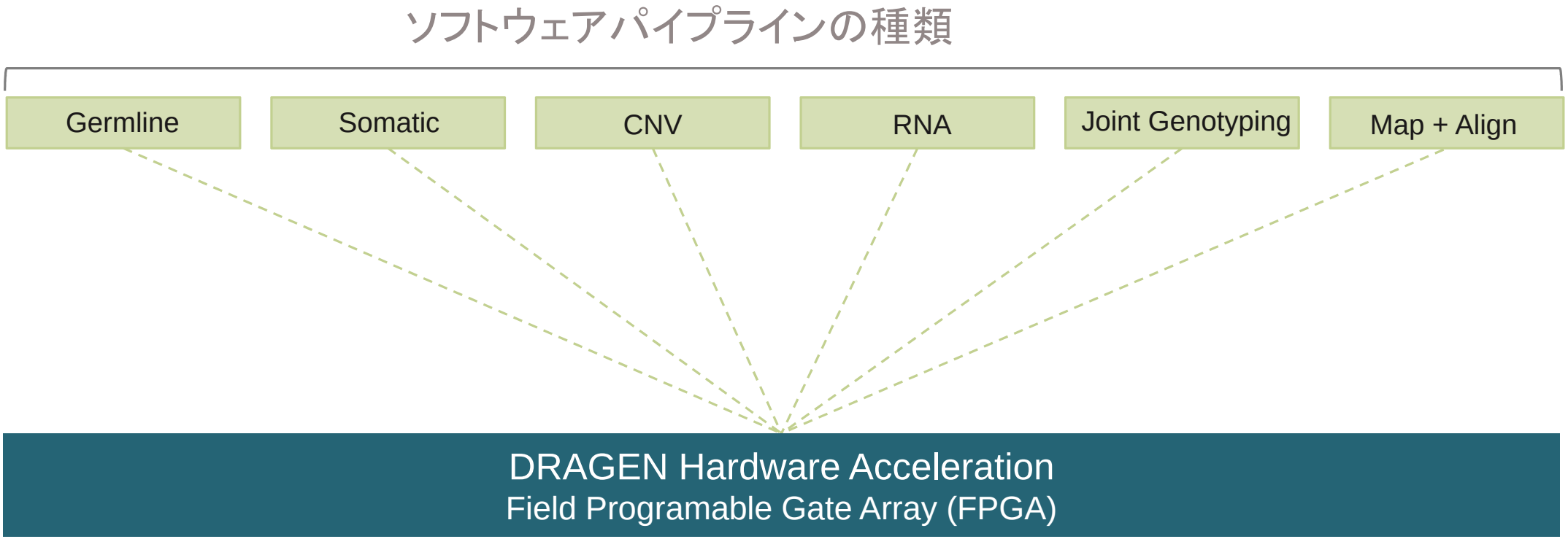
illumina®

陸の竜？ 空の竜？
両方あります． オンサイトローカルサーバ と クラウドソリューション



illumina®

DRAGEN™ の二次解析パイプライン *2019/6現在



illumina®

DRAGEN 日本語 フライヤー (定価入り)

http://bit.ly/draFlyer

超高速でコスト効率の高い、
NGS データの二次解析

DRAGEN Solutions

DRAGEN onsite　BaseSpace Sequence Hub

DRAGEN™（Dynamic Read Analysis for GENomics）は、Field-programmable gate array（FPGA）技術を用いた、超高速な次世代シーケンサー（NGS）データ解析ソリューションです。精確性を犠牲にすることなく、迅速、柔軟、コスト効率高く二次解析を実行します。

- 30x ヒトゲノムを 25 分で解析、業界最速レベルのスピード
- 小さなバリアントを高い感度と特異度でコールする精確性
- 必要に応じて、オンサイトサーバーとクラウドを選べる柔軟性

超高速
オンサイトの DRAGEN プラットフォームを使うことで、従来の CPU ベースのシステムでは 15 時間以上を要する 30x カバレッジでのヒト全ゲノム解析を、わずか 25 分間で処理することができます（図 1）。

高い精確性
DRAGEN プラットフォームは、世界で認められたアルゴリズムに基づき、主要なアプリケーションで高い解析感度と特異度を実現します。Precision FDA valiant calling challenge 2018 において、全ゲノムバリアントコールの精確性測定の 6 項目中 5 項目で最高評価を獲得しました[3]。

オンサイトとクラウドを選べる柔軟性
オンサイトサーバーを使えば、解析とデータ保管をローカルで実施できます。従来のコンピュートインスタンス 40 個に相当する処理を実行可能で、ハードウェア初期投資やメンテナンス等のコスト削減につながります。BaseSpace™ Sequence Hub（Amazon Web Services（AWS）をホストとして利用）を使えば、クラウド上でよりフレキシブルで拡張しやすい解析が実施できます。

ギネス世界記録

19.5 時間で最速癌全ゲノム診断
Rady Children's Institute Genomics Medicine

1000 人のゲノムデータ解析 2.5 時間
Childrens Hospital of Philadelphia

図 1：DRAGEN プラットフォームは、世界最速を記録した 2 つのゲノムデータ解析においても利用されました[1,2]

ポピュレーションシーケンス　シーケンスセンター
NICU　アクリゲノム　メタゲノム解析

図 2：DRAGEN プラットフォームは、生命科学のさまざまなアプリケーションで利用できます

最新の情報はウェブから　jp.illumina.com/DRAGEN

illumina

全ゲノム、エクソーム、トランスクリプトームからターゲットパネルまで、さまざまなアプリケーションに対応

表 1. DRAGEN 解析パイプライン

Illumina DRAGEN Germline Pipeline	BCL から VCF までの一貫した NGS データ解析。 advanced error model calibration により精確性を向上。 Illumina Expansion Hunter によるリピート伸長検出とジェノタイピングを実行
Illumina DRAGEN Somatic Pipeline	腫瘍サンプルの体細胞バリアント検出用に。 tumor-only と tumor-normal モードを提供
Illumina DRAGEN RNA Gene Fusion Detection Pipeline	スプライスジャンクションの探索、アラインメント、遺伝子融合検出を含むトランスクリプトーム解析を実施
Illumina DRAGEN Joint Genotyping/Population Pipeline	複数ゲノムの情報からジョイントバリアントコールを実施。 精確性を損なわずに迅速に数千サンプルまでスケールが可能
Illumina DRAGEN CNV Pipeline	エクソーム / ゲノム、生殖細胞 / 体細胞。 いずれにも対応するコピー数多型（CNV）解析。 様々なレベルのフィルターで、最終コール前に偽陽性を軽減
Illumina DRAGEN MAP + Align Pipeline	エクソーム / ゲノム、DNA/RNA。 いずれにも対応する超高速マッピングとアラインメント
Illumina DRAGEN Methylation Pipeline*	Reference-only のアラインメントや Annotation-assisted など 複数のモードを搭載予定

* 2019 年中にクラウドパイプラインとしてのみ提供予定。

製品情報 [*1]

カタログ番号	製品名	希望販売価格 (円)*	備考
オンサイトサーバー用 DRAGEN ライセンス（有効期限 1 年間）			
20027361	DRAGEN CS Lvl 1 1 year License	2,520,000	Level 1：スループット 100T base 分の シーケンスデータ解析用
20027362	DRAGEN CS Lvl 2 1 year License	5,990,000	Level 2：スループット 250T base 分の シーケンスデータ解析用
20027363	DRAGEN CS Lvl 3 1 year License	11,400,000	Level 3：スループット 500T base 分の シーケンスデータ解析用
20027364	DRAGEN CS Lvl 4 1 year License	20,200,000	Level 4：スループット 1,000T base 分の シーケンスデータ解析用
20027365	DRAGEN CS Lvl 5 1 year License	32,800,000	Level 5：スループット 2,000T base 分の シーケンスデータ解析用
オンサイトサーバー			
20027360	DRAGEN Bio-IT Platform 1 年保証付	4,370,000	FPGA 内蔵の DRAGEN サーバー [*2]
20031995	DRAGEN System Installation	500,000	イルミナスタッフによる据付を希望される場合のオプション
年間保守情報			
20032797	DRAGEN Server Advance Exchange Support Plan	551,000	1 年間の保証期間終了後のサポートプラン

*1. BaseSpace Sequence Hub 上での DRAGEN ご利用料金は、BaseSpace Sequence Hub 内の料金体系に従います。
*2. ラックマウント型、CentOS、CLI 環境です。解析予定の NGS データ量に合わせた DRAGEN ライセンスの同梱が必要です。

※ 価格に消費税は含まれません。
※ 2019 年 4 月 25 日現在の価格です。

参照ウェブサイト

1. Bio IT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. October 23, 2017. www.bio-itworld.com/2017/10/23/childrens-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed.aspx.Accessed September 19,2018.
2. The San Diego Union Tribune. Rady Children's Institutesets Guinness world record. February 12, 2018. www.sandiegouniontribune.com/news/health/sd-no-rady-record-20180209-story.html. Accessed September 19,2018.
3. PrecisionFDA Hidden Treasures Warm Up. https://precision.fda.gov/challenges/1/view/results Accessed September 14,2018.

(裏面)

illumina

19.5 時間での最速遺伝子診断

Ultra-Rapid Whole Genome Diagnosis for Critically Ill Newborns in 19.5 Hours

緊急性に応えます.

二次解析工程 **15 時間 → 25 分に！**

DRAGEN

サンプルプレップ → シーケンシング（NovaSeq6000） → 三次解析

19.5 時間

Guinness World Records® designation for Fastest Genetic Diagnosis for Genomic Medicine

米国 Rady Children's Institute が ギネスで 最速認定を更新

illumina®

Children’s Hospital of Philadelphiaの例

大量データでも一括で.　　大量再解析などに有効！

最速で 1,000 全ゲノムシーケンス解析を完了

1000WHG

2時間25分

DRAGEN

Guinness World Records® designation for Fastest Analysis of 1,000 Genomes

Children’s Hospital of Philadelphia (CHOP)が最速1000ゲノム解析でギネス認定

illumina®

製品データシートおよびアプリケーションノート

http://bit.ly/draAppNote201906

http://bit.ly/draDataSheet2019

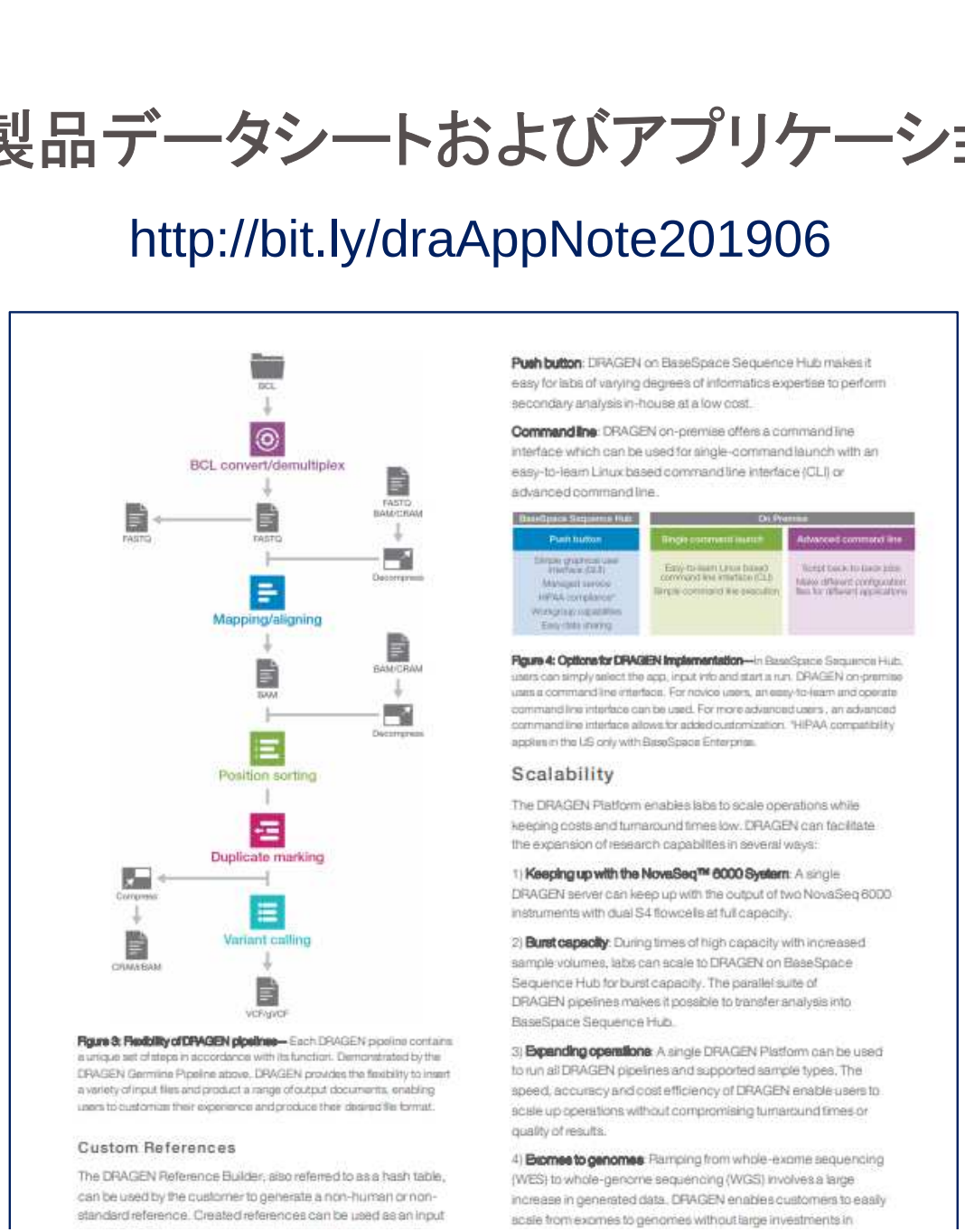


Figure 3: Flexibility of DRAGEN pipelines— Each DRAGEN pipeline contains a unique set of steps in accordance with its function. Demonstrated by the DRAGEN Germline Pipeline above, DRAGEN provides the flexibility to insert a variety of input files and product a range of output documents, enabling users to customize their experience and produce their desired file format.

Custom References

The DRAGEN Reference Builder, also referred to as a hash table, can be used by the customer to generate a non-human or non-standard reference. Created references can be used as an input

Push button: DRAGEN on BaseSpace Sequence Hub makes it easy for labs of varying degrees of informatics expertise to perform secondary analysis in-house at a low cost.

Command line: DRAGEN on-premise offers a command line interface which can be used for single-command launch with an easy-to-learn Linux based command line interface (CLI) or advanced command line.

Figure 4: Options for DRAGEN Implementation— In BaseSpace Sequence Hub, users can simply select the app, input info and start a run. DRAGEN on-premise uses a command line interface. For novice users, an easy-to-learn and operate command line interface can be used. For more advanced users, an advanced command line interface allows for added customization. *HIPAA compatibility applies in the US only with BaseSpace Enterprise.

Scalability

The DRAGEN Platform enables labs to scale operations while keeping costs and turnaround times low. DRAGEN can facilitate the expansion of research capabilities in several ways:

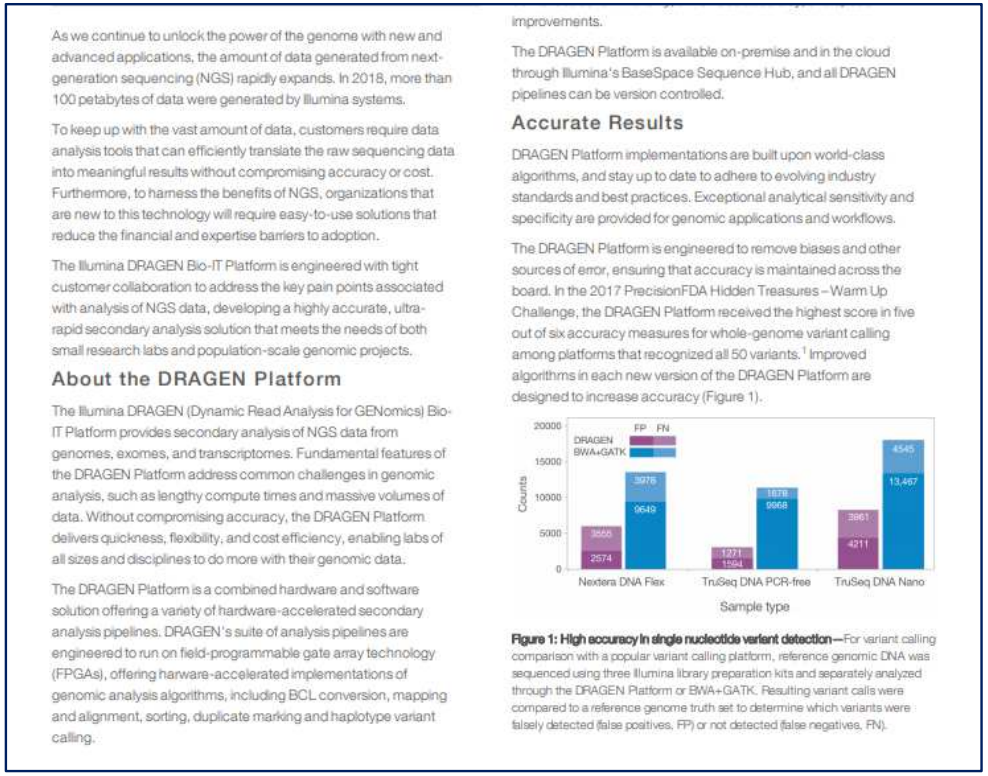
1) **Keeping up with the NovaSeq™ 6000 System**: A single DRAGEN server can keep up with the output of two NovaSeq 6000 instruments with dual S4 flowcells at full capacity.

2) **Burst capacity**: During times of high capacity with increased sample volumes, labs can scale to DRAGEN on BaseSpace Sequence Hub for burst capacity. The parallel suite of DRAGEN pipelines makes it possible to transfer analysis into BaseSpace Sequence Hub.

3) **Expanding operations**: A single DRAGEN Platform can be used to run all DRAGEN pipelines and supported sample types. The speed, accuracy and cost efficiency of DRAGEN enable users to scale up operations without compromising turnaround times or quality of results.

4) **Exomes to genomes**: Ramping from whole-exome sequencing (WES) to whole-genome sequencing (WGS) involves a large increase in generated data. DRAGEN enables customers to easily scale from exomes to genomes without large investments in

仕組みや精度について概説の資料



As we continue to unlock the power of the genome with new and advanced applications, the amount of data generated from next-generation sequencing (NGS) rapidly expands. In 2018, more than 100 petabytes of data were generated by Illumina systems.

To keep up with the vast amount of data, customers require data analysis tools that can efficiently translate the raw sequencing data into meaningful results without compromising accuracy or cost. Furthermore, to harness the benefits of NGS, organizations that are new to this technology will require easy-to-use solutions that reduce the financial and expertise barriers to adoption.

The Illumina DRAGEN Bio-IT Platform is engineered with tight customer collaboration to address the key pain points associated with analysis of NGS data, developing a highly accurate, ultra-rapid secondary analysis solution that meets the needs of both small research labs and population-scale genomic projects.

About the DRAGEN Platform

The Illumina DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT Platform provides secondary analysis of NGS data from genomes, exomes, and transcriptomes. Fundamental features of the DRAGEN Platform address common challenges in genomic analysis, such as lengthy compute times and massive volumes of data. Without compromising accuracy, the DRAGEN Platform delivers quickness, flexibility, and cost efficiency, enabling labs of all sizes and disciplines to do more with their genomic data.

The DRAGEN Platform is a combined hardware and software solution offering a variety of hardware-accelerated secondary analysis pipelines. DRAGEN's suite of analysis pipelines are engineered to run on field-programmable gate array technology (FPGAs), offering harware-accelerated implementations of genomic analysis algorithms, including BCL conversion, mapping and alignment, sorting, duplicate marking and haplotype variant calling.

improvements.

The DRAGEN Platform is available on-premise and in the cloud through Illumina's BaseSpace Sequence Hub, and all DRAGEN pipelines can be version controlled.

Accurate Results

DRAGEN Platform implementations are built upon world-class algorithms, and stay up to date to adhere to evolving industry standards and best practices. Exceptional analytical sensitivity and specificity are provided for genomic applications and workflows.

The DRAGEN Platform is engineered to remove biases and other sources of error, ensuring that accuracy is maintained across the board. In the 2017 PrecisionFDA Hidden Treasures – Warm Up Challenge, the DRAGEN Platform received the highest score in five out of six accuracy measures for whole-genome variant calling among platforms that recognized all 50 variants.[1] Improved algorithms in each new version of the DRAGEN Platform are designed to increase accuracy (Figure 1).

Figure 1: High accuracy in single nucleotide variant detection— For variant calling comparison with a popular variant calling platform, reference genomic DNA was sequenced using three Illumina library preparation kits and separately analyzed through the DRAGEN Platform or BWA+GATK. Resulting variant calls were compared to a reference genome truth set to determine which variants were falsely detected (false positives, FP) or not detected (false negatives, FN).

illumina®

BaseSpace ブログ記事

http://bit.ly/draBSblog

BaseSpace Informatics Blog
by Illumina

Home About Subscribe to RSS

May 29, 2019
by scatreux
in DRAGEN, Features, Releases
Leave a comment

Somatic Pipeline Improvements with DRAGEN v3.3

by Severine Catreux – Associate Director, Bioinformatics FPGA Development

Significant accuracy gains and speed improvements with DRAGEN v3.3, released April 2019

The DRAGEN engineering and bioinformatics team is excited to announce a new DRAGEN release, v3.3. The second of several releases scheduled for 2019, DRAGEN v3.3 contains improvements across the many pipeline offerings now supported by the DRAGEN platform. This includes accuracy improvements in the germline and somatic pipelines, new features (e.g. CNV DeNovo calling and RNA quantification) and speed gains (Somatic T/N, BCL conversion).

Please see DRAGEN v3.3 Release Notes for more details. This blog highlights the significant updates to the DRAGEN Somatic Pipeline for small variants, that are part of the v3.3 release.

As one of DRAGEN's core pipelines, the DRAGEN Somatic Pipeline for small variants is utilized by cancer research institutes around the globe. Expanding on the existing functionality, accuracy and speed of the DRAGEN Somatic Pipeline, the v3.3 release placed a high focus on the somatic tumor/normal WGS mode, producing step-function improvements in both accuracy and speed.

Accuracy Improvements:

During the development cycle for v3.3, the DRAGEN engineering and bioinformatics teams took a deep dive into the DRAGEN Somatic Pipeline tumor/normal mode, strengthening the existing algorithm for accuracy improvements. Specific improvements were made in the genotyping module,

SEARCH

BASESPACE SUITE

- BaseSpace Clarity LIMS
- BaseSpace Sequence Hub
- BaseSpace Variant Interpreter
- BaseSpace Cohort Analyzer
- BaseSpace Correlation Engine

RECENT POSTS

- Somatic Pipeline Improvements with DRAGEN v3.3
- Enabling Cancer Interpretation At Scale For The Genomics England 100K Genomics Project
- Doing more with DRAGEN™ v3.2.8
- New Sequence Quality Metrics in BaseSpace™ Sequence Hub
- Putting Your Privacy First

ARCHIVES

Select Month

CATEGORIES

tumor and normal control, to allow for differences in library preparation between the tumor sample and the control sample.

These changes are precursors to further accuracy improvements planned for the DRAGEN v3.4 release, specifically in the area of liquid tumor support, where tumor-in-normal contamination will be taken into account.

Accuracy gains of DRAGEN 3.3 over previous DRAGEN versions (3.2) as well as other pipelines (GATK4 MuTect2 and Strelka2) are shown in the plot below. Gains are measured for both SNVs and indels on most datasets.

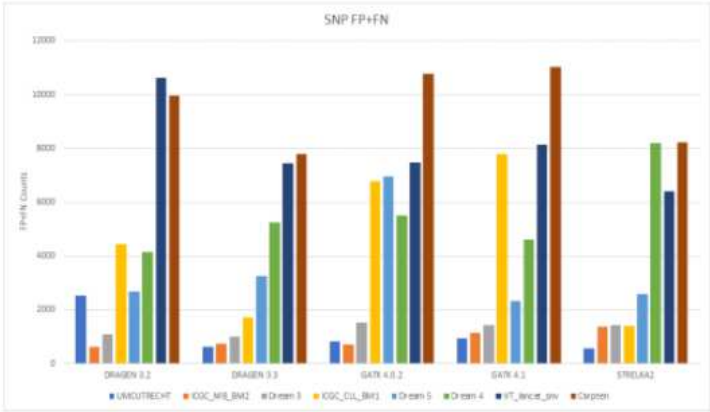


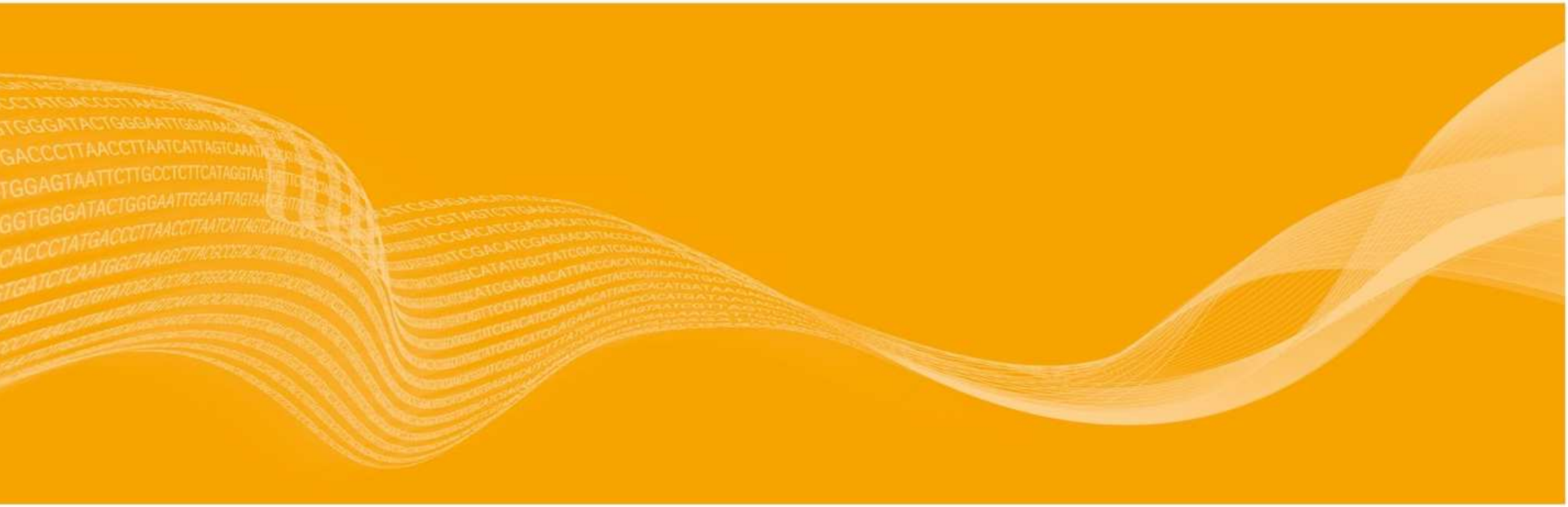
Figure 1: Comparison of False-Positives (FP) and False-Negatives (FN) between GATK4, Strelka2, DRAGEN 3.2 and DRAGEN 3.3. Lower values are better.

illumina®


ご覧いただきまして、誠にありがとうございました.
A New Era of Informatics with Illumina

Questions?

ご興味いただけましたら
担当営業もしくは以下へお声がけください.
contactJPN@illumina.com



既にご利用頂いているお客様のサポートお問合わせは
テクニカルサポート(techsupport@illumina.com) を引き続きご利用ください.

https://jp.illumina.com/company/contact-us.html