

DRAGEN™ ではじめる NGSデータ解析： クラウドで簡単・高速・高精度に！

篠崎 良仁

テクニカルアプリケーションサイエンティスト

2023年 8月31日

illumina®

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

M-JP-00168 | © 2023 Illumina, Inc. All rights reserved.



本ウェビナーの目的

■ 次世代シーケンサー (NGS) データの解析におけるボトルネック

出力されたシーケンスデータを、

いかに**正確かつ効率的に、解釈可能な情報に変換**するか



■ 期待される解析ソリューションの条件

- ✓ 正確性 & 効率性が高い
- ✓ コスト・技術・環境条件での障壁が低い
- ✓ 経験レベルが異なるユーザーに幅広く対応



上記に応える DRAGEN についてご紹介

DRAGEN Bio-IT Platform: <https://jp.illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html>

本ウェビナーの内容

- 1 DRAGENとは
- 2 DRAGENの解析性能のご紹介
- 3 クラウド解析プラットフォーム
BaseSpace™ Sequence Hub
(BSSH) 版 DRAGEN のご紹介
- 4 全ゲノム / エクソーム解析用
BSSH DRAGEN アプリの使い方
- 5 本ウェビナーのまとめ

進化を続ける高い性能

クラウド環境での解析方法

DRAGENとは



ILLUMINA DRAGEN (Dynamic Read Analysis for GENomics)

= 高い正確性と高速性を兼ね備えた NGS データの二次解析ツール

ライブラリー
調製



一次解析
シーケンシング
塩基配列の決定



二次解析
解釈可能な
情報への変換



三次解析
解釈および
レポート作成



主な解析パイプライン

リード QC・前処理

DNA

RNA

アンプリコン

メチレーション

シングルセル

ヒト全ゲノムの変異
(バリエーション) 解析の例

一次解析



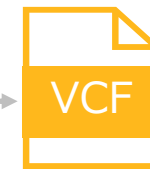
塩基配列

マッピング &
アラインメント



二次解析

バリエーション
コール



バリエーション

二次解析を
30分以内で実現*

三次解析へ



遺伝性疾患



腫瘍学



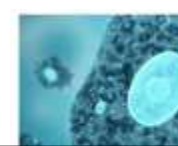
集団遺伝学



ファーマコゲノミクス



感染症学



細胞分子生物学



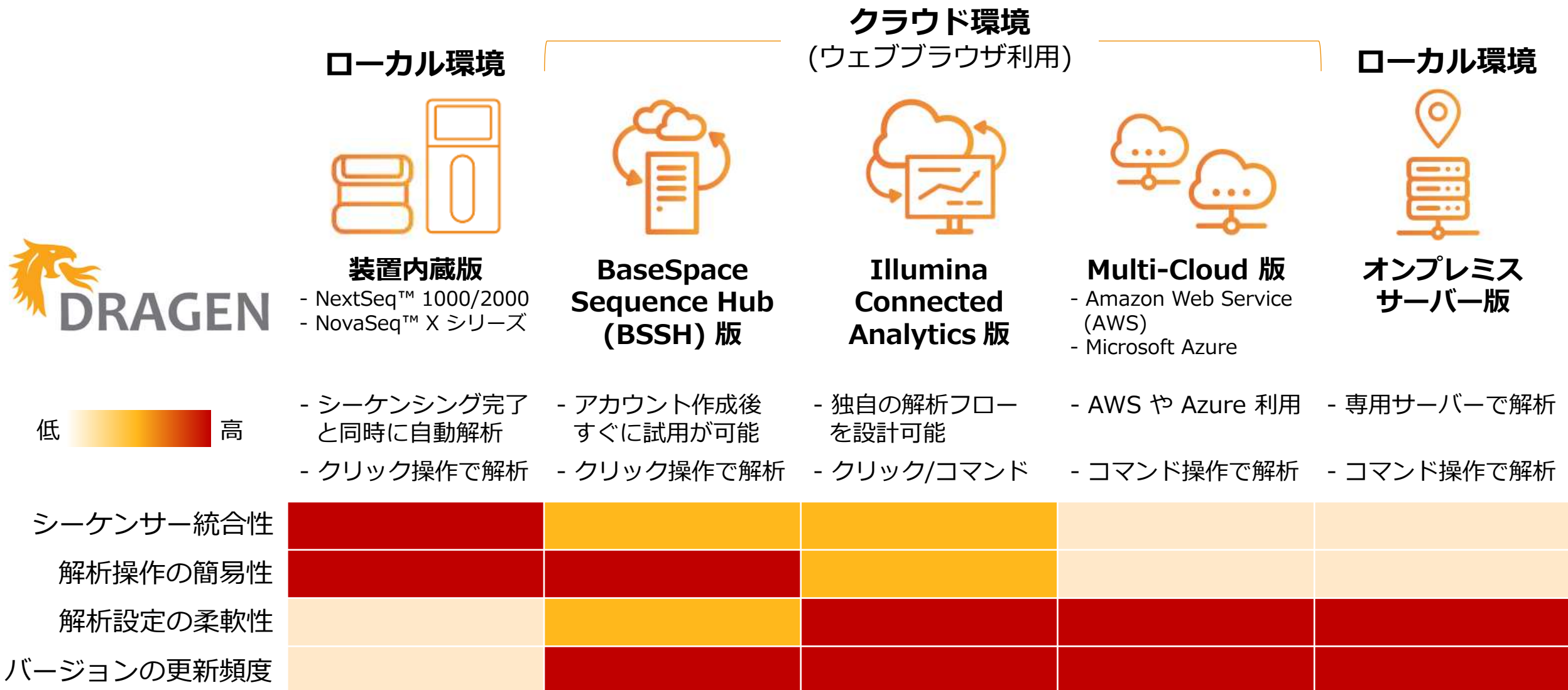
アグリゲノミクス

illumina

*HG002 from [PrecisionFDA truth challenge V2](#) run with DRAGEN v4.0 on DRAGEN v4 server, all callers

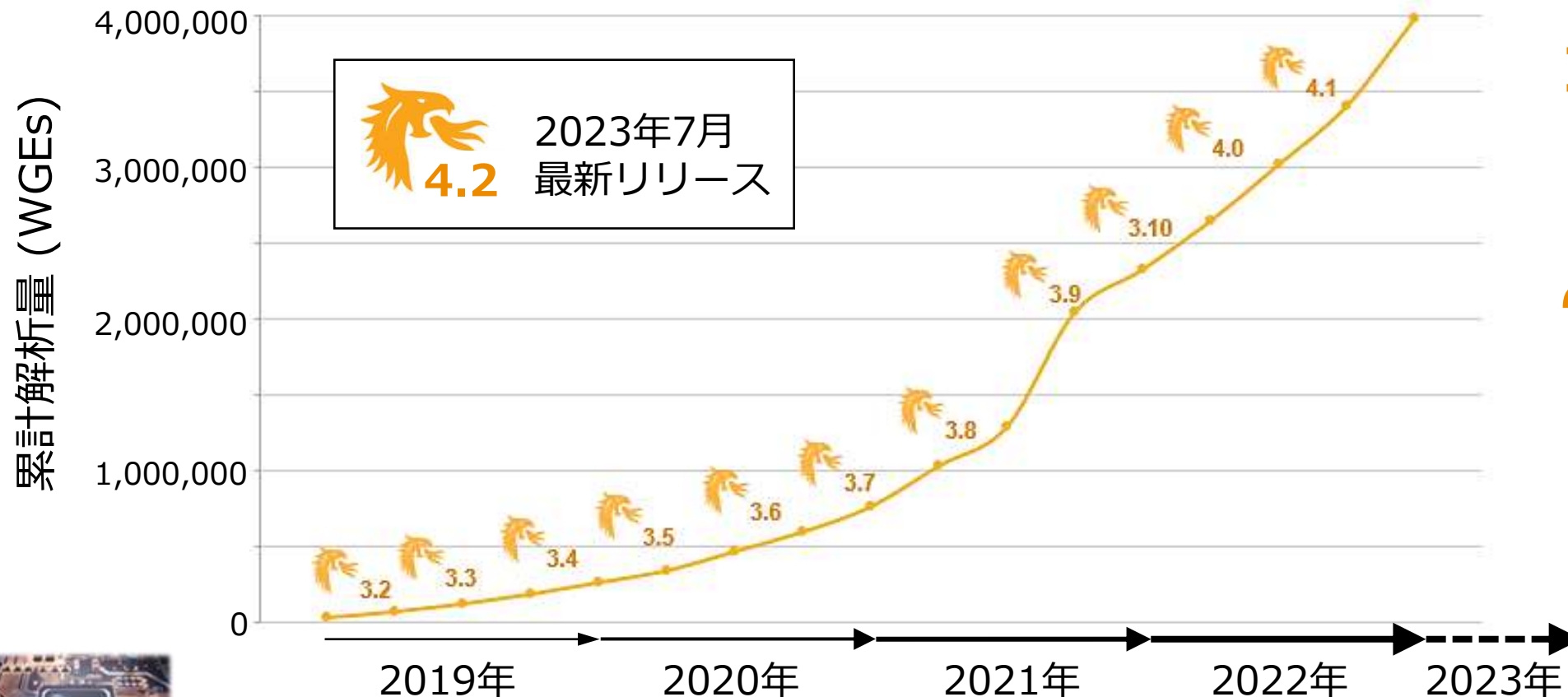
For Research Use Only. Not for use in diagnostic procedures.

最適な解析プラットフォームの選択が可能



強い成長を続けるDRAGENの解析実績

DRAGENで解析された Whole Genome Equivalents (WGEs) 量



125%

直近3年間の解析量の
平均増加率

~400万

累計解析 WGE 量

1000+ 機関

45か国



代表的なDRAGEN利用文献リスト

DRAGEN Publications: <https://developer.illumina.com/news-updates/dragen-publications>

DRAGENの解析性能の ご紹介



高精度なバリエント検出 (バリエントコール)



バリエントコールの精度を競う
コンペティション

“0.01%” の差が
さらなる精度向上に貢献

CMRGs⁴ の Germline バリエントコールで
偽陽性+偽陰性 が25%減少する例も⁵

生殖細胞系列 (Germline) バリエント

[Truth Challenge V2](#)¹

2020年8月受賞



先天的

マッピング困難領域 & 全ベンチマーク領域
→ 最も高いバリエントコール精度²

2 - 検出感度 (Recall) & 陽性的中率 (Precision) に基づく F1スコア評価

体細胞系列 (Somatic) バリエント

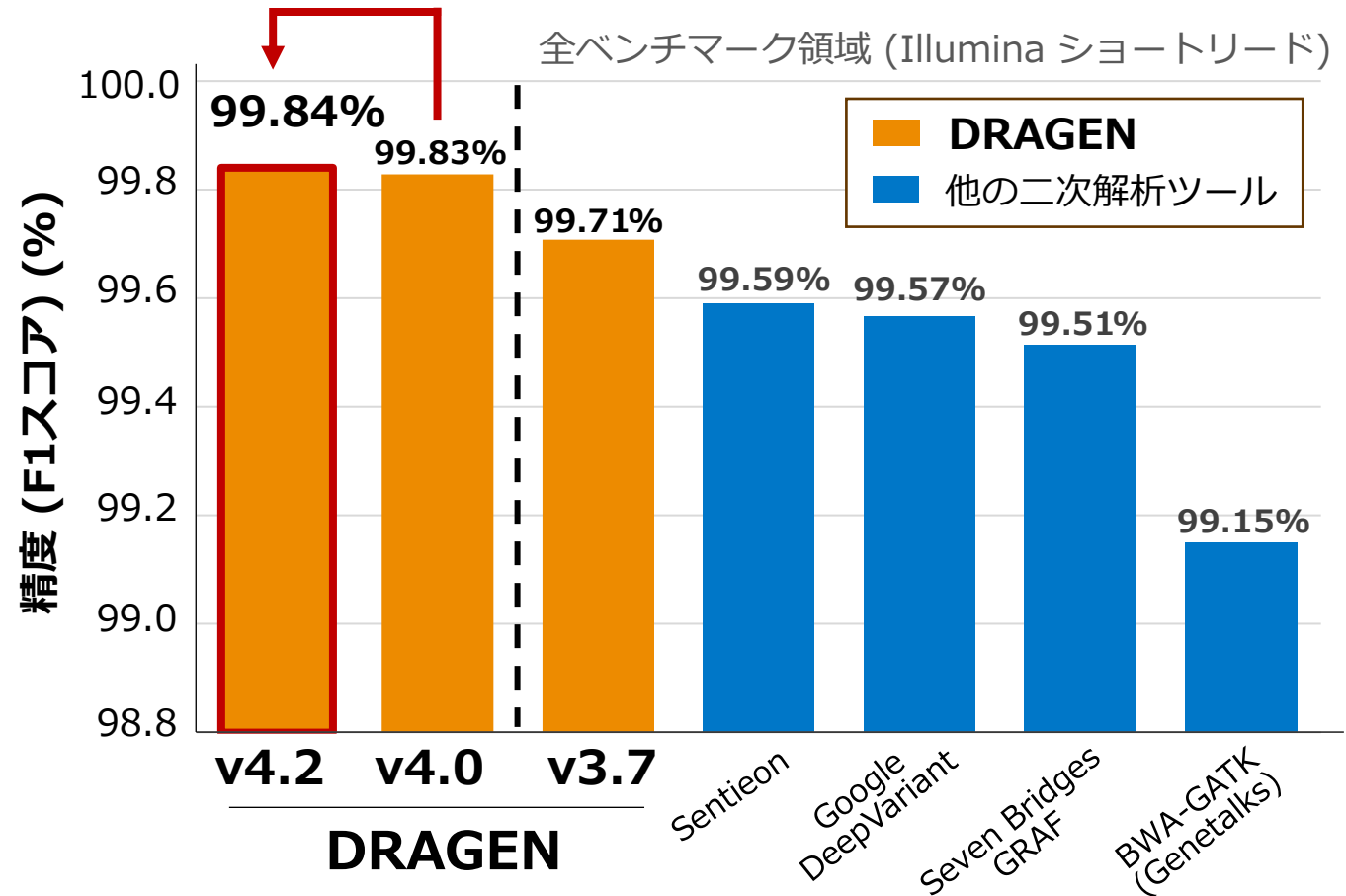
[NCTR Indel Calling from
Oncopanel Sequencing Data](#)³

2022年9月受賞



後天的

ベンチマーク用 がん遺伝子パネル X
→ 最も高い Indel バリエントコール精度²



4 - Challenging Medically Relevant Genes

(バリエント解析が困難かつ医学的に重要なベンチマーク遺伝子群)

5 - [リリースノート \(v4.2.4\)](#)

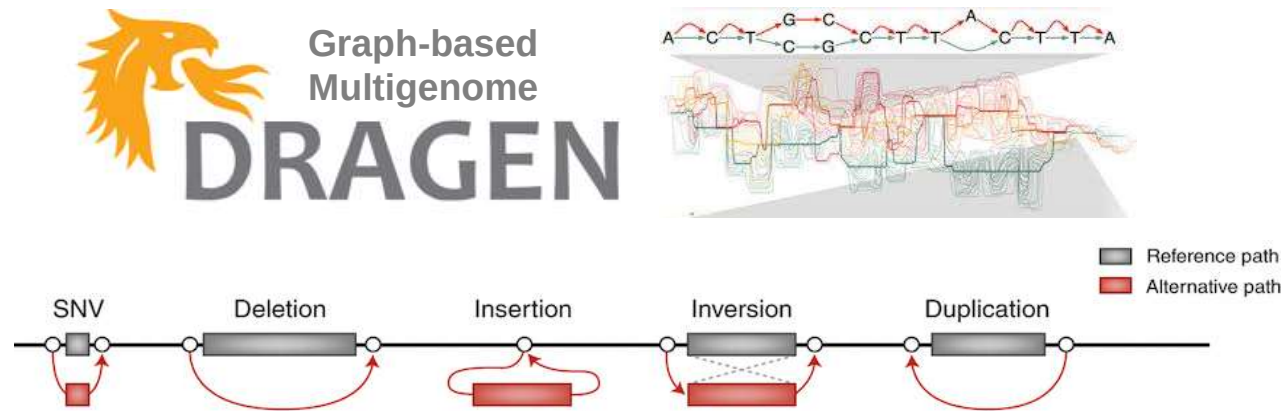
1 - <https://precision.fda.gov/challenges/10>

3 - <https://precision.fda.gov/challenges/21>

DRAGENの高精度なバリエントコールを実現した特徴的技術

グラフリファレンスゲノム¹ 3.7~

→ 既知の多型やハプロタイプを含むヒトゲノム参照配列



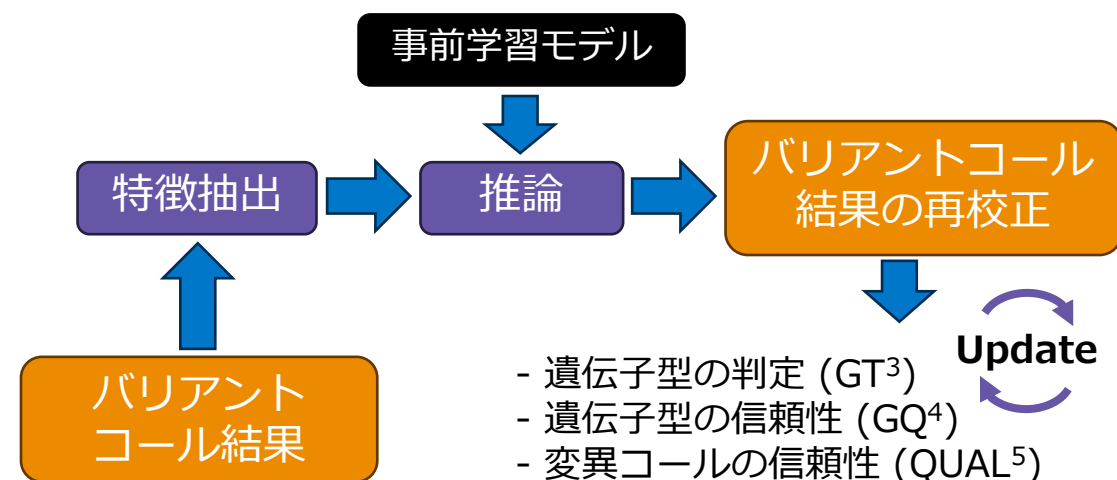
- マッピングが困難な領域² のバリエントコール精度を大幅に向上
- 一般的なリファレンス (hg38, hg19, GRCh37d5, T2T-CHM13)、アラインメント (BAM) やバリエントコール (VCF) のフォーマットと互換

¹ - マルチゲノムとも呼ばれる

² - 多型性が高い領域、類似性の高いセグメント重複領域、リピート配列領域など

機械学習 3.9~

→ Germline バリエントコールの精度を向上



- 事前学習モデルを利用してバリエントコールを再校正
- 一般的なリファレンスとグラフリファレンスのいずれと合わせてもバリエントコールの精度を向上

3 - Genotype

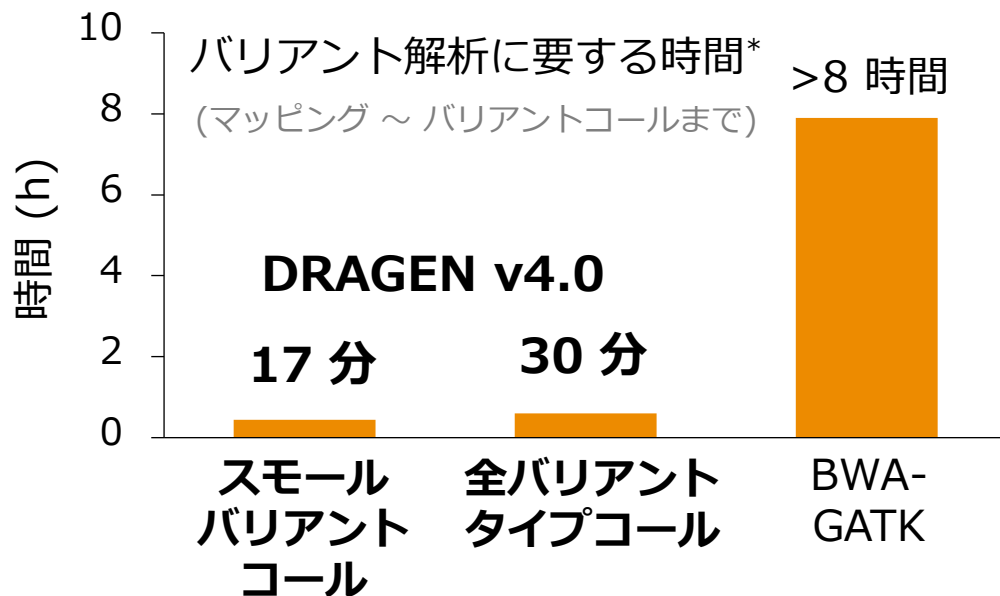
4 - Genotype quality

5 - Quality

ハイスループット解析を可能とする高速性 & データ圧縮性能



超高速な二次解析を実現

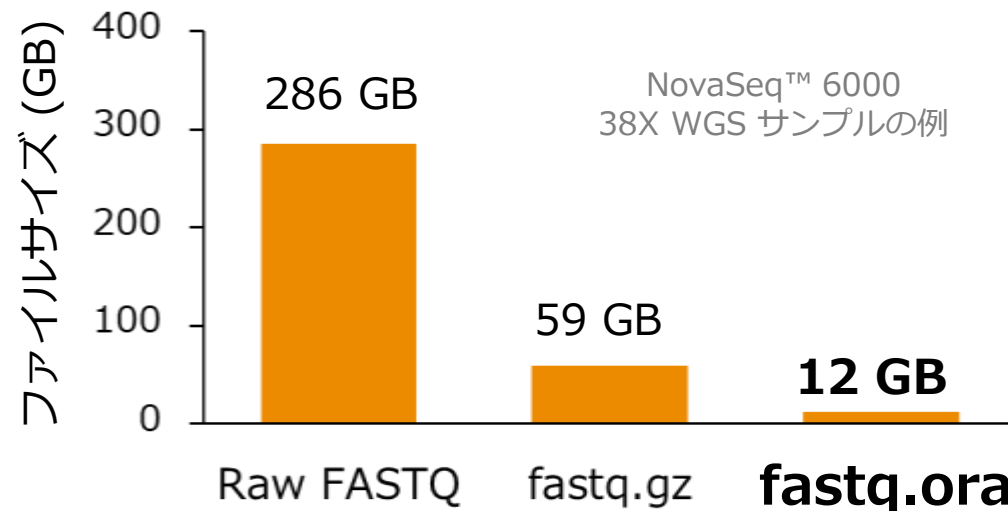


BWA-GATK よりも16倍以上高速

*[PrecisionFDA Truth Challenge V2](#) の HG002 (34.46X) の解析結果例。
DRAGEN は v4サーバーで解析 (U200 FPGAカード + 機械学習オン)。
BWA GATK 4.1.4.0 は 2x Intel Xeon Gold 6126 (48 threads) で解析
(394 GB RAM + 2TB NVME SSD)。並列化のために BCBIO を使用。



ロスレスなFASTQの
高圧縮率フォーマット “ORA”
(オプション)



- gz 形式の ~1/5 サイズ & 同等以上の圧縮・解凍速度
- 解凍ツールの無償提供

ストレージやデータ転送に係る運用コストの削減

ORA技術の詳細:

<https://www.illumina.com/science/genomics-research/articles/design-ora-lossless-genomic-compression.html>

最新版 DRAGEN v4.2 におけるアップデート (2023年7月リリース)



正確性

- スモールバリエーション (SNV/Indel) 、コピー数バリエーション (CNV)、構造バリエーション (SV) のコールの精度向上
- CMRG 領域のカバー率が向上し、新たに5つのターゲットコーラー¹ が追加
- 5つの新たな Star Allele コーラー² によるファーマコゲノミクス解析精度の向上
- **Telomere-to-Telomere (T2T) CHM13 v2.0 リファレンスのサポート**

1 – CYP21A2, HBA, SMN silent carrier variant, LPA, RH

2 – BCHE, ABCG2, NAT2, F5, UGT2B17

包括性

- 性染色体の解析サポートの強化
- Explify™ 解析パイプライン (RPIP³ & UPIP⁴) の統合
- バルクRNA QC メトリクスの強化
- DRAGEN Germline 解析における高感度モードの実装

3 – Respiratory Pathogen ID/AMR Enrichment Panel Kit

4 – Urinary Pathogen ID/AMR Enrichment Panel Kit

効率性

- ジョイントジェノタイピングの高速化
- マッピング/アラインメント が有効な場合の ORA 圧縮速度の向上

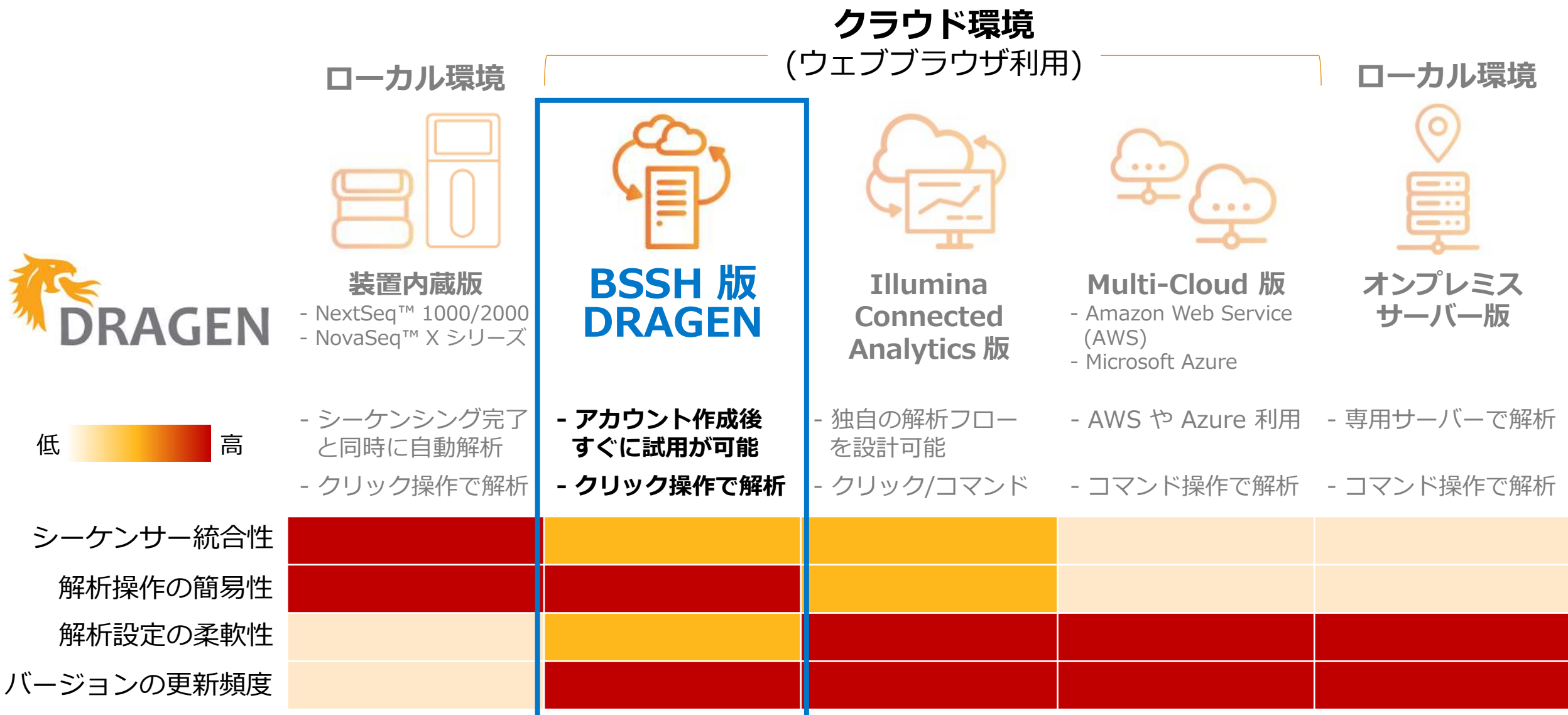
詳細は下記をご参照ください。

- ◆ ウェビナー録画 > [Illumina DRAGEN 4.2: Enhanced machine learning, new targeted callers, and more](#)
- ◆ ユーザーガイド > [Illumina DRAGEN Bio-IT Platform v4.2 Online Help](#), [リリースノート \(v4.2.4\)](#)

クラウド解析プラットフォーム BaseSpace Sequence Hub (BSSH) 版 DRAGEN のご紹介



ハードウェアの制限なく、手軽に使い始められるクラウド BSSH 版

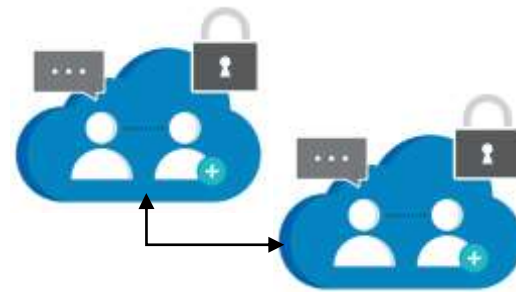


クラウド解析プラットフォーム BaseSpace Sequence Hub (BSSH) とは

配列情報の
アップロード



アプリによる解析



安全なデータ共有



DRAGENを含む100以上の解析アプリ
1ヶ月間の無償トライアルが可能¹



クリック操作での解析フロー



低い導入コスト (年間 69,600円² から)



Amazon Web Service (AWS) 上に
構築された安全性の高い環境

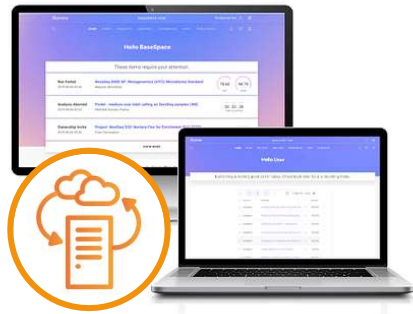
参考 (日本語ウェビナー)

[BaseSpace Sequence Hub \(BSSH\) の基本: NGSデータのアップロードから解析まで](#)

1 - 有効期限1カ月間の 250 iCredits を含む

2 - Professionalアカウント (500 iCredits を含む)
年間ライセンス希望販売価格 (税抜) (2023年8月現在)

BSSH DRAGEN アプリを用いた解析の流れ



Win・Macなどの端末の
ウェブブラウザからログイン

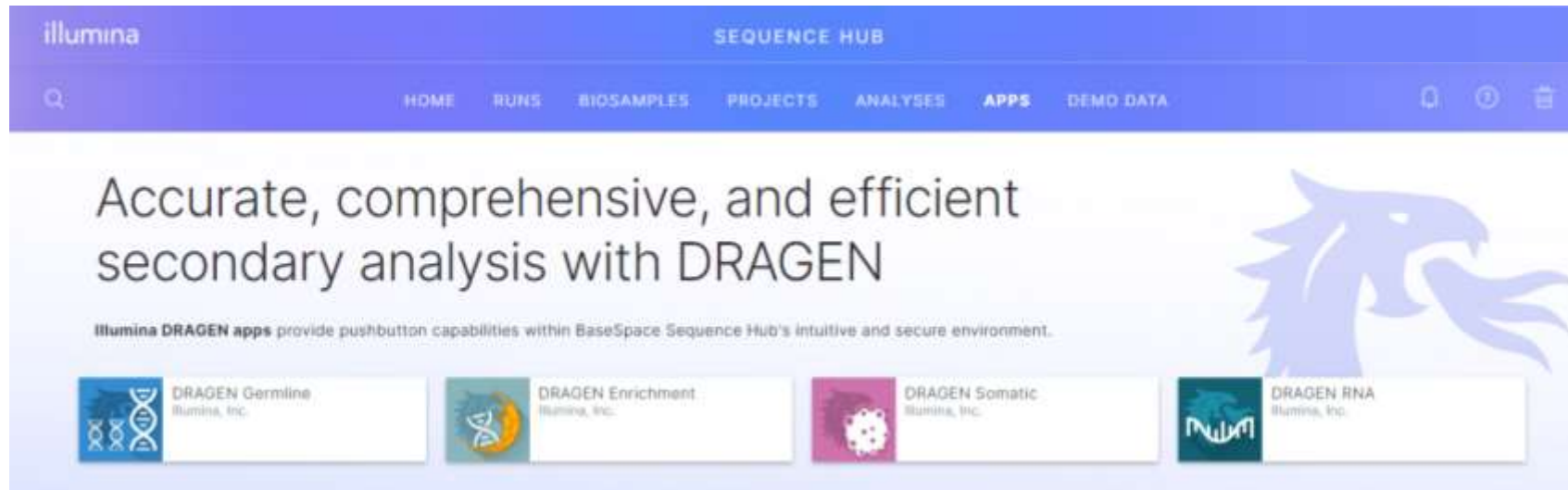


解析アプリを選択して起動



- アップロード済みの解析サンプルの選択
- 解析パラメータの設定

↓ 解析実行



など

出力されたQCレポートや
解析結果の確認

代表的な BSSH DRAGEN アプリ

リードQC・前処理

 **DRAGEN
FastQC+MultiQC**

FastQC

 **DRAGEN
FASTQ Toolkit**

- アダプタートリム
- クオリティトリム
- ダウンサンプリング など

その他

-  Metagenomics
-  COVID Lineage
-  RNA Pathogen Detection
-  Imputation

他8以上
(2023年8月現在)

illumina®

DNA

本ウェビナーで解説: 全ゲノム / エクソーム 解析用



**DRAGEN
Germline**

Germline バリアントコール



**DRAGEN
Enrichment**
(Germline & Somatic)

Somatic バリアントコール



**DRAGEN
Somatic**



**DRAGEN
Joint Genotyping**

Joint Calling

- Pedigree mode (トリオ解析) - Population mode



**DRAGEN
PopGen**

アンプリコン



**DRAGEN
Amplicon**

Germline/Somatic バリアントコール

遺伝子 融合検出/発現定量

DNA アンプリコン

メチレーション



**DRAGEN
Methylation Pipeline**

バイサルファイトシーケンシング解析

RNA アンプリコン

RNA

[DRAGEN RNA App](#)
[解説日本語ウェビナー](#)



**DRAGEN
RNA**

スモールバリエントコール

遺伝子 融合検出/発現定量



**DRAGEN
Differential Expression**

遺伝子発現差解析

シングルセル



**DRAGEN
Single Cell RNA**

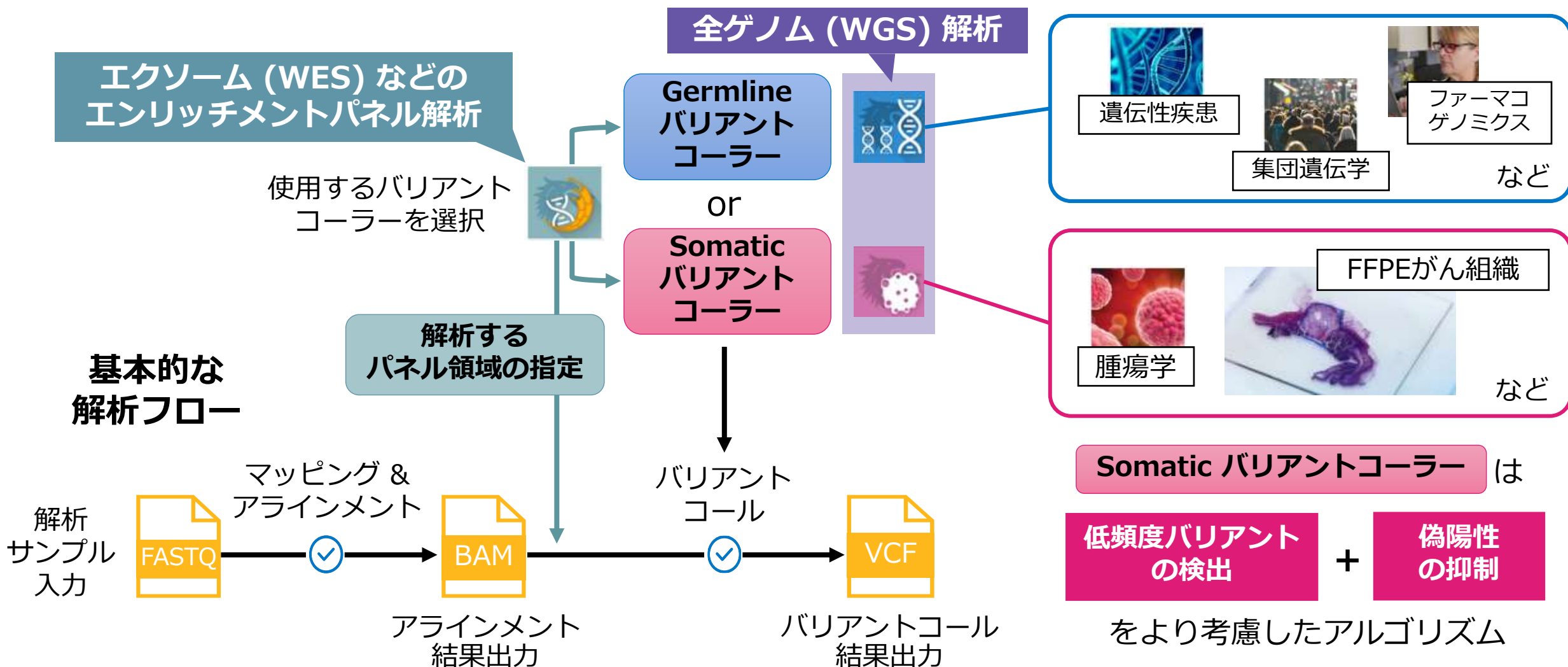
発現マトリックスの出力



**DRAGEN
Single Cell ATAC**

クロマチンアクセシビリティのピークコール

DRAGEN Germline / Somatic / Enrichment アプリの使い分け

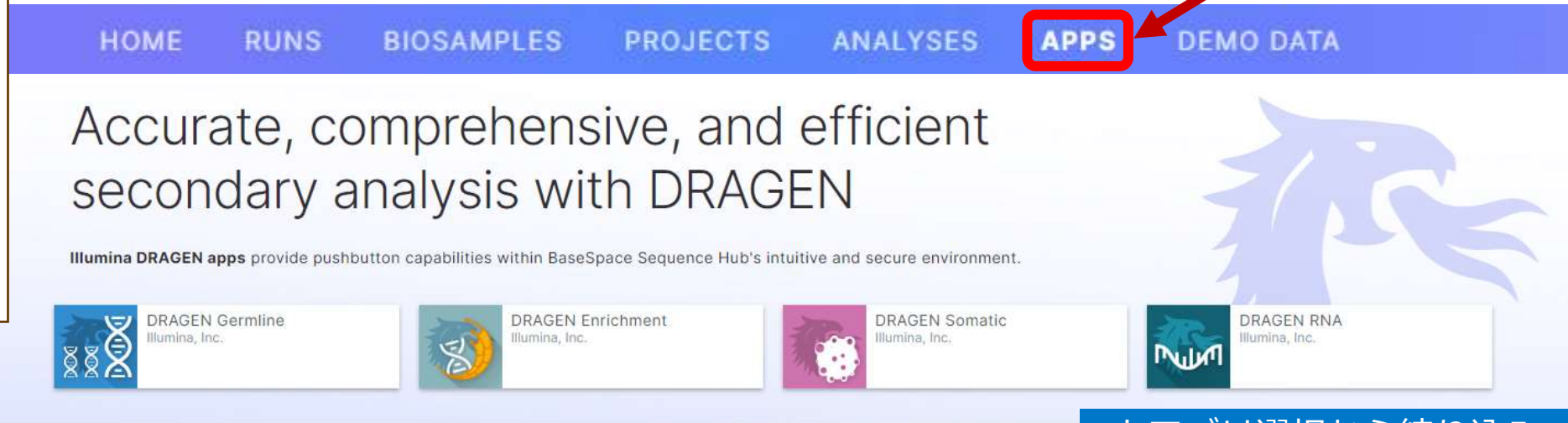


全ゲノム / エクソーム解析用 BSSH DRAGEN アプリの 使い方



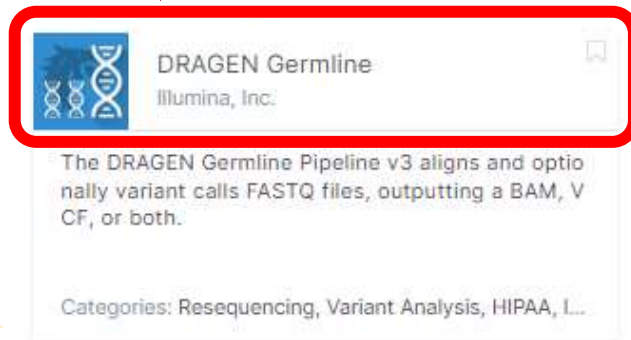
BSSH DRAGEN アプリの起動方法

① アプリの一覧ページに移動



Win・Macなどの端末の
ウェブブラウザからログイン

② 選択してアプリの
トップページに移動



Germline

文字検索から
絞り込み

下記3つは旧バージョン (v3.2 - 3.7) の
アプリとなりますので、ご注意ください。



- DRAGEN Germline **Pipeline**
- DRAGEN Somatic **Pipeline**
- DRAGEN RNA **Pipeline**

最新版は **Pipeline** が付かないアプリ

カテゴリ選択から絞り込み

DRAGEN (28)

All Categories

ChIP-Seq (4)

De Novo Assembly (9)

Differential Expression (18)

Dragen (28)

Gene Fusion Detection (11)

HIPAA (36)

illumina

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

DRAGEN アプリのトップページ (DRAGEN Germline の例)



DRAGEN Germline

Illumina, Inc.

Bookmark this app

DRAGEN Germline Online Help

Help

解析パラメータの説明など

Pricing

Compute cost 5.00 iCredits

Version

per node hour

4.2.4



アプリのバージョン

解析コスト単価
(ノード × 時間あたり)

DRAGEN Germline is an accurate and efficient end-to-end (FASTQ to VCF) secondary analysis solution for whole genome, whole exome, and targeted panel NGS data. This app takes input files in FASTQ, BAM, and CRAM format. Files may be decompressed, go through Map/Align/Sort, and go through variant calling.

仕様の概要

最新バージョン
をお知らせ



A newer version of this app (v4.2.4) is available.

Mapping

リファレンス一覧

- Human (UCSC hg19 Alt-Masked v3)
- Human (UCSC hg19 Alt-Masked v3, Graph Enabled)
- Human (UCSC hg38 Alt-Masked v3)
- Human (UCSC hg38 Alt-Masked v3, Graph Enabled) (default)
- Human (Ensembl GRCh37)
- Human (Ensembl hs37d5: GRCh37 + decoys)
- Human (Ensembl hs37d5: GRCh37 + decoys, Graph Enabled)
- Human (CHM13v2 Alt-Masked v3)
- Human (CHM13v2 Alt-Masked v3, Graph Enabled)
- Custom

クリックでアプリ起動

LAUNCH APPLICATION

Variant Calling

バリエーションコール
の仕様

- Small Variants (Single Nucleotide Variants and Indels)
- Copy Number Variants
- Structural Variants
- Short Tandem Repeats
- Star Allele Caller – identifies star alleles and metabolism status of 16 Tier-1 PGx genes
- Targeted Callers – for CYP2D6, CYP2B6, SMN, GBA, and HLA typing

Input Files

入力ファイル

- FASTQ
- BAM or CRAM
- Custom reference (optional)
- Target BED (optional) (Contigs in bed file must match the chosen reference)
- Additional optional input files

Output Files

出力ファイル

- Aligned reads (if enabled) – BAM or CRAM
- Small variant calls – VCF or gVCF
- Structural variants (if enabled) – VCF
- Copy number variants (if enabled) – GFF3 and VCF
- Tandem repeats (if expansion hunter is enabled) – BAM and VCF
- Variant annotations (if enabled) – JSON
- Targeted callers (if enabled) – TSV
- Star alleles (if enabled) – JSON, TSV, and VCF
- Metrics – CSV

デモデータ

Example Data Set

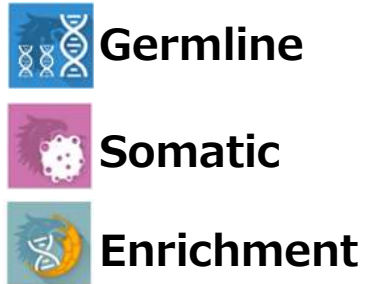
You can view sample data sets and run outputs here: <https://basespace.illumina.com/s/vpeTqSWN8LrN>

解析サンプルと結果の例

illumina®

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

解析パラメータの設定 - 1/4 : ① 解析名の入力、② 保存先の選択

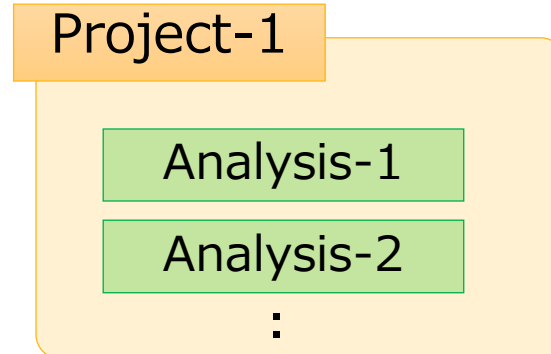


① 任意の解析名を入力

Analysis Name ⓘ Analysis-1

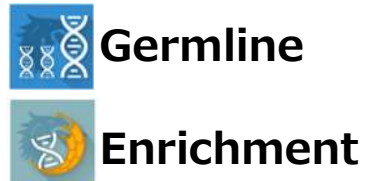
② 結果の保存先プロジェクト* を選択（作成）

Save Results To ⓘ
SELECT PROJECT Project-1



*BSSHでは解析 (Analysis) 結果を
プロジェクト (Project) 単位で管理

解析パラメータの設定 - 2/4 : ③ サンプルの選択



③ 解析サンプルを選択

Input FASTQs ⓘ
SELECT BIOSAMPLE(S):

事前にアップロードされたファイルを指定

Input BAMs ⓘ Input CRAMs ⓘ
SELECT DATASET FILE(S) **SELECT DATASET FILE(S)**

他の解析のアラインメント結果を入力することも可能



Tumor サンプルと Normal サンプル (オプション) を区別して選択

Tumor Biosample ⓘ
SELECT BIOSAMPLE(S):

- ・ 腫瘍 (Tumor) サンプルのみ選択
= **Tumor-only** 解析

Normal Biosample (optional) ⓘ
SELECT BIOSAMPLE(S):

- ・ 同一個人の正常/非腫瘍部 (Normal) サンプルを追加選択することで、Germlineバリエーションや実験的ノイズの除去に利用可能
= **Tumor-Normal (T/N)** 解析

より精度の高い結果が期待

解析パラメータの設定 - 3/4 : ④ リファレンス & 解析領域の選択

-  Germline
-  Somatic
-  Enrichment

④ リファレンスゲノムの選択

Reference 
Human (UCSC hg38 Alt-Masked v3, Graph Enabled) 

Custom DRAGEN Reference File 



DRAGEN Reference Builder
Edico Genome Inc.

ヒト以外などカスタムリファレンスは
あらかじめ別途作成が必要*

- Human (UCSC hg38 Alt-Masked v3, Graph Enabled) 
- Human (UCSC hg19 Alt-Masked v3)
- Human (UCSC hg19 Alt-Masked v3, Graph Enabled)
- Human (UCSC hg38 Alt-Masked v3)
- Human (UCSC hg38 Alt-Masked v3, Graph Enabled)**
- Human (Ensembl GRCh37)
- Human (Ensembl hs37d5: GRCh37 + decoys)
- Human (Ensembl hs37d5: GRCh37 + decoys, Graph Enabled)
- Human (CHM13v2 Alt-Masked v3)
- Human (CHM13v2 Alt-Masked v3, Graph Enabled)
- Custom

*参考 (日本語ウェビナーPDF資料: 31-36ページ目)

-  Enrichment

+ 解析 (パネル) 領域の選択

Targeted Regions 

Custom BED (select BED file below)

Custom Target BED File 

カスタムの場合、リファレンスに対応した
領域指定 (BED) ファイルを用意

- Illumina Exome Panel v1.2 (CEX) 
- Illumina Exome Panel v1.2 (CEX)
- TruSight Cancer v1.0
- TruSight Cardio v1.0
- TruSight Inherited Disease v1.0
- TruSight One v1.1
- TruSight One Expanded v2.0
- TruSight Hereditary Cancer v2.0
- Twist Bioscience® for Illumina Exome 2.0 Plus Panel
- Twist Bioscience® for Illumina Exome 2.0 Plus with Mitochondrial Panel
- Custom BED (select BED file below)**

解析パラメータの設定 - 4/4 : オプション設定 (例) ~ ⑤ 解析実行

その他オプションはデフォルト、または設定変更して解析



Germline

Expansion Hunter

Expansion Hunter ⓘ

☐ Enable Expansion Hunter

リピート伸長変異の検出
(PCR-free WGS データ推奨)

デフォルトではオフ



Somatic

UMI Settings

UMI ⓘ

☐ Enable UMI

UMI を利用した PCR バイアスなどの
データ補正を実施

デフォルトではオフ



Enrichment

Small Variant Caller ⓘ

Switch between Germline or Somatic
small variant calling

☒ Germline

☐ Somatic

デフォルトでは
Germline
バリエーションコーラー

各設定の説明については
 ⓘ マーク または
 トップページの HELP 参照

LAUNCH APPLICATION



⑤ クリックで解析を実行

Germline 解析

デフォルト

スモールバリエーション (SNV/Indel) コール

主要オプション

コピー数バリエーション (CNV) コール
構造バリエーション (SV) コール
リピート伸長変異の検出
ターゲットコール
- 遺伝性疾患関連変異
- HLA typing
- Star allele など

Somatic 解析

デフォルト

スモールバリエーション (SNV/Indel) コール

主要オプション

コピー数バリエーション (CNV) コール
構造バリエーション (SV) コール
分子バーコード (UMI) 補正
バイオマーカー オプション
- 相同組換え修復欠損 (HRD)
- 腫瘍遺伝子変異量 (TMB)
- マイクロサテライト不安定性 (MSI)

SNV, single nucleotide variant; SV, structural variant; CNV, copy number variant; UMI, unique molecular identifier; HRD, homologous recombination deficiency; TMB, tumor mutation burden; MSI, microsatellite instability

解析結果の確認： SUMMARY – 基本情報、解析状況、コストなど



Germline



Somatic



Enrichment

Analysis: Analysis-1

Project [Project-1](#)

SUMMARY

REPORTS

INPUTS

FILES

General Info

Name Analysis-1
Application DRAGEN Enrichment | Version: 4.2.4
Date Started 2023-07-22 19:22
Date Completed 2023-07-22 19:46
Duration 23 minutes 53 seconds
Compute Charge 2.00 iCredits
Session Type Multi-Node
Node Count 1 Node Complete
Size 4.51 GB
Status Complete
Delivery None



Enrichment

エクソームパネル 1サンプル
(およそ10 Gbase) の解析例

2 iCredits = 278円

(1 iCredit = 139円* 換算)

*希望販売価格 (税抜) (2023年8月現在)

解析に要した時間および
コスト (iCredits)

解析完了前の状態

← Running ← Queued ← Initializing

解析完了はメールで通知

解析結果の確認：REPORTS - 基本的な数値情報

-  Germline
-  Somatic
-  Enrichment



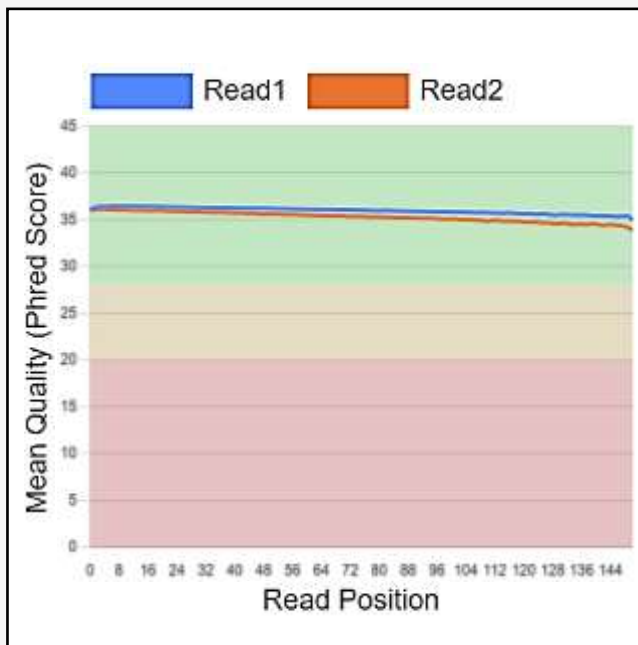
 dragen-reports.html
FILES からダウンロード可能



FastQC、インサートサイズ、カバレッジなどの
メトリクス情報の確認・評価 (QC) に利用

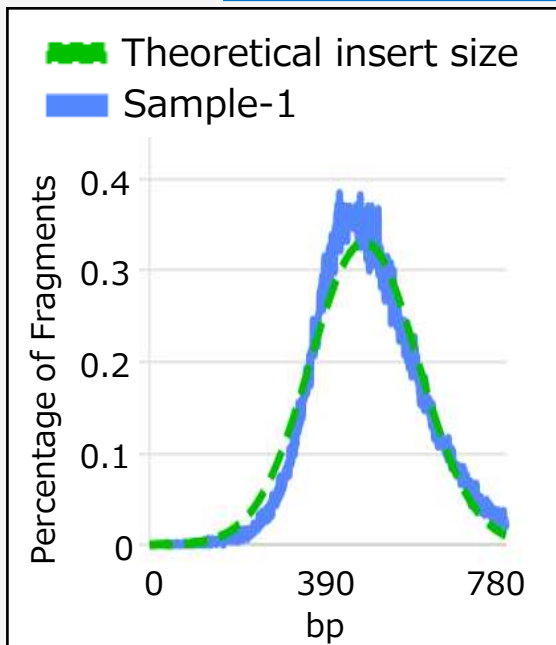
DRAGEN-FastQC

ベース
クオリティ



Mapping

予想された
インサートサイズ



Coverage

カバレッジ

Coverage Metrics

Metric	Value
Average Coverage	42.65
Uniformity of Coverage	93.88%
Mean / Median Ratio	0.96
% >=1x	94.93%
% >=10x	93.81%
% >=20x	93.13%
% >=50x	26.05%

アプリやオプション設定に
応じたレポートフォーマット

例：



Enrichment

Enrichment

特有のQC項目

Read Level Enrichment

Aligned Reads	% Aligned	On-Target Reads
100,519,920	78.36%	76,185,333
% Enrichment	Padded On-Target Reads	% Padded Enrichment
75.79%	81.97%	

リードの
濃縮効率

解析結果の確認：INPUTS - 解析パラメータの確認



Germline



Somatic



Enrichment

SUMMARY REPORTS **INPUTS** FILES

リファレンス

Reference

HUMAN (UCSC HG38 ALT-MASKED V3, GRAPH ENABLED)

サンプル

Select Biosamples

FASTQ

1

Input FASTQs

FASTQ (Library ...	Biosample	Project
NA12878-PCRF45...	NA12878-PCRF45...	NovaSeq 6000 SP...
NA12878-PCRF45...	NA12878-PCRF45...	NovaSeq 6000 SP...

オプション設定

☐ Enable HBA Genotyping

Rh Caller

☒ Enable Rh caller

LPA Caller

☐ Enable LPA caller

Advanced Settings

Duplicate Marking

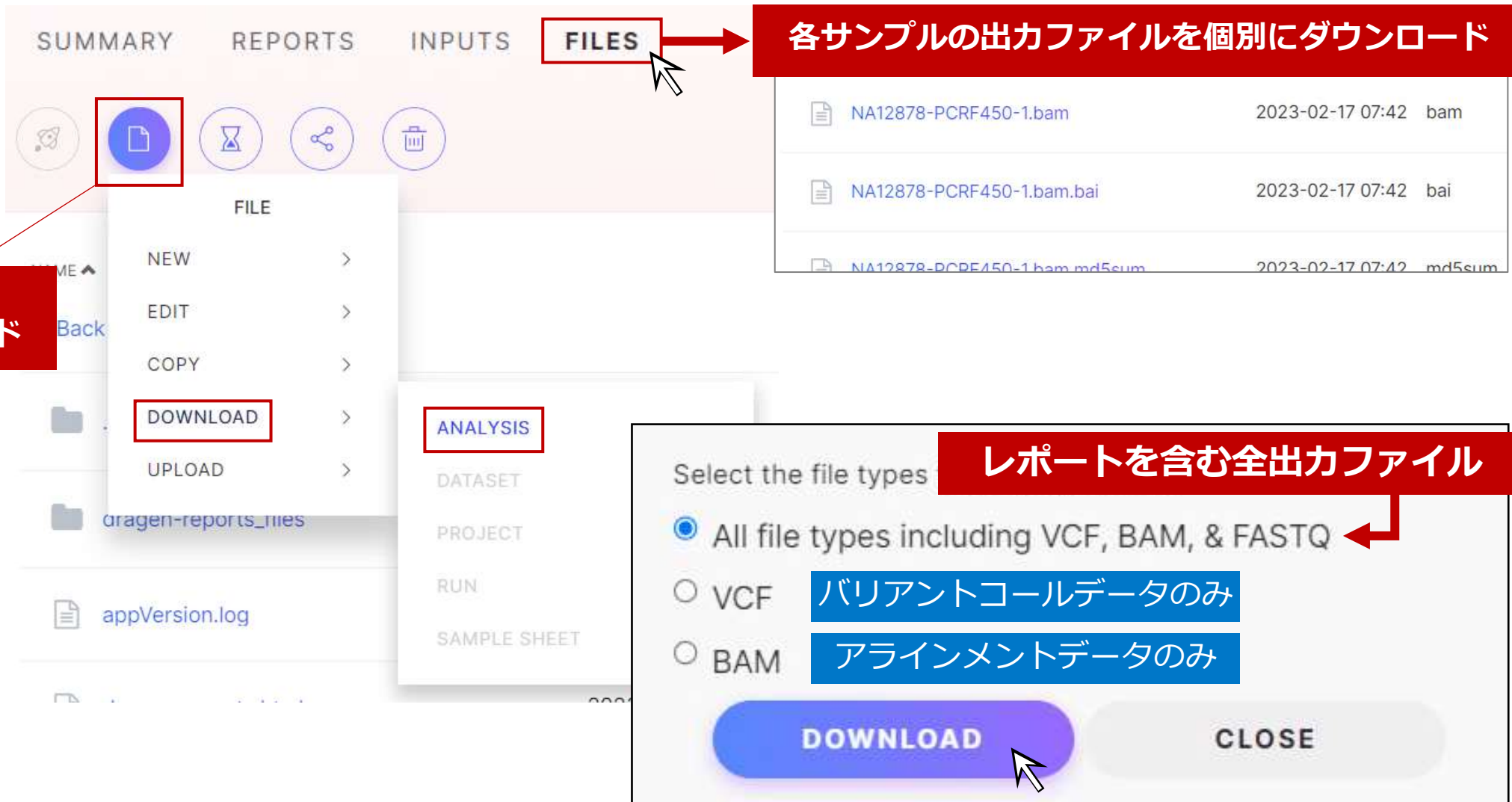
☒ Enable Duplicate Marking

解析に用いたサンプル、リファレンス、オプションなどのパラメータ設定が確認可能

解析結果の確認： FILES - 出力ファイルのダウンロード

-  Germline
-  Somatic
-  Enrichment

出力ファイルを
まとめてダウンロード



SUMMARY REPORTS INPUTS **FILES**

FILE

- NEW
- EDIT
- COPY
- DOWNLOAD**
- UPLOAD

ANALYSIS

DATASET

PROJECT

RUN

SAMPLE SHEET

appVersion.log

dragen-reports_files

NA12878-PCRF450-1.bam 2023-02-17 07:42 bam

NA12878-PCRF450-1.bam.bai 2023-02-17 07:42 bai

NA12878-PCRF450-1.bam.md5sum 2023-02-17 07:42 md5sum

Select the file types

- ☒ All file types including VCF, BAM, & FASTQ
- ☐ VCF
- ☐ BAM

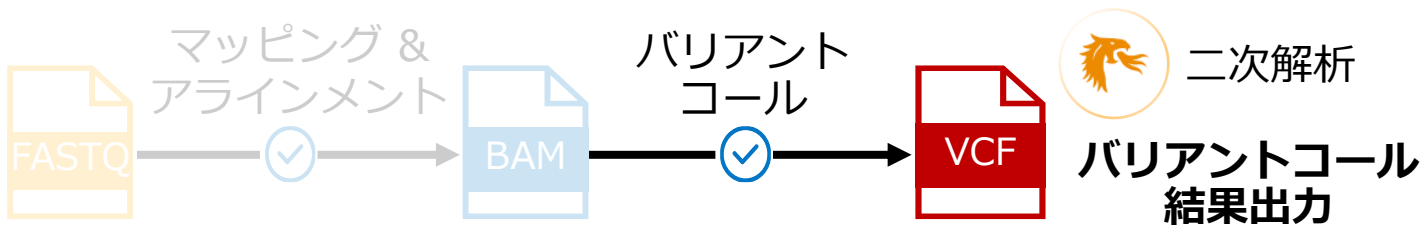
DOWNLOAD CLOSE

レポートを含む全出力ファイル

バリエーションコールデータのみ

アラインメントデータのみ

主要な出力ファイル: VCF - 1/2



● バリアントコール結果ファイル (VCF / gVCF)

バリアント位置のみ

バリアントのない位置も含む

バリアントの位置、配列変化、フィルタリング結果などの情報が記載されたテキスト形式のファイル

VCF ファイルの内容例*

条件を満たし通過した場合はPASS;
 不通過の場合は満たされない条件

	染色体	位置		配列変化		フィルタリング結果		サンプルごとの情報		
	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	Sample_NA12878
バリアント1	chr1	55926	.	T	C	61.32	PASS	AC=2;	GT:AD:AF:DP:F1R2:F	1/1:0,30:1.0000:30:0,9
バリアント2	chr1	111735	.	C	A	16.27	PASS	AC=1;	GT:AD:AF:DP:F1R2:F	0/1:20,41:0.6721:61:8
バリアント3	chr1	147048	.	G	A	5.88	LowDepth	AC=2;	GT:AD:AF:DP:F1R2:F	1/1:0,1:1.0000:1:0,1:0

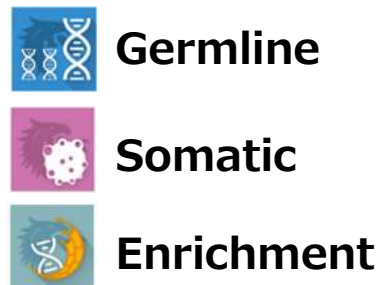
GT = 遺伝子型; 1/1 = ホモ;
 0/1 = ヘテロ; 0/0 = ホモ Ref (gVCF)

AD = アリル深度 (REF, ALT)

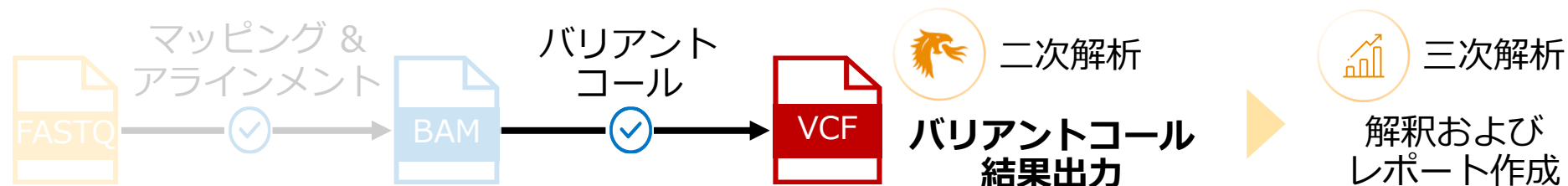
AF = バリアント アリル頻度

DP = リード深度

*Germline アプリ解析におけるスモールバリアントコールの結果ファイル (.hard-filtered.vcf.gz) を解凍して閲覧



主要な出力ファイル: VCF - 2/2



● バリアントコール結果ファイル (VCF / gVCF)

三次解析では

VCFのアノテーションによりバリアントの
機能的・病的な意義などの解釈を行う

VARIANT INTERPRETER (2023年8月現在 新規登録の受付停止中)
> <https://variantinterpreter.informatics.illumina.com/>
- BSSH クラウド上のアノテーションツール

illumina | Emedgene **NEW**
> <https://jp.illumina.com/products/by-type/informatics-products/emedgene.html>

人工知能 (AI) を Germline バリアントの解釈に利用

バリアントコール

DRAGEN などの二次解析による VCF 出力
(Emedgene 内部でも実行可能)

解釈

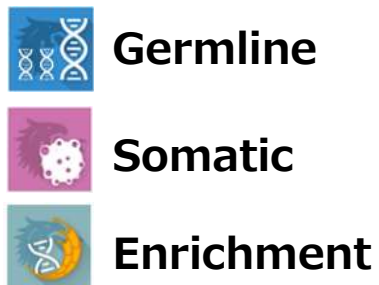
AI 主導型の自動化されたワークフロー

レポート

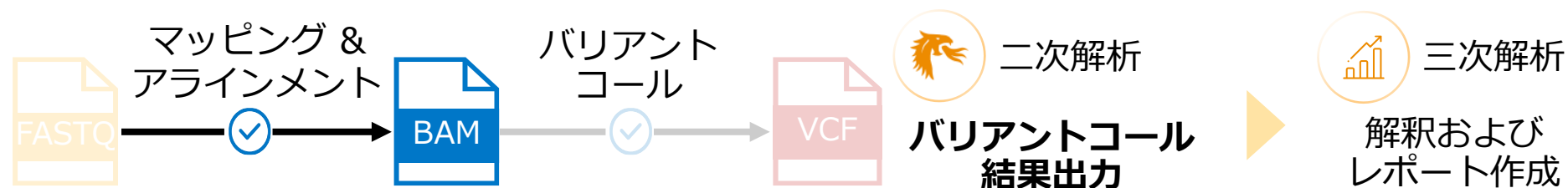
カスタマイズ可能なレポート生成



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



主要な出力ファイル：BAM



● アライメント結果ファイル (BAM / CRAM*)

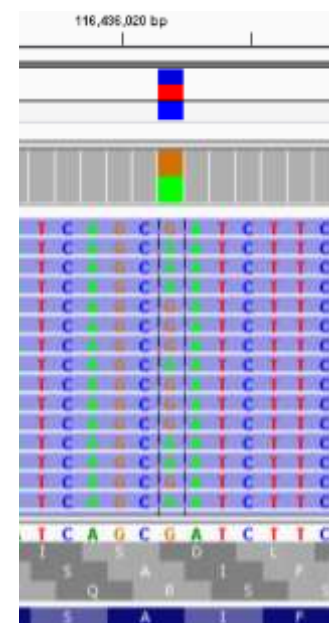
*高圧縮版

リファレンスにアラインメントされたリード情報を含むバイナリー形式のファイル
(多くの場合インデックスファイル bai / crai とセットで使用)

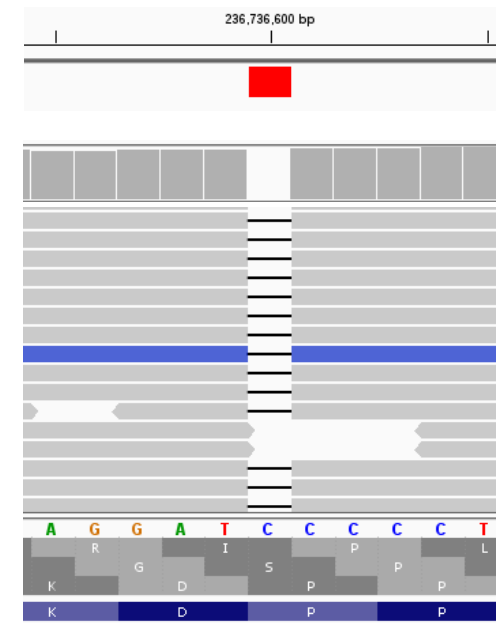
三次解析では

バリエントのコール領域に
アラインメントされた
リード配列の可視化に利用

主要な可視化ツール:
Integrative Genomics Viewer (IGV)



例：ヘテロ接合 SNV



ホモ接合 欠失

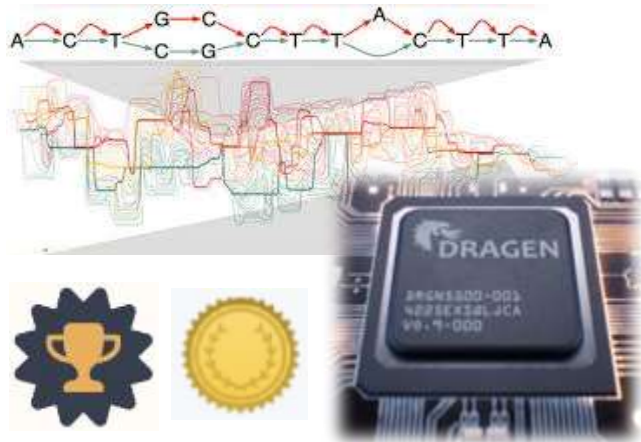
本ウェビナーのまとめ





本ウェビナーのまとめ:

DRAGENはNGS生データの解析を強力にサポート



- グラフリファレンスや機械学習に牽引された非常に高精度かつ高速なバリエーション解析性能
- 全ゲノム、エクソーム、トランスクリプトーム、メチローム、シングルセルなどの解析パイプラインをローカル & クラウド のマルチプラットフォームで提供



全ゲノム解析



Germline



Somatic

エクソーム / エンリッチメントパネル解析



Enrichment

- BSSH クラウド版 DRAGEN では簡単・高速・高精度な解析が可能であり、NGS データ解析を始める方や、DRAGEN の性能を試してみたい方にもおすすめ

Appendix

いずれも2023年8月時点の情報

BSSH DRAGEN アプリ解析における消費 iCredit 目安 (1サンプル当たり)

アプリケーション	コスト	詳細
ヒト全ゲノムシーケンス	7 iCredit	サンプル：30X hWGS サイズ：120Gb アプリ：DRAGEN Germline 入力：FASTQ 出力：VCF、BAM/CRAM
エクソームシーケンス	2 iCredit	サンプル：エクソーム サイズ：10Gb アプリ：DRAGENエンリッチメント 入力：FASTQ 出力：VCF、BAM/CRAM
トランスクリプトーム	2 iCredit	サンプル：RNA-Seq サイズ：50Mリード アプリ：DRAGEN RNA 入力：FASTQ 出力：融合遺伝子、トランスクリプト

[BaseSpace Sequence Hub 製品ページ参照](#)

BSSH アプリにおける解析パラメータの設定: 2種類の解析サンプルの選択方法

Input Data

Please select your preferred method for input data to launch the analysis ⓘ

- Biosamples - Launch with FASTQs across multiple sequencing runs

BaseSpace Sequence Hub上のサンプル管理単位

> 複数ランにまたがるSampleを1つの「**BioSample**」に統合可能
(単一ランのみでもOK)

Biosample 1

をサンプル単位として選択する方法

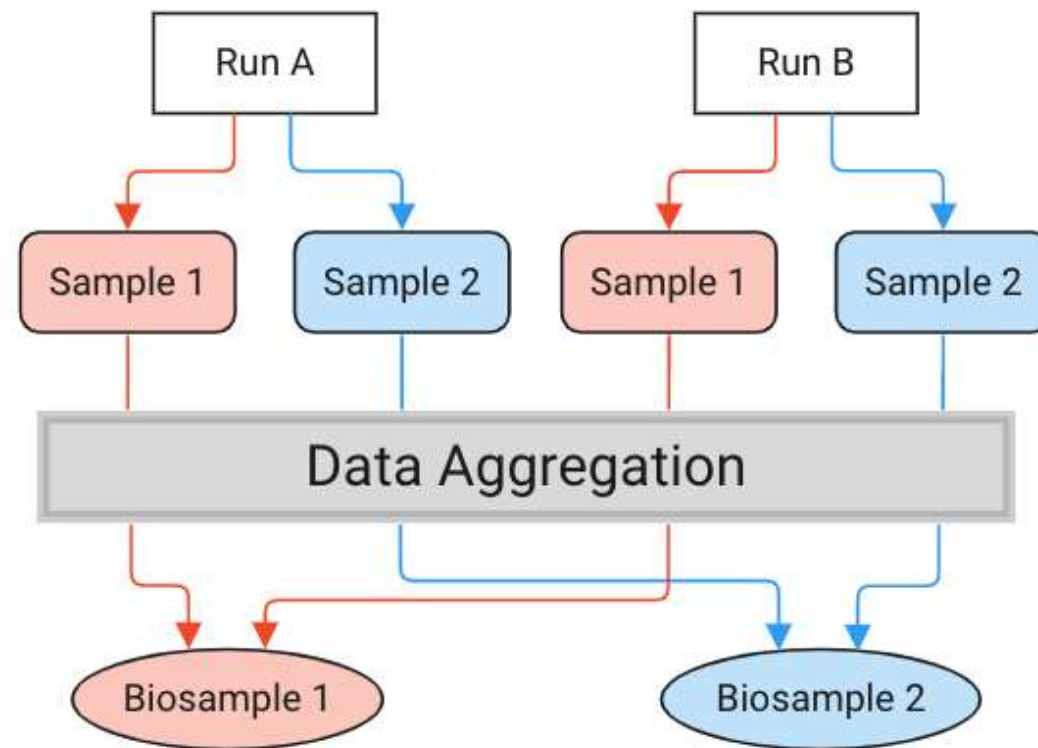
- Samples - Launch with FASTQs from a single sequencing run

Sample Sheet上のサンプル管理単位

> 単一ランに由来する「**Sample**」(ラン毎に区別)

Sample 1

をサンプル単位として選択する方法



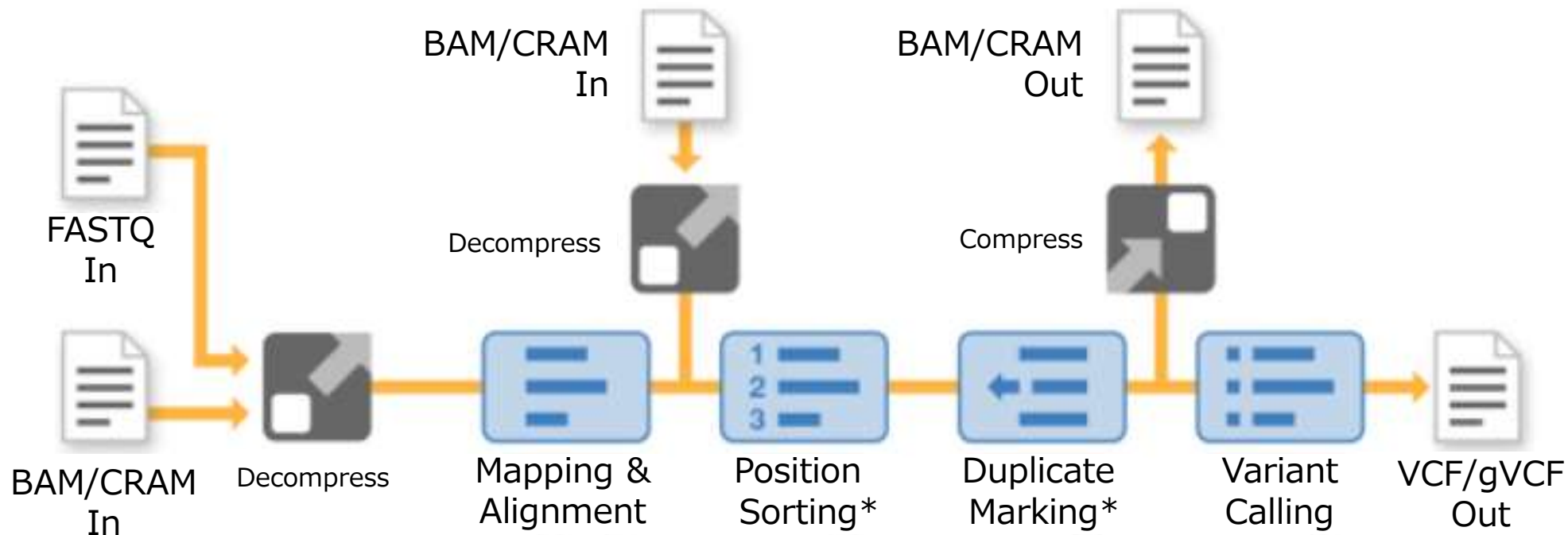
BaseSpace Sequence Hub上での
BioSample と Sample の取り扱い概念図

BioSample と Sample についての詳細は下記参照 (日本語ウェビナー)

[BaseSpace Sequence Hub \(BSSH\) の基本: NGSデータのアップロードから解析まで](#)



DRAGEN Germline の解析フロー詳細



*オプション

DRAGEN Germline は FASTQ、BAM、または CRAM ファイルを入力とする。入力された圧縮ファイルは解凍され、Map/Align/Sort を経て、バリエントコールが行われる。CNV コール、SV コール、Expansion Hunter はオプションで有効化できる。SMA、CYP2D6、CYP2B6、HLA、GBA のようなターゲットとなる追加コーラーも有効化できる。バリエントコール後にVCF、gVCF、BAM/CRAM ファイルが出力される。



DRAGEN Germline v4.2.4 の主な出力ファイル (一部)

DRAGENレポート

dragen-reports.html

DRAGENレポートファイル

アラインメント

*.bam / .bai

アラインメント結果出力 / インデックスファイル

フィルタリング済みバリエーションコール

*.hard-filtered.vcf.gz / .tbi

*.hard-filtered.gvcf.gz / .tbi

[> Germline Variant Small Hard Filtering: フィルタリング条件 \(Germline\)](#)

フィルタリング情報が付与されたVCFバリエーションコール出力 / インデックスファイル

フィルタリング情報が付与されたgVCFバリエーションコール出力 / インデックスファイル

QC Metrics

*.mapping_metrics.csv

*.vc_metrics.csv

*.vc_hethom_ratio_metrics.csv

[> QC Metrics and Coverage/Callability Reports: メトリクス情報のフォーマット](#)

Mapping Metrics

Variant Calling Metrics

Per Contig Het/Hom Ratio

Coverage Reports

*_coverage_metrics.csv

*_fine_hist.csv

*_hist.csv

*_overall_mean_cov.csv

*_contig_mean_cov.csv

*.ploidy_estimation_metrics.csv

[> Coverage/Callability Reports Over Custom Regions: カバレッジレポートのフォーマット](#)

Coverage Metrics Report

Fine Histogram Coverage Report

Histogram Coverage Report

Overall Mean Coverage Report

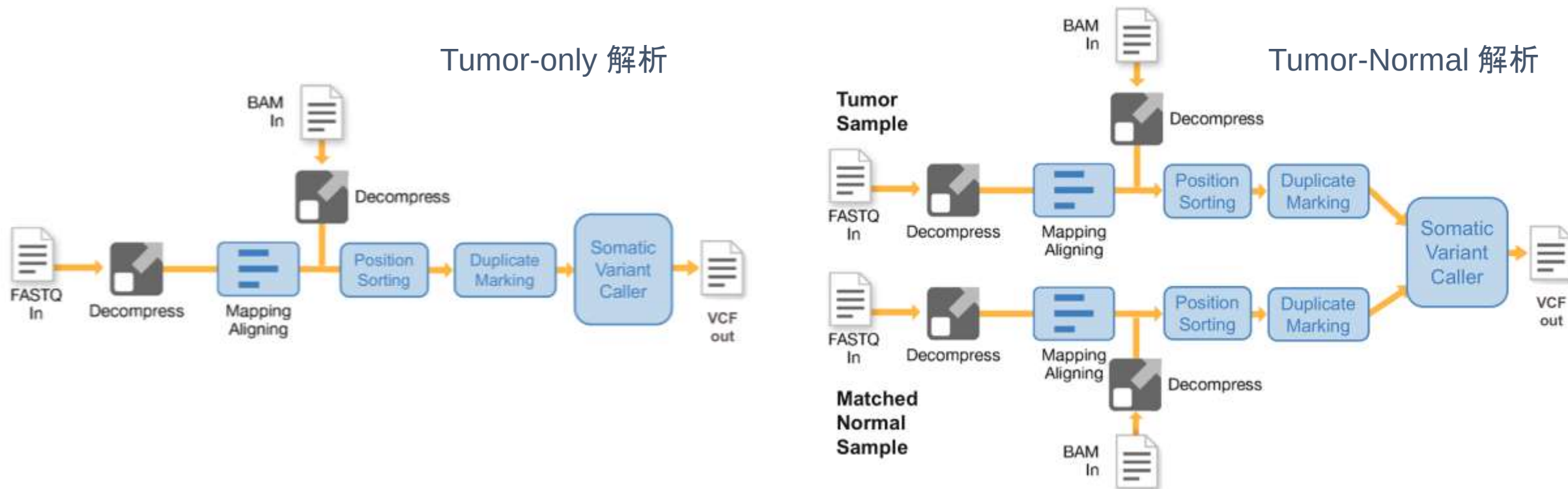
Per Contig Mean Coverage Report

Predicted Ploidy Report

[Illumina DRAGEN Bio-IT Platform Documentation: Online Help 参照](#)



DRAGEN Somatic の解析フロー詳細



DRAGEN Somatic は、腫瘍 (Tumor) サンプルに低アレル頻度で存在する可能性のある体細胞変異を同定する。Tumor と Normal のペア (Tumor-Normal) や Tumor のみ (Tumor-only) のシーケンスデータを解析できる。Normal サンプルが存在する場合は、生殖細胞系列 (germline) バリエーションやシステムティックなシーケンスアーティファクトの存在するポジションでのバリエーションコールを回避するために使用される。Germline 解析とは異なり、Somatic 解析フローでは Tumor サンプルの ploidy を仮定しないため、低頻度のアレルを高感度に検出できる。



DRAGEN Somatic v4.2.4 の主な出力ファイル (一部)

DRAGENレポート

dragen-reports.html

DRAGENレポートファイル

アラインメント

*.bam / .bai

アラインメント結果出力 / インデックスファイル

フィルタリング済みバリエーションコール

*.hard-filtered.vcf.gz / .tbi

[> Post Somatic Calling Filtering: フィルタリング条件 \(Somatic\)](#)

フィルタリング情報が付与されたVCFバリエーションコール出力 / インデックスファイル

QC Metrics

*.mapping_metrics.csv

[> QC Metrics and Coverage/Callability Reports: メトリクス情報のフォーマット](#)

*.vc_metrics.csv

Mapping Metrics

Variant Calling Metrics

Coverage Reports

*_coverage_metrics.csv

[> Coverage/Callability Reports Over Custom Regions: カバレッジレポートのフォーマット](#)

*_fine_hist.csv

[> Somatic Metrics Report: Tumor-Normal 解析における出力レポート情報](#)

*_hist.csv

Coverage Metrics Report (_tumor / _normal)

*_overall_mean_cov.csv

Fine Histogram Coverage Report (_tumor / _normal)

*_contig_mean_cov.csv

Histogram Coverage Report (_tumor / _normal)

*.ploidy_estimation_metrics.csv

Overall Mean Coverage Report (_tumor / _normal)

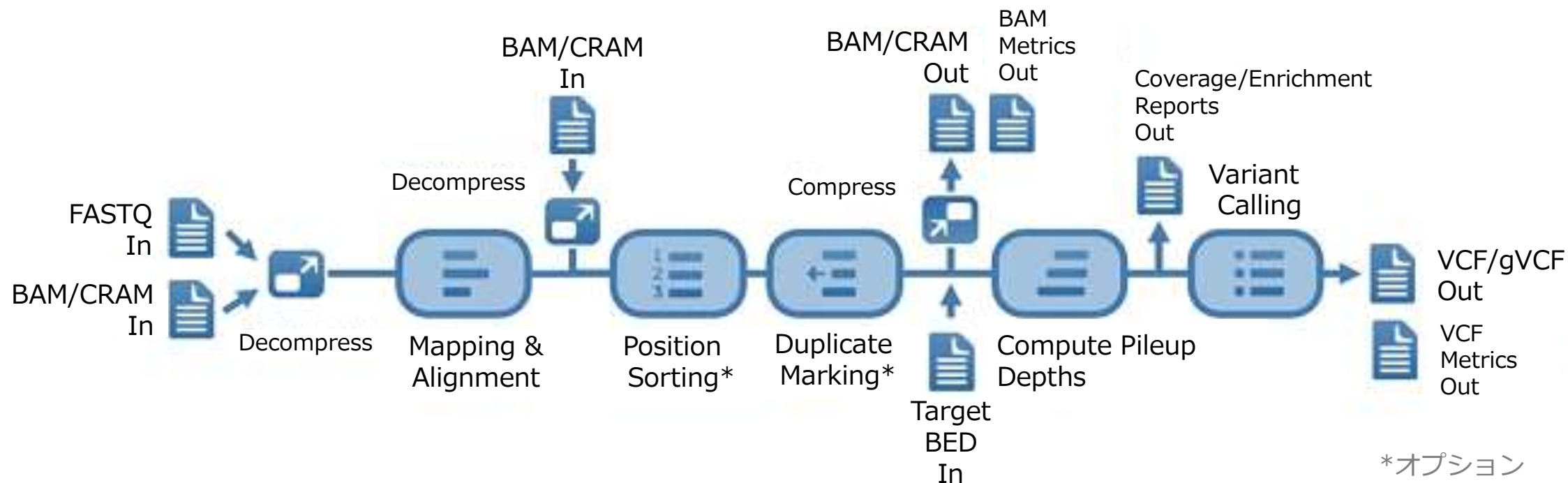
Per Contig Mean Coverage Report (_tumor / _normal)

Predicted Ploidy Report (Tumor-only 解析の場合はTumor サンプル、Tumor-Normal 解析の場合は Normal サンプルについて出力)

[Illumina DRAGEN Bio-IT Platform Documentation: Online Help 参照](#)



DRAGEN Enrichment の解析フロー詳細



DRAGEN Enrichment は Germline または Somatic モードで解析可能で、それぞれ生殖細胞系列のスモールバリエーションまたは体細胞系列 (低頻度) バリエーションのコールを実行する。CNV コールとSV コールはオプションで有効化できる。メトリクス情報は、入力された解析領域指定 (BED) ファイルに基づく。



DRAGENレポート

dragen-reports.html

DRAGENレポートファイル

アラインメント

*.bam / .bai

アラインメント結果出力 / インデックスファイル

フィルタリング済みバリエーションコール

*.hard-filtered.vcf.gz / .tbi
*.hard-filtered.gvcf.gz / .tbi

[> Germline Variant Small Hard Filtering: フィルタリング条件 \(Germline\)](#)
[> Post Somatic Calling Filtering: フィルタリング条件 \(Somatic\)](#)

フィルタリング情報が付与されたVCFバリエーションコール出力 / インデックスファイル
フィルタリング情報が付与されたgVCFバリエーションコール出力 / インデックスファイル
(Germline バリエーションコーラー選択の場合にデフォルトで出力)

QC Metrics

*.mapping_metrics.csv
*.vc_metrics.csv
*.vc_hethom_ratio_metrics.csv

[> QC Metrics and Coverage/Callability Reports: メトリクス情報のフォーマット](#)

Mapping Metrics
Variant Calling Metrics
Per Contig Het/Hom Ratio

Coverage Reports (BED領域について)

*_coverage_metrics.csv
*_fine_hist.csv
*_hist.csv
*_overall_mean_cov.csv
*_contig_mean_cov.csv
*_ploidy_estimation_metrics.csv

[> Coverage/Callability Reports Over Custom Regions: カバレッジレポートのフォーマット](#)

Coverage Metrics Report
Fine Histogram Coverage Report
Histogram Coverage Report
Overall Mean Coverage Report
Per Contig Mean Coverage Report
Predicted Ploidy Report

[Illumina DRAGEN Bio-IT Platform Documentation: Online Help 参照](#)

ウェビナー本編は以上です。
ご清聴ありがとうございました。
