

イルミナでAmpliSeqパネル解析をはじめよう 【情報解析編】初めての方でも大丈夫、クラウドを用いた簡単クリック情報解析

寺倉 伸治

シニアフィールドアプリケーションサイエンティスト

May/30/2018



プレゼンテーションを始める前に

- 本資料で紹介しているイルミナ試薬の価格は、2018年4月改定版の試薬消耗品価格表を参照しています。
- 本プレゼンテーションはのちほど録画とスライドをイルミナHPで公表します。
- 本スライドで紹介する試薬は、研究用試薬となります。

イルミナウェビナー「イルミナでAmpliSeqパネル解析をはじめよう」シリーズのご案内



AmpliSeqの製品紹介
ライブラリー調製

4月25日

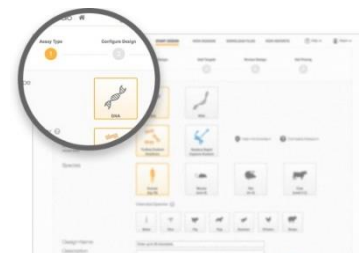
- AmpliSeq製品群の紹介の他、ライブラリー調製を成功させるためのワークフローと実験時のTipsをお伝えします。



AmpliSeqの
情報解析

5月30日

- AmpliSeqの情報解析の流れから出力結果まで、がんパネルを例に紹介します。



AmpliSeqの
カスタムデザイン

6月末予定

- AmpliSeqのカスタムデザインの手順を、実際に遺伝子やホットスポットなどを例に紹介します。

イルミナの過去ウェビナーは以下で動画とスライドを公開しています
<https://jp.illumina.com/events/webinar.html>

AmpliSeq for Illumina ワークフロー

4月のウェビナーで説明済み



1日目

ライブラリー調製

5 ~ 7時間

- ハンズオン1.5時間
- 10ステップ

2日目

シーケンス

約17 ~ 32時間*

- イルミナデスクトップシーケンサーに対応
- 試薬充填済みカートリッジ
- タッチパネル式操作
- 固定パネルでは 2x151 bp

*シーケンサーごとに異なる

3日目

解析

- ローカルおよびクラウドベースの解析
- ✓ Local Run Manager
- ✓ BaseSpace Sequence Hub
- ✓ BaseSpace Variant Interpreter

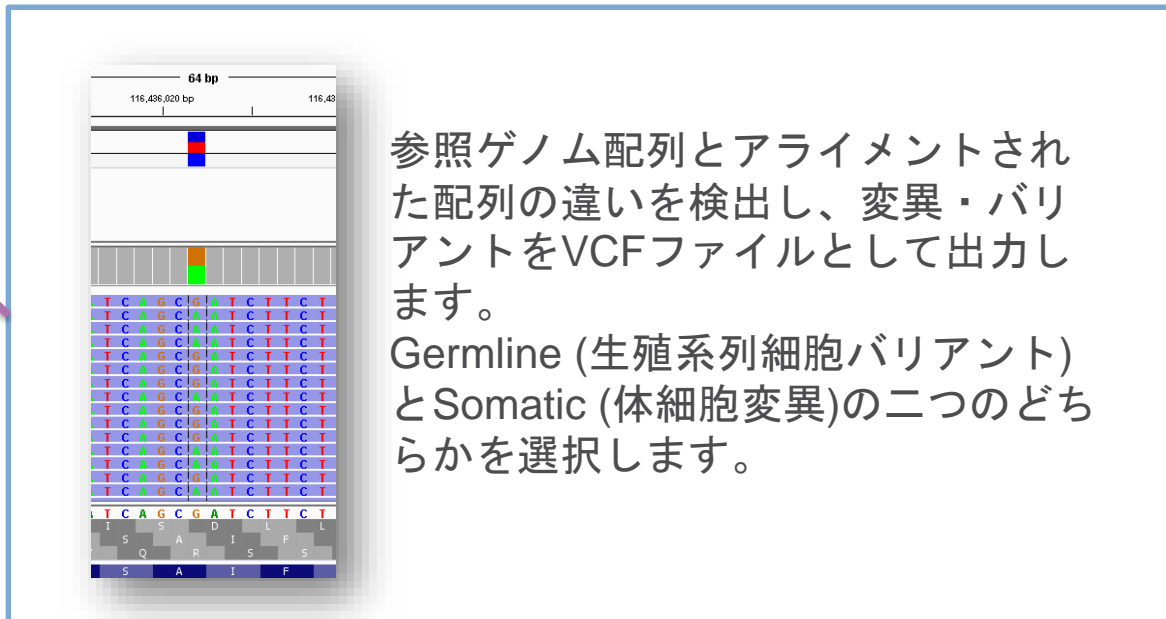
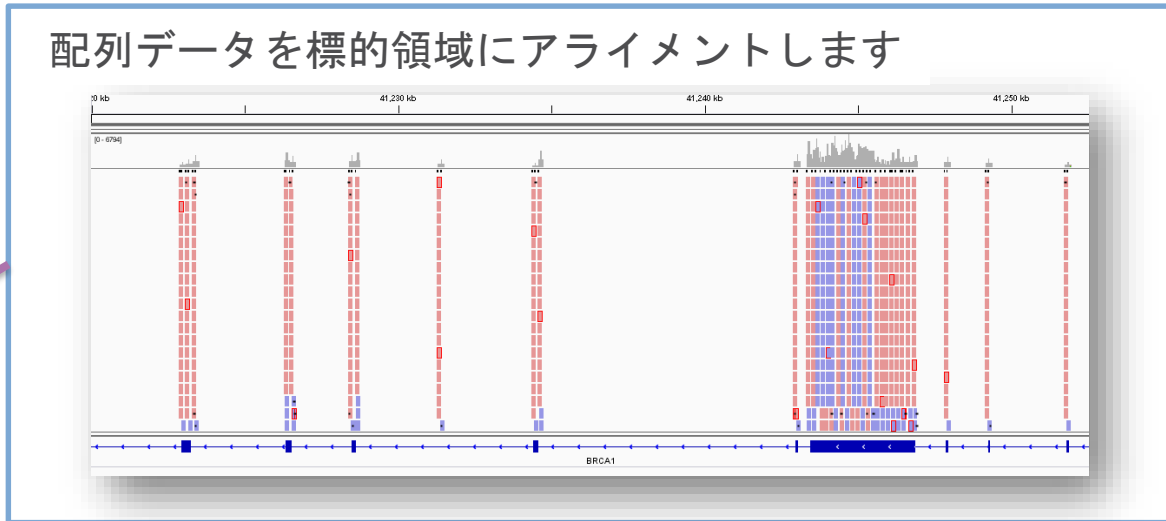
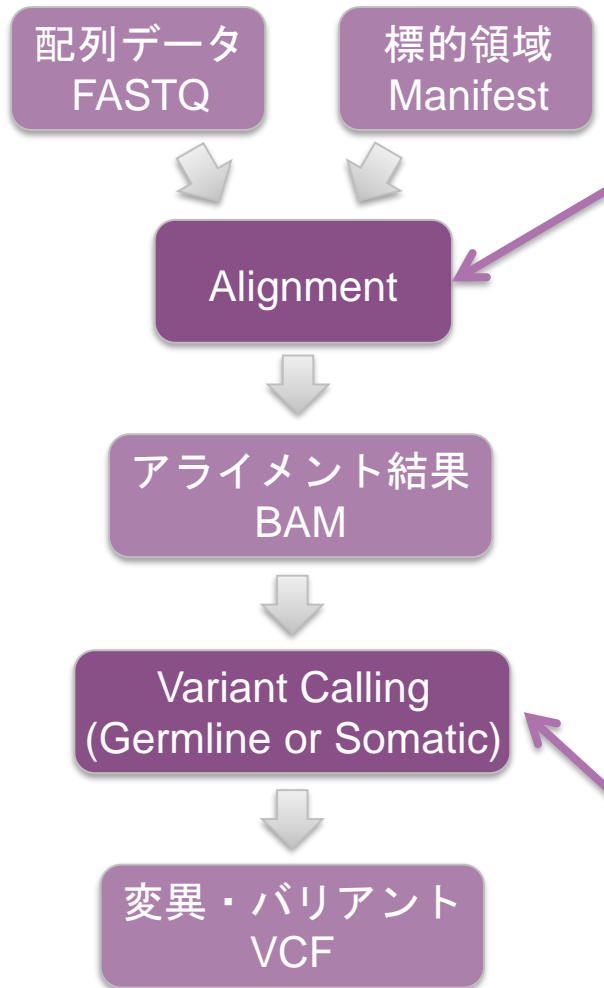
今回のセッションの流れ

- 情報解析の実行
- 変異解析に必要なカバレッジ
- 情報解析結果の確認
- 変異解析結果へのアノテーションとフィルタリング

情報解析の実行 Local Run Managerと BaseSpace App



AmpliSeq DNA Ampliconの解析ワークフロー



AmpliSeqの情報解析の実行にはLocal Run ManagerとBaseSpaceの二つのオプション

Local Run Manager



シーケンサー装置のPC
もしくは、お客様のWindows PC

Pros

- ✓ シーケンサー装置にすでにインストール済み
- ✓ シーケンス終了後そのまま情報解析を実行
- ✓ お客様のWindows PCにもインストール可能 (Local Run Manager Off-Instrumentと呼ぶ)
- ✓ インターネット接続は不要

Cons

- ✓ 一部の複雑な情報解析が実行不可

BaseSpace Sequence Hub



クラウド環境

- ✓ シーケンサーからDataをアップロード可能
- ✓ ブラウザを通してどこでも結果閲覧が可能
- ✓ データシェアが容易
- ✓ CNVといった複雑な解析にも対応
- ✓ PCやソフトウェアのメンテナンスを行う必要がない

- ✓ 情報解析の実行は有償
- ✓ クラウド環境へのデータアップロードが必要

MiSeq・MiniSeq・NextSeqについては、Local Run ManagerはOff Instrumentのみしか現状は対応していません (2018年5月)。年内にリリース予定です。

Local Run Manager Off Instrumentについて

- Local Run Manager Off Instrumentはお客様準備のPCに解析ソフトウェアをインストールし、Localで情報解析を行います。
- 必要なPCのスペック
 - Computing Requirements
 - ≥ 16 GB RAM (メモリ不足だと解析がAbortしてしまいます)
 - ≥ 1 TB of disk space (**少なくとも100 GB以上は必要**)
 - 64-bit quad core processor (2.8 GHz or higher) with the AVX instruction set (recommended) or Intel Core i7-210QE 2.10 GHz (or equivalent) processor with the AVX instruction set (minimum).
 - Display resolution of 1024 x 768 or higher
 - Software Requirements
 - Windows 7 Pro 64-bit with Service Pack 1 (English-US) or Windows 10 (Personal Edition versions are not supported *Windowsのバージョンを確認ください。
 - Microsoft .NET 4.5.2 or higher
 - Google Chrome 46.0.2490 (最新のものであればOK!)
 - 以下のプログラムがあるとインストールできない場合があります。
 - PostgreSQL 9.3、Rabbitmq 3.5.1、Erlang/OTP 17、C++ 2012 64-bit Redistributable、C++ 2013 64-bit Redistributable、C++ 2015 64-bit Redistributable、Illumina Run copy service、Internet Information Services、Local Run Manager

AmpliSeq for illuminaの情報解析ツール

AmpliSeq DNA

AmpliSeq RNA

解析目的

SNP
Short INDEL
(塩基置換
短い挿入・欠損)

CNV
(コピー数異常)

Fusion Gene
(遺伝子融合)

Gene Expression
(遺伝子発現)

Local Run
Manager

DNA Amplicon
Analysis Module

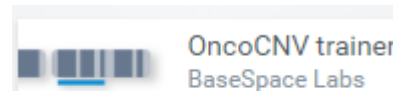
対応していない

RNA Amplicon Analysis Module

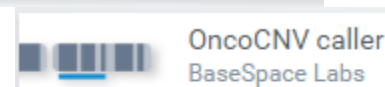
BaseSpace



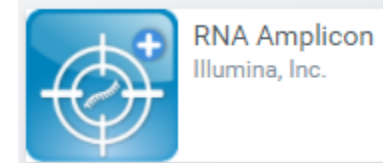
DNA Amplicon
Illumina, Inc.



OncoCNV trainer
BaseSpace Labs



OncoCNV caller
BaseSpace Labs



RNA Amplicon
Illumina, Inc.

情報解析に必要なもの

【Local Run Managerの場合】

- Run Folderそのもの
- Manifest File (標的を示す位置情報ファイル)
 - 固定パネルの場合は、イルミナテクニカルサポート(techsupport@illumina.com)にお問い合わせください、HPからは配布しておりません。
 - カスタムパネルの場合は、DesignStudioよりダウンロード可能です。

【BaseSpace Sequencing Hubの場合】

- FASTQファイル、もしくはシーケンサーをBaseSpaceにつなぎ、データをアップロードする。
- Manifest File (標的を示す位置情報ファイル)
 - 固定パネルの場合は、選択肢より選択することが可能です。
 - カスタムパネルの場合は、DesignStudioよりダウンロードし、BaseSpaceにアップロードする必要があります。

カスタムパネルのManifest Fileは DesignStudioよりダウンロードが可能です

BaseSpace DESIGNSTUDIO SEQUENCING

START DESIGN VIEW DESIGNS VIEW REPORTS ? Help Shinji

View Designs Project Id: [] Customer Email: [] Clear Q Search: []

Legend: Dual Pool

AmpliSeq DNA	Nextera Rapid Capture	TruSeq Custom Amplicon	TruSeq Targeted RNA							
Type	Design ID	Solution ID	Design Name	Amplicon Size	Amplicon Count	Coverage		Added	Actions	
Gene	[]	[]	[]	175	1277 Amplicons	99%	[]	04/30/2018	Select	
Gene	[]	[]	[]	175	1190 Amplicons	95%	[]	04/30/2018	Select	
Gene	[]	[]	[]	175	1255 Amplicons	96%	[]	04/28/2018	Select	
Gene	[]	[]	[]	175	1161 Amplicons	92%	[]	04/27/2018	Select	
Gene	[]	[]	[]	275	16 Amplicons	99%	[]	04/24/2018	Select	
Gene	[]	[]	[]	275	15 Amplicons	94%	[]	04/24/2018	Select	
Gene	[]	[]	[]	275	15 Amplicons	94%	[]	04/18/2018	Select	
Gene	[]	[]	[]	275	524 Amplicons	99%	[]	03/28/2018	Select	
Gene	[]	[]	[]	275	414 Amplicons	98%	Orderable	grch38.p2	03/27/2018	Select
Gene	[]	[]	[]	275	632 Amplicons	98%	Orderable	grch38.p2	03/27/2018	Select

Select

- Copy
- Transfer
- Delete
- Download
- Manifest

BaseSpace DNA Ampliconアプリ



DNA Amplicon
Illumina, Inc.

- **Reference Genome**は以下を使用することが可能です。
 - Human, UCSC hg19
 - Human, NCBI GRCh38Decoy
 - Chicken, Ensembl Galgal4
 - Cow, Ensembl UMD3.1
 - Dog, Ensembl CanFam3.1
 - Maize, Ensembl AGPv3
 - Mouse, UCSC mm9
 - Pig, Ensembl Sscrofa10.2
 - Rat, UCSC rn4
 - Rice, Ensembl IRGSP-1.0
 - Sheep, Ensembl Oar_v3.1
 - Soybean, Ensembl Gm01

BaseSpaceでのDNA Ampliconの解析実行方法

使用するアプリケーションを選択

The screenshot shows the BaseSpace interface with the 'APPS' menu highlighted in red. Below the menu, a grid of application tiles is displayed. The 'DNA Amplicon' tile is also highlighted in red. A red callout box contains the following text:

BaseSpaceのAPPSより実行するアプリケーションを選択します。
AmpliSeq DNAであれば”DNA Amplicon”を、
AmpliSeq RNAであれば”RNA Amplicon”を、
選択してください。

BaseSpace DNA Amplicon パラメーター入力 (1)



DNA Amplicon v1.1.0
Illumina, Inc.

Configuration

Analysis Name ⓘ

DNA Amplicon HD200 & HD729

Analysis name: デフォルトでアプリケーションと日付が入りますが、解析目的を書いておくと後で見直す際に便利。

Save Results To ⓘ

Select Project

1805_AmpliSeq_Cancer_HotSpot-v2-Demo

Save Results to: 解析結果を保存するプロジェクトを選択します。

Biosample(s) ⓘ

Select Biosample(s):

HD200-10ng-E02A-B10: Unknown

HD200-10ng-E02A-B02: Unknown

HD729-10ng-E02A-C11: Unknown

HD729-10ng-E02A-D05: Unknown

Biosample(s): 解析を実行するサンプル(FASTQ配列)を選択します。一度に384検体まで実行することが可能です。

BaseSpace DNA Amplicon パラメーター入力 (2)

Custom Genome (Optional) ⓘ

Select Dataset File(s)

Custom Genome (Optional): BaseSpaceに登録のないゲノムを解析する場合に、事前にUploadしたGenome配列を選択。

Targeted Amplicons ⓘ

AmpliSeq for Illumina Cancer HotSpot Panel v2 (20171207)

Targeted Amplicons: リストから使用した製品を選択します。カスタムパネルは”Custom Panels”を選択してください。

Custom Manifest File ⓘ

Select Dataset File(s)

Custom Manifest File: カスタムパネルの場合は、事前にアップロードしたManifestを選択する。

Targeted Amplicons ⓘ

Custom Panels (select manifest file below)

Select Panel

AmpliSeq for Illumina Cancer HotSpot Panel v2 (20171207)
AmpliSeq for Illumina Comprehensive Cancer Panel (20171207)
AmpliSeq for Illumina Exome Panel (20171207)
TruSeq Amplicon - Cancer Panel (AFP1)
TruSeq Custom Amplicon Control Pool 1 (ACP1)
TruSeq Custom Amplicon Control Pool 3 (ACP3)
TruSight Myeloid Sequencing Panel
Custom Panels (select manifest file below)

Aligner ⓘ

- BWA
- TruSeq Amplicon Aligner

Aligner: AmpliSeqの場合はBWAを選択する。

Variant Caller ⓘ

- Germline
- Somatic

Variant Caller: 対象としている変異に対して選択。生殖細胞系列の変異は”Germline”にチェック、がんといった体細胞変異は”Somatic”を選択する。

Somatic Variant Frequency Threshold (Percentage) ⓘ

5

Somatic Variant Frequency Threshold (Percentage): 体細胞変異の場合、指定したパーセンテージ以下の割合の変異にはVCFファイルのFILTERにLowVariantFreqのフラグが立ちます。Default設定は5、1~30まで入力可能

Variant Caller Depth Filter ⓘ

10

Variant Call Depth Filter: 変異部位のリードのデプスが指定よりも低い場合にはCallしない。Defaultは10、10~10,000まで入力可能

BaseSpace DNA Amplicon パラメーター入力 (3)

Genotypes of Interest VCF (Optional) ⓘ

Select Dataset File(s)

Indel Realignment ⓘ

On
 Off

Annotation ⓘ

RefSeq
 Ensembl
 None

Launch Application

Indel Realignment: ONを選択すると、Hygeaというイルミナのツールで、INDEL部位のアライメントの補正を行う。INDELの検出が正確になることが期待できるので、ONを選択すると良い。

Annotation: ヒトの場合のみとなるが、アノテーションをRefSeq、Ensemble、もしくは実施しないを選択できる。

Launch Application: パラメーターの設定に問題がないことを確認したら、クリックを押すと解析がスタートします。

Variant Callerの主なコール条件

GermlineとSomatic Variant Callerは違いがあるので、気を付けましょう。
(頻度が低いがんの体細胞変異はGermline Variant Callerでは検出できない、
モザイクの場合にはSomatic Variant Callerを使用するなど)

● Germline Variant Caller

- 20%以下のバリエントは、Callしない。Variant Qualityが20未満は、Callしない。
- 20~70%の頻度のバリエントはヘテロ接合としてコールする。70%の頻度を越えたバリエントはホモ接合としてコールする。
- 設定したVariant Call Depth Filter以下のDepth(Defaultは10)の部位ではCallしない。

● Somatic Variant Caller

- Somatic Variant Frequency Filterで設定した数値未満の頻度の変異・バリエントは、VCFファイルのFILTERにLowVariantFreqのフラグがたちます。
- 設定したVariant Call Depth Filter以下のDepth(Defaultは10)の部位ではCallしない。



どちらもイルミナ開発のPiscesというVariant Callerを用いています。

<https://github.com/Illumina/Pisces/wiki>

illumina®

BaseSpace DNA Amplicon 情報解析コスト

General Info

Name	DNA Amplicon 96 samples analysis
Application	DNA Amplicon Version: 1.1.0
Date Started	Wed, May 2, 2018 3:44 PM
Date Completed	Wed, May 2, 2018 4:06 PM
Duration	22 minutes 36 seconds
Compute Charge	15.00 iCredits
Session Type	Multi-Node
Node Count	16 Nodes Complete
Size	2.12 GB
Status	Complete
Delivery	--

Cancer HotSpot v2パネルの96検体を解析すると、22 min.で解析は終了し、解析コストは15 iCredits*となりました。*1 iCredit = 120円 (2018年5月現在)となります。

解析コストは、解析を実施するシーケンス量などに依存して変わります。
たくさん解析すれば、その分だけ解析コストは上がります。

Local Run Managerも同様に情報解析の実行が可能です。

Create Run DNA AMPLICON

Import Sample Sheet

Run Name*
TST15-2

Run Settings

Library Prep Kit*

Index Reads*

Module-Specific Settings

Variant Caller* Somatic ?

Variant Caller Depth Filter* 10 ?

Variant Frequency* 0.05 ?

Aligner* BWA Whole-Genome ?

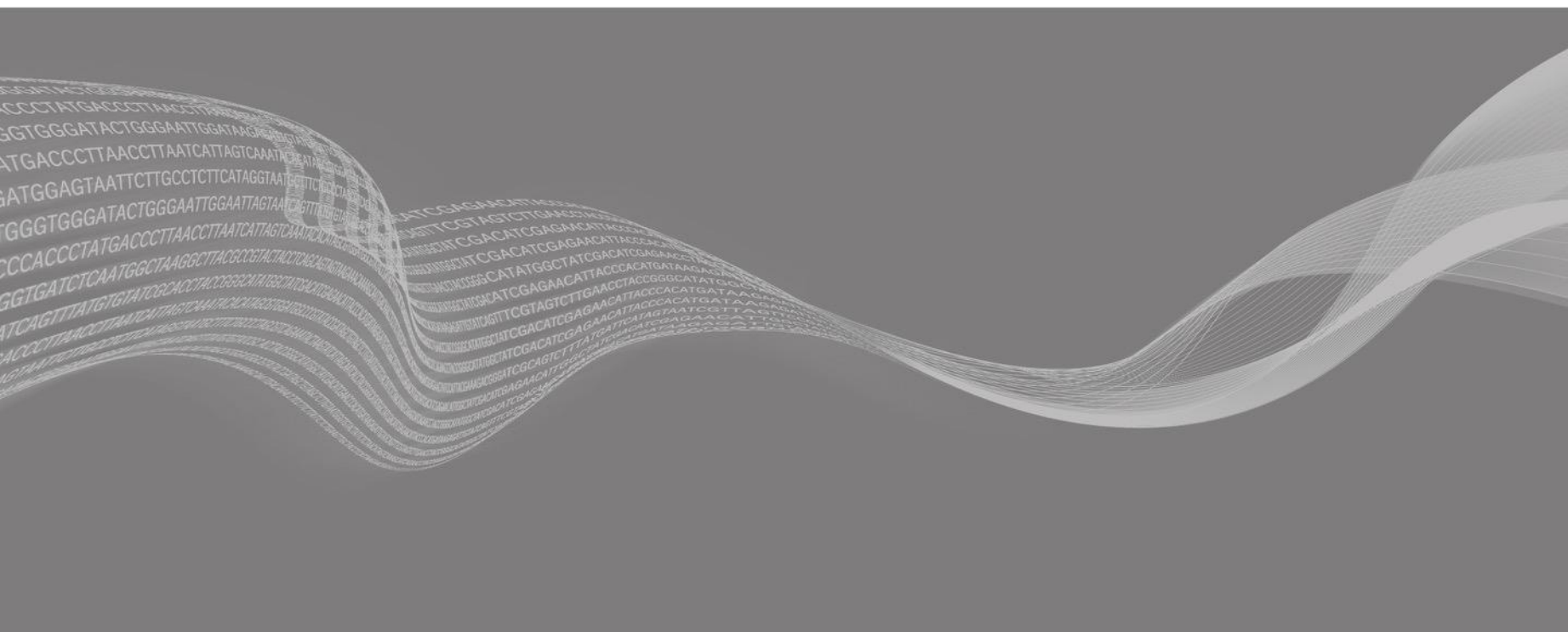
Genotype of interest VCF (Optional) + Add a Genotype VCF File

Import VCF ?

Annotation RefSeq Ensembl None ?

BaseSpaceと同様にクリック操作で情報解析を実行可能です。
Local Run Managerでは、Indel Realignmentのオプションが実行できません(2018年5月現在)。

変異解析に必要なカバレッジ



がんといった体細胞変異の検出に必要なカバレッジ

- 癌体細胞変異は、生殖細胞系の正常なレファレンスゲノムには存在しません。このため、実験ノイズおよび混在する正常細胞のため、新たな体細胞変異を正確に検出することは困難です。新たな体細胞変異を検出するためには、極めてカバレッジの高いシーケンスが必要です。

がんといった体細胞変異の検出に必要なカバレッジ

頻度5%の変異・バリエーションを仮定した場合の例、
カバレッジをあげれば上げるほど、P-Valueは小さくなる（より信頼性が上がる）。

表1：異なるカバレッジにおけるSNPのQスコア

	コールされた レファレンス塩基	変異塩基 コール	ミスコールの予想値 (1%エラー率*)	カバレッジ	P 値	Q スコア
変異無し	100	0	1	100	1	0
SNP頻度5%	95	5	1	100	0.004	24
SNP頻度5%	190	10	2	200	0.000046	43
SNP頻度5%	475	25	5	500	1.60×10^{-10}	98
SNP頻度5%	4750	250	50	5000	3.33×10^{-16}	155

P値 = SNPが偽陽性である確率

Qスコア = 高値であるほど、高い信頼性

Somatic Varinat Callerは、ポワソンモデルに基づいて、SNPに関するQスコアを算出します。例えば、3段目に示す、頻度5%のSNPに関しては、100xカバレッジにおいて、エラー率を保守的に1%と仮定した場合、Qスコアが24、P値が0.004となります（偽陽性である可能性が4/1000）。

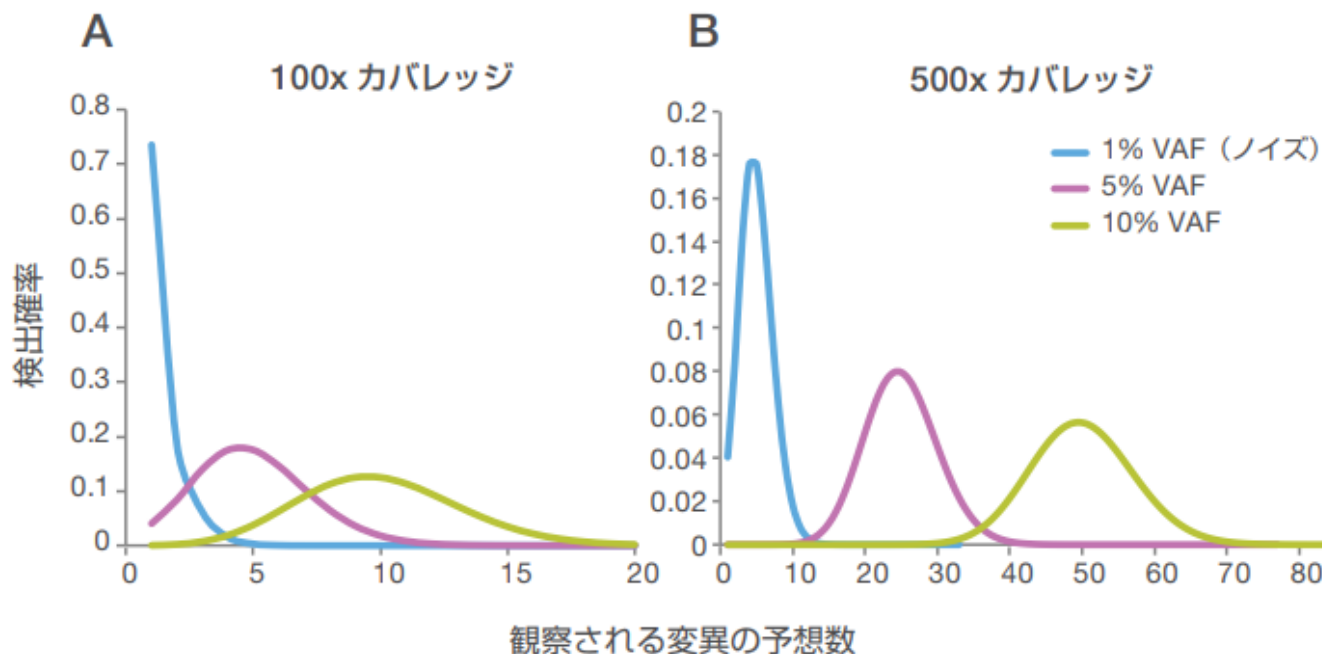
*より高い信頼性のコールを実現するため、Somatic Variant Callerでは、保守的なエラー率を採用しています。

Somatic Variant Caller, illumina Technote: Sequencing Software

[https://jp.illumina.com/content/dam/illumina-](https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/technote_somatic_variant_caller-j.pdf)

[marketing/apac/japan/documents/pdf/technote_somatic_variant_caller-j.pdf](https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/technote_somatic_variant_caller-j.pdf)

がんといった体細胞変異の検出に必要なカバレッジ



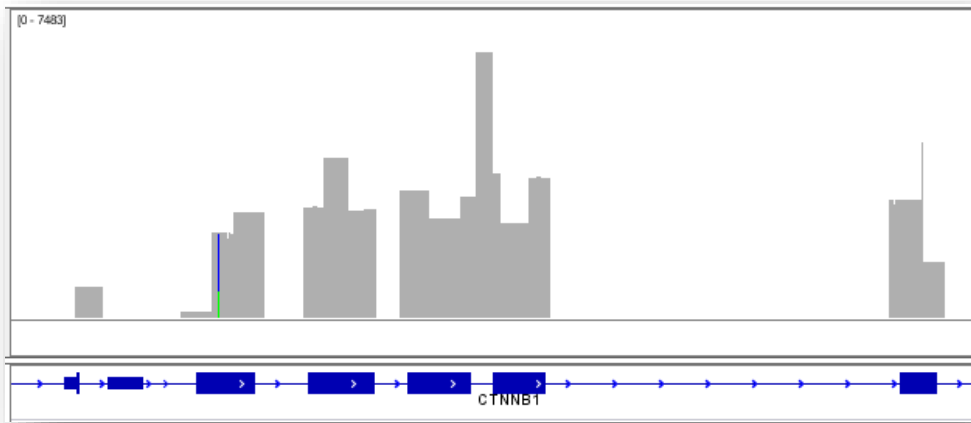
100×カバレッジ（A）では、5% VAFと10% VAFの間にはかなりのオーバーラップが存在するため、頻度が5%に満たない変異を高信頼性でコールすることは不可能です。これに対し、カバレッジを500×以上に増加させると（パネルB）、頻度が5%に満たない変異でも高い信頼性でコールすることが可能となります。

X500以上のカバレッジであれば、5%程度の変異を十分に検出することが可能です。
⇒ Liquid Biopsyを材料とした、1%以下といった非常に低い頻度の変異を検出するには、一般的に Unique Molecular Identifier (UMI)を用いたアプリケーションが要求されます。

Somatic Variant Caller, illumina Technote: Sequencing Software
https://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/technote_somatic_variant_caller-j.pdf

がん体細胞変異の検出に必要なカバレッジ

- 標的領域の広い範囲でX500以上のカバレッジを得るために、平均でX2,500カバレッジ程度までのシーケンスを行うことを推奨しています。



標的配列によって、リードのカバレッジにはバラつきが生じてしまうので、余裕を持ったシーケンスデザインが必要になる。

- たとえば、Amplicon数が1,000のカスタムパネルを構築した場合、検体当たり250万クラスター(1,000X2,500)までのシーケンスを行うようにする。
- カタログ製品(固定パネル)は、各シーケンサープラットフォームでシーケンスできる検体数のガイダンスがあります。

がん体細胞変異の検出を目的としたカスタムパネルの場合は、最初のシーケンスを厚めにシーケンスを行い、その結果を見て、目的に適したカバレッジを設定するとよい。

カタログ製品の各シーケンサープラットフォームでのランあたりの解析検体数 (Data Sheetより抜粋) (1)

表 2: AmpliSeq for Illumina *BRCAPanel*に推奨されるイルミナのシーケンサー

装置	ラン当たりのサンプル数	ラン時間
体細胞変異研究		
iSeq™ 100システム	12	17時間
MiniSeq™システム (中出力)	24	17時間
MiniSeqシステム (高出力)	80	24時間
MiSeqシステム (v2ケミストリー Nano)	3	17時間
MiSeqシステム (v2ケミストリー Micro)	12	19時間
MiSeqシステム (v2ケミストリー)	48	24時間
MiSeqシステム (v3ケミストリー)	80	32時間
生殖細胞系列変異研究		
iSeq™ 100システム	96	17時間
MiSeqシステム (v2ケミストリー Nano)	32	17時間
MiSeqシステム (v2ケミストリー Micro)	96	19時間
最大96サンプルをマルチプレックス可能		

表 3: AmpliSeq for Illumina Cancer Hotspot Panel v2に推奨されるイルミナシーケンサー

装置	ラン当たりのサンプル数	ラン時間
iSeq™ 100システム	16	17時間
MiniSeqシステム (中出力)	32	17時間
MiniSeqシステム (高出力)	96	24時間
MiSeqシステム (v2ケミストリー Nano)	4	17時間
MiSeqシステム (v2ケミストリー Micro)	16	19時間
MiSeqシステム (v2ケミストリー)	60	24時間
MiSeqシステム (v3ケミストリー)	96	32時間

表 2: AmpliSeq for Illumina Comprehensive Panel v3に推奨されるイルミナシーケンサー

装置	ラン当たりのサンプル数 (DNAおよびRNA混合)	ラン時間
MiniSeqシステム (高出力)	3	24時間
MiSeqシステム (v3ケミストリー)	3	32時間
NextSeqシステム (中出力)	16	26時間
NextSeqシステム (高出力)	48	29時間

カタログ製品の各シーケンサープラットフォームでのランあたりの解析検体数 (Data Sheetより抜粋) (2)

表 2: AmpliSeq for Illumina Comprehensive Cancer Panelに推奨されるイルミナシーケンサー

装置	ラン当たりのサンプル数	ラン時間
NextSeqシステム (中出力)	4	26時間
NextSeqシステム (高出力)	12	29時間

表 3: AmpliSeq for Illumina Focus Panelに推奨されるイルミナのシーケンサー

装置	ラン当たりのサンプル数 (DNAとRNAの合計)	ラン時間
iSeq™ 100システム	8	17時間
MiniSeqシステム (中出力)	16	17時間
MiniSeqシステム (高出力)	48	24時間
MiSeqシステム (v2ケミストリー-Micro)	8	19時間
MiSeqシステム (v2ケミストリー)	30	24時間
MiSeqシステム (v3ケミストリー)	48	32時間

Table 2: Illumina Sequencing System Recommended for Use with the AmpliSeq for Illumina Exome Panel

Sequencing System	Exomes per Run	Run Time
NextSeq System (high output)	8	29 hours

表 3: AmpliSeq for Illumina Immune Response Panelに推奨されるイルミナシーケンサー

装置	ラン当たりのサンプル数	ラン時間
MiniSeq™システム (中出力)	8	17時間
MiniSeqシステム (高出力)	24	24時間
MiSeq™システム (v2ケミストリー)	16	24時間
MiSeqシステム (v3ケミストリー)	24	32時間
NextSeqシステム (中出力)	96	26時間

表 2: AmpliSeq for Illumina Transcriptome Human Gene Expression Panelで推奨されるイルミナのシーケンサー

装置	ラン当たりのサンプル数	ラン時間
MiniSeq™システム (高出力)	3	24時間
MiSeq™システム (v3ケミストリー)	3	32時間
NextSeqシステム (中出力)	12	26時間
NextSeqシステム (高出力)	40	29時間

Sequencing Coverage Calculatorの紹介

- Sequencing Coverage Calculatorを用いれば、カタログパネルの、各シーケンサープラットフォームのシーケンス条件を調べることも可能です。

http://jp.support.illumina.com/downloads/sequencing_coverage_calculator.html

Sequencing Coverage Calculator

Thank you for using the Illumina coverage calculator.

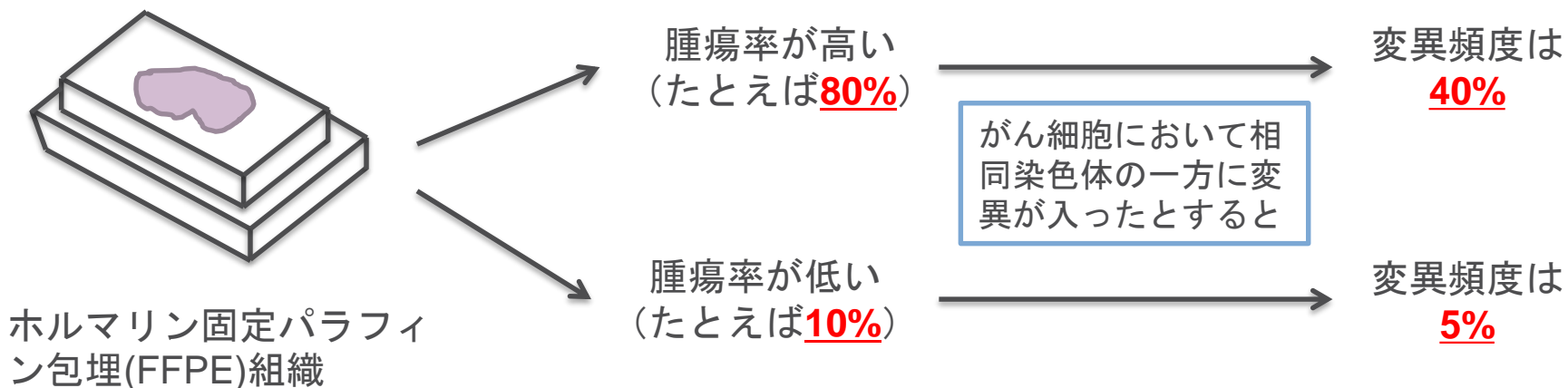
The results were calculated based on: **Recommended samples per flow cell or lane.** [Explain the calculations](#)

Application or product:	AmpliSeq BRCA Panel for Illumina (germline)
Genome or region size:	0.026 Mbases
Read length:	150
Recommended number of clusters (on target):	33,333

	iSeq 100
Run type	i1 Reagents
Clusters	4,000,000 per flow cell
Exceeds maximum read length?	Does not exceed maximum (2x300)
Number of units (flow cell or lane)	0.01 flow cells
Samples per unit (flow cell or lane)	96 / flow cell
	iSeq 100 i1 Reagents

がんといった体細胞変異の検出に必要なカバレッジ

あくまで仮定ですが、



腫瘍率によっては、それほど高いカバレッジは要求されないかもしれません。また、腫瘍率からおおよそその変異頻度が推察できるので、解析の結果、得られた変異頻度と矛盾がないかを確認すると良い。

生殖細胞系列でのバリエーション検出に必要なカバレッジ

- 体細胞変異5%程度の検出にはX2,500程度までのカバレッジが必要ですが、生殖細胞系列のバリエーション検出の場合、X500程度までのカバレッジが得られるようにシーケンスすれば、ヘテロ接合を取りこぼしなくスクリーニングすることが可能です。



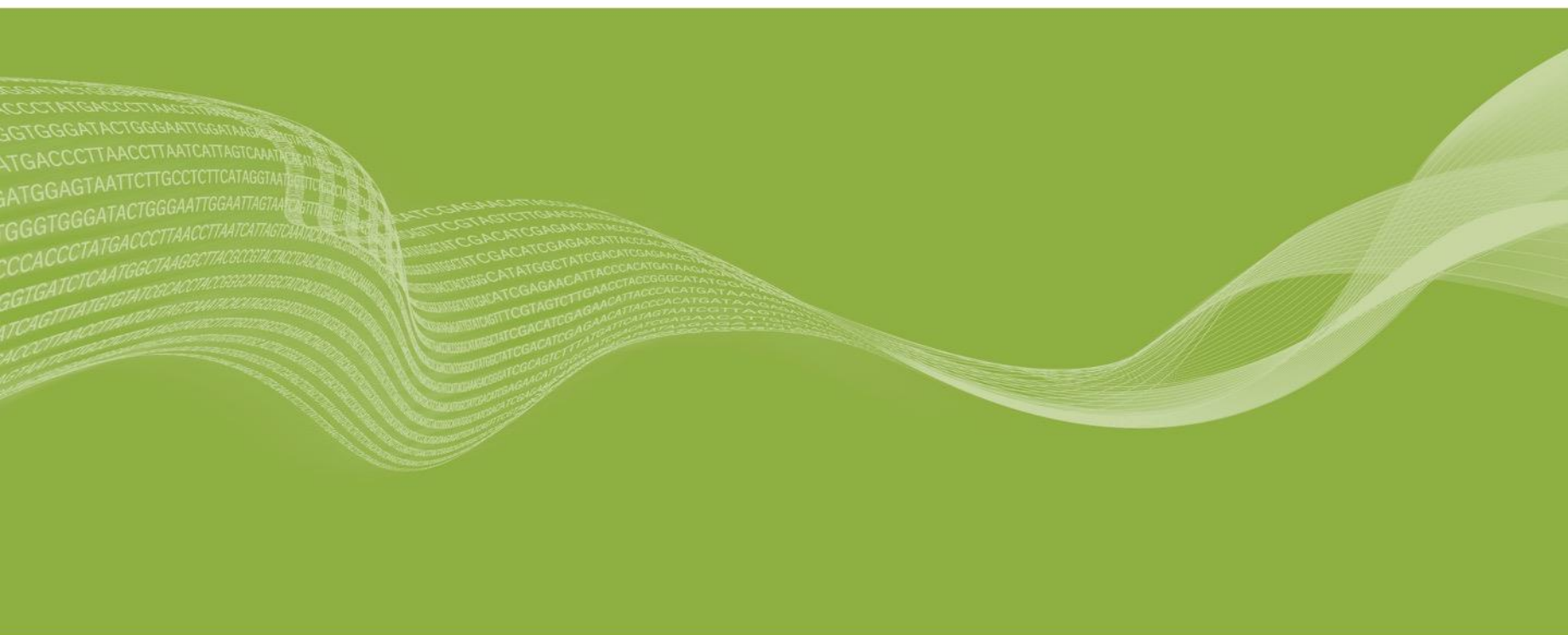
イルミナ®のデモデータのカタログパネルのカバレッジは以下のようになっています

対象	製品名	平均カバレッジ
がん体細胞 変異	Cancer Hotspot v2 Panel	2,500~3,300程度
	Comprehensive Cancer Panel	1,300~2,500程度
体細胞変異	Exome Panel	300~450程度

イルミナ®のデモデータは、BaseSpaceで公開をしています、閲覧にはアカウントの取得が必要になります(無償)。

<https://basespace.illumina.com/datacentral>

情報解析結果の確認 実験がうまくいったか？

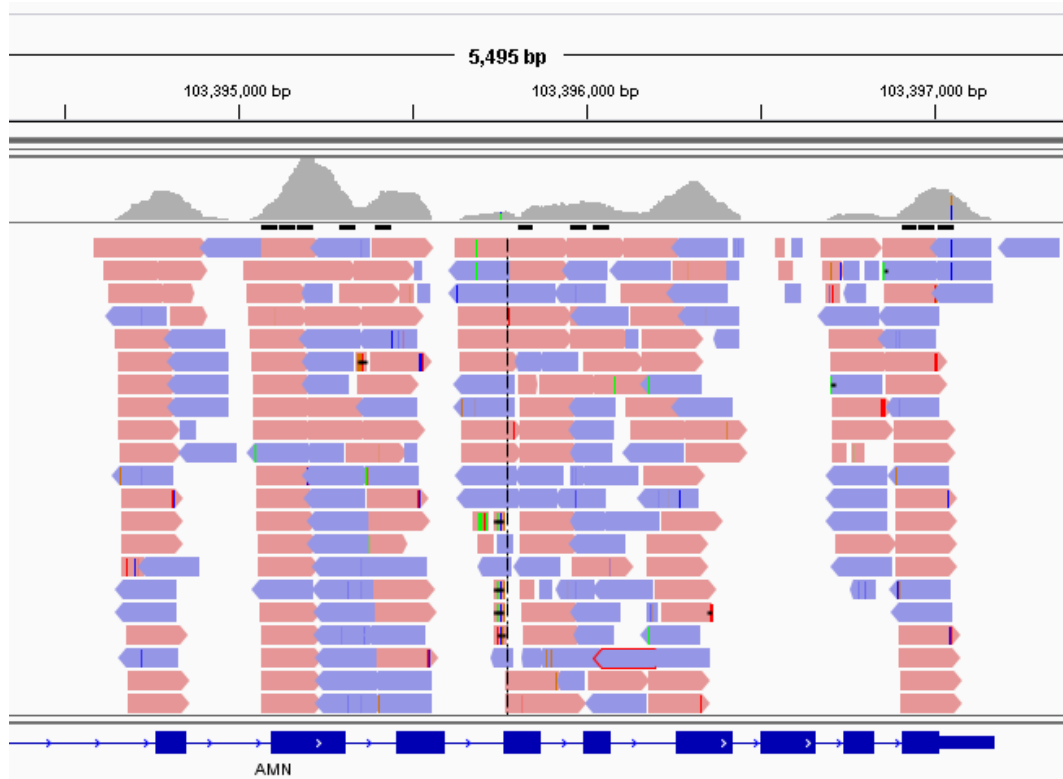


実験がうまくいったか確認することは3つ

On target Rate、Coverage Depth、Coverage Uniformity

On Target Rate (%):

読んだデータのうちどれだけ標的領域にアライメントできたか？



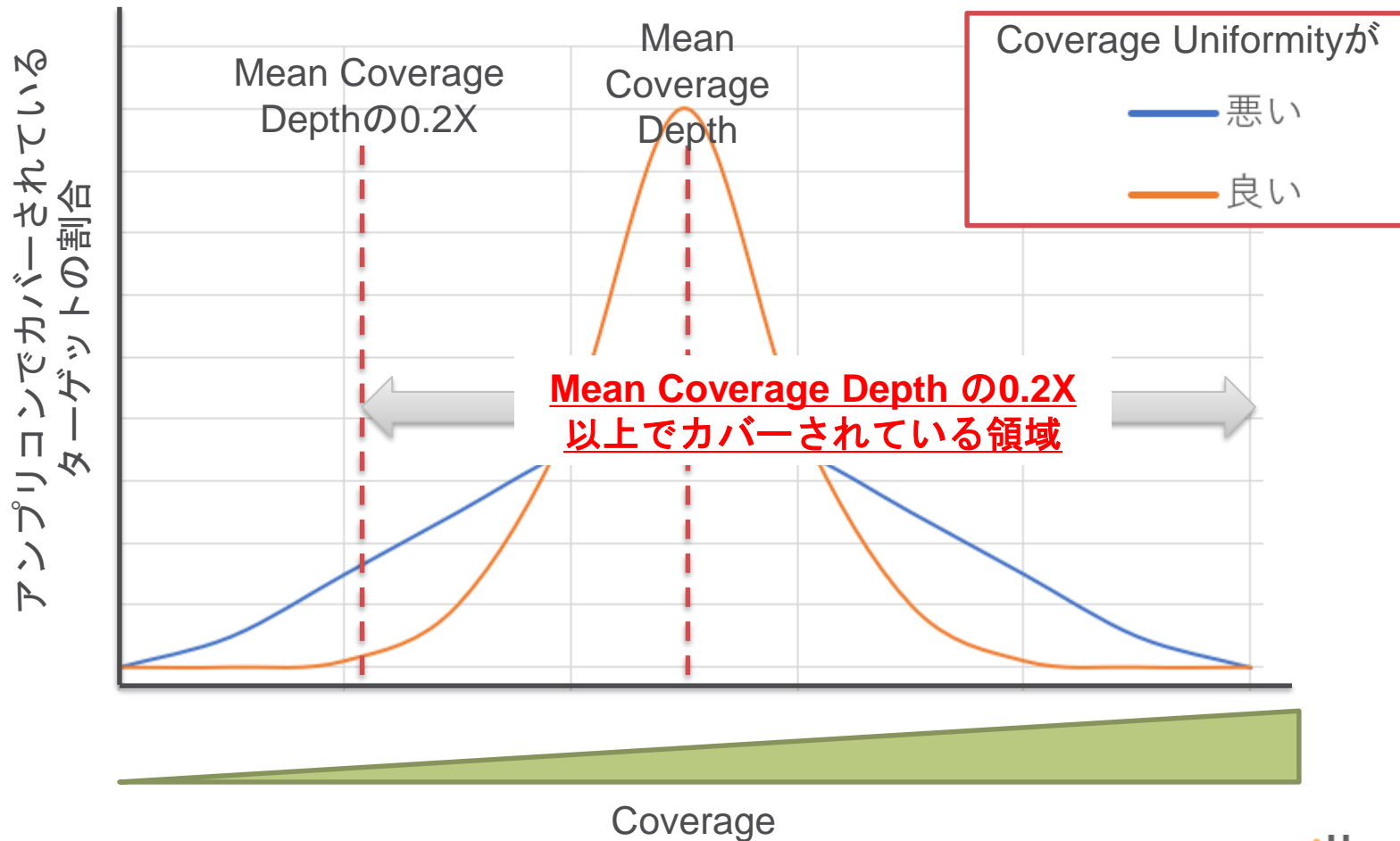
Coverage Depth:

標的領域をどれぐらいの厚みで読めたか？

Coverage uniformity:

標的領域が均一な厚さで読めたか？

Coverage Uniformityのスコアは、 Mean Coverage Depthの0.2X以上でカバーされている領域割合で示される。



BaseSpace DNA Amplicon解析結果の確認

解析結果に問題がなかったかを確認しましょう

Reportタブを選択し、Summaryを選択してください

AGGREGATE SUMMARY

Aggregate Summary Metrics

SAMPLE INFORMATION ⁱ

Sample	Sample Name	Total PF Reads	Percent Q30 Bases	Percent On-target Aligned Reads	Autosome Call Rate
1	HD200-10ng-E02A-B10	592,058	95.46%	97.85%	99.97%
2	HD729-10ng-E02A-D05	679,966	95.13%	97.85%	99.94%
3	HD729-10ng-E02A-C11	704,438	95.18%	98%	99.93%
4	HD200-10ng-E02A-B02	461,070	95.53%	97.70%	99.96%

解析したすべての検体の解析サマリーが表示されます



次のスライドからどこを確認するかを示します。

BaseSpace DNA Amplicon解析結果の確認

On target Rate、Coverage DepthとUniformity

Read Level Statistics i

Sample	Sample Name	Total Aligned Reads (R1/R2)	Percent Aligned Reads (R1/R2)	Overall Percent Aligned Reads
1	HD200-10ng-E02A-B10	287,721 / 286,054	97.19% / 96.63%	96.91%
2	HD729-10ng-E02A-D05	329,381 / 327,399	96.88% / 96.3%	96.59%
3	HD729-10ng-E02A-C11	341,089 / 339,486	96.84% / 96.38%	96.61%
4	HD200-10ng-E02A-B02	223,789 / 222,677	97.08% / 96.6%	96.84%

Show entries

Percent Aligned Reads (R1/R2): On Target Rateとも言われるスコアになります。標的領域にアライメントされたReadの割合*を示します。*BaseではなくReadで確認したほうがよい

COVERAGE SUMMARY

Sample	Sample Name	Amplicon Mean Coverage Depth	Uniformity of coverage (Pct > 0.2*mean)
1	HD200-10ng-E02A-B10	2712.3	99.52
2	HD729-10ng-E02A-D05	3104.6	100
3	HD729-10ng-E02A-C11	3222	100
4	HD200-10ng-E02A-B02	2109.1	100

Amplicon Mean Coverage Depth: カバレッジデプスの平均値を示しています。体細胞変異の高いカバレッジデプスが必要です。

Uniformity of Coverage (Pct > 0.2*mean): 平均カバレッジデプスのx0.2以上のデプスで覆われている、標的領域の割合を示しています。

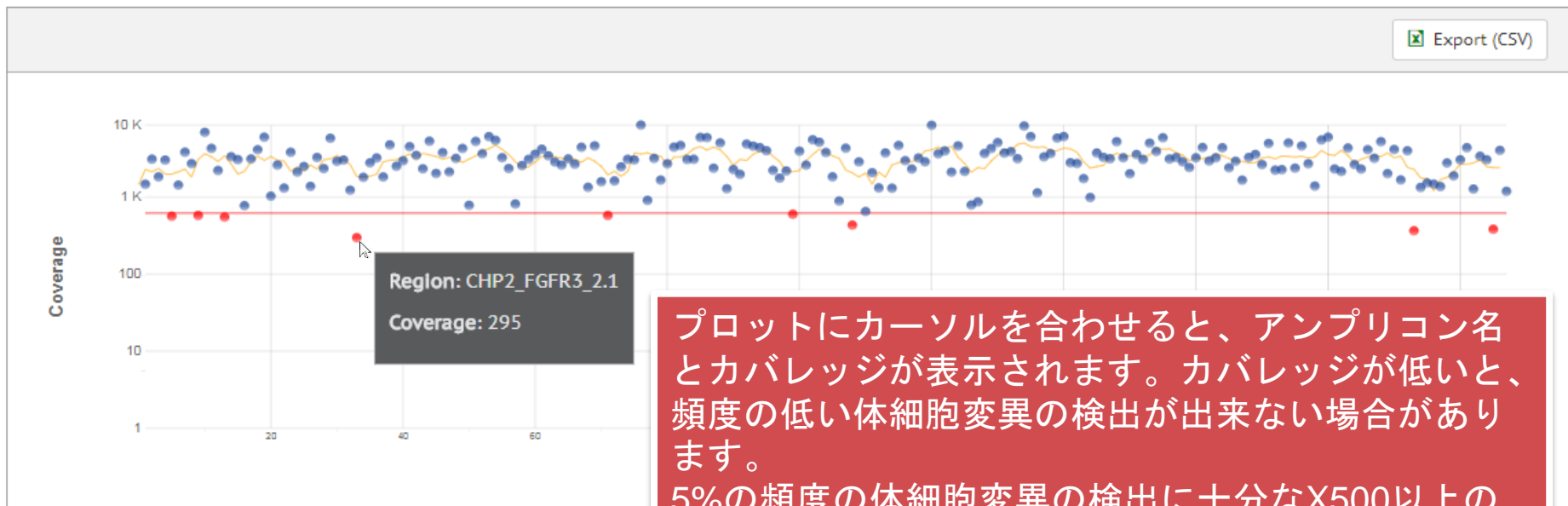
BaseSpace DNA Amplicon解析結果の確認

On target Rate、 Coverage Depthと Uniformity

COVERAGE SUMMARY

Amplicon Mean Coverage	Uniformity of Coverage (Pct > 0.2*mean)
3078.7	95.65

Coverage by Amplicon Region



カタログパネル製品のスペック

パネル製品	On Target Ratio (%)/ Percent Aligned Reads	Uniformity of Coverage (Pct > 0.2*mean)
BRCA Panel	>80%	>95%
Cancer Hotspot Panel v2	>80%	>95%
Focus Panel	>80%	>95%
Comprehensive Panel v3	>85%	>95%
Comprehensive Cancer Panel	>90%	>90%
Exome Panel	>80%	>90%

*各製品のデータシートを参考

BaseSpace DNA Amplicon解析結果の確認

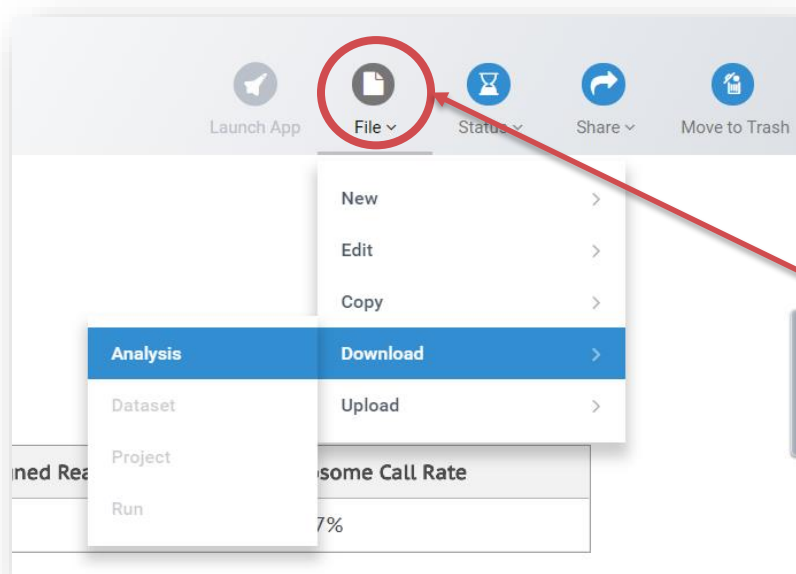
On target Rate、Coverage DepthとUniformity

- 期待した結果が得られなかった場合

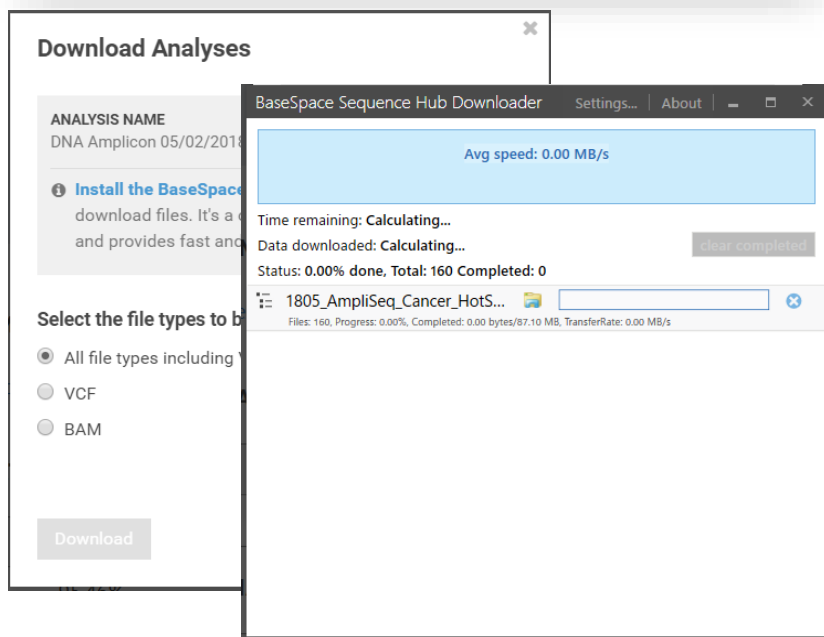
- Percent Aligned ReadsやUniformity of Coverage といった数値が低い場合には、PCR条件を確認する。
- Amplicon Mean Coverage Depthが低い場合には、シーケンス量が十分かを確認する。
- サンプルの分解度やサンプル量にあったPCR条件であるかを確認する。(サンプル量や分解度に合わせる形でPCRサイクル数をReference Guideに従い変更します。)

標的領域(GC% richなど)によっては、解析が難しい場合もあります。

BaseSpaceでの解析結果ファイルの一括ダウンロード



画面右上で、File > Download > Analysisと進むと、解析結果ファイル (BAMファイル、VCFファイルを含む) を一括ダウンロードすることが可能です。



容量が非常に大きくなることもあるので、専用のダウンローダー (BaseSpace Sequence Hub Downloader) を用いてダウンロードを実行します。

個々のファイルもダウンロード可能です。

1805_AmpliSeq_Cancer_HotSpot-v2-Demo
Analysis: DNA Amplicon 05/02/2018 2:25:52

SUMMARY REPORTS INPUTS **FILES**

Launch App File Status Share Move to Tra

Files

NAME		
..		
.basespace		
HD729-10ng-E02A-D05_S4.annotations.json.gz	gz	81.67 KB
HD729-10ng-E02A-D05_S4.bam	bam	20.32 MB
HD729-10ng-E02A-D05_S4.bam.bai	bai	1.49 MB
HD729-10ng-E02A-D05_S4.baseCoverage.csv	csv	12.02 KB
HD729-10ng-E02A-D05_S4.coverage.csv		0 KB
HD729-10ng-E02A-D05_S4.coverage.plot.csv		KB
HD729-10ng-E02A-D05_S4.genome.vcf.gz		10 KB
HD729-10ng-E02A-D05_S4.genome.vcf.gz.tbi		KB
HD729-10ng-E02A-D05_S4.report.html	html	827.17 KB
HD729-10ng-E02A-D05_S4.report.pdf	pdf	709.37 KB
HD729-10ng-E02A-D05_S4.summary.csv	csv	2.83 KB

contact us

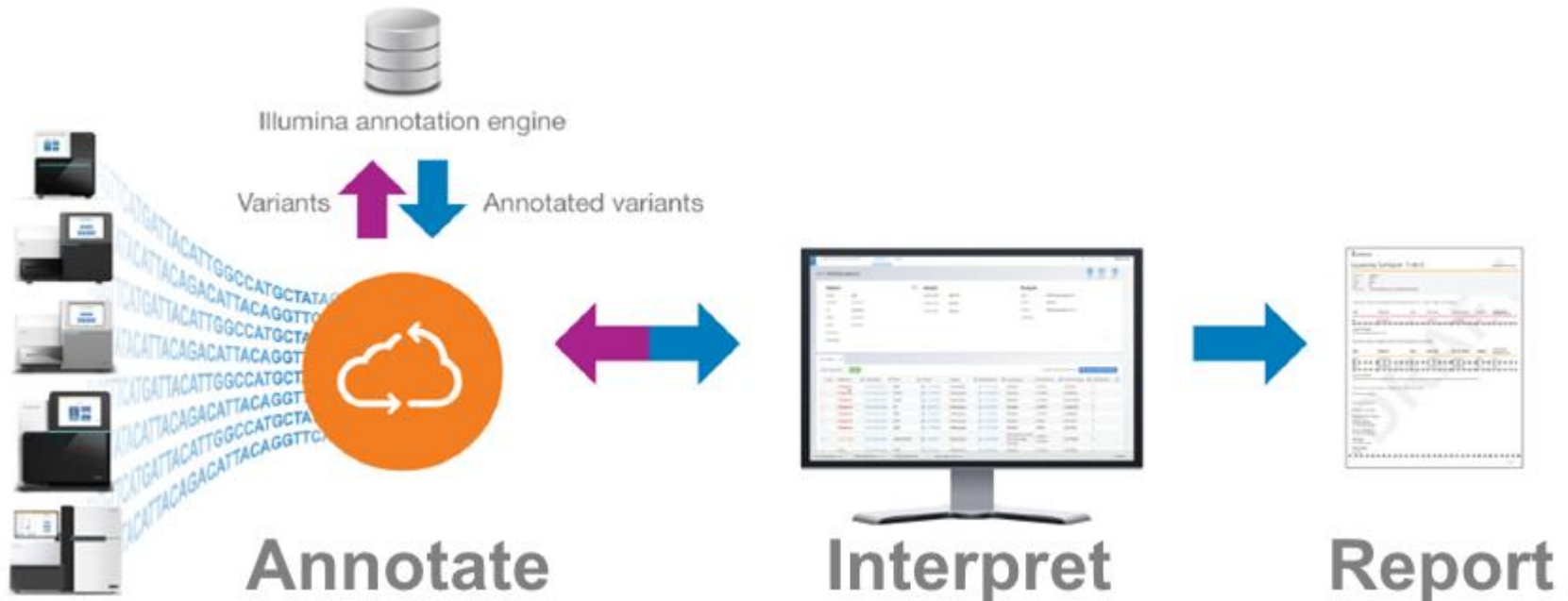
左上のFILEのタブから、個々の出力されたファイルの確認とダウンロードを行うことも可能です。

出力ファイルが表示され、VCFファイルなど必要なファイルだけのダウンロードも可能です。

変異解析結果へのアノテーションとフィルタリング



変異・バリエーションへのアノテーションとフィルタリング



- さまざまなデータベースを用いて正確な注釈を提供する
- 生物学的に重要な変異・バリエーションを同定する
- 解釈とレポートの合理化

変異・バリエーションへのアノテーションとフィルタリング

基本的に、VCF fileには、参照配列と異なるゲノム位置と塩基情報の記載のみ
(イルミナの解析ソフトはいくつかの項目に対しアノテーションを実施します)

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT
Chr2	48030639	.	A	C	94.00	PASS	DP=12703	GT:GQ:VF:NL:SB 0/1:94:0.0322:20:-40.7136:



【アノテーションの実施】

- ✓ タンパク質の機能的影響予測
- ✓ 人種集団におけるアレル頻度
- ✓ 疾患やがん体細胞変異データベース(ClinVarやCOSMIC)との比較
- ✓ 臨床試験データの情報 (Variant Interpreterのみ)



【フィルタリング】

アノテーション情報をもとに、原因と考えられる変異・バリエーションの絞り込みを行う。

VariantStudioとVariant Interpreterの紹介

VariantStudio



VariantStudio™

お客様のWindows PC

Variant Interpreter



クラウド環境

Pros

- ✓ 専用のソフトウェアを用いて結果を閲覧し、多検体閲覧も非常に楽。
- ✓ インターネット接続は不要。

- ✓ ブラウザを通してどこでも結果閲覧が可能。
- ✓ 臨床試験データの情報が閲覧できる。

Cons

- ✓ データベースが最新のものとは限らない
- ✓ Variant Interpreterで閲覧できる、臨床試験データの情報が閲覧できない。

- ✓ 現在は無償で提供しているが、将来的に有償の予定。
- ✓ 多検体をまとめて閲覧するのは不得手。

VariantStudioでのアノテーションとフィルタリング

The screenshot displays the Illumina VariantStudio 3.0 interface. The left sidebar contains a 'Filters' panel with categories: General, Variant, Gene, Consequence, Population Frequency, Cross Sample Subtraction, Family Based, Custom, and Classification. The 'Variant' filter is highlighted with a red box, and 'Consequence' and 'Population Frequency' are highlighted with purple and yellow boxes respectively. The main area shows three filter configuration panels:

- COSMIC Databaseに登録のある変異への絞り込み** (Filtering by variants registered in the COSMIC Database):
 - Only variants with Cosmic annotation
 - where matches mutant allele
 - where not matches mutant allele
- タンパク質の機能への影響** (Impact on protein function):
 - Show only variants that are:
 - Missense
 - Polyphen "damaging"
 - SIFT "deleterious"
 - Frameshift
 - Stop gained
 - Stop lost
 - Initiator codon
 - In-frame insertion
 - In-frame deletion
 - Splice

- 集団内頻度での絞り込み** (Filtering by population frequency):
- Thousand Genomes**
 - Global Frequency < 5.00 %
 - American Pop Frequency < 5.00 %
 - E. Asian Pop Frequency < 5.00 %
 - S. Asian Pop Frequency < 5.00 %
 - African Pop Frequency < 5.00 %
 - European Pop Frequency < 5.00 %
- ExAC**
 - Global Frequency < 5.00 %
 - American Pop Frequency < 5.00 %
 - E. Asian Pop Frequency < 5.00 %
 - S. Asian Pop Frequency < 5.00 %
 - African Pop Frequency < 5.00 %
 - Eur (Fin) Pop Frequency < 5.00 %
 - Eur (NF) Pop Frequency < 5.00 %

At the bottom left, the text 'フィルタリング' (Filtering) is written in red. The status bar at the bottom shows 'Sample: HD200-rep1-001_54 Genes, Variants: (40, 110) -> (40, 110)'. The Illumina logo is in the bottom right corner.

VariantStudioでのアノテーションとフィルタリング

The screenshot displays the Illumina VariantStudio 3.0 software interface. The top menu bar includes Home, Annotation & Classification, Reports, and Help. The ribbon contains various actions like Open, Save, Import VCF, Add Variants to Sample, Import Folder, View All Samples, Remove Sample, Filter Favorites, and Table Options. The main window is titled 'Gene View' and shows a table of variants. A 'Filters' sidebar on the left lists categories like General, Variant, Gene, Population Frequency, and Classification. A large blue circle highlights the 'Filters' sidebar and the 'Apply Filters =>' and 'Clear Filters' buttons. A large white circle highlights the variant table. Overlaid text in Japanese reads '絞り込みフィルター' (Filtering Filter) and '変異部位一覧とアノテーション情報' (Variant List and Annotation Information). The variant table has columns: Gene, Variant, Chr, Coordinate, Classification, Type, Genotype, Exonic, Filters, Quality, GQX, Inherited From, Alt Variant Freq, Read Depth, Alt Read Depth, Allelic Depths, and Custom Annotation. The table shows variants for genes like MPL, NRAS, IDH1, ERBB4, VHL, and CTNNB1. At the bottom, the status bar shows 'Sample: HD701-rep1_55.genome.vcf Genes, Variants: (39, 134) -> (39, 134)'. The bottom right corner features the Illumina logo.

Gene	Variant	Chr	Coordinate	Classification	Type	Genotype	Exonic	Filters	Quality	GQX	Inherited From	Alt Variant Freq	Read Depth	Alt Read Depth	Allelic Depths	Custom Annotation
MPL	G>G/T	1	43815030	...	snv	het	yes	LowV...	100	100		2	9785	193	9590,...	
MPL	G>G/T	1	43815050	...	snv	het	yes	LowV...	28	28		1.3	9748	128	9608,...	
NRAS	AG>AG/A	1	115256618	...	deletion	het	yes	PASS	49	49		12.5	48	6	42,6	
NRAS	G>G/A	1	115256621	...	snv	het	no	LowGQ	25	25		8.3	48	4	44,4	
NRAS	G>G/A	1	115256622	...	snv	het	no	LowGQ	26	26		8.9	45	4	38,4	
NRAS	G>G/T	1	115256622	...	snv	het	no	LowGQ	20	20		6.7	45	3	38,3	
NRAS	G>G/A	1	115256623	...	snv	het	no	PASS	81	81		20.5	44	9	35,9	
IDH1	G>G/A	2	209113192	...	snv	het	yes	LowV...	100	100		4.6	33984	1549	32430,...	
ERBB4	TAAAAA>...	2	212578379	...	deletion	het	yes	R8	100	100		5.9	19167	1140	18027,...	
ERBB4	TAAAA>T...	2	212578379	...	deletion	het	yes	R8	100	100		49.2	19203	9447	9756,...	
ERBB4	TAAA>TA...	2	212578379	...	deletion	het	yes	no	100	100		28.4	19299	5475	13824,...	
ERBB4	TAA>TAA/T	2	212578379	...	deletion	het	yes	R8	100	100		7.6	19369	1474	17895,...	
ERBB4	TA>TA/T	2	212578379	...	deletion	het	yes	LowV...	84	84		1.4	19385	280	19105,...	
VHL	A>A/T	3	10183870	...	snv	het	yes	LowV...	100	100		2	7947	162	7783,...	
VHL	C>C/T	3	10191488	...	snv	het	yes	LowV...	100	100		2.1	35571	743	34821,...	
CTNNB1	C>C/A	3	41266101	...	snv	het	yes	PASS	100	100		36.7	16045	5885	10156,...	
CTNNB1	CCTT>CC...	3	41266133	...	deletion	het	yes	PASS	100	100		8.8	15988	1406	14582,...	

VariantStudioを用いたがん体細胞変異のフィルタリングの例

Quality Filter

- コールされたバリアントのうち、品質が低いバリアントを除きます。

Population Frequency Filter

- あるポピュレーションにおいて、頻度が1%以上で報告されているバリアントを除きます。

Allele Frequency Filter

- 体細胞変異の頻度が低いバリアントを除きます。頻度が低いバリアントは、偽陽性の可能性があります、

COSMIC annotation Filter

- がんの体細胞変異データベースであるCOSMICに登録のあるバリアントのみに絞り込みを行います。

機能的影響

- ミスセンス・ナンセンス変異やタンパク質の機能予測に絞り込みを行います。

Primary Site · Histology

- 興味のある組織（や器官）でCOSMICで報告があるかで、絞り込みを行います。



VariantStudio™

専用のソフトウェアを用いてクリック操作で、簡単に上記の絞り込み操作を行うことが可能です。

illumina®

Variant Interpreterでのアノテーションとフィルタリング

SNV / INDELS

APPLY VARIANT FILTERS or Select Saved Filter Apply Saved Filter Showing 131 of 131 variants.

VARIANT	GENE	CONSEQUENCE	ASSOCIATIONS	POPULATION FREQ	METRICS	CUSTOM ANNOTATIONS	INTERPRETATION																																
chr1:115256530 rs121913254 REF: G ALT: T UCSC View in IGV COSMIC Variant Details	NRAS	Missense NM_002524.4 c.181C>A p.(Gln61Lys) Exon: 3 In Silico Predictions S P	<table border="1"><thead><tr><th></th><th>Cases</th><th>MyKB</th><th>BSKN</th></tr></thead><tbody><tr><td>Predictive</td><td>—</td><td>—</td><td>ON</td></tr><tr><td>Mol. Class</td><td>—</td><td>—</td><td>✓</td></tr><tr><td>Prognostic</td><td>✓</td><td>—</td><td>✓</td></tr><tr><td>Clinical Trials</td><td>—</td><td>—</td><td>✓</td></tr></tbody></table>		Cases	MyKB	BSKN	Predictive	—	—	ON	Mol. Class	—	—	✓	Prognostic	✓	—	✓	Clinical Trials	—	—	✓		<table border="1"><thead><tr><th>Filters</th><th>PASS</th></tr></thead><tbody><tr><td>Variant Read ...</td><td>0.1026</td></tr><tr><td>Alt Allele Depth</td><td>1382</td></tr><tr><td>Total Read D...</td><td>13466</td></tr><tr><td>GQX</td><td>100</td></tr><tr><td>Quality Score</td><td>100</td></tr></tbody></table>	Filters	PASS	Variant Read ...	0.1026	Alt Allele Depth	1382	Total Read D...	13466	GQX	100	Quality Score	100		<input type="checkbox"/> Add to Report CASE INTERPRETATION Comments (0)
	Cases	MyKB	BSKN																																				
Predictive	—	—	ON																																				
Mol. Class	—	—	✓																																				
Prognostic	✓	—	✓																																				
Clinical Trials	—	—	✓																																				
Filters	PASS																																						
Variant Read ...	0.1026																																						
Alt Allele Depth	1382																																						
Total Read D...	13466																																						
GQX	100																																						
Quality Score	100																																						

Variant

バリエーションの概要、リンク、Curated Contents

Gene

OMIMへのリンク

Consequence

変異・バリエーションとEnsembleへのリンク

Associations

MyKB: My Knowledge Base network
BSKN: BaseSpace Knowledge Network

Metrics

変異・バリエーションの品質情報 (Depth、QUALITY)

Variant Interpreterでのアノテーションとフィルタリング

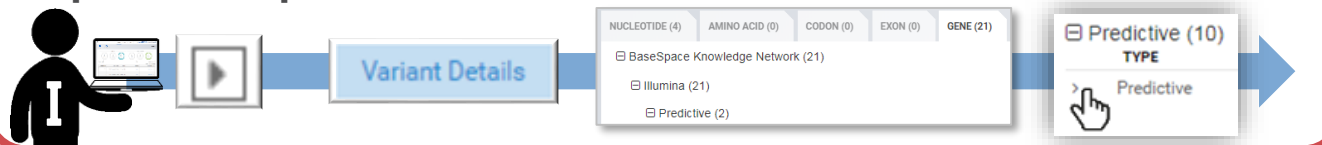
The screenshot displays the Variant Interpreter interface. At the top, there's a filter section with 'SNV / INDELS' selected and 'APPLY VARIANT FILTERS' button. Below, a table shows variant details for 'chr1:115256530' (G > T) in the NRAS gene, with a Missense consequence (p.(Gln61Lys)). A 'Predictive (10)' dropdown is open, showing a table of predictive annotations:

TYPE	GENOME BUILD	PHENOTYPE	DRUG	DIRECTION	VALIDATION LEVEL	SUBMITTED BY	LAST UPDATED	STATUS
Predictive	GRCh37	Malignant melanoma	Vemurafenib	Reduced Response	Case Reports	Illumina Inc.	Aug 25, 2016	Enabled

Below the table, there are sections for 'Details' and 'Evidence'. The 'Evidence' section includes a 'Publication (3)' with 'Publication PMID: 22395615' and a summary: 'In an in vitro study, a M249 R4 cell line expressing BRAF V600E (a known sensitizing mutation to vemurafenib) and NRAS Q61K co-mutations demonstrated acquired resistance to vemurafenib treatment. Resistance was determined by assessing cell survival.'

外部リソースへのリンク、論文のリンク、エビデンスサマリーといった臨床試験の情報

Explore Interpretation Tools



VariantStudioとVariant Interpreter 使用方法はウェビナーを参考に

- **BaseSpace Variant Interpreter (Beta!)、ことはじめ【イルミナiSchool 初級】**
 - https://jp.illumina.com/events/webinar/2017/webinar_170628_j.html
- **バージョンアップしたデスクトップ版VariantStudio v3.0変異解析ソフトウェアの徹底活用法【イルミナiSchool 中級】**
 - https://jp.illumina.com/events/webinar/2017/webinar_170524_j.html

イルミナの過去ウェビナーは以下で動画とスライドを公開しています
<https://jp.illumina.com/events/webinar.html>

まとめ (1)

● 情報解析の実行

- 標的領域を示すManifestファイルといった必要なファイルを準備する。
- クラウド環境で使用するBaseSpace、もしくはLocalのWindows PCで使用するLocal Run Managerを用いて、クリック操作で簡単に情報解析を実行ができます。
- がんの体細胞変異解析と生殖細胞系列のバリエーション解析では使用するVariant Callerが異なる。

● 変異解析に必要なカバレッジ

- がんの体細胞変異解析で5%頻度の変異を検出するにはX500カバレッジが必要十分。標的領域の広い範囲で、X500カバレッジを達成するには、標的領域のX2,500程度までを読むとよい。
- 生殖細胞系列のバリエーション解析は、標的領域のX500程度までを読むとよい。

まとめ (2)

- **情報解析結果の確認**

- Percent Aligned Reads (On Target Rate)、Uniformity of Coverage、Amplicon Mean Coverage Depthといった値を確認し、必要なカバレッジが得られたか、ライブラリー調製に問題がなかったかを確認します。

- **変異解析結果へのアノテーションとフィルタリング**

- Windows PCにインストールして使用するVariantStudioと、クラウド環境で使用するVariant Interpreterの二つを用いて解析を行うことが可能です。