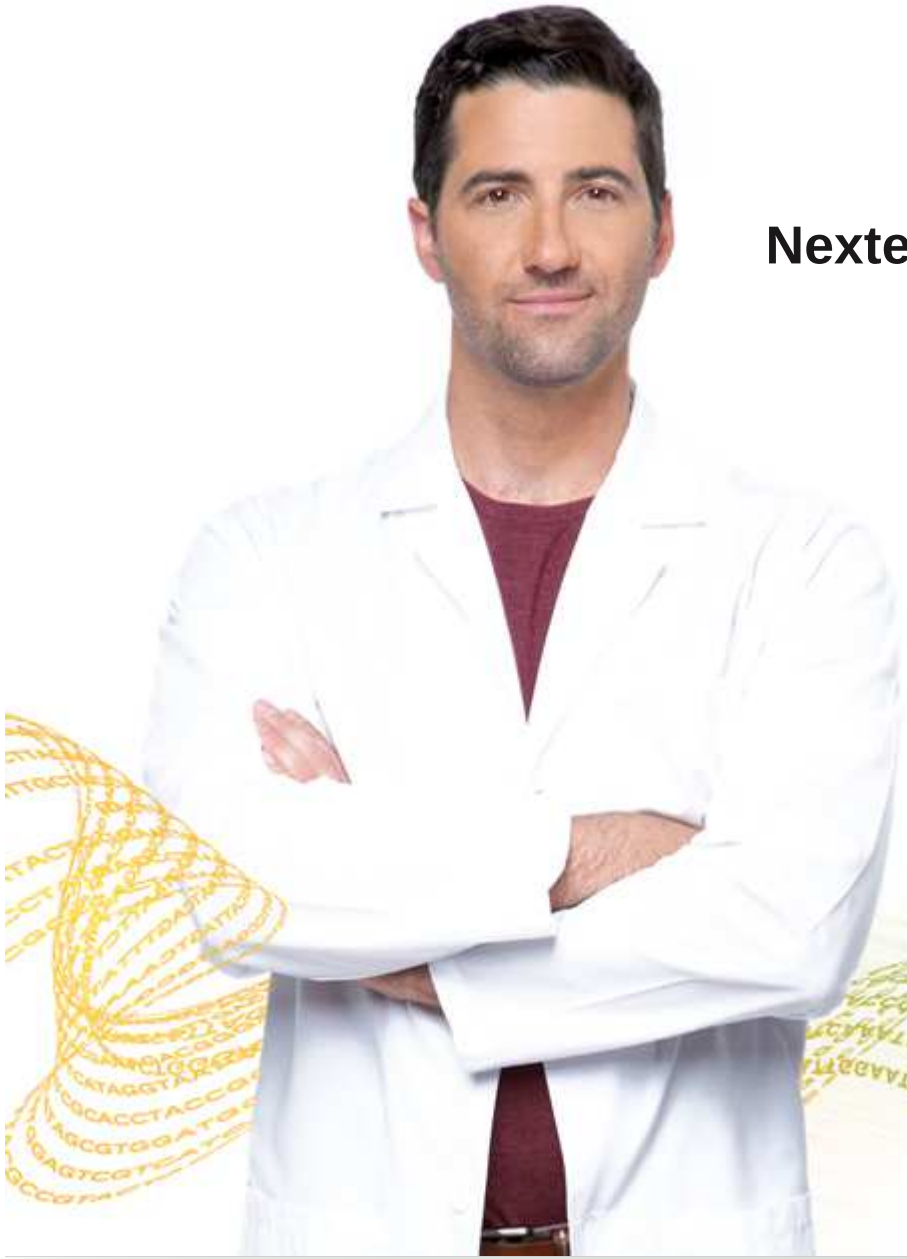
2015年5月15日
イルミナサポートウェビナー



Nextera Rapid Capture Exomeキットを用いたエクソームシーケンス - ドライ編 –

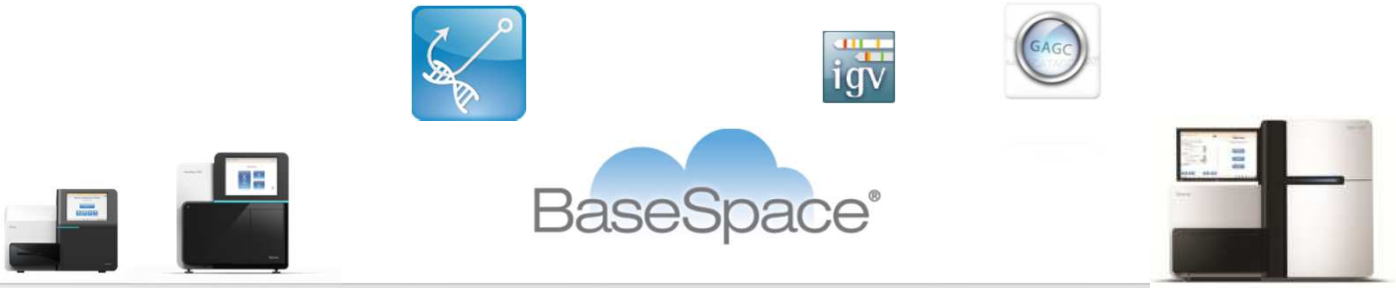
BaseSpaceで行うかんたんNGSデータ解析
< Enrichment アプリ >

イルミナ株式会社
バイオインフォマティクス
サポートサイエンティスト
癸生川絵里 (Eri Kibukawa)

illumina®

本日の内容

- エクソームシーケンス解析 概要
- BaseSpace のエクソーム解析 アプリ
- BaseSpace 実行例
  - BaseSpaceでNetera Rapid Captureエクソームのデモデータを取り込む
  - BaseSpaceでEnrichmentアプリの実行結果をみる
  - BaseSpaceでEnrichmentアプリを実行する



illumina®

DNA リシーケンシング、エクソーム 解析
典型的な解析ワークフロー

＜データフォーマット＞

BCL からの変換 → FASTQ

↓

アライメント → BAM

↓

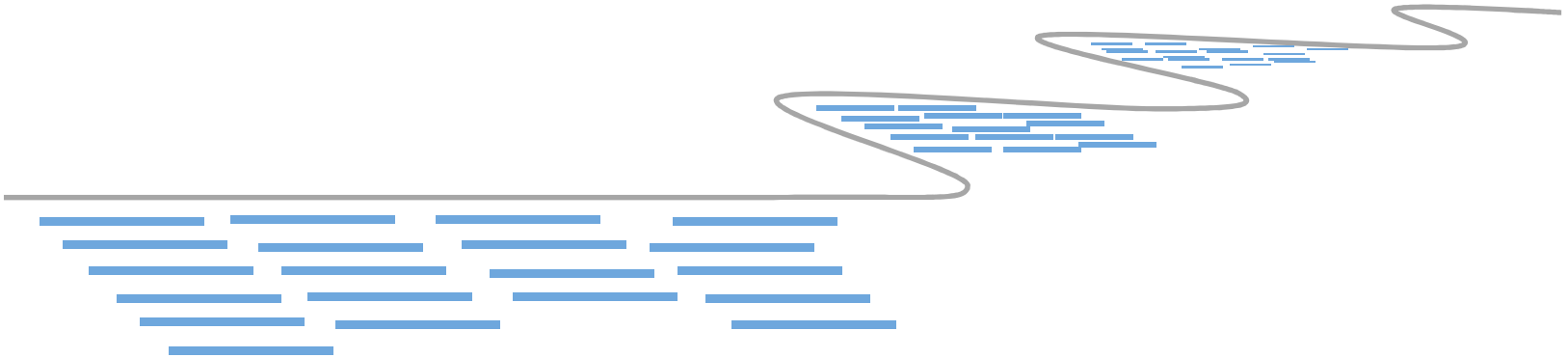
変異解析 → VCF

↓

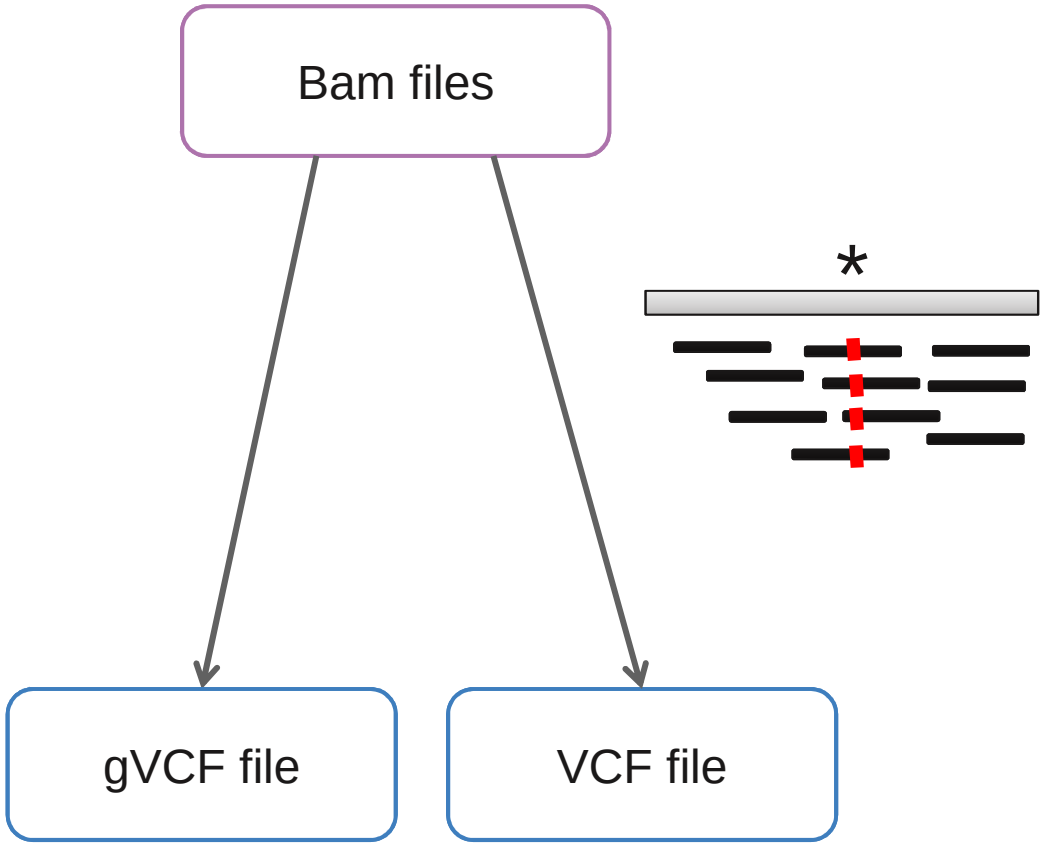
アノテーションやフィルタリングによる結果の絞込み → 表形式ファイルや やpdf等一般レポートの形

アライメント

- 通常 genomic DNA サンプル　(全ゲノム, 濃縮系, アンプリコン 等)
- リファレンスゲノム配列に対して,リードをアライメント(マップ)



変異コール

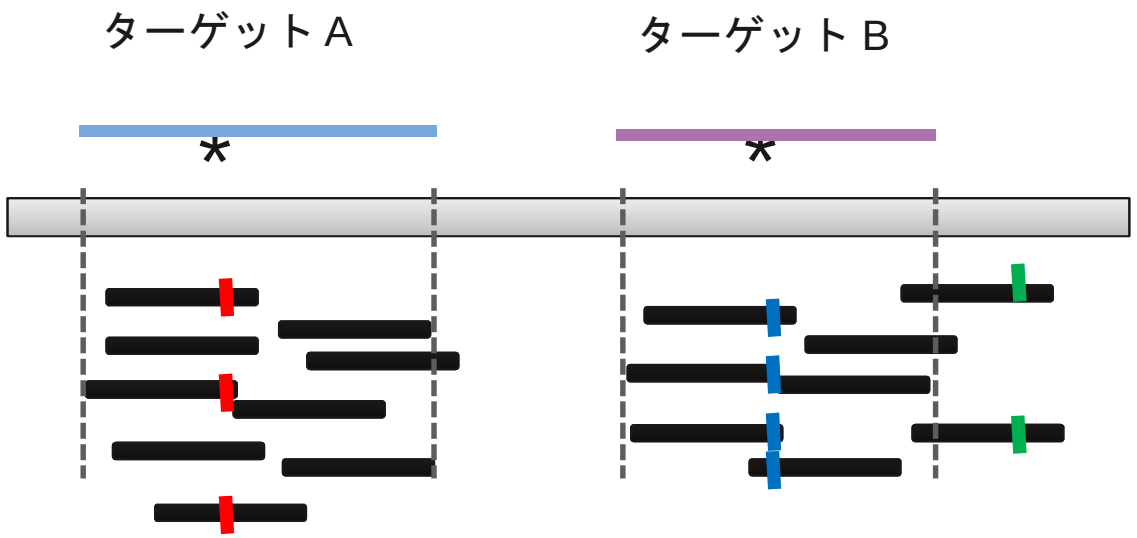


- サンプルとリファレンス間の違いについて検出するための統計的検定
- 主要な検討材料;
  - Quality score
  - Frequency
  - Depth
  - Strand bias
- 結果をVCF出力

VCF, genomeVCF(gvcf) につきましては、
サポートウェビナー　2013/11/01
MiSeq Reporterアップデート　をご参考いただけます。
http://www.illuminakk.co.jp/events/webinar_japan.ilmn

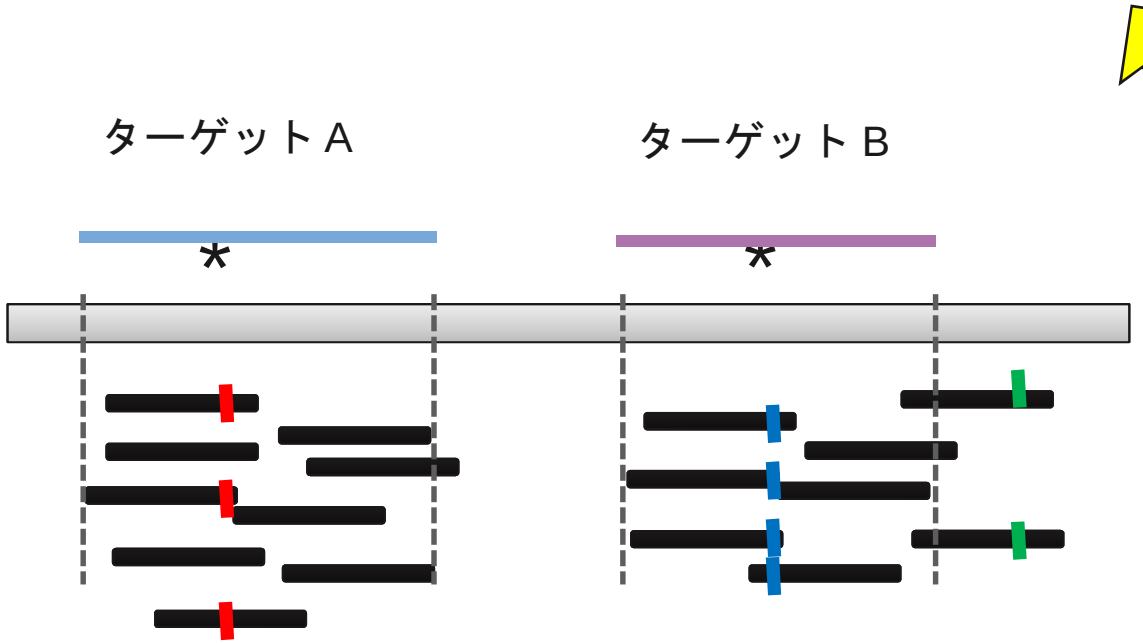
全ゲノム (WGS) と エクソーム？

Exome・・・設計したエクソン領域(ターゲット領域)を解析対象としている



全ゲノム (WGS) と エクソーム？

Exome・・・設計したエクソン領域(ターゲット領域)を解析対象としている



設計した
ターゲット領域
について記した
ファイルを、
マニフェスト
ファイルと呼ぶ

上流工程のEnrichmentサンプル調整のご説明は、
前回サポートウェビナー ウェット編 2015/04/24 をご参考下さい。
http://www.illuminakk.co.jp/events/webinar_japan.ilmn

マニフェストファイル

- ターゲットとして設計した配列領域を指定したポジションファイル
- Enrichment ワークフローではマニフェストファイルにどのExon セット(Exome)を解析対象とするかを定義している
- タブ区切りテキスト形式ファイルのため、notepadなどで開くことができる
- “マニフェストファイル”は総称で、ワークフローによりフォーマットの違いがある
- Enrichmentワークフローの場合, イルミナから提供されるファイルの拡張子は **.txt**

- **ユーザが独自に設計したターゲット領域のポジションファイルは、カスタムマニフェスト、と呼ぶ。**
  ※BaseSpace Enrichment App v2より解析対応

解析を行う領域の指定、Enrichment がどの程度できたか(濃縮効率)など計算の要となるのがこのファイル

製品によりターゲットは異なる

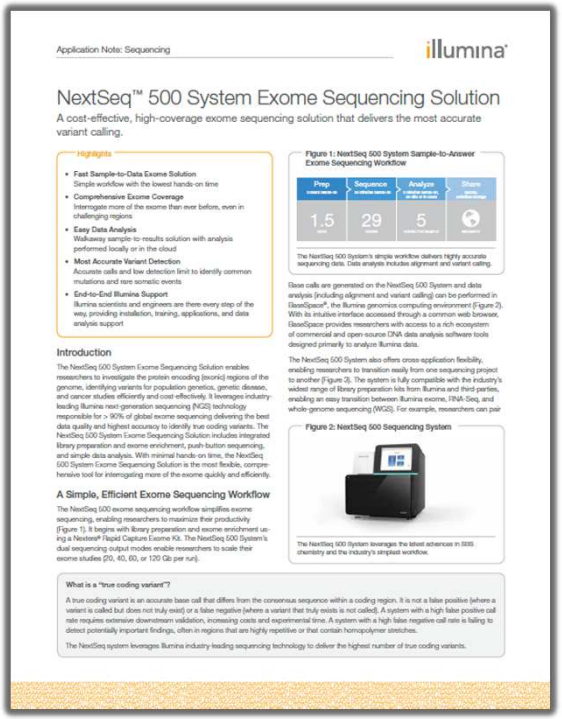
	Nextera Rapid Capture Exome	Nextera Rapid Capture Expanded Exome
ターゲットサイズ	~37 Mb	~62 Mb
コンテンツ	エクソン	エクソン, UTRs, miRNA
必要なシーケンス	>4 Gb	>8 Gb
濃縮時のプーリング	12 plex	
ゲノムDNAスタート量	50 ng	
ハンズオン時間	1.5 日 (5 時間)	
バッチサイズ	～96 エクソーム	

※Manfest v1.1

このほか、TruSightパネル製品などにも
それぞれのターゲットに応じたマニフェストがある

エクソームシーケンシングの資料

- テクニカルノート、Appノートやデータシート
  - Exome Sequencing with NextSeq 500 System Data Sheet
  - **Read Length Optimization for NRC Exome Data Tech Note　☆解析**


Application Note: Sequencing

illumina

NextSeq™ 500 System Exome Sequencing Solution

A cost-effective, high-coverage exome sequencing solution that delivers the most accurate variant calling.

Highlights

- Fast Sample-to-Data Exome Solution
  Simple workflow with the lowest hands-on time
- Comprehensive Exome Coverage
  Interrogate more of the exome than ever before, even in challenging regions
- Easy Data Analysis
  Walkaway sample-to-results solution with analysis performed locally or in the cloud
- Most Accurate Variant Detection
  Accurate calls and low detection limit to identify common mutations and rare somatic events
- End-to-End Illumina Support
  Illumina scientists and engineers are there every step of the way, providing installation, training, applications, and data analysis support

Figure 1: NextSeq 500 System Sample-to-Answer Exome Sequencing Workflow

The NextSeq 500 System's simple workflow delivers highly accurate sequencing data. Data analysis includes alignment and variant calling.

Base calls are generated on the NextSeq 500 System and data analysis (including alignment and variant calling) can be performed in BaseSpace®, the Illumina genomics computing environment (Figure 2). With its intuitive interface accessed through a common web browser, BaseSpace provides researchers with access to a rich ecosystem of commercial and open-source DNA data analysis software tools designed primarily to analyze Illumina data.

The NextSeq 500 System also offers cross-application flexibility, enabling researchers to transition easily from one sequencing project to another (Figure 3). The system is fully compatible with the industry's widest range of library preparation kits from Illumina and third-parties, enabling an easy transition between Illumina exome, RNA-Seq, and whole-genome sequencing (WGS). For example, researchers can pair

Figure 2: NextSeq 500 Sequencing System

The NextSeq 500 System leverages the latest advances in SBS chemistry and the industry's simplest workflow.

Introduction

The NextSeq 500 System Exome Sequencing Solution enables researchers to investigate the protein encoding (exonic) regions of the genome, identifying variants for population genetics, genetic disease, and cancer studies efficiently and cost-effectively. It leverages industry-leading Illumina next-generation sequencing (NGS) technology responsible for > 90% of global exome sequencing delivering the best data quality and highest accuracy to identify true coding variants. The NextSeq 500 System Exome Sequencing Solution includes integrated library preparation and exome enrichment, push-button sequencing, and simple data analysis. With minimal hands-on time, the NextSeq 500 System Exome Sequencing Solution is the most flexible, comprehensive tool for interrogating more of the exome quickly and efficiently.

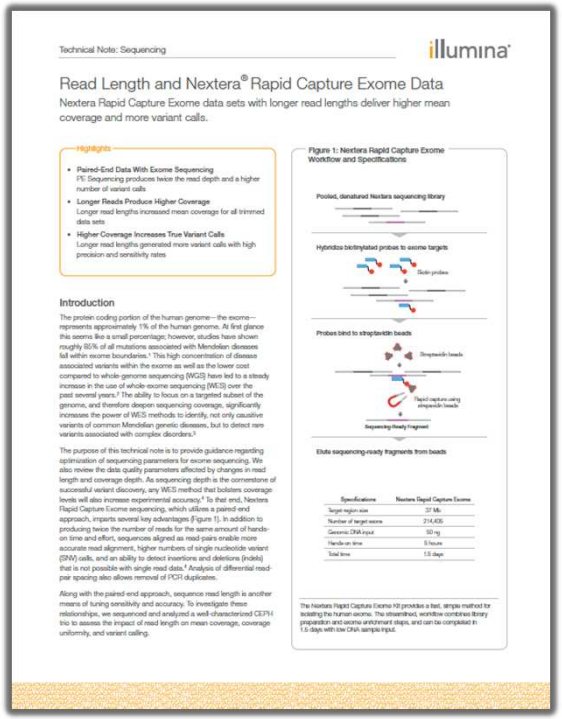
A Simple, Efficient Exome Sequencing Workflow

The NextSeq 500 exome sequencing workflow simplifies exome sequencing, enabling researchers to maximize their productivity (Figure 1). It begins with library preparation and exome enrichment using a Nextera® Rapid Capture Exome Kit. The NextSeq 500 System's dual sequencing output modes enable researchers to scale their exome studies (20, 40, 60, or 120 Gb per run).

What is a "true coding variant"?

A true coding variant is an accurate base call that differs from the consensus sequence within a coding region. It is not a false positive (where a variant is called but does not truly exist) or a false negative (where a variant that truly exists is not called). A system with a high false positive call rate requires extensive downstream validation, increasing costs and experimental time. A system with a high false negative call rate is failing to detect potentially important findings, often in regions that are highly repetitive or that contain homopolymer stretches.

The NextSeq system leverages Illumina industry-leading sequencing technology to deliver the highest number of true coding variants.


Technical Note: Sequencing

illumina

Read Length and Nextera® Rapid Capture Exome Data

Nextera Rapid Capture Exome data sets with longer read lengths deliver higher mean coverage and more variant calls.

Highlights

- Paired-End Data With Exome Sequencing
  PE Sequencing produces twice the read depth and a higher number of variant calls
- Longer Reads Produce Higher Coverage
  Longer read lengths increased mean coverage for all trimmed data sets
- Higher Coverage Increases True Variant Calls
  Longer read lengths generated more variant calls with high precision and sensitivity rates

Figure 1: Nextera Rapid Capture Exome Workflow and Specifications

Specifications	Nextera Rapid Capture Exome
Target region size	37 Mb
Number of target exons	214,405
Genomic DNA input	50 ng
Hands-on time	5 hours
Total time	1.5 days

The Nextera Rapid Capture Exome Kit provides a fast, simple method for isolating the human exome. The streamlined, workflow combines library preparation and exome enrichment steps, and can be completed in 1.5 days, with low DNA sample input.

Introduction

The protein coding portion of the human genome—the exome—represents approximately 1% of the human genome. At first glance this seems like a small percentage; however, studies have shown roughly 85% of all mutations associated with Mendelian diseases fall within exome boundaries.[1] This high concentration of disease associated variants within the exome as well as the lower cost compared to whole-genome sequencing (WGS) have led to a steady increase in the use of whole-exome sequencing (WES) over the past several years.[2] The ability to focus on a targeted subset of the genome, and therefore deepen sequencing coverage, significantly increases the power of WES methods to identify, not only causative variants of common Mendelian genetic diseases, but to detect rare variants associated with complex disorders.[3]

The purpose of this technical note is to provide guidance regarding optimization of sequencing parameters for exome sequencing. We also review the data quality parameters affected by changes in read length and coverage depth. As sequencing depth is the cornerstone of successful variant discovery, any WES method that bolsters coverage levels will also increase experimental accuracy.[4] To that end, Nextera Rapid Capture Exome sequencing, which utilizes a paired-end approach, imparts several key advantages (Figure 1). In addition to producing twice the number of reads for the same amount of hands-on time and effort, sequences aligned as read-pairs enable more accurate read alignment, higher numbers of single nucleotide variant (SNV) calls, and an ability to detect insertions and deletions (indels) that is not possible with single read data.[5] Analysis of differential read-pair spacing also allows removal of PCR duplicates.

Along with the paired-end approach, sequence read length is another means of tuning sensitivity and accuracy. To investigate these relationships, we sequenced and analyzed a well-characterized CEPH trio to assess the impact of read length on mean coverage, coverage uniformity, and variant calling.

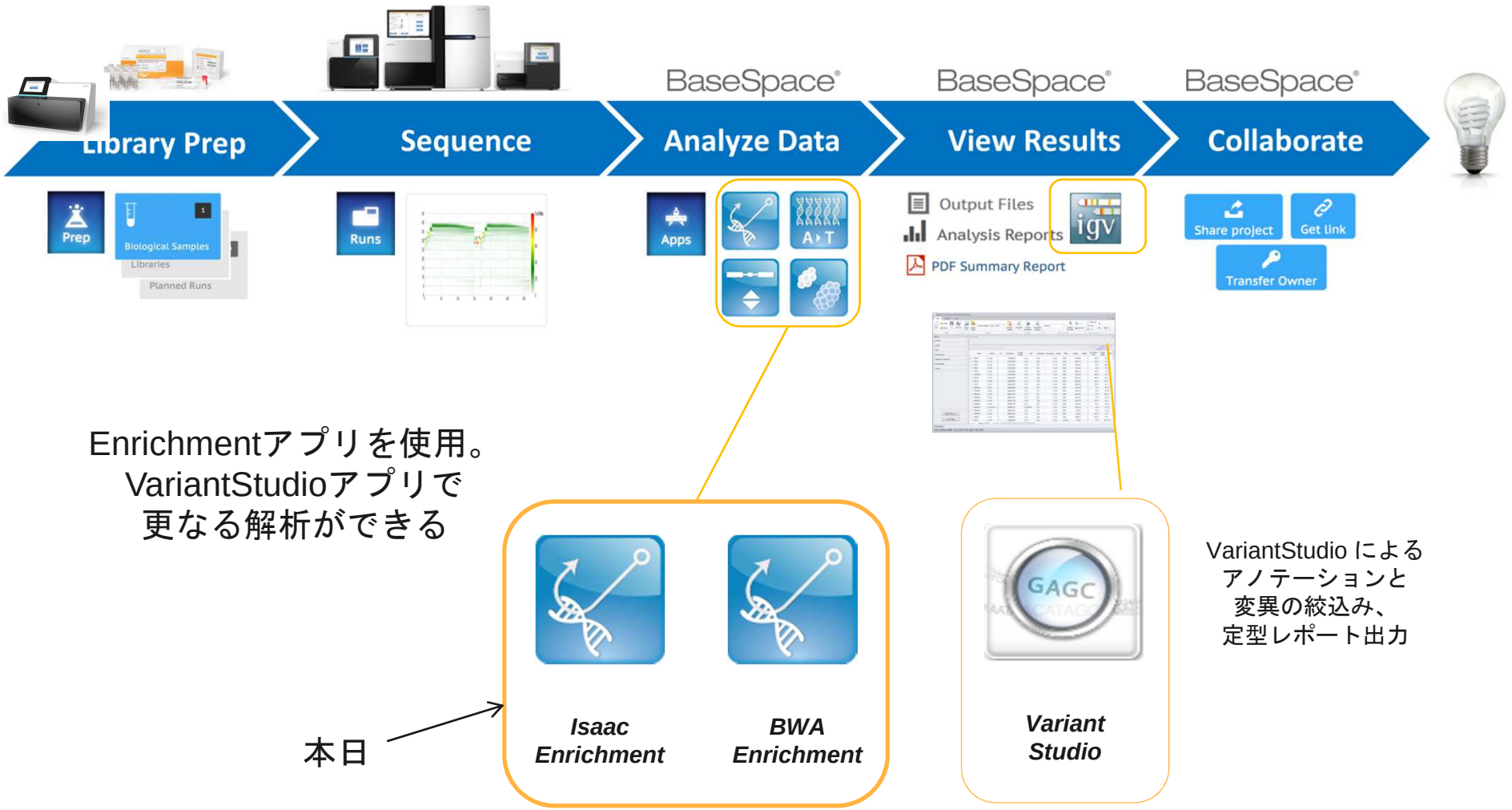
http://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-nrc-exome-read-length.pdf
http://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote-nextseq-exome.pdf
http://www.illumina.com/documents/products/datasheets/datasheet_nextera_rapid_capture_exome.pdf

illumina

本日の内容

- エクソームシーケンス解析 概要
- BaseSpace のエクソーム解析 アプリ
- BaseSpace 実行例 （デモ）
  - BaseSpaceでNetera Rapid Captureエクソームのデモデータを取り込む
  - BaseSpaceでEnrichmentアプリの実行結果を読む
  - BaseSpaceでEnrichmentアプリを実行する

BaseSpace による エクソーム解析



illumina

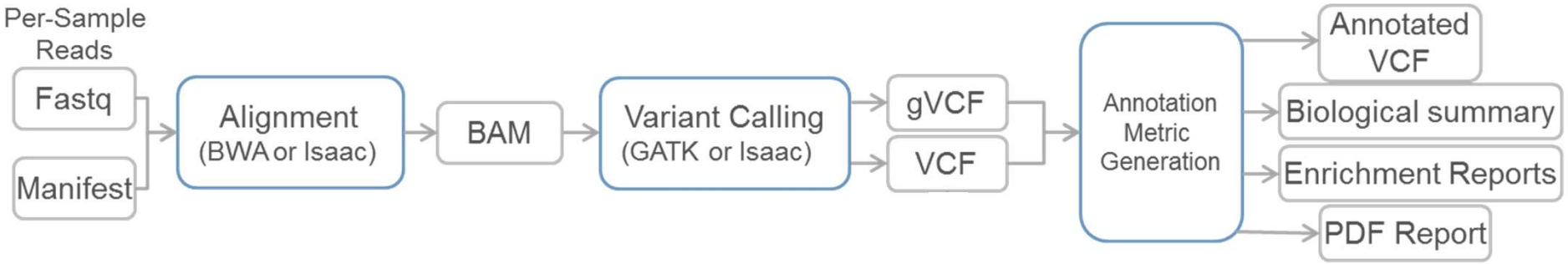
BaseSpace Enrichment アプリ

Biological Application	App Name	App Icon	App Description
Exome	BWA Enrichment		• BWA によるアライメントとGATK による変異コール 業界での使用実績の累積がある • Nextera Rapid Capture 製品と TruSight パネル※向け • v2からカスタムマニフェストに対応 • 濃縮に関するメトリクスの計算 • カバレッジデプスやフラグメント長等 • SNPやsmall Indelsの検出とその 内訳
	Isaac Enrichment (アイザック)		• ISAAC(イルミナ製のプログラム)によるアライメントと変異コール bwa/gatk と同程度の感度と特異度で５倍程度速い • Nextera Rapid Capture 製品と TruSight パネル※向け • v2からカスタムマニフェストに対応 • 濃縮に関するメトリクスの計算 • カバレッジデプスやフラグメント長等 • SNPやsmall Indelsの検出とその内訳

※一部他の専用アプリを使用するものもございます

・２種類をどちらもフリーで何回でも使用可能
・hg19のみ対応

エクソーム解析のワークフロー

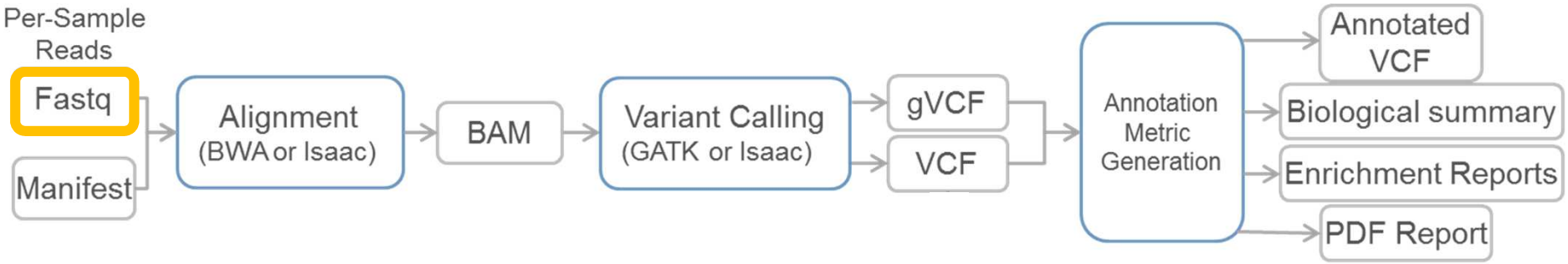


illumina®

BaseSpace エクソーム解析
入力 – FASTQ

Isaac Enrichment

BWA Enrichment




```
@HWI-BRUNOP20X:994:B809UWABXX:1:1101:13501:2240 1:N:0:CTTGTA
TGAAACCAGTGTTCTTAATTGGCATTTTACACACACACACACAGAATTTAAAAAAAAAATCAAAGG
+
=55>7;?::BDADDD@EE88DCD?DFFEFFECBE6666BB=B;<;<-34:;<CB51>=BBEE>EE?
@HWI-BRUNOP20X:994:B809UWABXX:1:1101:13660:2247 1:N:0:CTTGTA
CCAAACATTAAGTAACTCTTAAAATGGCACACAGGTTTTAAAGCTATTGGTTTTTCCTTCCTAACT
+
FFEDFBGEGGGGDFGEFFFFGGDF=FBFFFGGGE7CEEDEFBFBFGEEGF@FCDDFDFFEGFEAGF
@HWI-BRUNOP20X:994:B809UWABXX:1:1101:13966:2183 1:N:0:CTTGTA
TTGGGTAACTTGAATATAACATGGCTCCCTTGCTGTAAGCAAATGTTTTAGAGCTGAATTTTTCCT
+
HHHHHEHHHHHHFHHHHHHHHHHHHHHHHHHHGGFHHHHHHHHHHFHHHFHEHHFHEHHHHFHHHF
```

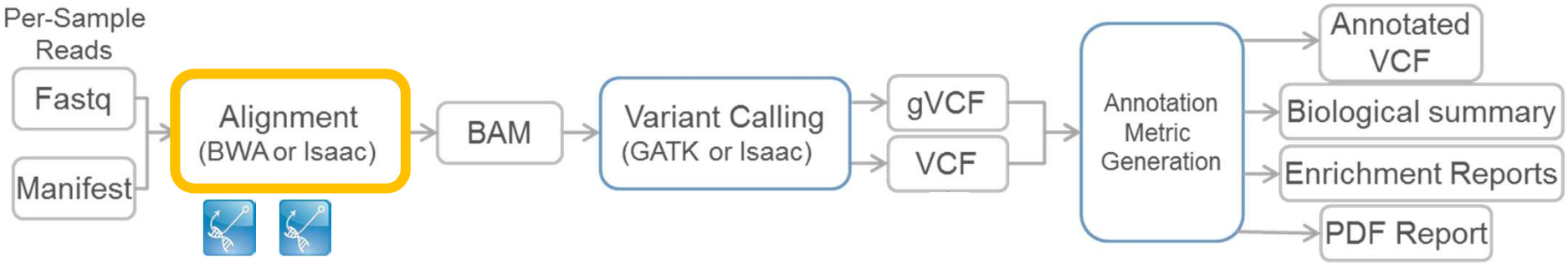

illumina®

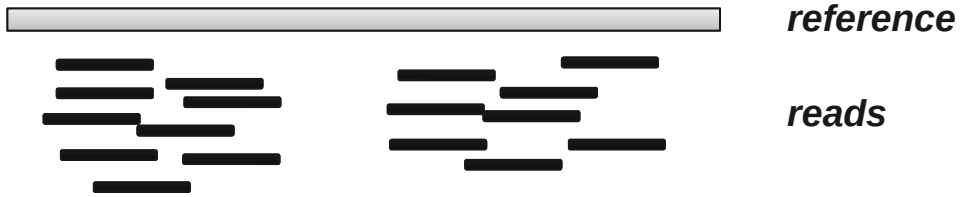
BaseSpace エクソーム解析

解析工程 – アライメント

Isaac Enrichment

BWA Enrichment





reference

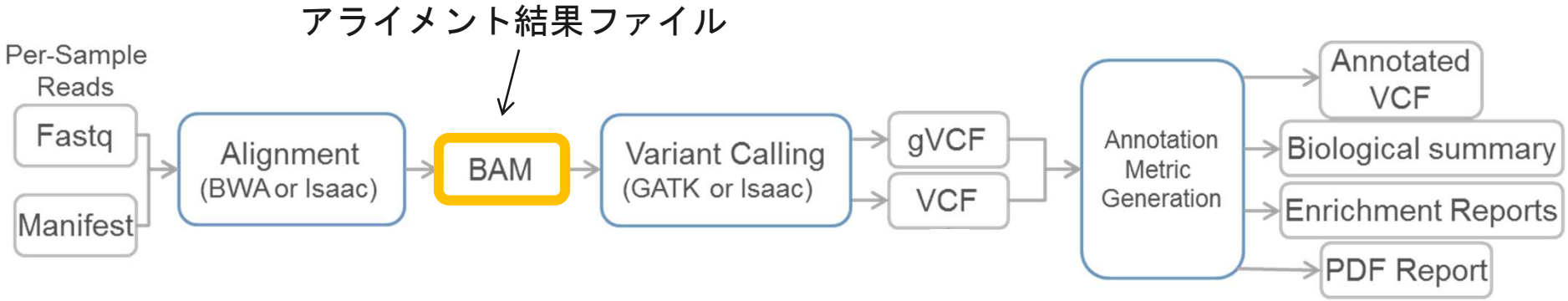
reads

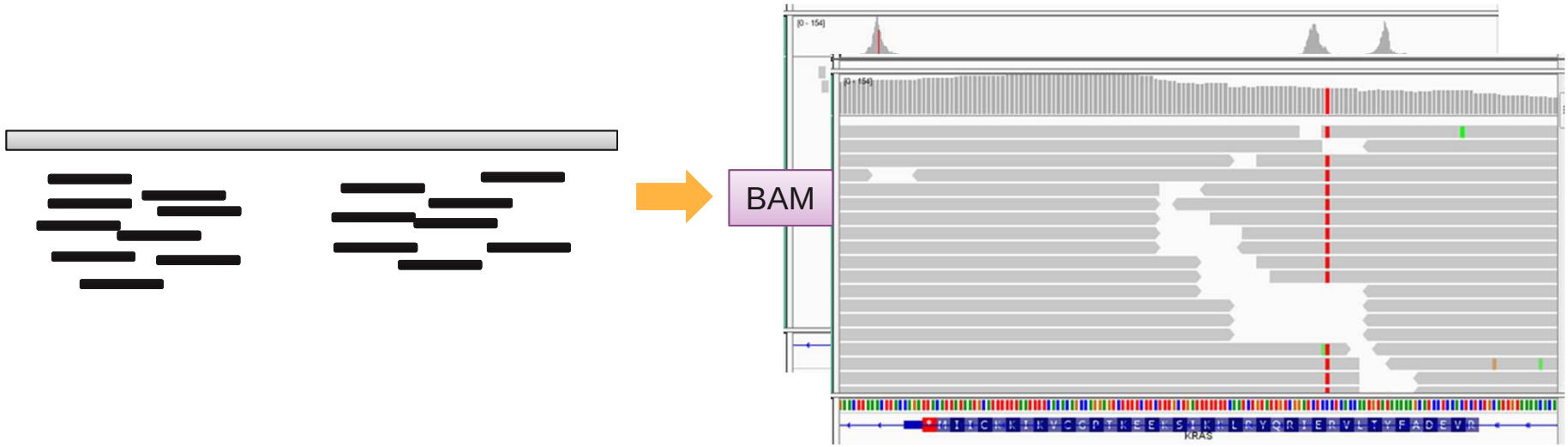
illumina®

BaseSpace エクソーム解析

結果 – アライメント – BAMファイル

Isaac Enrichment ***BWA Enrichment***





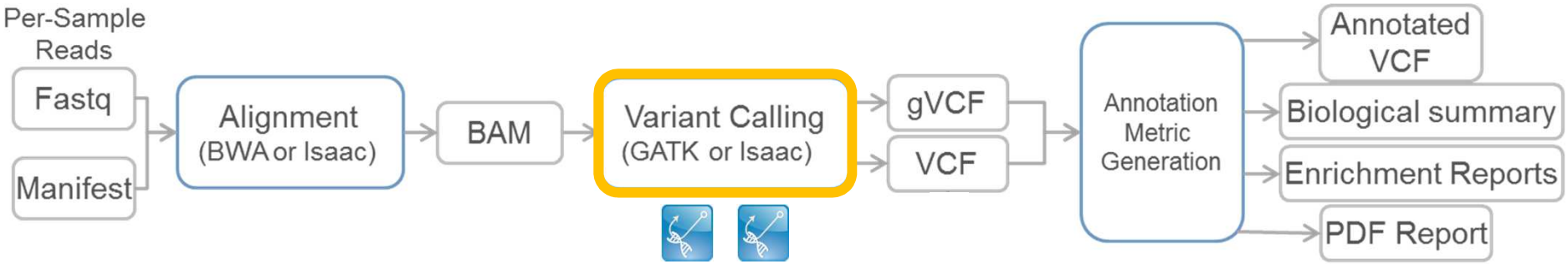
IGVでアライメント結果bamファイルを可視化

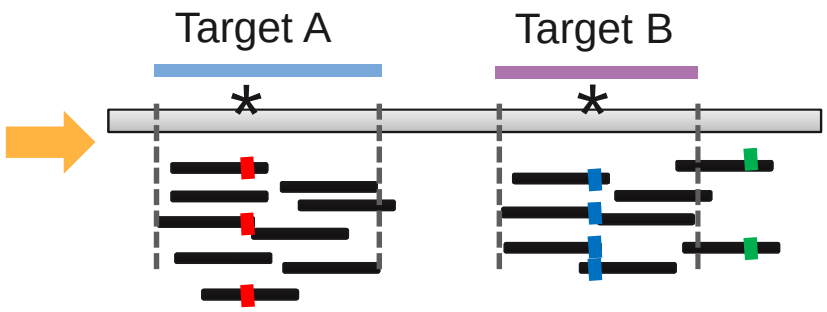
BaseSpace エクソーム解析

解析工程 – 変異コール

Isaac Enrichment

BWA Enrichment





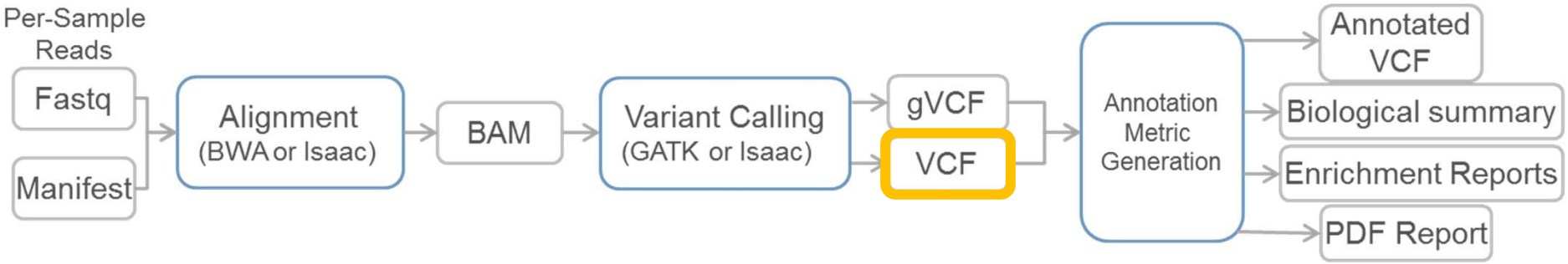
illumina®

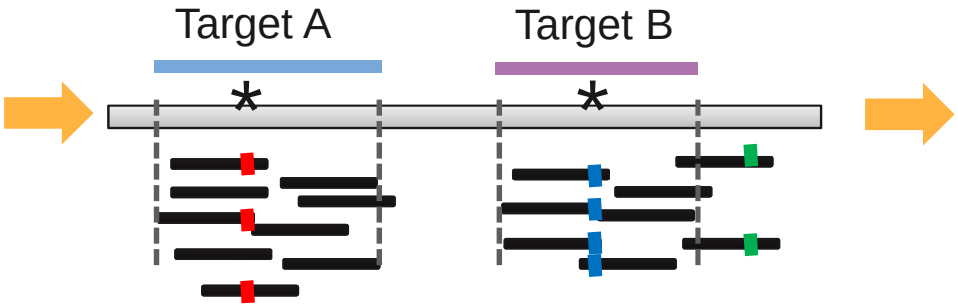
BaseSpace エクソーム解析

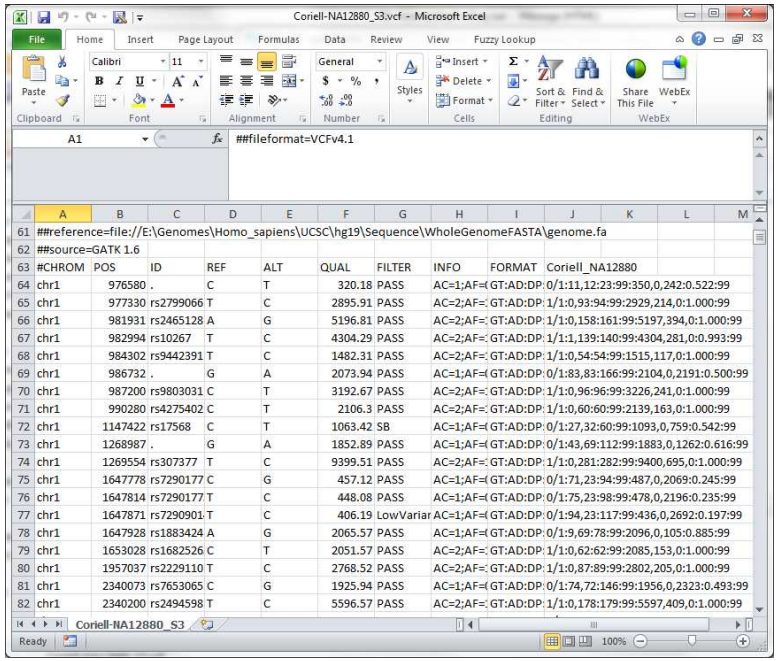
結果 – 変異コール – VCF

Isaac Enrichment

BWA Enrichment





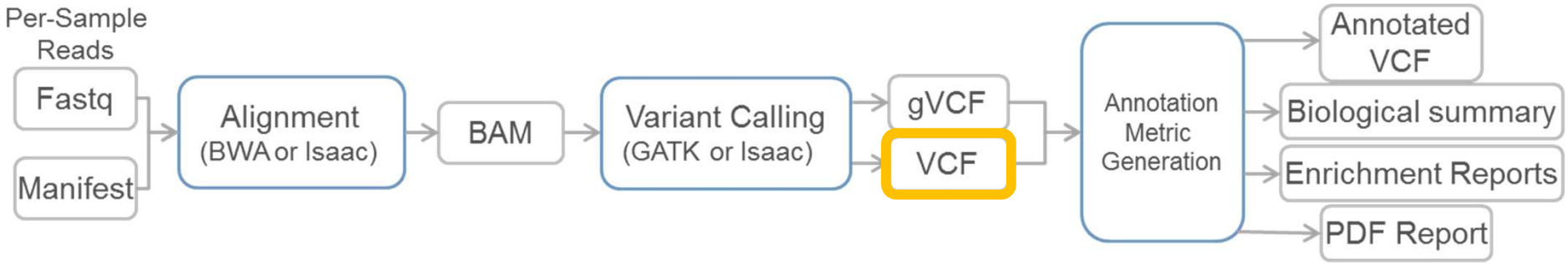


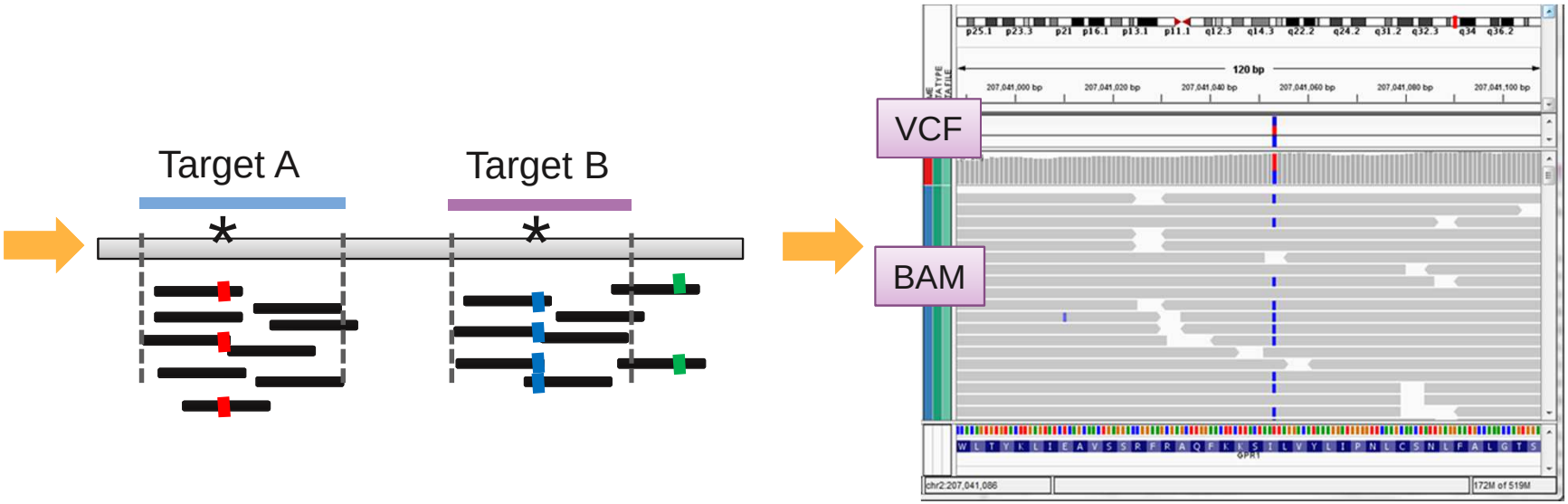
illumina®

BaseSpace エクソーム解析

結果 – 変異コール – VCF

Isaac Enrichment ***BWA Enrichment***





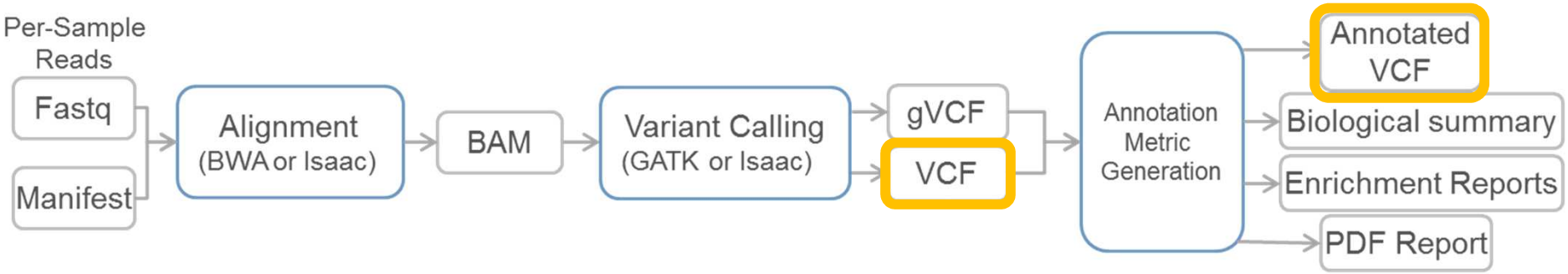
illumina®

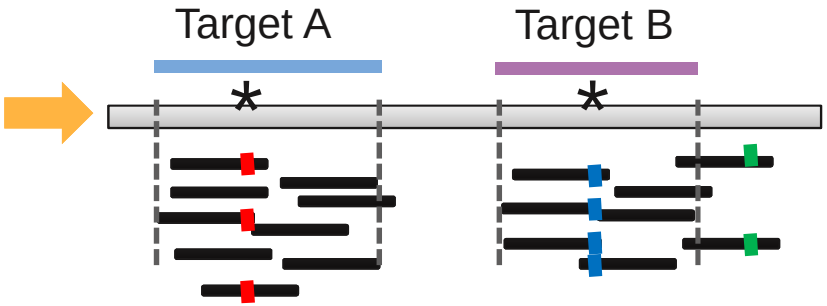
BaseSpace エクソーム解析

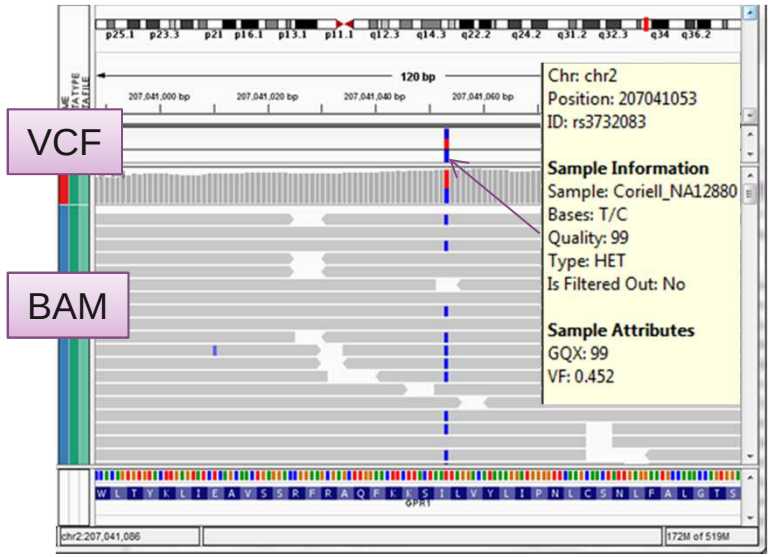
結果 – 変異コール – VCF

Isaac Enrichment

BWA Enrichment



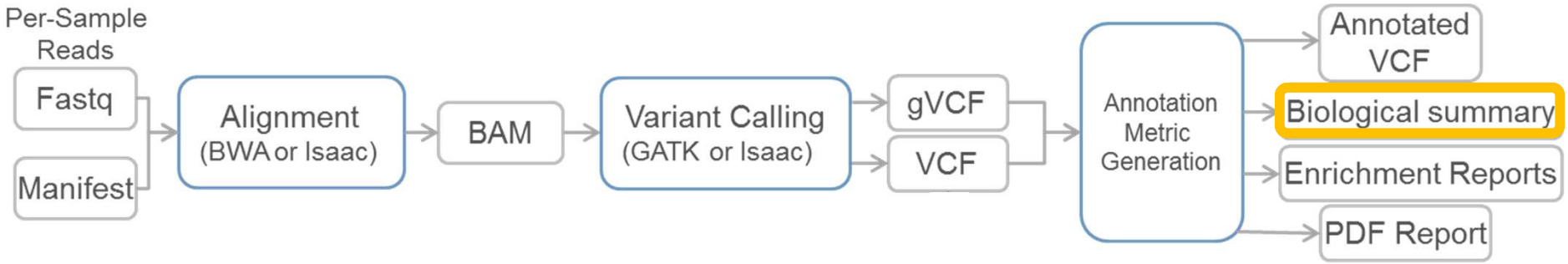


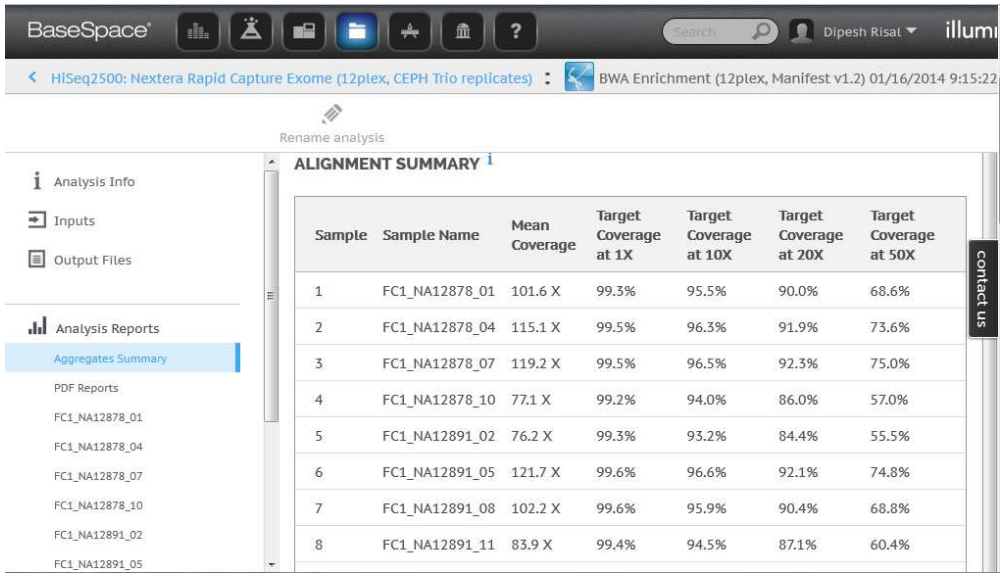


illumina®

BaseSpace エクソーム解析

結果 – Biological サマリ レポート





Sample	Sample Name	Mean Coverage	Target Coverage at 1X	Target Coverage at 10X	Target Coverage at 20X	Target Coverage at 50X
1	FC1_NA12878_01	101.6 X	99.3%	95.5%	90.0%	68.6%
2	FC1_NA12878_04	115.1 X	99.5%	96.3%	91.9%	73.6%
3	FC1_NA12878_07	119.2 X	99.5%	96.5%	92.3%	75.0%
4	FC1_NA12878_10	77.1 X	99.2%	94.0%	86.0%	57.0%
5	FC1_NA12891_02	76.2 X	99.3%	93.2%	84.4%	55.5%
6	FC1_NA12891_05	121.7 X	99.6%	96.6%	92.1%	74.8%
7	FC1_NA12891_08	102.2 X	99.6%	95.9%	90.4%	68.8%
8	FC1_NA12891_11	83.9 X	99.4%	94.5%	87.1%	60.4%

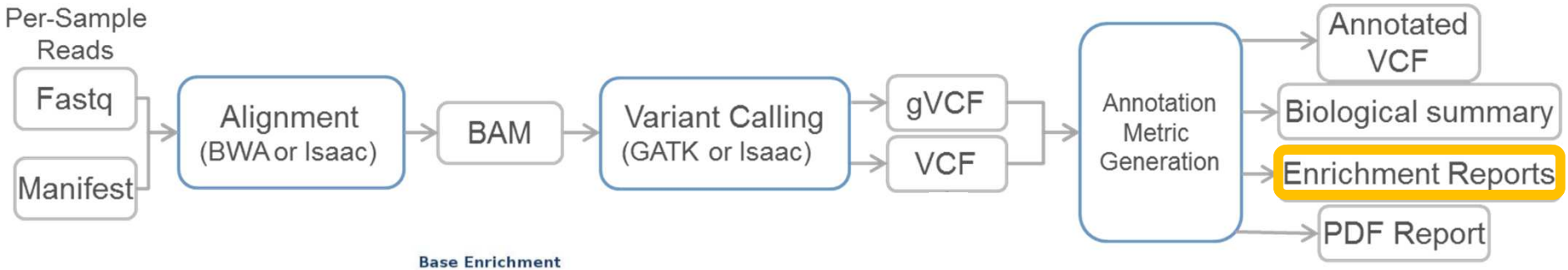
Variants by Sequence Context

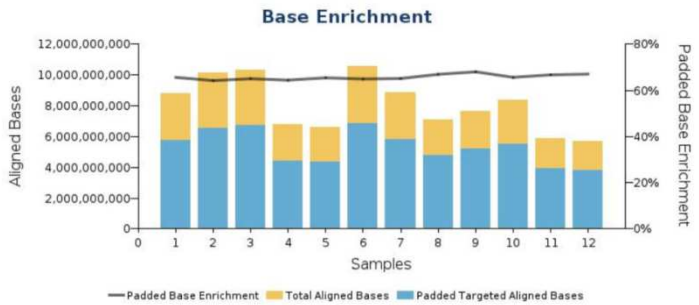
	SNVs	Insertions	Deletions
Number in Genes	32,157	1,046	1,228
Number in Exons	22,949	305	295
Number in Coding Regions	21,641	216	205
Number in UTR Regions	1,308	89	90
Number in Splice Site Regions	1,310	99	152

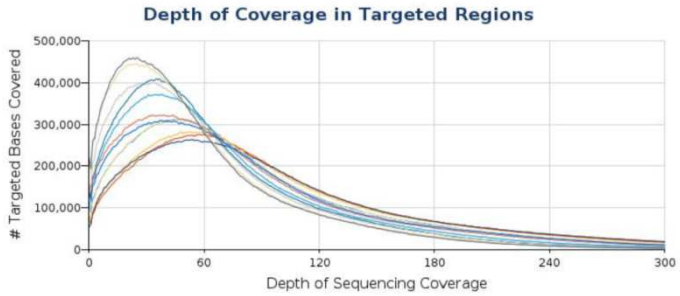
Variants by Consequence

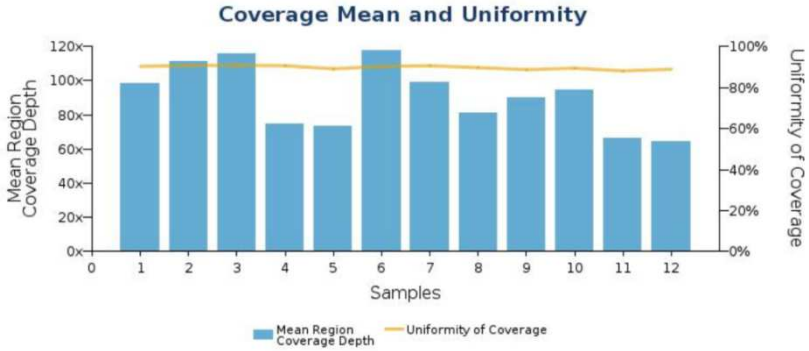
	SNVs	Insertions	Deletions
Frameshift	-	104	82
Non-synonymous	10,464	112	122
Synonymous	11,076	-	-
Stop Gained	76	0	0
Stop Lost	24	0	0

BaseSpace エクソーム解析
結果 – Enrichment　レポート

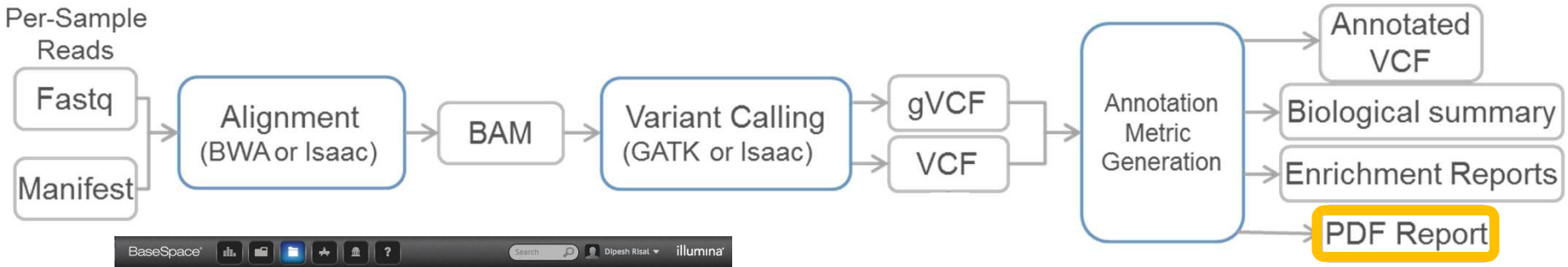


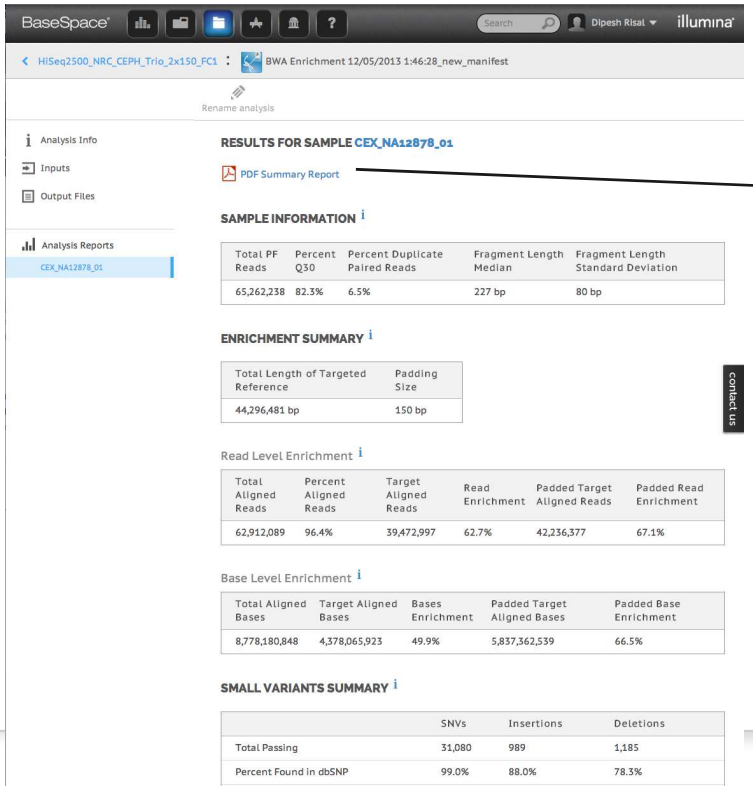






BaseSpace エクソーム解析
結果ファイル – PDF レポート



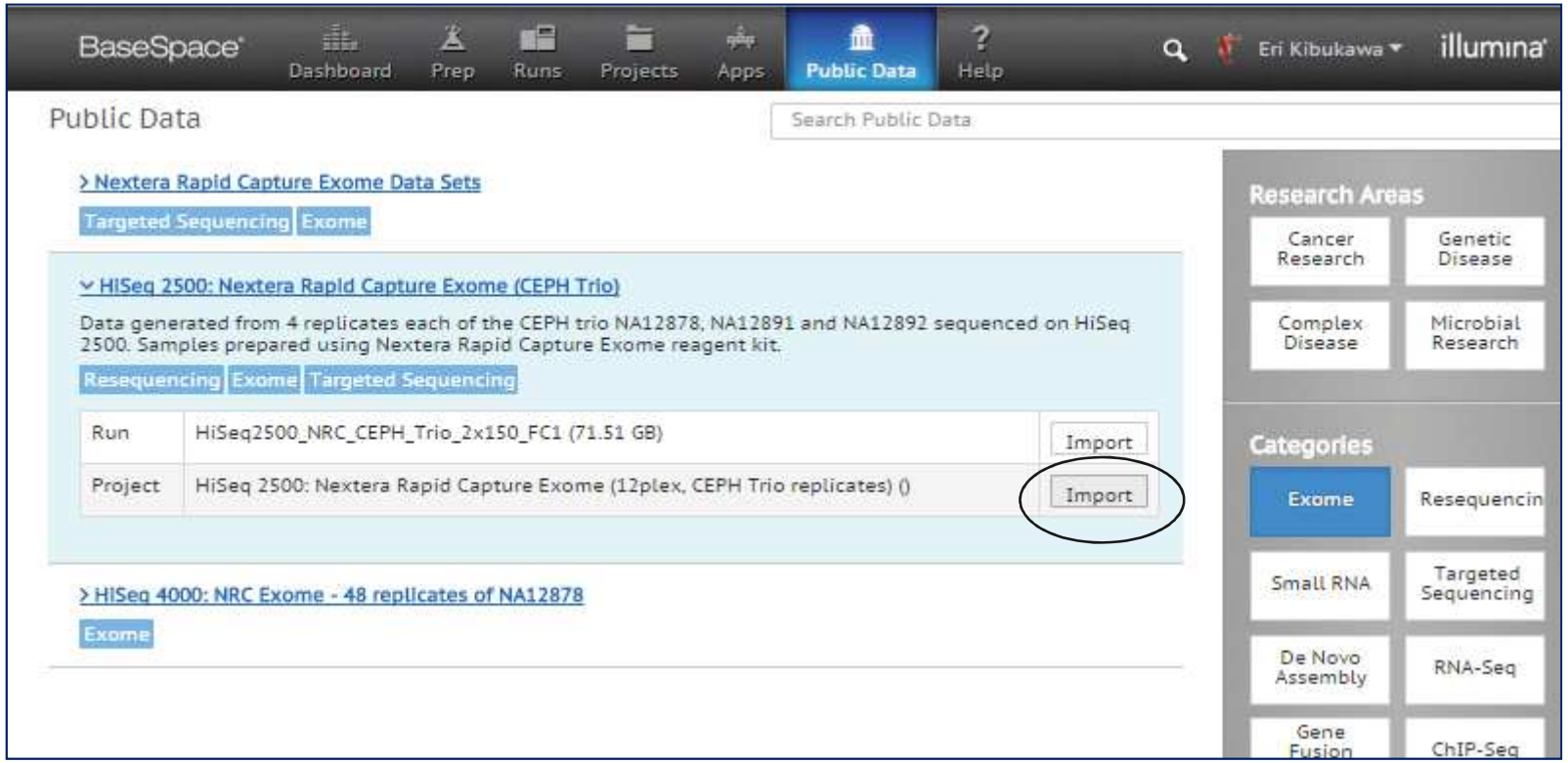


illumina

本日の内容

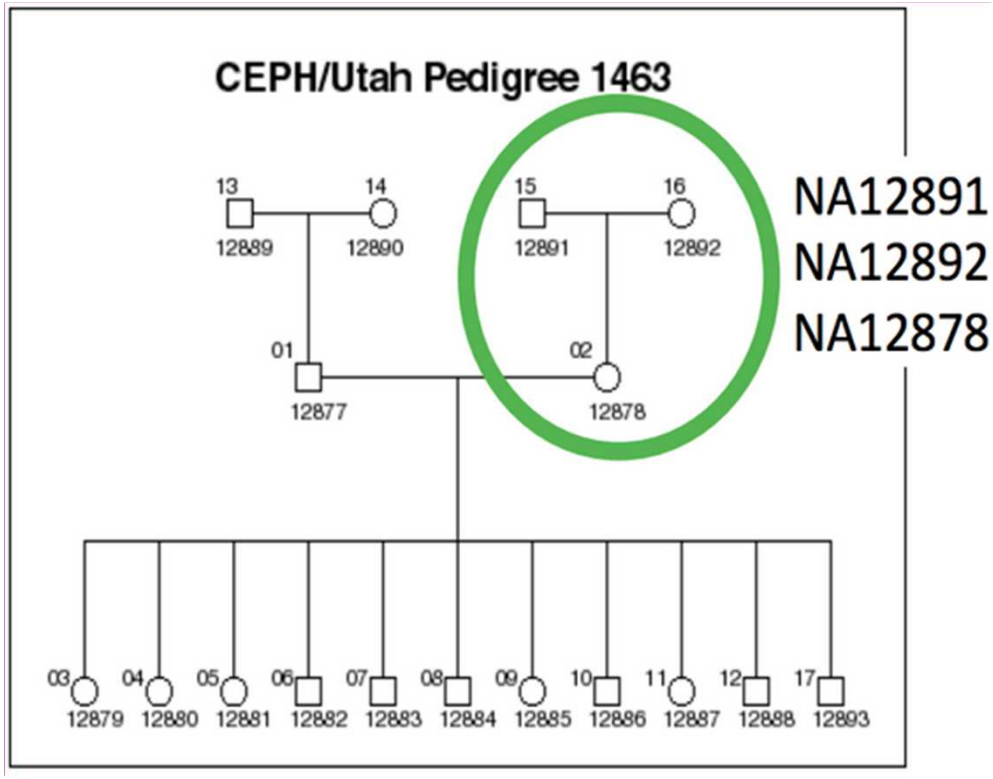
- エクソームシーケンス解析 概要
- BaseSpace のエクソーム解析 アプリ
- BaseSpace 実行例 （デモ）
  - BaseSpaceでNetera Rapid Captureエクソームのデモデータを取り込む
  - BaseSpaceでEnrichmentアプリの実行結果を読む
  - BaseSpaceでEnrichmentアプリを実行する

本日のデモデータ: Public Dataから Nextera Rapid Capture Exomeのデータ解析結果をみる



あるいは以下シェアリンクURLからImport
https://basespace.illumina.com/s/xs2HUjmXiRCk

本日のデモデータ: BaseSpace にある CEPH Trio(母父子) データによるエクソーム解析



デモデータ情報:

- Nextera rapid capture エクソームを使用
- 各4レプリケート
- 150bp のペアードエンド
- HiSeq 2500でシーケンス

Coriell	catalog.coriell.org/0/sections/Collections/NIGMS/CEPHResources.aspx
HapMap	www.hapmap.org
1Kゲノム	www.1000genomes.org
プラチナゲノム	www.illumina.com/platinumgenomes

デモ (動画からご参考下さい)

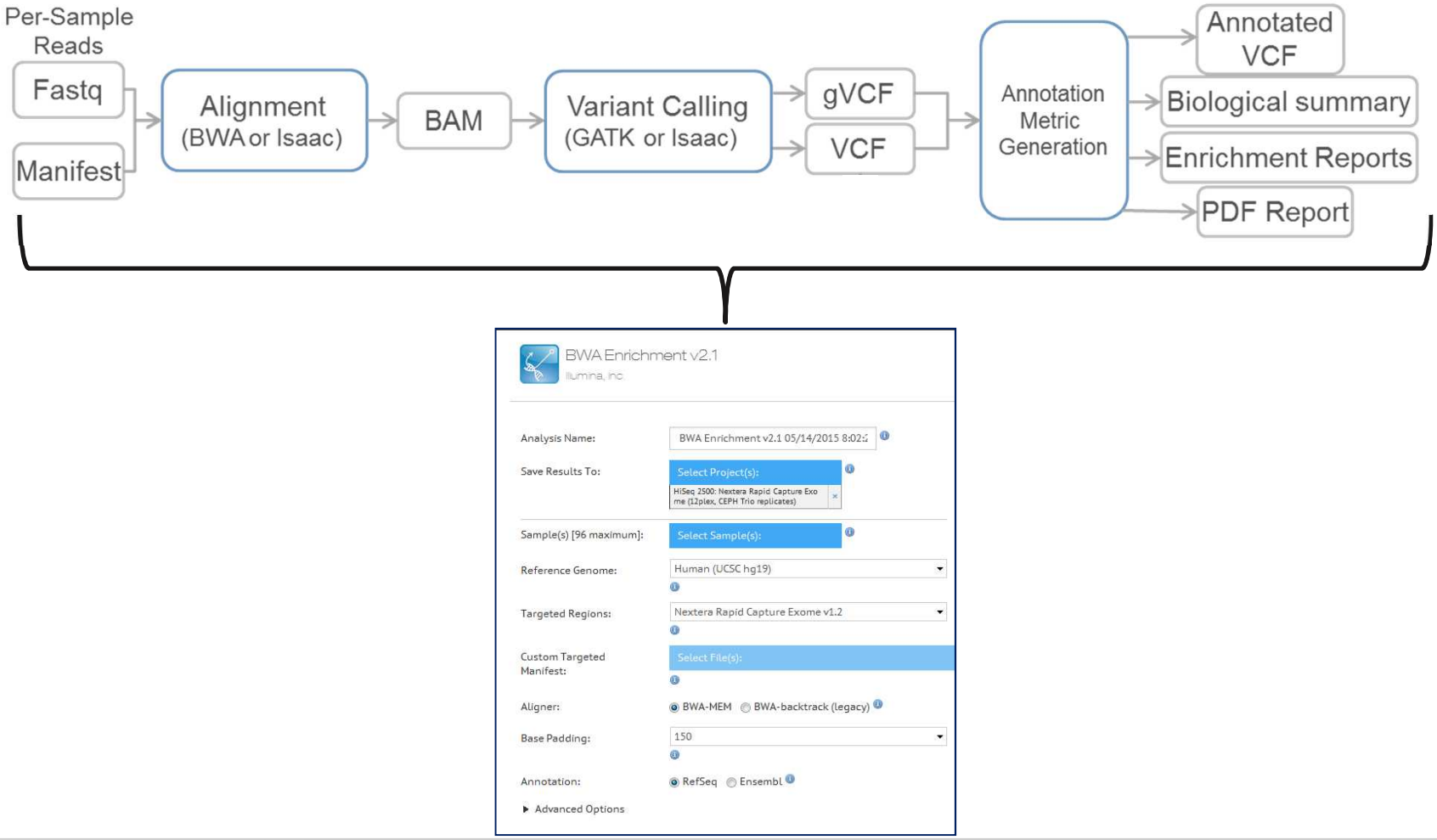
BaseSpaceでNetera Rapid Captureエクソームのデモデータを取込む
BaseSpaceでEnrichmentアプリの実行結果を読む
BaseSpaceでEnrichmentアプリを実行する

https://www.youtube.com/watch?v=twYrwlLXHzw

BaseSpace App によるエクソーム解析

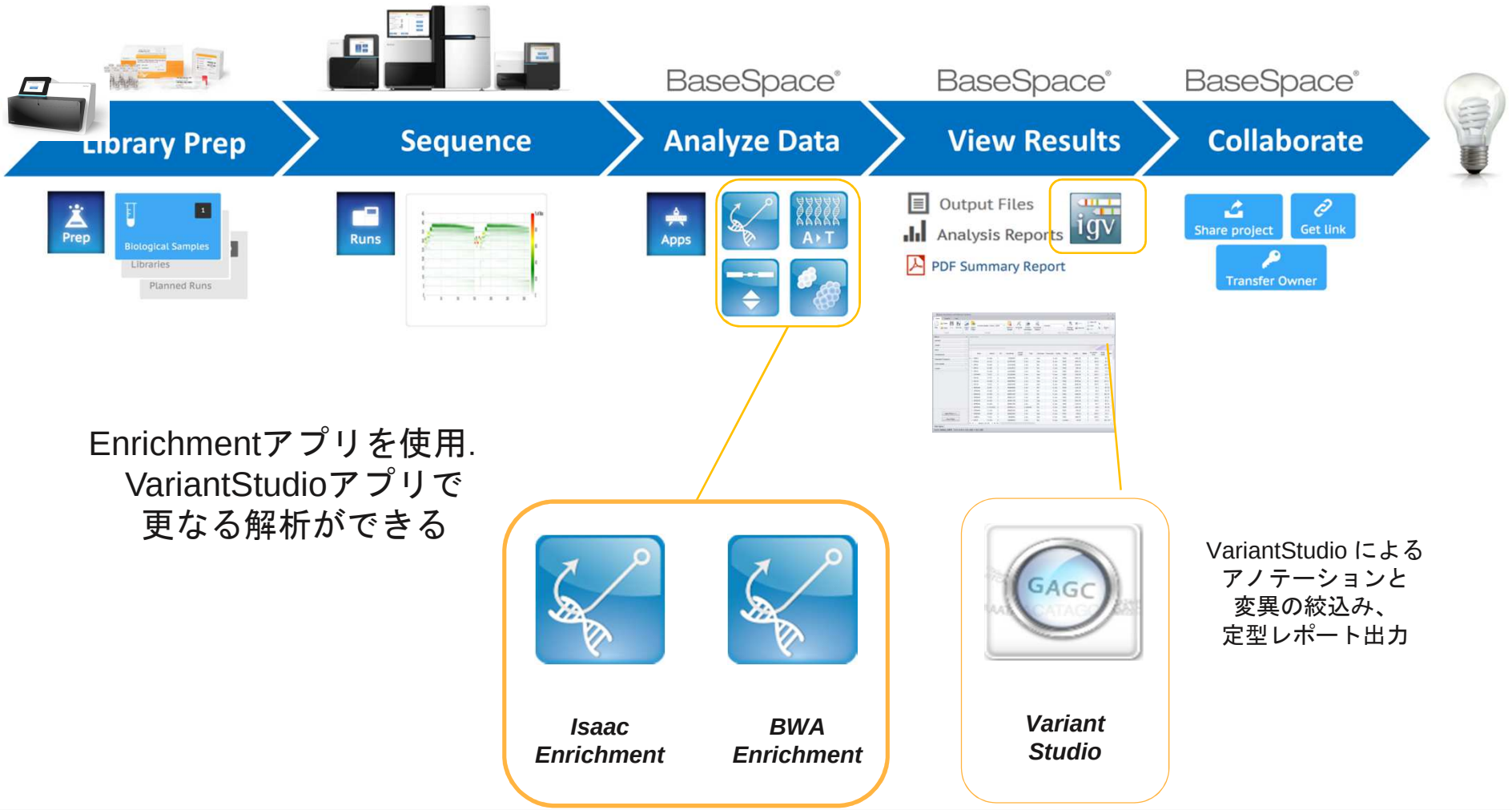
BWA Enrichment

Isaac Enrichment



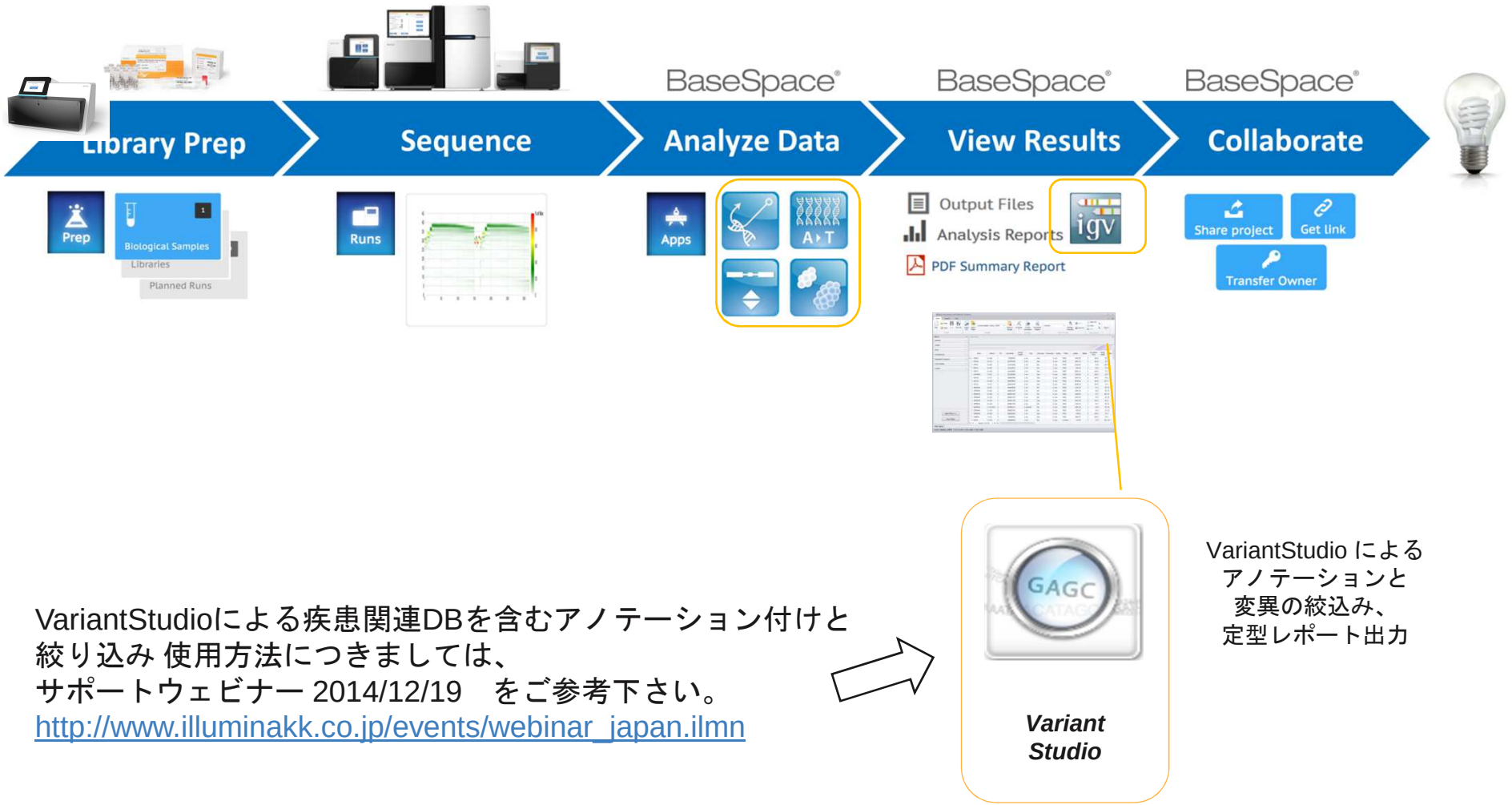
illumina®

BaseSpace による エクソーム 解析



illumina

BaseSpace による エクソーム解析



VariantStudioによる疾患関連DBを含むアノテーション付けと絞り込み 使用方法につきましては、
サポートウェビナー 2014/12/19　をご参考下さい。
http://www.illuminakk.co.jp/events/webinar_japan.ilmn

VariantStudio による
アノテーションと
変異の絞込み、
定型レポート出力

illumina

詳細は各アプリの Instruction ガイドをご確認頂けます

BWA Enrichment

Isaac Enrichment

VariantStudio

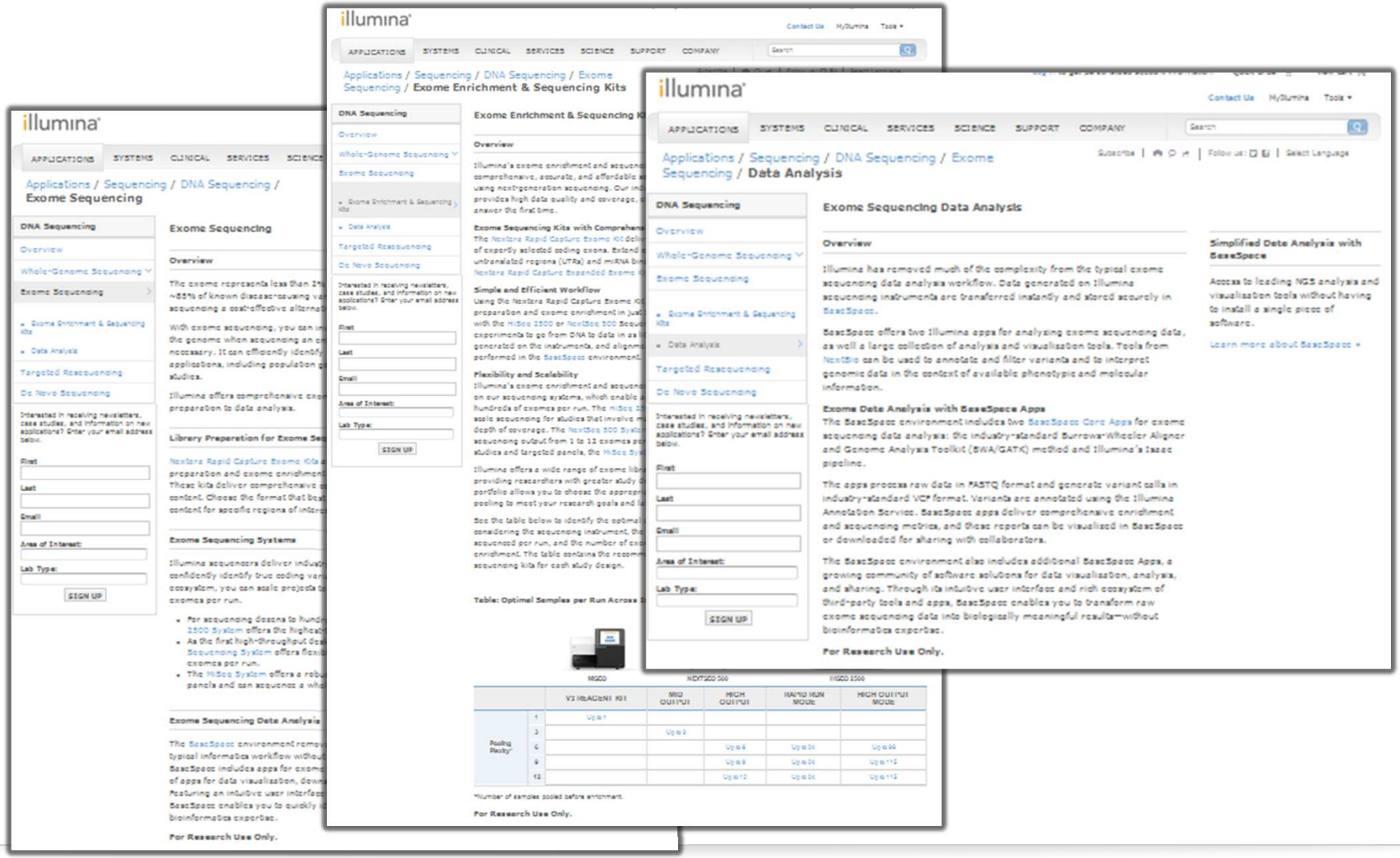
http://support.illumina.com/downloads/basespace_core_apps_user_guides.html
http://support.illumina.com/sequencing/sequencing_software/variantstudio.html

illumina

エクソームシーケンシング web ページ

Appendix.

http://www.illumina.com/applications/sequencing/dna_sequencing/exome-sequencing.html



illumina

ご清聴ありがとうございました

本日セッション終了後のご質問は、
techsupport@illumina.comにお問い合わせください

illumina®