

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)

© 2014 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, GAIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, ForenSeq, MiSeq, MiSeqDx, MiSeqFGx, NeoPrep, Nextera, NextBio, NextSeq, Powered by Illumina, SeqMonitor, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

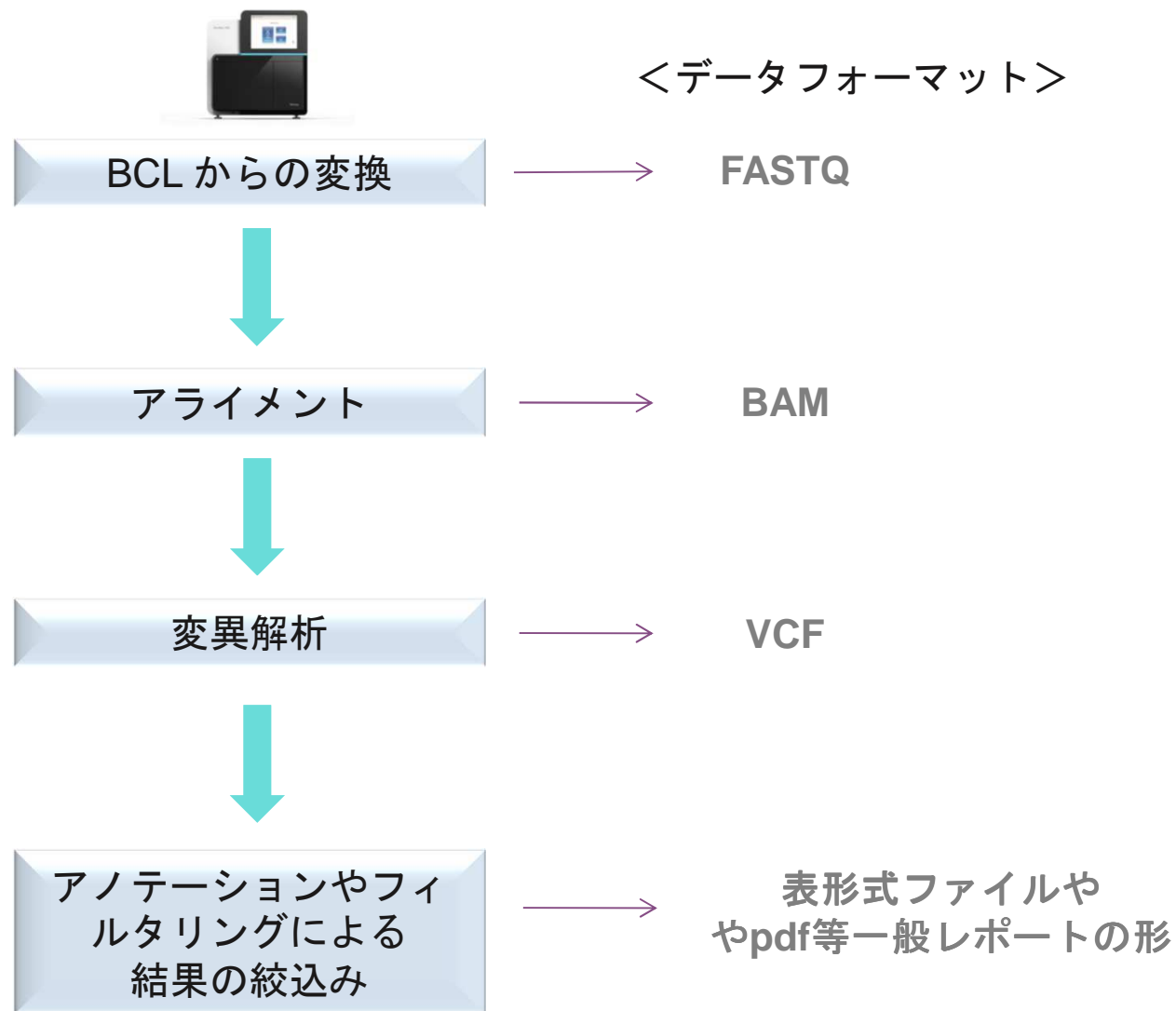
illumina®

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

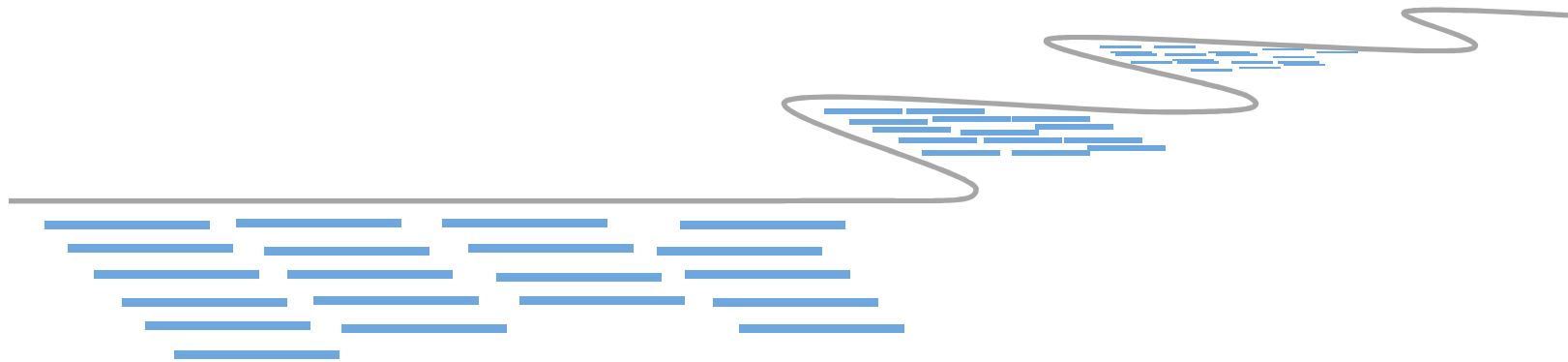


DNA リシーケンシング、エクソーム解析 典型的な解析ワークフロー

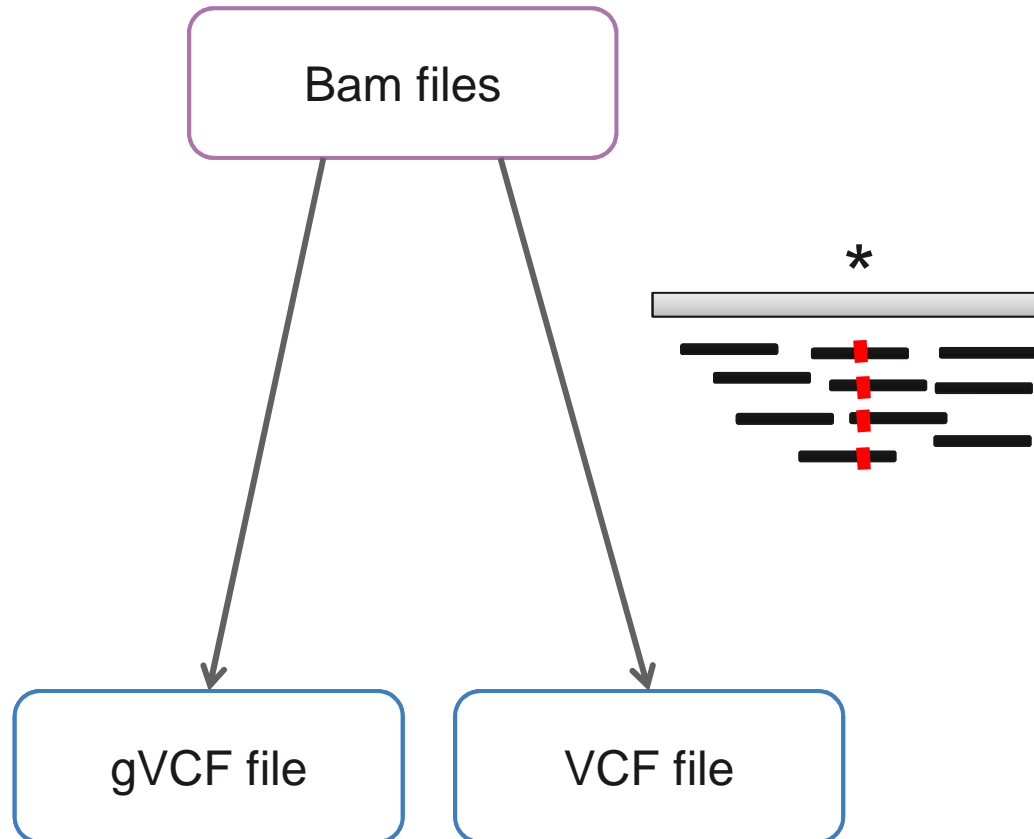


アライメント

- ▶ 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

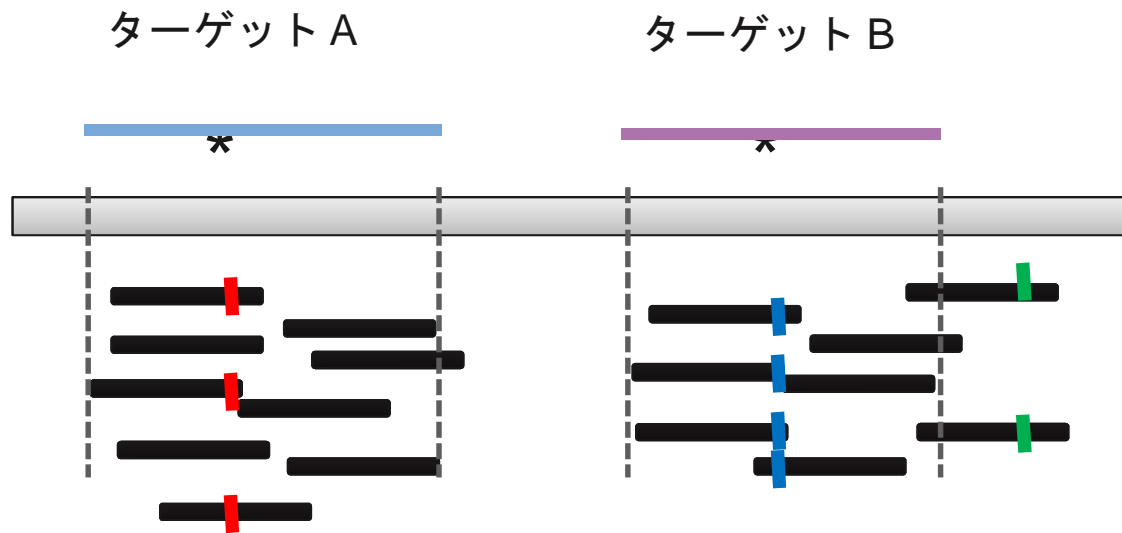


- サンプルとリファレンス間の違いについて検出するための統計的検定
- 主要な検討材料;
 - Quality score
 - Frequency
 - Depth
 - Strand bias
- 結果をVCF出力

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.illumina.com/japan/events/webinar_japan.ilmn

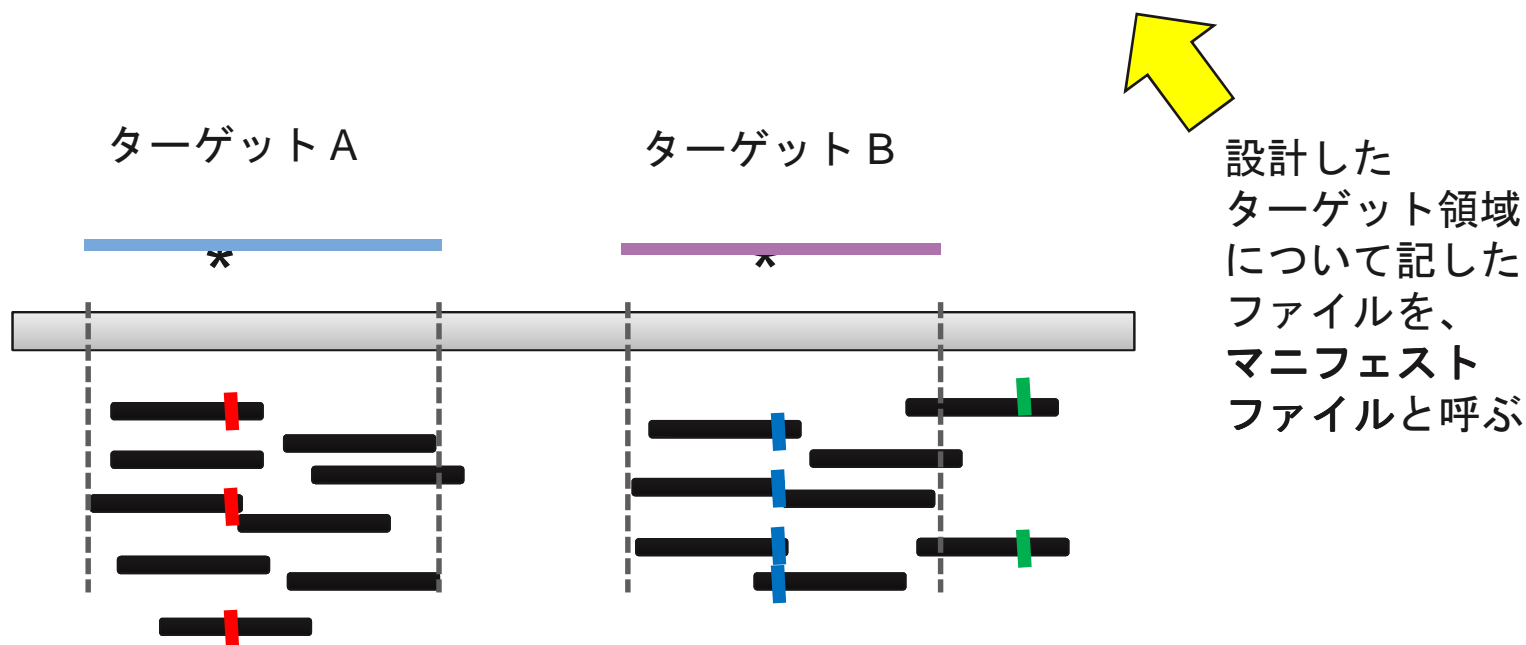
全ゲノム (WGS) とエクソーム?

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



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.illumina.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	

※Manifest 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 Burnina Support**
Burnina scientists and engineers are there every step of the way, providing installation, training, applications, and data analysis support

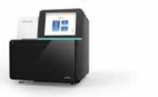
Figure 1: NextSeq 500 System Sample-to-Answer Exome Sequencing Workflow

Prep	Sequencing	Analysis	Share
1.5	29	5	1
hours	hours	hours	hours

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 the cloud or on the Burnina premises 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 Burnina data.

Figure 2: NextSeq 500 Sequencing System



The NextSeq 500 System integrates the latest advances in 300x 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 Burnina 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 input mode enables researchers to pool their exome studies (20, 40, 60, or 120 Chx 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 Burnina 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 provides better read depth and a higher number of variant calls
- **Longer Reads Produce Higher Coverage**
Longer read lengths increased mean coverage for all inbred data sets
- **Higher Coverage Increases True Variant Calls**
Longer read lengths generated more variant calls with high precision and sensitivity rates

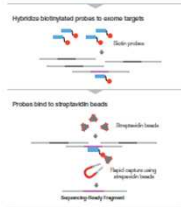
Introduction

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 80% of all mutations associated with Mendelian diseases fall within exome boundaries.¹ The 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 decrease 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 balances 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 CEP110 to assess the impact of read length on mean coverage, coverage uniformity, and variant calling.

Figure 1: Nextera Rapid Capture Exome Workflow and Specifications



The workflow includes: Project, deconvoluted Nextera sequencing library; Hybridize biotinylated probes to exome targets; Wash away non-bound probes; Probes bind to oligonucleotide beads; Wash away unbound beads; Rapid capture using oligonucleotide beads; Sequencing library fragments from beads; Elute sequencing-ready fragments from beads.

Specification	Nextera Rapid Capture Exome
Target region size	27 Mb
Number of target genes	214,026
Capture DNA input	50 ng
Hybridize time	5 hours
Total time	1.5 days

The Nextera Rapid Capture Exome V2 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 Chx sample input.

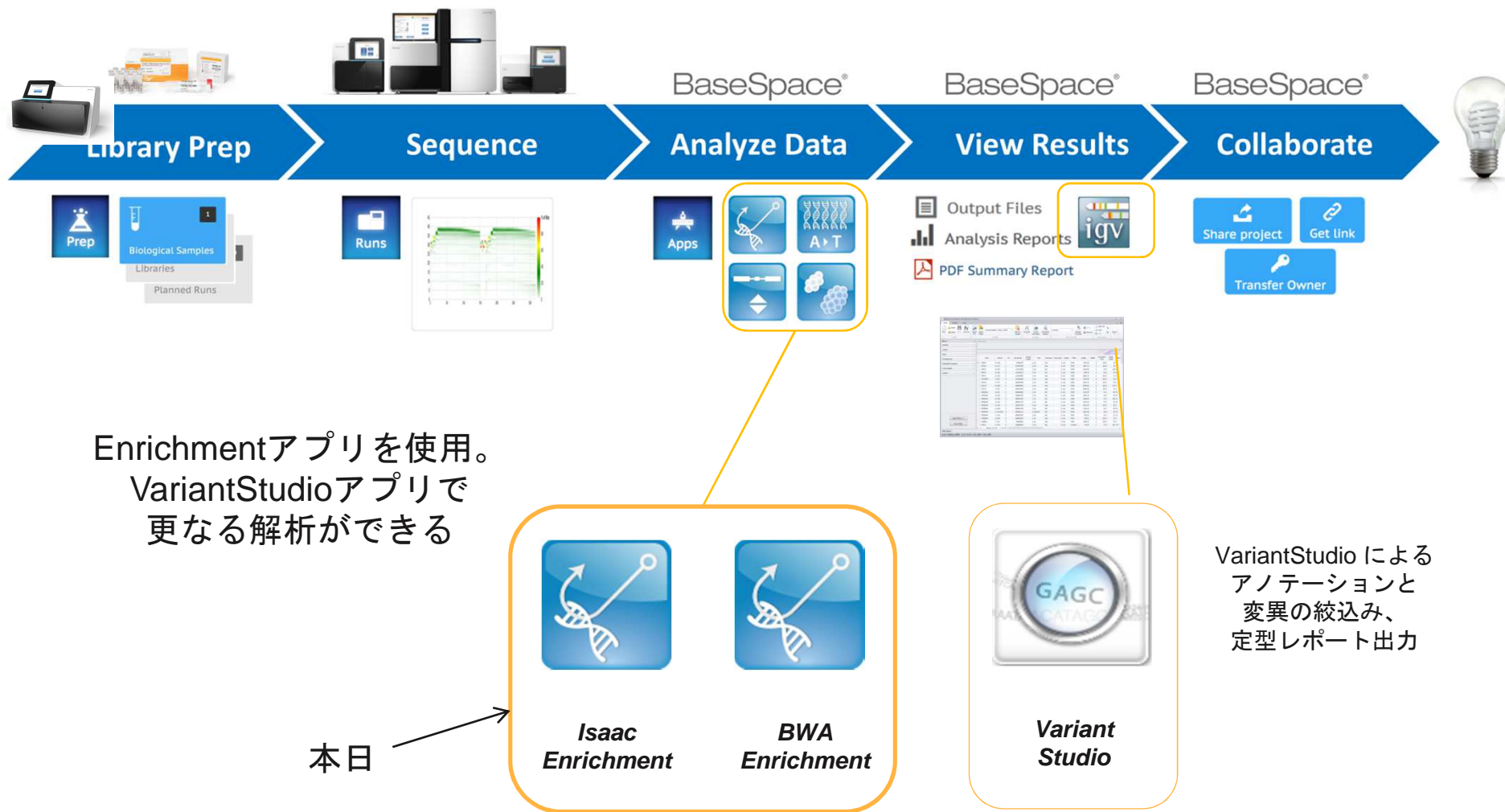
<http://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-nrc-exome-read-length.pdf>
<http://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote-nextseq-exome.pdf>
http://www.illumina.com/documents/products/datasheets/datasheet_nextera_rapid_capture_exome.pdf

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



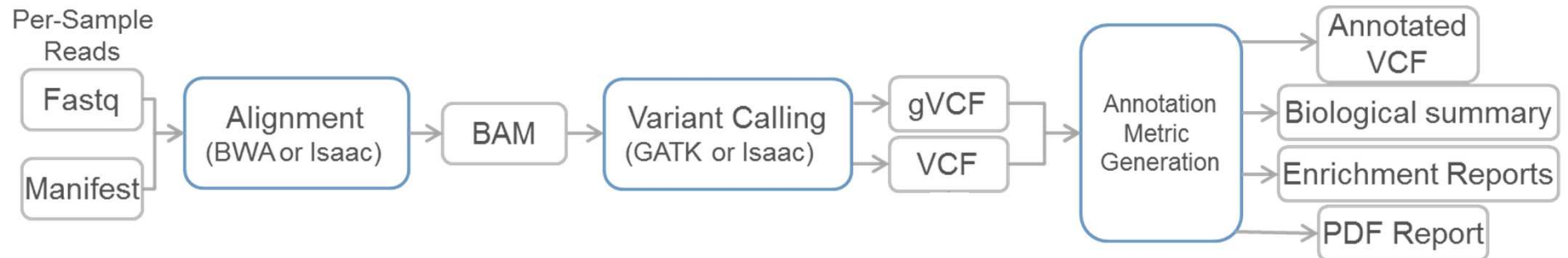
BaseSpace Enrichment アプリ

Biological Application	App Name	App Icon	App Description
Exome	BWA Enrichment		<ul style="list-style-type: none"> • 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)		<ul style="list-style-type: none"> • 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



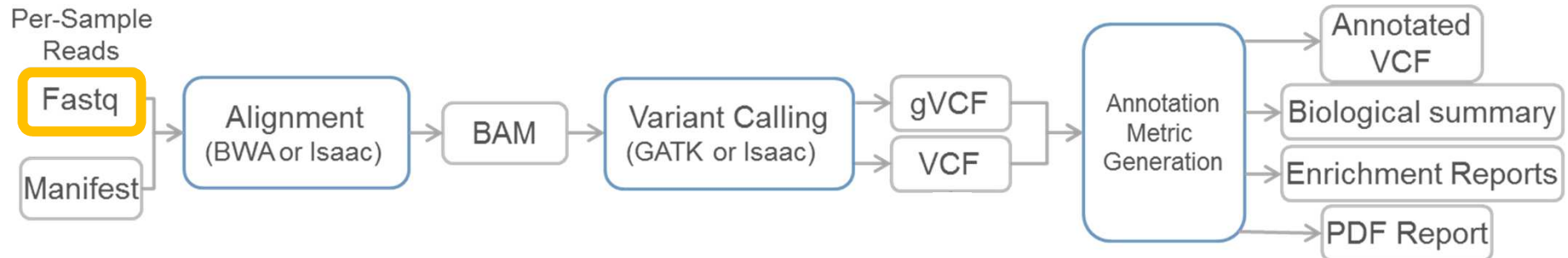
BaseSpace エクソーム解析 入カ - FASTQ



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
    CCAAACATTAAGTAACTCTTAAAATGGCACACAGGTTTTAAAGCTATTGGTTTTTCCTTCTAACT
    +
    FFEDFBGEGGGGDFGEFFFFGGDF=FBFFFGGGE7CEEDEFBFBFGEEGF@FCDDFDFFEGFEAGF
    @HWI-BRUNOP20X:994:B809UWABXX:1:1101:13966:2183 1:N:0:CTTGTA
    TTGGGTA ACTTGAATATAACATGGCTCCCTTGCTGTAAGCAAATGTTTTAGAGCTGAATTTTCT
    +
    HHHHHEHHHHHHFHHHHHHHHHHHHHHHHHHHHGGFHHHHHHHHHHFHHHFHEHHFHEHHHHFHHHF
  
```

BaseSpace エクソーム解析

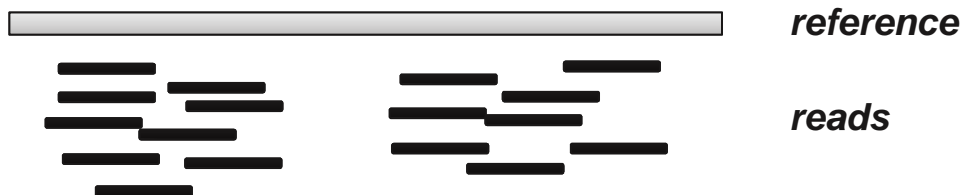
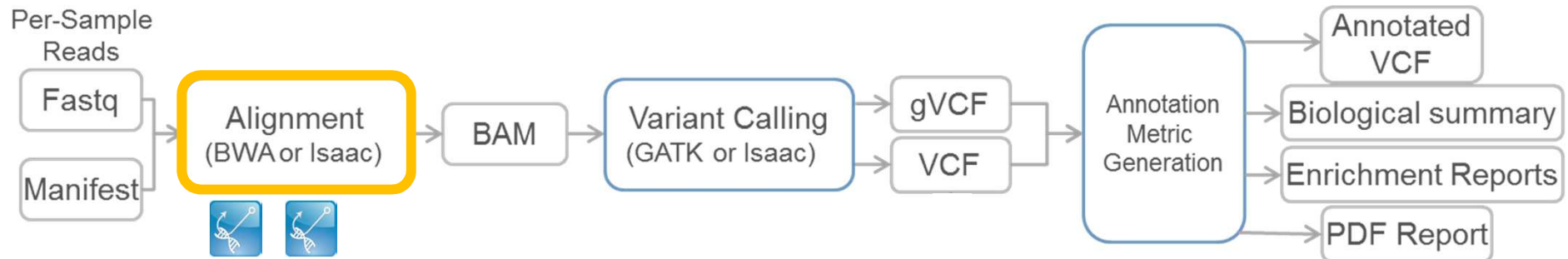
解析工程 – アライメント



Isaac
Enrichment



BWA
Enrichment



BaseSpace エクソーム解析

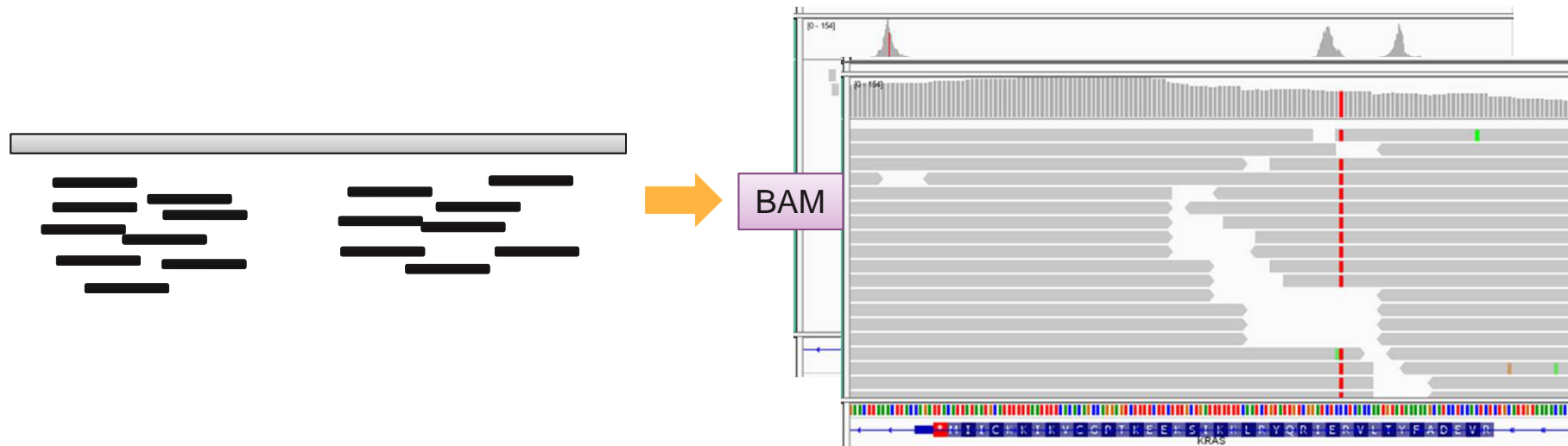
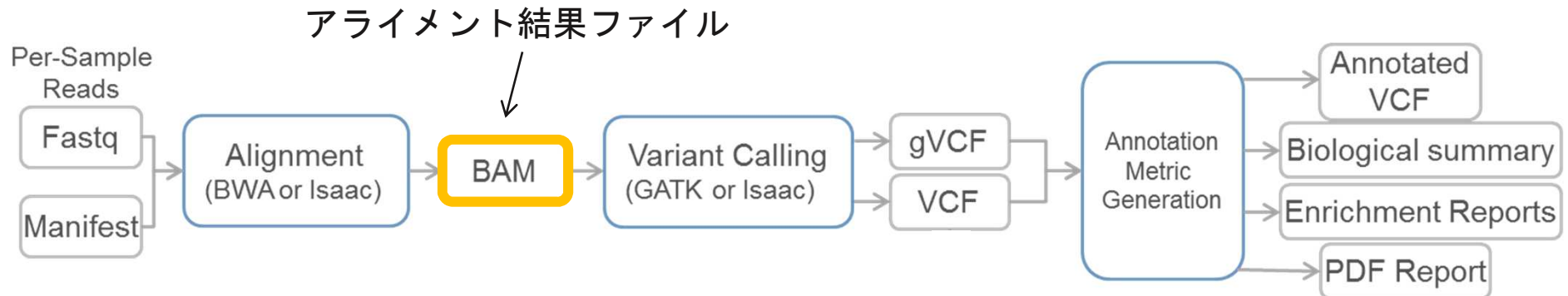
結果 – アライメント – BAMファイル



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

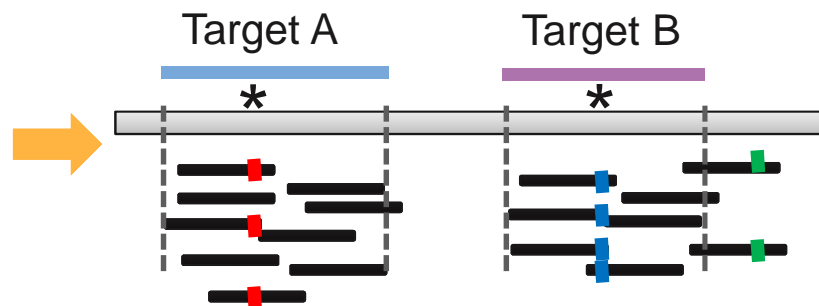
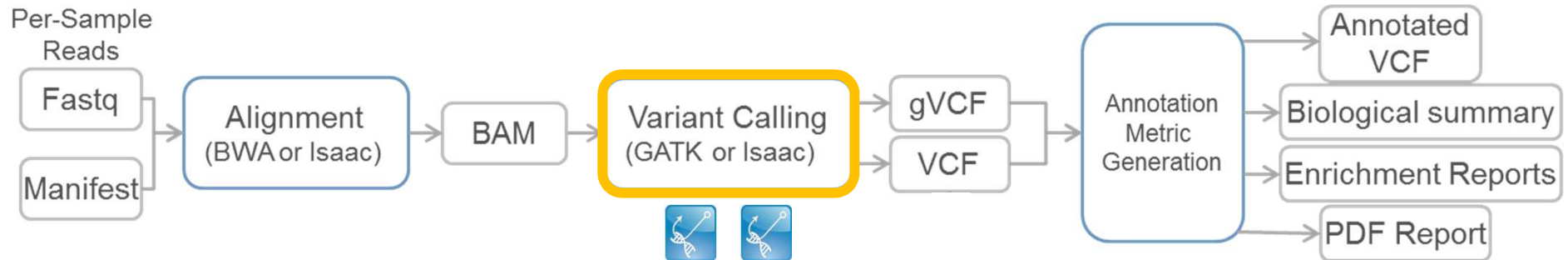
解析工程 - 変異コール



Isaac
Enrichment



BWA
Enrichment



BaseSpace エクソーム解析

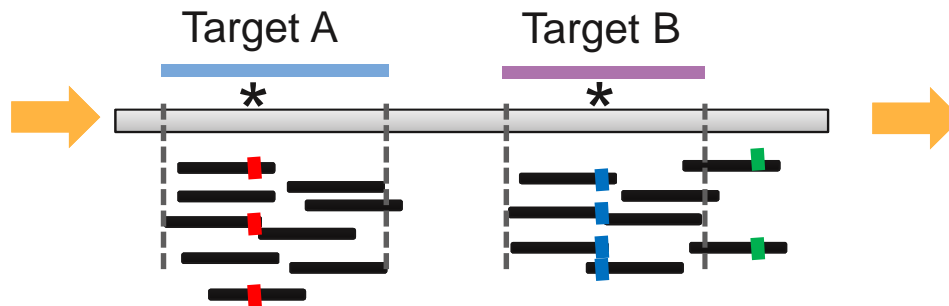
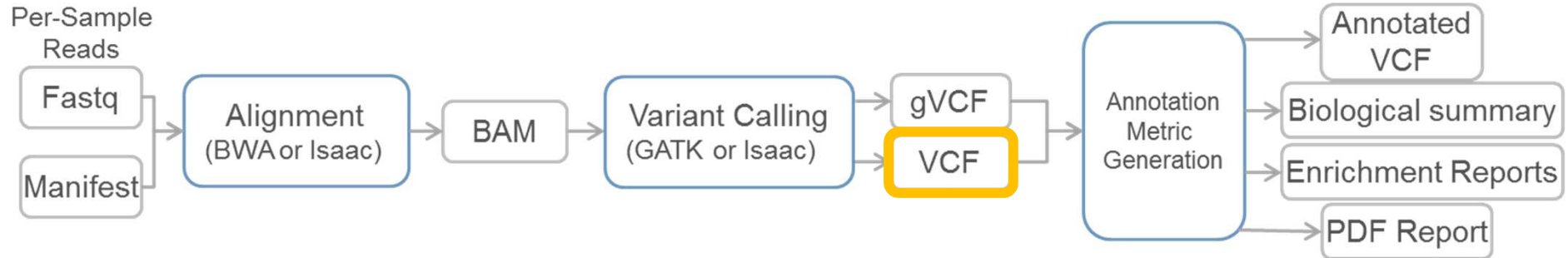
結果 - 変異コール - VCF



Isaac
Enrichment



BWA
Enrichment



Coriell-NA12880_S3.vcf - Microsoft Excel

##reference=file:///E:/Genomes/Homo_sapiens/UCSC/hg19/Sequence/WholeGenomeFASTA/genome.fa
##source=GATK 1.6
##fileformat=VCFv4.1

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	Coriell_NA12880
chr1	976580	.	C	T	320.18	PASS	AC=1;AF=;GT:AD:DP:0/1:11,12:23:99:350,0,242:0.522:99		
chr1	977330	rs2799066	T	C	2895.91	PASS	AC=2;AF=;GT:AD:DP:1/1:0,93:94:99:2929,214,0:1.000:99		
chr1	981931	rs2465128	A	G	5196.81	PASS	AC=2;AF=;GT:AD:DP:1/1:0,158:161:99:5197,394,0:1.000:99		
chr1	982994	rs10267	T	C	4304.29	PASS	AC=2;AF=;GT:AD:DP:1/1:1,139:140:99:4304,281,0:0.993:99		
chr1	984302	rs9442391	T	C	1482.31	PASS	AC=2;AF=;GT:AD:DP:1/1:0,54:54:99:1515,117,0:1.000:99		
chr1	986732	.	G	A	2073.94	PASS	AC=1;AF=;GT:AD:DP:0/1:1:83,83:166:99:2104,0,2191:0.500:99		
chr1	987200	rs9803031	C	T	3192.67	PASS	AC=2;AF=;GT:AD:DP:0/1:0,96:96:99:3226,241,0:1.000:99		
chr1	990280	rs4275402	C	T	2106.3	PASS	AC=2;AF=;GT:AD:DP:1/1:0,60:60:99:2139,163,0:1.000:99		
chr1	1147422	rs17568	C	T	1063.42	SB	AC=1;AF=;GT:AD:DP:0/1:27,32:60:99:1093,0,759:0.542:99		
chr1	1268987	.	G	A	1852.89	PASS	AC=1;AF=;GT:AD:DP:0/1:43,69:112:99:1883,0,1262:0.616:99		
chr1	1269554	rs307377	T	C	9399.51	PASS	AC=2;AF=;GT:AD:DP:1/1:0,281:282:99:9400,695,0:1.000:99		
chr1	1647778	rs7290177	C	G	457.12	PASS	AC=1;AF=;GT:AD:DP:0/1:71,23:94:99:487,0,2069:0.245:99		
chr1	1647814	rs7290177	T	C	448.08	PASS	AC=1;AF=;GT:AD:DP:0/1:75,23:98:99:478,0,2196:0.235:99		
chr1	1647871	rs7290901	T	C	406.19	LowVariat	AC=1;AF=;GT:AD:DP:0/1:94,23:117:99:436,0,2692:0.197:99		
chr1	1647928	rs1883424	A	G	2065.57	PASS	AC=1;AF=;GT:AD:DP:0/1:9,69:78:99:2096,0,105:0.885:99		
chr1	1653028	rs1682526	C	T	2051.57	PASS	AC=2;AF=;GT:AD:DP:1/1:0,62:62:99:2085,153,0:1.000:99		
chr1	1957037	rs2229110	T	C	2768.52	PASS	AC=2;AF=;GT:AD:DP:1/1:0,87:89:99:2802,205,0:1.000:99		
chr1	2340073	rs7653065	C	G	1925.94	PASS	AC=1;AF=;GT:AD:DP:0/1:74,72:146:99:1956,0,2323:0.493:99		
chr1	2340200	rs2494598	T	C	5596.57	PASS	AC=2;AF=;GT:AD:DP:1/1:0,178:179:99:5597,409,0:1.000:99		

BaseSpace エクソーム解析

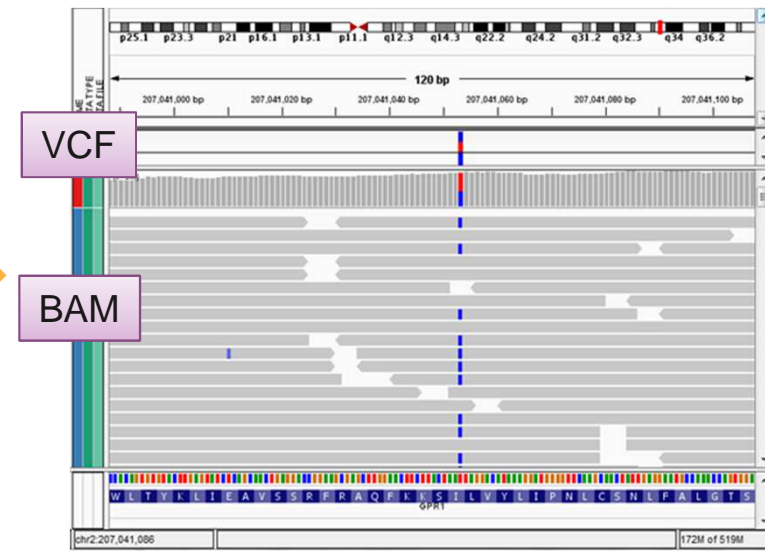
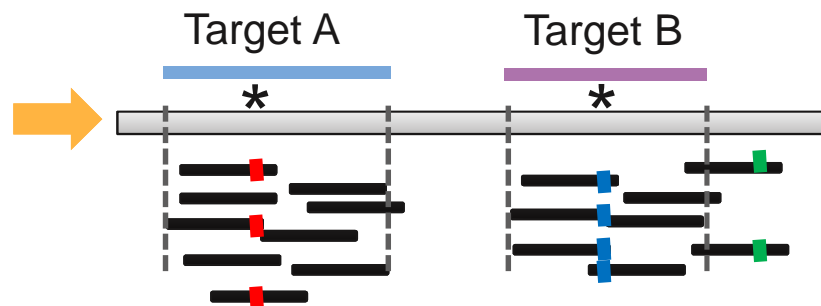
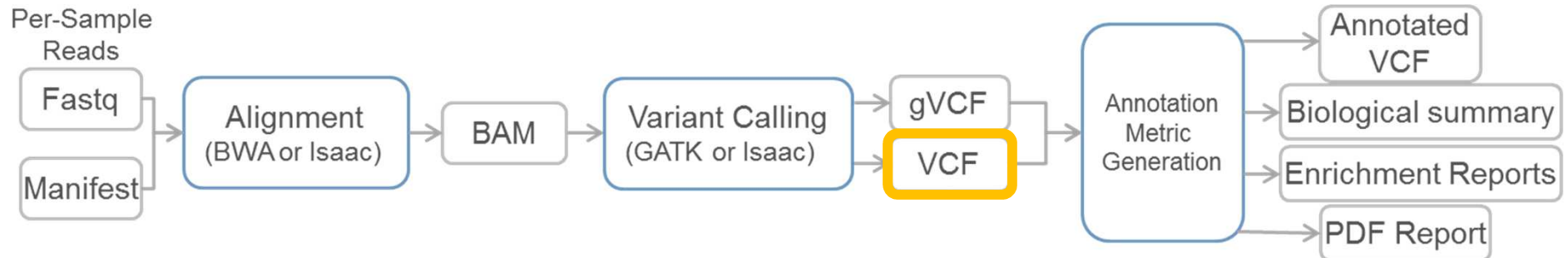
結果 - 変異コール - VCF



Isaac
Enrichment



BWA
Enrichment



BaseSpace エクソーム解析

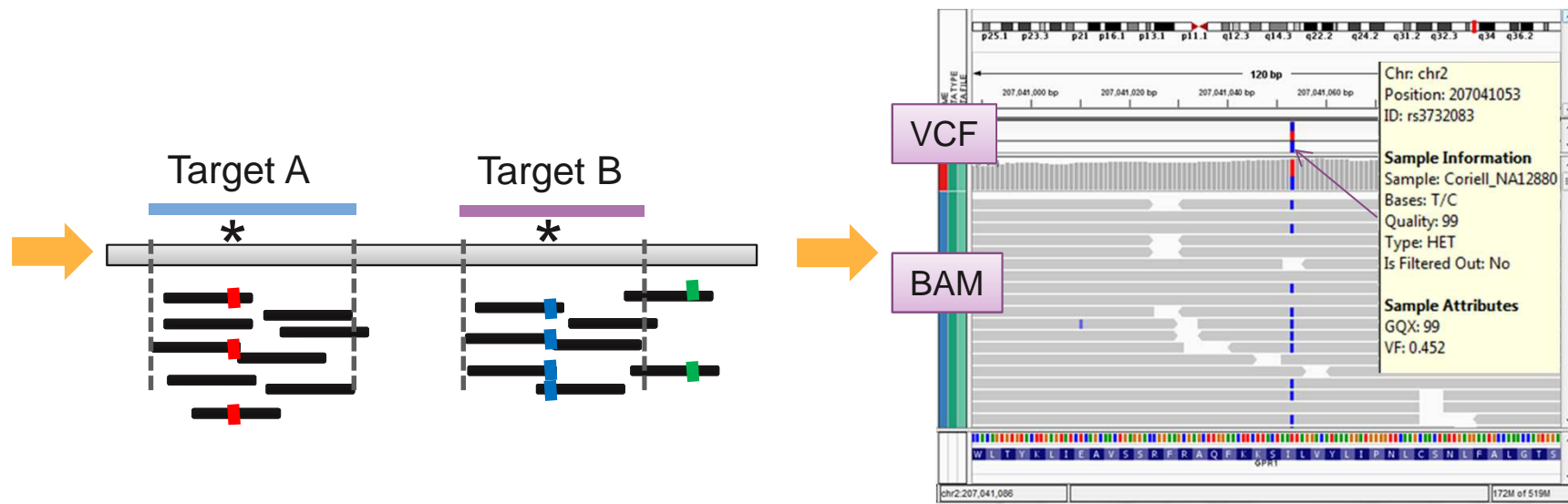
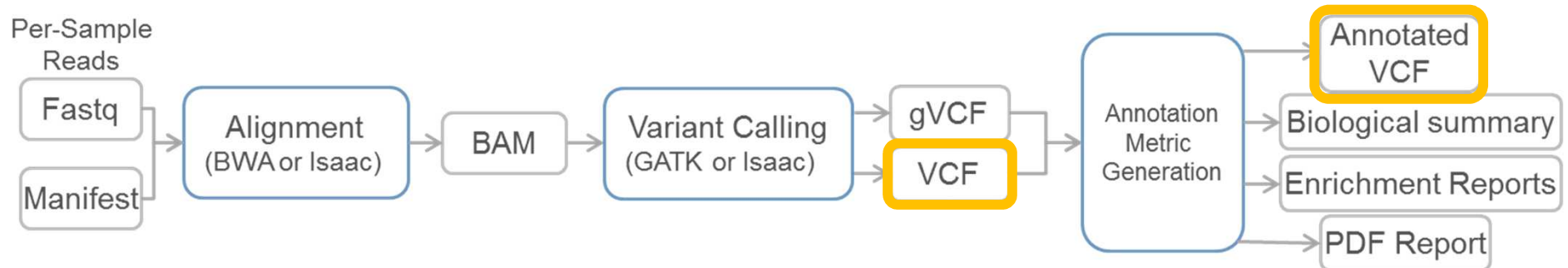
結果 - 変異コール - VCF



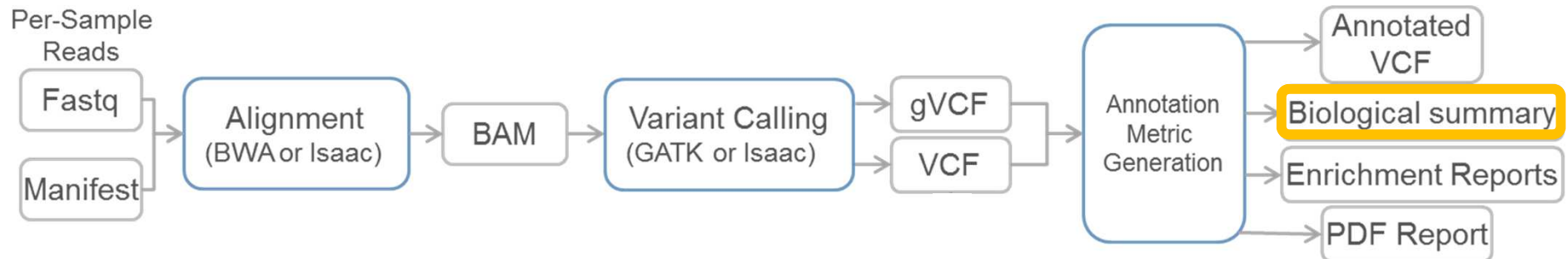
Isaac
Enrichment



BWA
Enrichment



BaseSpace エクソーム解析 結果 – Biological サマリ レポート



BaseSpace
HiSeq2500: Nextera Rapid Capture Exome (12plex, CEPH Trio replicates) : BWA Enrichment (12plex, Manifest v1.2) 01/16/2014 9:15:22

Analysis Info

Inputs

Output Files

Analysis Reports

Aggregates Summary

PDF Reports

FC1_NA12878_01

FC1_NA12878_04

FC1_NA12878_07

FC1_NA12878_10

FC1_NA12891_02

FC1_NA12891_05

FC1_NA12891_08

FC1_NA12891_11

ALIGNMENT SUMMARY 1

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%

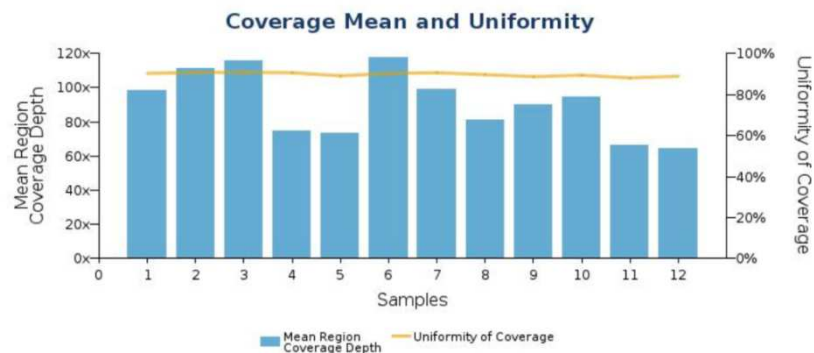
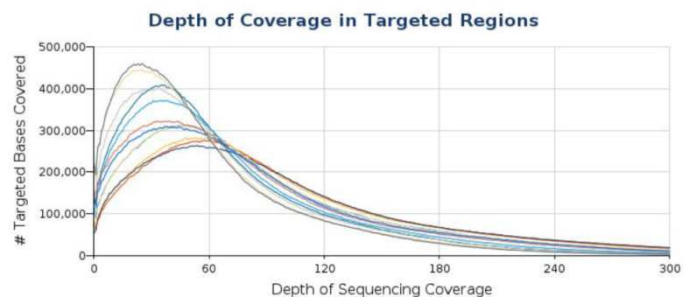
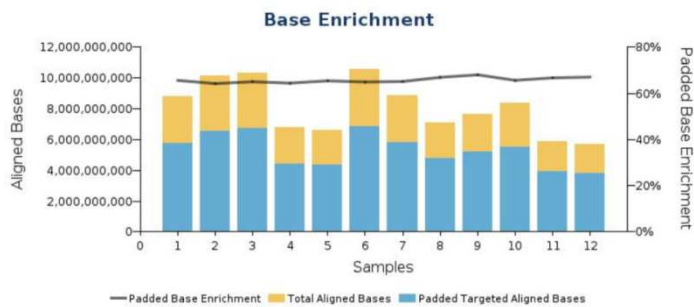
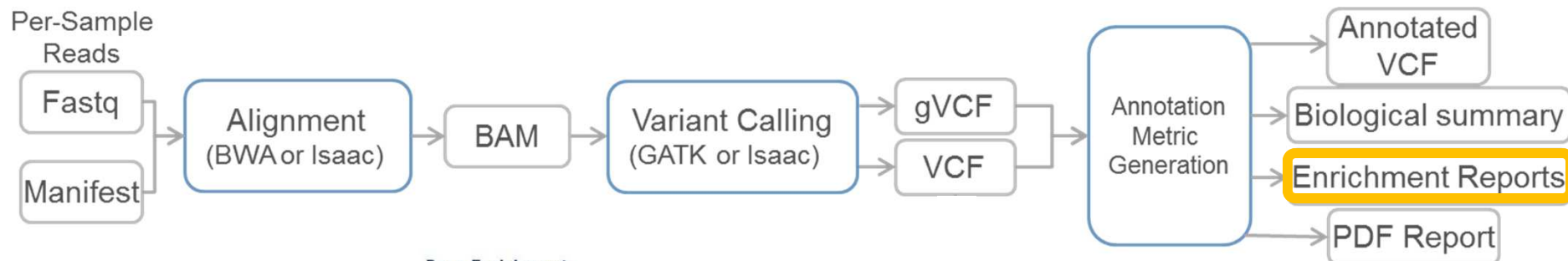
Variants by Sequence Context

	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

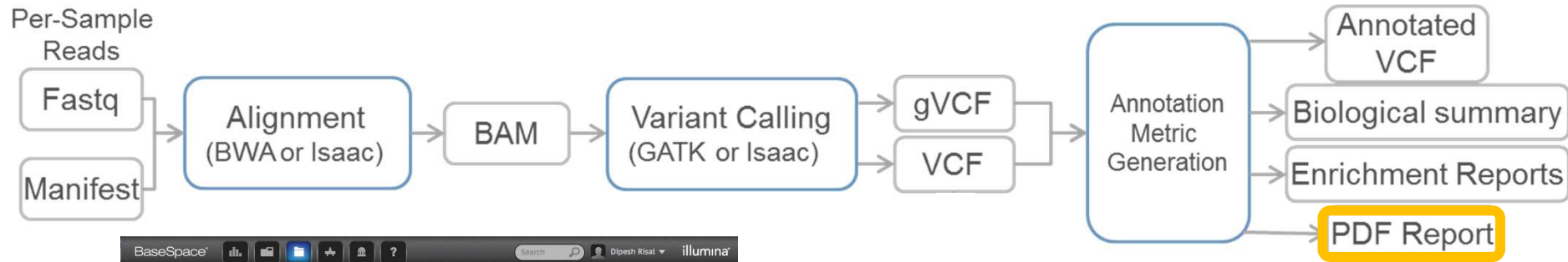
Variants by Consequence

	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



BaseSpace Dipesh Risal | illumina

HISeq2500_NRC_CEPH_Trio_2x150_FC1 : BWA Enrichment 12/05/2013 1:46:28_new_manifest

Rename analysis

Analysis Info | Inputs | Output Files | Analysis Reports

RESULTS FOR SAMPLE CEX_NA12878_01

PDF Summary Report

SAMPLE INFORMATION

Total PF Reads	Percent Q30	Percent Duplicate Paired Reads	Fragment Length Median	Fragment Length Standard Deviation
65,262,238	82.3%	6.5%	227 bp	80 bp

ENRICHMENT SUMMARY

Total Length of Targeted Reference	Padding Size
44,296,481 bp	150 bp

Read Level Enrichment

Total Aligned Reads	Percent Aligned Reads	Target Aligned Reads	Read Enrichment	Padded Target Aligned Reads	Padded Read Enrichment
62,912,089	96.4%	39,472,997	62.7%	42,236,377	67.1%

Base Level Enrichment

Total Aligned Bases	Target Aligned Bases	Bases Enrichment	Padded Target Aligned Bases	Padded Base Enrichment
8,778,180,848	4,378,065,923	49.9%	5,837,362,539	66.5%

SMALL VARIANTS SUMMARY

	SNVs	Insertions	Deletions
Total Passing	31,080	989	1,185
Percent Found in dbSNP	99.0%	88.0%	78.3%

illumina®

Enrichment Sequencing Report

Sample: Coriell_NA12880

Analysis Name: Isaac Enrichment 04/14/2014 8:21:14

Report Date: 04/13/2014 23:23:14 (UTC)

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

The screenshot shows the BaseSpace Public Data interface. The top navigation bar includes 'BaseSpace', 'Dashboard', 'Prep', 'Runs', 'Projects', 'Apps', 'Public Data', and 'Help'. The user is logged in as 'Eri Kibukawa'. The main content area is titled 'Public Data' and features a search bar. Below the search bar, there are links for 'Nextera Rapid Capture Exome Data Sets' and 'Targeted Sequencing Exome'. A specific data set is highlighted: 'HiSeq 2500: Nextera Rapid Capture Exome (CEPH Trio)'. The description states: 'Data generated from 4 replicates each of the CEPH trio NA12878, NA12891 and NA12892 sequenced on HiSeq 2500. Samples prepared using Nextera Rapid Capture Exome reagent kit.' Below the description, there are tabs for 'Resequencing', 'Exome', and 'Targeted Sequencing'. A table lists the data set details:

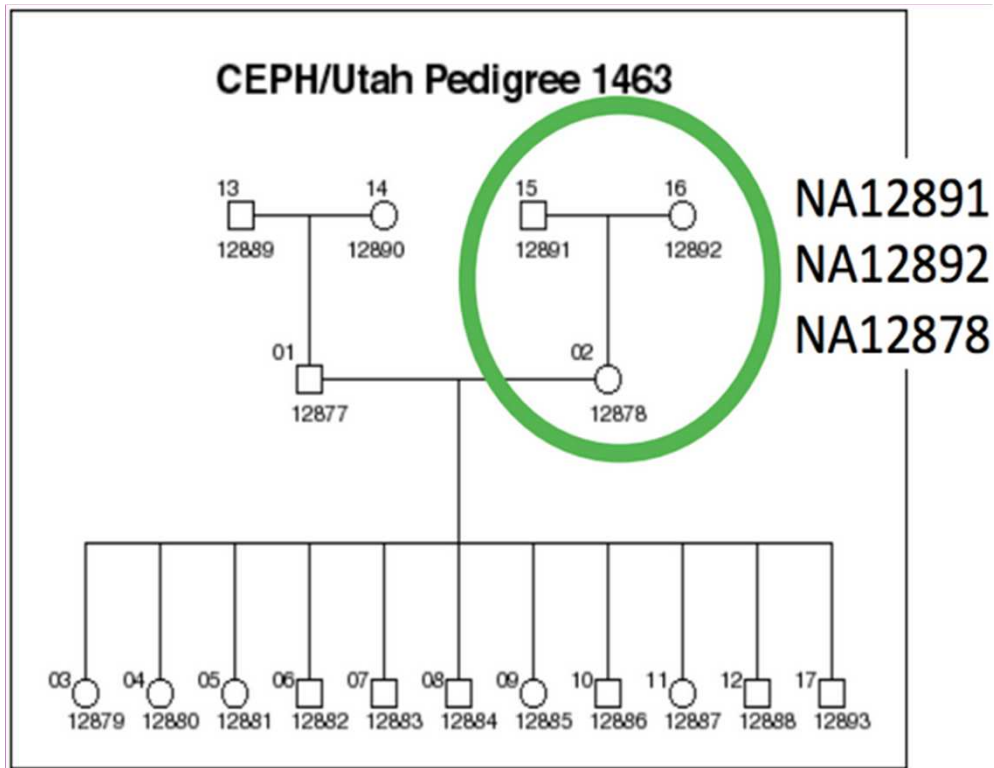
Run	HiSeq2500_NRC_CEPH_Trio_2x150_FC1 (71.51 GB)	Import
Project	HiSeq 2500: Nextera Rapid Capture Exome (12plex, CEPH Trio replicates) ()	Import

The 'Import' button for the Project row is circled in red. To the right of the main content, there are two panels: 'Research Areas' with buttons for 'Cancer Research', 'Genetic Disease', 'Complex Disease', and 'Microbial Research'; and 'Categories' with buttons for 'Exome', 'Resequencing', 'Small RNA', 'Targeted Sequencing', 'De Novo Assembly', 'RNA-Seq', 'Gene Fusion', and 'ChIP-Seq'.

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
HapMap
1Kゲノム
プラチナゲノム

catalog.coriell.org/0/sections/Collections/NIGMS/CEPHResources.aspx
www.hapmap.org
www.1000genomes.org
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>

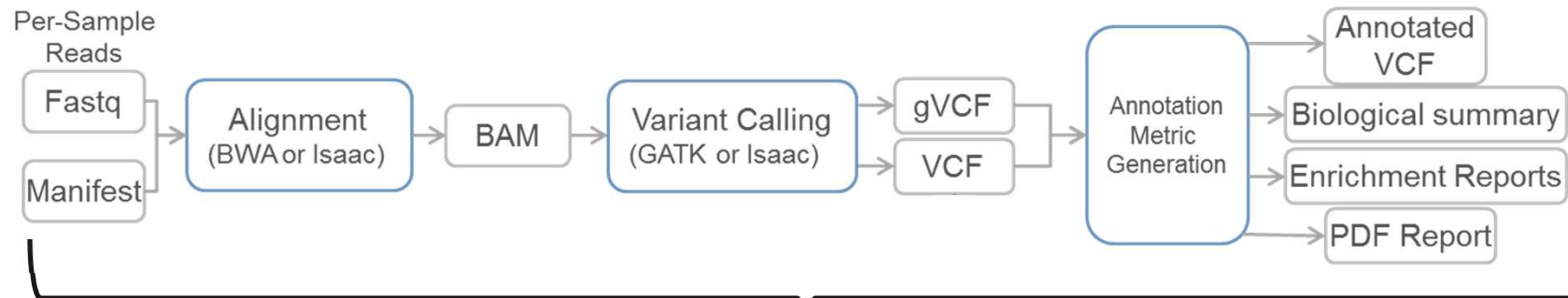
BaseSpace App によるエクソーム解析



BWA
Enrichment



Isaac
Enrichment



BWA Enrichment v2.1
illumina, inc

Analysis Name: BWA Enrichment v2.1 05/14/2015 8:02:4

Save Results To: Select Project(s):
HiSeq 2500: Nextera Rapid Capture Exome (12plex, CEPH Trio replicates)

Sample(s) [96 maximum]: Select Sample(s):

Reference Genome: Human (UCSC hg19)

Targeted Regions: Nextera Rapid Capture Exome v1.2

Custom Targeted Manifest: Select File(s):

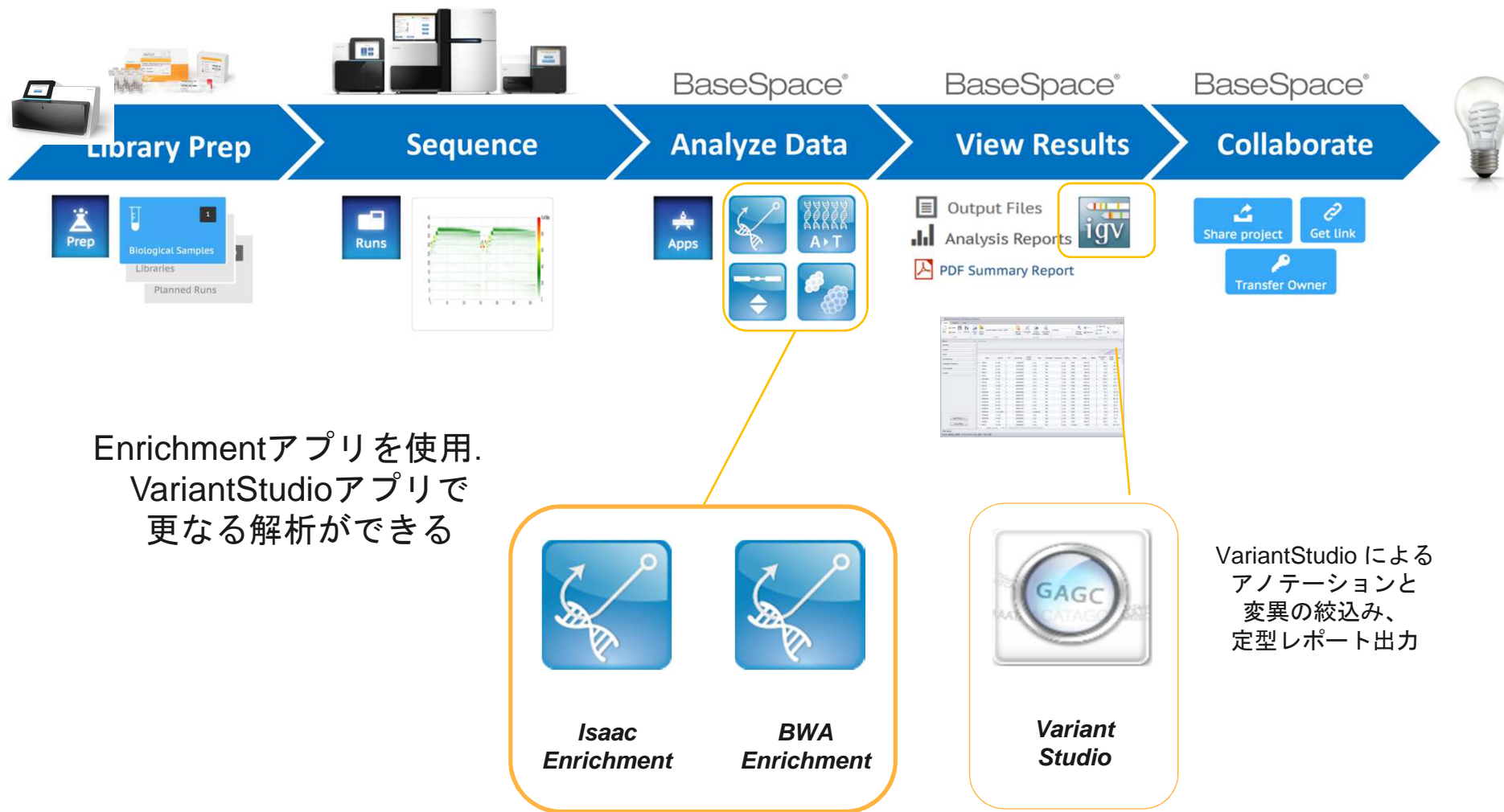
Aligner: BWA-MEM BWA-backtrack (legacy)

Base Padding: 150

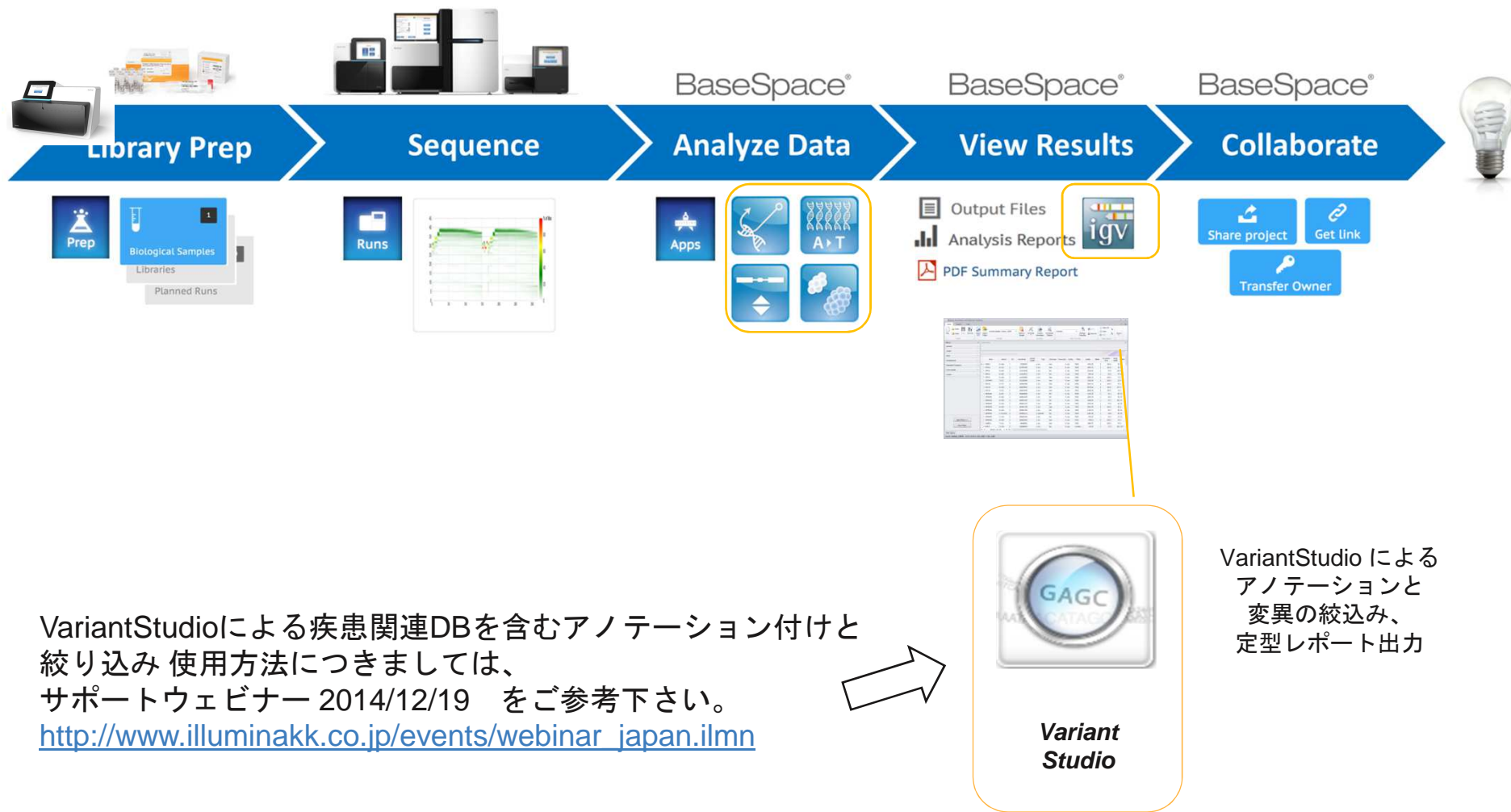
Annotation: RefSeq Ensembl

▶ Advanced Options

BaseSpace によるエクソーム解析



BaseSpace によるエクソーム解析



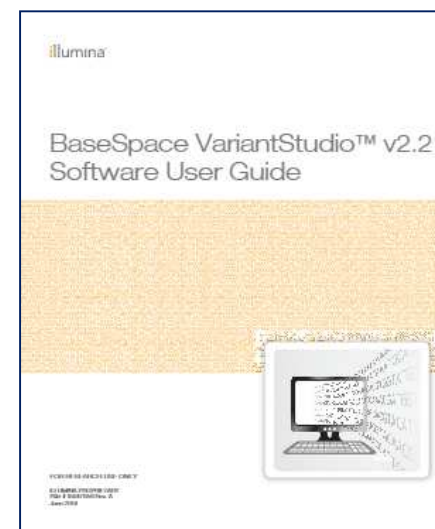
詳細は各アプリの Instruction ガイドをご確認頂けます



**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

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

The image displays three overlapping screenshots of the Illumina website's Exome Sequencing section. The top-left screenshot shows the 'Exome Sequencing' overview page, highlighting the benefits of exome sequencing and providing a sign-up form for interested parties. The top-right screenshot shows the 'Exome Sequencing Data Analysis' page, detailing the BaseSpace environment and the Exome Data Analysis with BaseSpace Apps. The bottom screenshot shows a table titled 'Table: Optimal Samples per Run Across' which provides technical specifications for different sequencing systems.

	MI20	MI250 500	MI250 1500
	V3 REAGENT KIT	MD OUTPUT	HIGH OUTPUT
Rolling Retry	1	Up to 1	
	3	Up to 3	
	5		Up to 5
	9		Up to 9
	12		Up to 12

Number of exomes pooled before enrichment.
For Research Use Only.

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