

イリミナソフトウェアを用いた NGSデータの解析におけるヒント

Dec 4, 2015



渡邊 大
イリミナ株式会社
テクニカルアプリケーションサイエンティスト

© 2015 Illumina, Inc. All rights reserved.
Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPPro, CytoChip, DesignStudio, Epicentre, ForenSeq, Genetic Energy, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiSeq, MiSeqDX, MiSeq FGx, NeoPrep, NextBio, Nextera, NextSeq, Powered by Illumina, 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 US and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

illumina®

本日の内容

- ▶ BaseSpaceの使用に関するお問い合わせ
 - fastq fileがダウンロードできない。
 - fastq fileがアップロードできない。
 - BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ
- ▶ Bcl conversionに関するお問い合わせ
 - 全ての塩基でNとコールされてしまう
 - 解析が途中で停止し、fastqが生成されない

本日の内容

▶ BaseSpaceの使用に関するお問い合わせ

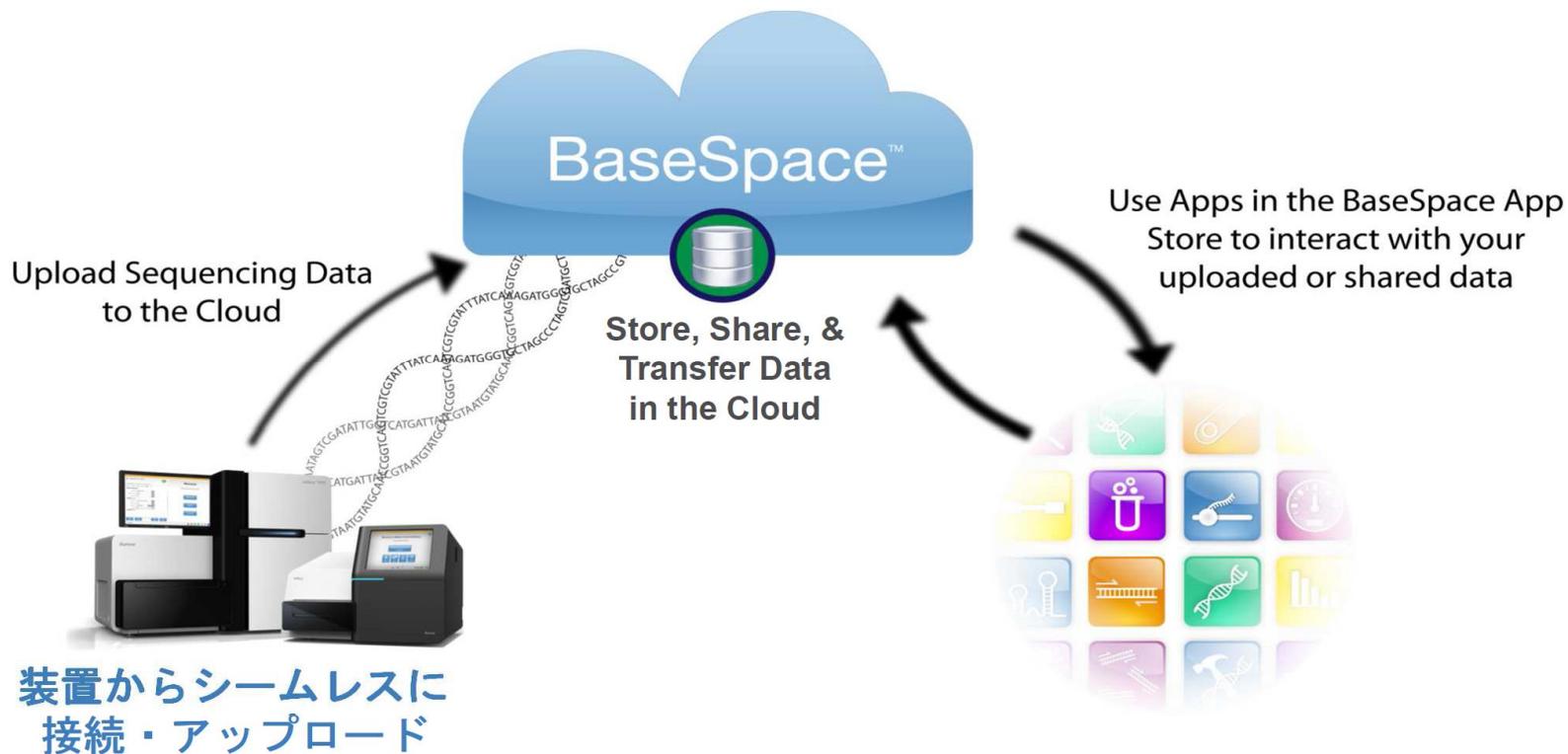
- fastq fileがダウンロードできない。
- fastq fileがアップロードできない。
- BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ

▶ Bcl conversionに関するお問い合わせ

- 全ての塩基でNとコールされてしまう
- 解析が途中で停止し、fastqが生成されない

BaseSpaceとは

Seamless upload, monitoring, storage, analysis, and collaboration



The Future: Cloud Solutions will be everywhere and specifically valuable to researchers

BaseSpaceとは



The Future: Cloud Solutions will be everywhere and specifically valuable to researchers

fastq fileがダウンロードできない。

- ▶ Basespaceの便利な機能の一つに、ランデータをクラウド上に共有し、共同研究者の方々にシェアできるというものがあります。
- ▶ fastq fileは、シーケンスの結果をクオリティと共に記載したファイル

```
@HISEQ:132:XXXX4IJGG:2:0202:9999:9999 1:N:0:5 [ヘッダー]
NAAGCATGCATCGCATGCATGCGTACCAGTTANNAAGANNNGCTNNNNNANCNAANNNGCATGCATGCANNGTACCCNNNTCCCTCTNCCNCNNACNNAAN [配列]
+
#0<FFFFFFFFFIIIIIIIIIIIBFIIIFIFIF##0<BF##00B#####0#0#07###IIIIIIIIIII'##' '0'0'###' '0' '' '#0' #'##0' ##00' [score]
```

- ▶ 共同研究者からシーケンス結果をシェアしてもらったが、fastqをダウンロードしようとするとき  が表示されて選択できない。

fastq fileがダウンロードできない

Download

BaseSpace utilizes a desktop client that allows for fast, reliable downloads transferred securely over SSL.

Install the BaseSpace downloader

The downloader requires a one-time installation. Click to install the downloader on this machine. If the downloader is already installed, you can skip this step.

[Install the Downloader](#) [Don't show me this again](#)

Select files to be downloaded

- All FASTQ files for this run ← アクティブにならない
- SAV file [Learn more about SAV](#)
- No analysis files are available for this run

[Download your files](#)

[Close](#)

Runのシェアだけしていませんか？

Projectのシェアをしてください。

fastq fileがダウンロードできない

BaseSpace® Dashboard Prep Runs Projects Apps Public Data Help

Projects : PhiX 2x250 : Samples List : PhiX

PhiX

Launch App

Sample name	PhiX
Date created	08/20/2013 4:36:41
Genome	Phix - Illumina
Paired end	Paired
Number of Clusters	18,306,599
Number of Reads	36,613,198
Read 1 length	251
Read 2 length	251
Size	7.20 GB

Origin

Graph List

Files

FASTQ Files

APP SESSIONS

MSR: Assembly Complete

Download Selected ← ここからダウンロード可能

	Name ▲	Size	Type	Status	Date
<input checked="" type="checkbox"/>	PhiX_S1_L001_R1_001.fastq.gz	3.52 GB	.gz	UploadCompl...	Aug 20, ...
<input checked="" type="checkbox"/>	PhiX_S1_L001_R2_001.fastq.gz	3.68 GB	.gz	UploadCompl...	Aug 20, ...

本日の内容

▶ BaseSpaceの使用に関するお問い合わせ

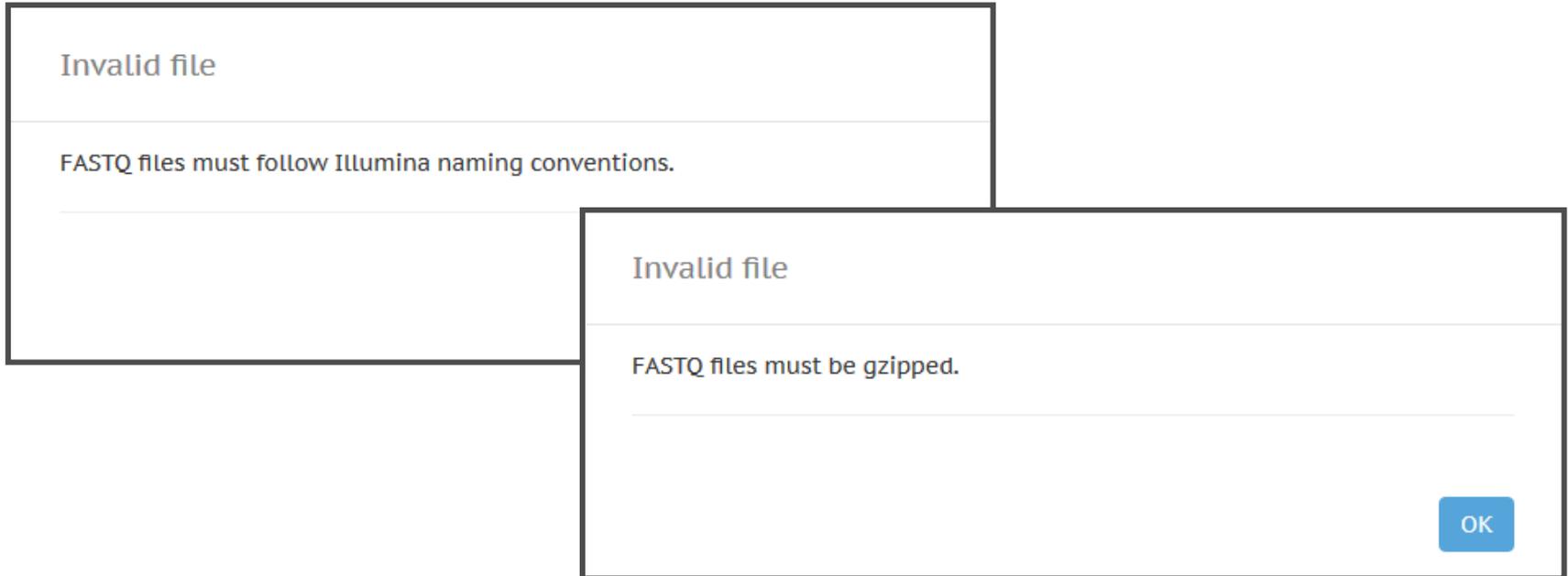
- fastq fileがダウンロードできない。
- fastq fileがアップロードできない。
- BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ

▶ Bcl conversionに関するお問い合わせ

- 全ての塩基でNとコールされてしまう
- 解析が途中で停止し、fastqが生成されない

fastq fileをアップロードできない

- ▶ BaseSpaceにfastqファイルなどをアップロードし、BaseSpace appsのinputとして使う、共同研究者の方々と共有する、ということも可能です。
- ▶ Projectsタブから新規プロジェクトを作成し、fastqをアップロードしようとしても、'Invalid file'などのエラーが出てアップロードが開始されない。



イルミナファイルの仕様に沿っているか、ご確認ください。

fastq fileをアップロードできない

- ▶ 下記のルールに沿っているか、確認ください。

- Confirming that the file name conforms to Illumina standards:
SampleName_SampleNumber_Lane_Read_FlowCellIndex.fastq.gz
SampleName_S#_L00#_R#_00#.fastq.gz

https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/basespace/basespace-user-guide-15044182-e.pdf

- ▶ また、ネットワーク環境が悪いことでアップロードができないこともあります。BaseSpace cloudの利用の際には、10Mbps以上のネットワーク環境をご用意ください。
- ▶ fastqのアップロードに関する詳細は、下記ウェビナーもご参考下さい

2015/09/04

解析に適したリード前処理を行うために

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

本日の内容

▶ BaseSpaceの使用に関するお問い合わせ

- fastq fileがダウンロードできない。
- fastq fileがアップロードできない。
- BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ

▶ Bcl conversionに関するお問い合わせ

- 全ての塩基でNとコールされてしまう
- 解析が途中で停止し、fastqが生成されない

BaseSpace appsとは

<イルミナコアアプリ>

				
16S Metagenomics	TopHat Alignment	Cufflinks Assembly & DE	RNA Express	Variant Studio
				
BWA Enrichment	Isaac Enrichment	Broad IGV	TruSeq Amplicon	Amplicon DS
				
BWA WGS	Isaac WGS	Tumor Normal	Long Read Assembly	Long Read Phasing

<イルミナラボアプリ>

					
FastQC	Kraken Metagenomics	NextBio Annotates	VCAT	NextBio Transporter	SRST2
					
Fastq Toolkit	Velvet Assembly	Picard Space	NCBI SRA Importer	Prokka	

他

他

<他社製アプリ>

			
SPAdes	Novo Align	Advaita	DNA Star
			
AB SCIEX	AB SCIEX	AB SCIEX	AB SCIEX
			
SWATH Atlas	MetaPhlan	n of One	My FLQ
			
Lo Feq	eGB	Genomatix	Genome Profiler
			
OncoMD	GeneTalk	PathGEN Dx	Melanoma Profiler
			
TUTE	DeepCheck HIV,HBV,HCV	Pedant	他

BaseSpace appsが使用できない

- ▶ Basespace appsを使用しようとしても入力の画面から進めなくなってしまう。



TopHat Alignment v1.0.0

Illumina, Inc.

Please review the validation errors below

App Session Name:

TopHat Alignment 11/12/2015 4:54:45

Save Results To:

Select Project(s):

KOtestforDW

Samples:

Select Sample(s):

Select All



TSNanodownsizehort



Stranded

Reference Genome:

Homo sapiens/hg19 (RefSeq)

使用するアプリのlimitationをご確認ください。

BaseSpace appsが使用できない

- ▶ Appsのlimitationはトップページから確認できる。



TopHat Alignment

Illumina, Inc. [Website](#) »

Free

version 1.0.0

Launch

CATEGORIES RNA-Seq, Gene Fusion Detection, Tumor Normal

License Info

[Privacy](#)
[EULA](#)

Description

The TopHat Alignment workflow performs the following functions

- Read mapping using the TopHat 2 aligner
- FPKM estimation of reference genes and transcripts using Cufflinks 2
- Variant calling (SNVs and small indels) with the Isaac Variant caller
- Optional fusion calling with TopHat-Fusion

Available reference genomes include

- Homo sapiens UCSC hg19 (RefSeq & Gencode gene annotations)
- Mus musculus UCSC mm10 (RefSeq gene annotations)
- Rattus norvegicus UCSC rn5 (RefSeq gene annotations)

Current limitations:

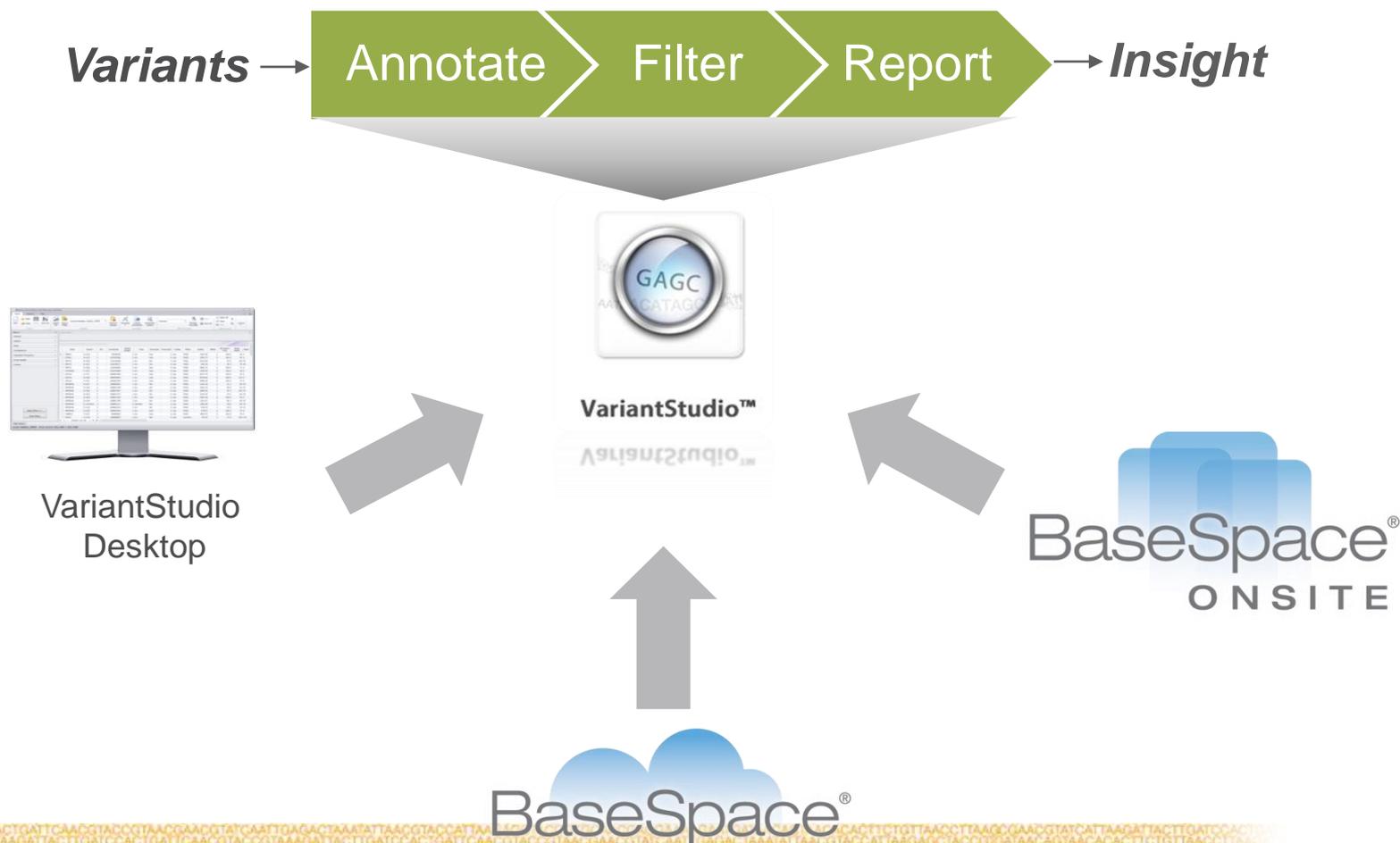
- Minimum and maximum of 100,000 and 400,000,000 reads per-sample, respectively
- Maximum of 2,000,000,000 read across all samples in a single analysis
- Minimum Read length 35bp and Maximum Read Length 125bp
- Paired-end reads are required for fusion detection

本日の内容

- ▶ BaseSpaceの使用に関するお問い合わせ
 - fastq fileがダウンロードできない。
 - fastq fileがアップロードできない。
 - BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ
- ▶ Bcl conversionに関するお問い合わせ
 - 全ての塩基でNとコールされてしまう
 - 解析が途中で停止し、fastqが生成されない

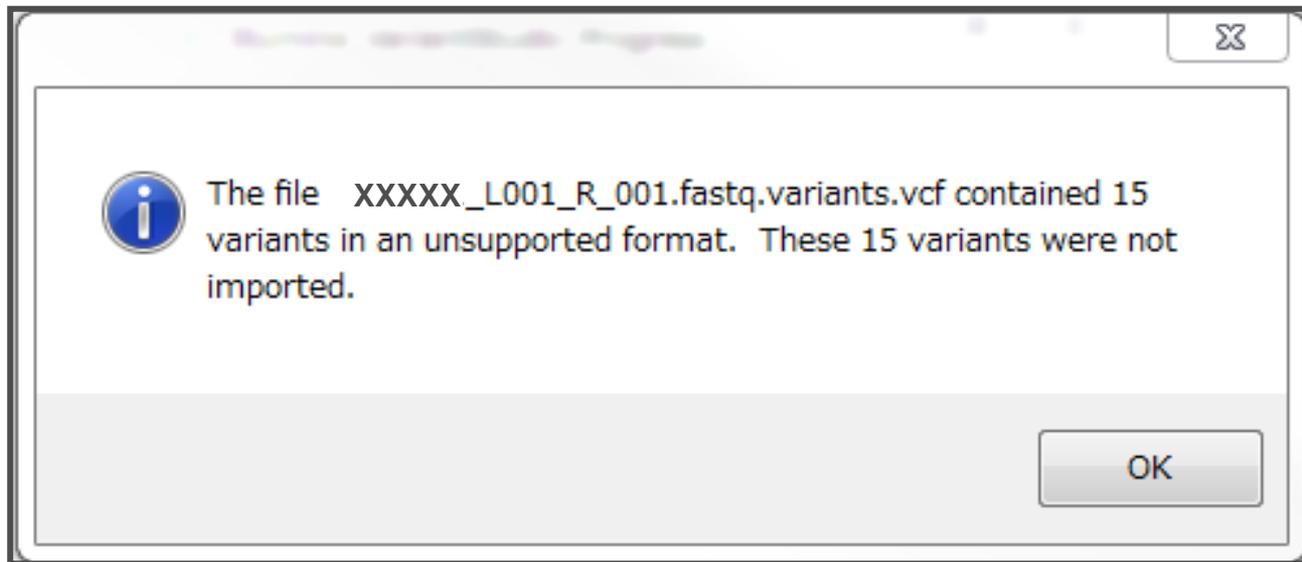
VariantStudioとは

- ▶ 変異解析のためのイルミナソフトウェア。
- ▶ リシーケンスappsで生成されるvcfファイルを読み込ませることで、コールした変異を解析できる。



VariantStudioにいくつかの変異が読み込めない

- ▶ VariantStudioにvcfファイルをインポートしようとする時、いくつかのファイルフォーマットが読み込めないとエラーがでる。



vcfファイルをテキストエディタなどで開き、
記載内容を確認してみてください。

VariantStudioにいくつかの変異が読み込めない

- ▶ 本お問い合わせは、vcfファイルを作成するのに、イルミナソフトウェア以外を使用している場合、またはカスタムのbedファイルを使用している場合に見られる場合が多くございます。
- ▶ イルミナソフトウェアで作成されたvcfファイルは以下の体裁になっています。まずは、この体裁に沿っているか、イルミナが使用しない記述などがいないか、ご確認ください（BaseSpaceのpublic dataをご参照下さい）。

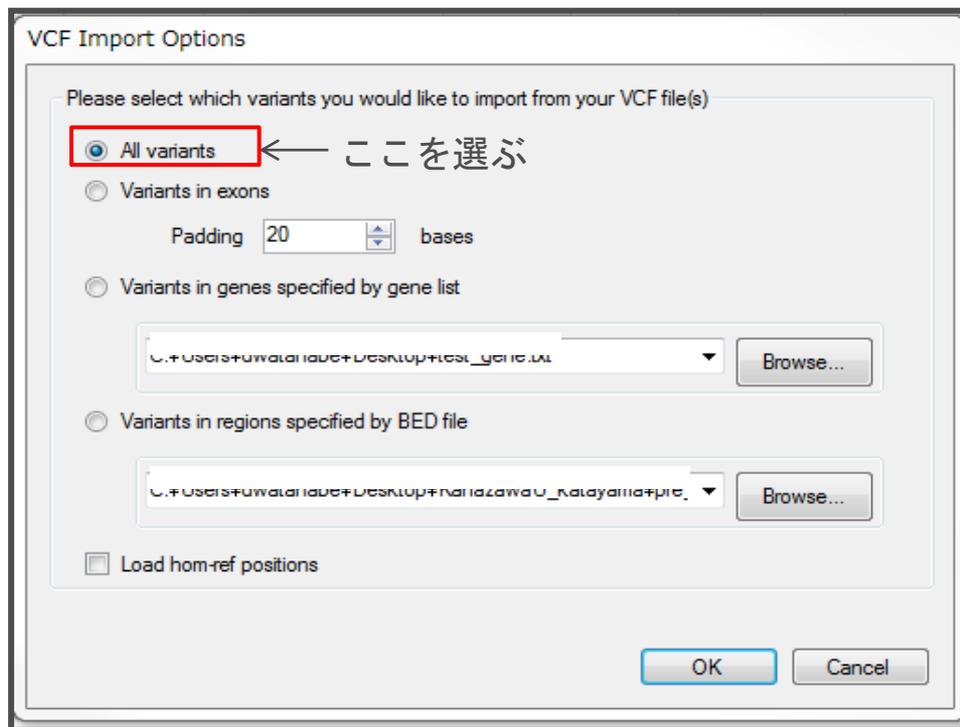
Example)

```
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT HDC749-Rep1-MixA
chr1 115256530 . G T 100 PASS DP=13952 GT:GQ:AD 0/1:100:12569
```

- ▶ 例えば、一度に全ての変異を読み込むのではなく、確認できる数ずつサブセットを作成して読み込ませることで、原因の記述を絞り込むことが可能です。

VariantStudioにいくつかの変異が読み込めない

- ▶ カスタムのbedファイルをご使用ではありませんか？
- ▶ もしお使いの場合は、vcfインポートの際に'All variants'をご選択頂き、エラーなく読み込めるかどうか、ご確認ください。
- ▶ これで読み込める場合は、vcfファイルとbedファイルの記載内容が一致しないことで起こるエラーの可能性がります。ご確認ください。



BaseSpace HelpCenterのご紹介

- ▶ BaseSpaceをご利用になるうえで、HelpCenterの情報も役立つと存じます。これまでにご紹介した内容も含め、詳細にご案内しております。

BaseSpace® HelpCenter [Sign in](#)

Welcome to our HelpCenter

Learn how to unlock the potential of Illumina's next-generation informatics ecosystem.

- New to BaseSpace?**
Get started and Explore BaseSpace >
- Articles**
Everything in one place; All our articles >
- BaseSpace Developers**
Start building the next generation of sequencing apps >

What's New

- Nov 2015 | [New and Updated Illumina Core Apps](#)
- Oct 2015 | [New BaseSpace Features and Tiers](#)
- Oct 2015 | [Interpret Protein-Coding and Splicing Variants with MutationForecaster](#)

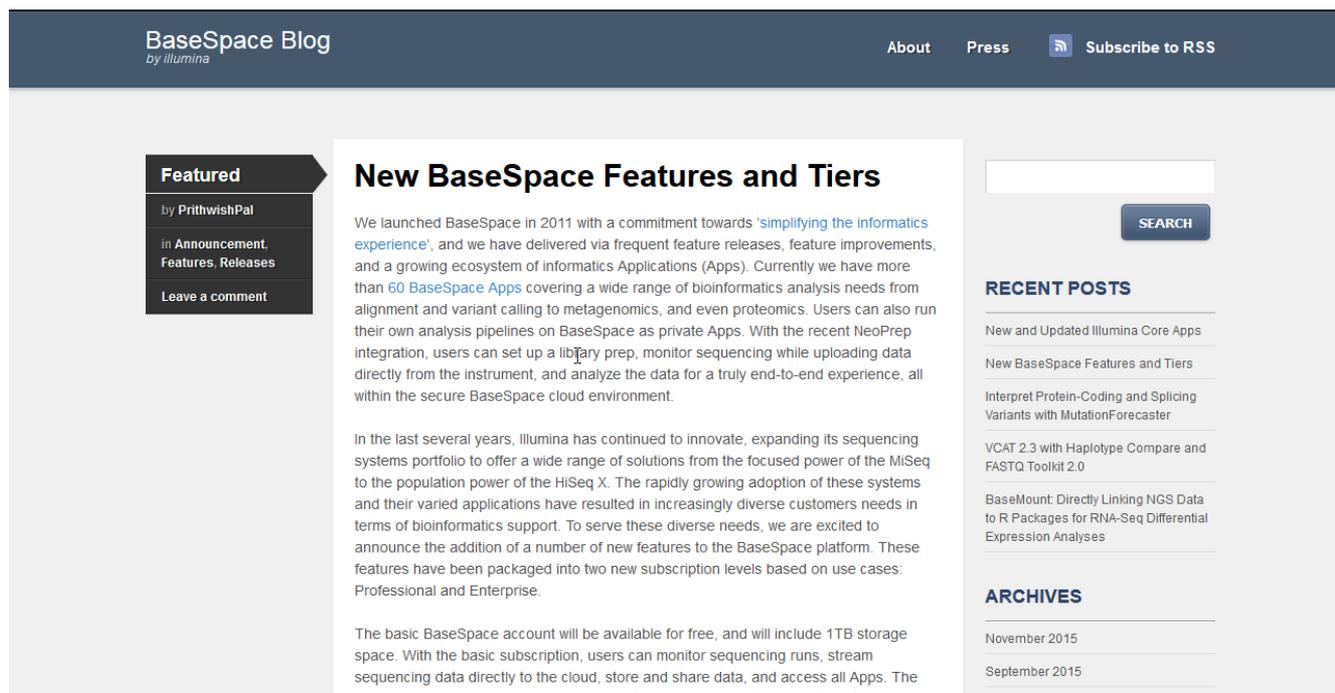
Popular Articles

- [How to share data using Email option](#)
- [How to transfer ownership](#)
- [How to delete/restore data](#)
- [How to combine samples](#)

- ▶ <https://help.basespace.illumina.com/>をご参照下さい。

BaseSpace blogのご紹介

- ▶ BaseSpaceをご利用になるうえで、こまめにup-to-dateな情報を得たい方は、BaseSpace blogがお勧めです。



The screenshot shows the BaseSpace Blog homepage. The header includes the logo 'BaseSpace Blog by illumina', navigation links for 'About' and 'Press', and a 'Subscribe to RSS' button. The main content area features a 'Featured' section with a post titled 'New BaseSpace Features and Tiers' by PrithwishPal. The post text discusses the company's commitment to simplifying the informatics experience and the addition of new features and subscription tiers. A search bar and a 'RECENT POSTS' section are visible on the right side of the page.

BaseSpace Blog
by illumina

About Press  Subscribe to RSS

Featured

by PrithwishPal

in Announcement, Features, Releases

Leave a comment

New BaseSpace Features and Tiers

We launched BaseSpace in 2011 with a commitment towards 'simplifying the informatics experience', and we have delivered via frequent feature releases, feature improvements, and a growing ecosystem of Informatics Applications (Apps). Currently we have more than 60 BaseSpace Apps covering a wide range of bioinformatics analysis needs from alignment and variant calling to metagenomics, and even proteomics. Users can also run their own analysis pipelines on BaseSpace as private Apps. With the recent NeoPrep integration, users can set up a library prep, monitor sequencing while uploading data directly from the instrument, and analyze the data for a truly end-to-end experience, all within the secure BaseSpace cloud environment.

In the last several years, Illumina has continued to innovate, expanding its sequencing systems portfolio to offer a wide range of solutions from the focused power of the MiSeq to the population power of the HiSeq X. The rapidly growing adoption of these systems and their varied applications have resulted in increasingly diverse customer needs in terms of bioinformatics support. To serve these diverse needs, we are excited to announce the addition of a number of new features to the BaseSpace platform. These features have been packaged into two new subscription levels based on use cases: Professional and Enterprise.

The basic BaseSpace account will be available for free, and will include 1TB storage space. With the basic subscription, users can monitor sequencing runs, stream sequencing data directly to the cloud, store and share data, and access all Apps. The early access Linux-based command-line interface, [BaseMount](#), will continue to be

SEARCH

RECENT POSTS

- New and Updated Illumina Core Apps
- New BaseSpace Features and Tiers
- Interpret Protein-Coding and Splicing Variants with MutationForecaster
- VCAT 2.3 with Haplotype Compare and FASTQ Toolkit 2.0
- BaseMount: Directly Linking NGS Data to R Packages for RNA-Seq Differential Expression Analyses

ARCHIVES

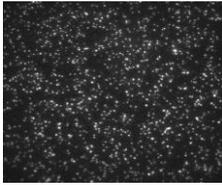
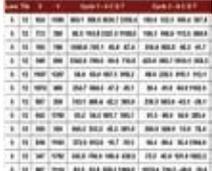
- November 2015
- September 2015

- ▶ <http://blog.basespace.illumina.com/>をご参照下さい。

本日の内容

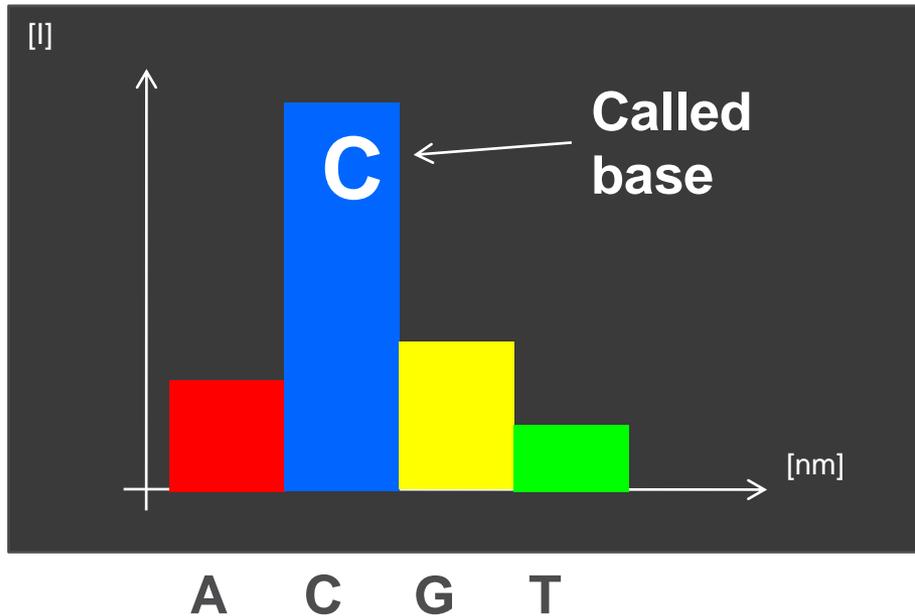
- ▶ BaseSpaceの使用に関するお問い合わせ
 - fastq fileがダウンロードできない。
 - fastq fileがアップロードできない。
 - BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ
- ▶ Bcl conversionに関するお問い合わせ
 - 全ての塩基でNとコールされてしまう
 - 解析が途中で停止し、fastqが生成されない

bcl conversionとは

Analysis Type	Software	Outputs
<p style="color: blue; font-size: 1.2em;">Sequencing</p>	 <p style="text-align: center;">ICS/RTA</p>	 <p style="text-align: center;">Images/TIFF files</p>
<p style="color: green; font-size: 1.2em;">Primary Analysis</p>	 <p style="text-align: center;">ICS/RTA</p>	<div style="display: flex; justify-content: space-around;"> <div data-bbox="1304 635 1516 806">  </div> <div data-bbox="1545 635 1767 806">  </div> </div> <p style="text-align: center;"> Intensities Base Calling bclファイル </p>
<p style="color: purple; font-size: 1.2em;">Secondary Analysis</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div data-bbox="691 928 975 1106">  </div> <div data-bbox="1014 978 1197 1149">  <p>HiSeq Analysis Software</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div data-bbox="753 1120 927 1256">  <p>CASAVA</p> </div> </div>	 <p style="text-align: center;"> Alignments and Variant Detection </p>

fastqが全てNと表記されてしまう

- ▶ シーケンスの結果、リードのクオリティが悪い、画像処理の結果、特定の塩基をコールできない場合などはNと表記される。



$$C = \frac{I_A}{I_A + I_B}$$

CHASTITY formula

- ▶ シーケンス結果は悪くなく、何らかの塩基をコールしているように見えるのに、結果は全てNになってしまう。

fastqが全てNと表記されてしまう

bcl2fastq2 Conversion Software

Guide

<code>--minimum-trimmed-read-length</code>	<p>Minimum read length after adapter trimming. bcl2fastq trims the adapter from the read down to the value of this parameter. If there is more adapter match below this value, then those bases are masked, not trimmed (replaced by N rather than removed).</p> <p><u>Default: 35</u></p>
<code>--mask-short-adapter-reads</code>	<p>This option applies when a read is trimmed to below the length specified by the <code>--minimum-trimmed-read-length</code> option (default of 35). These parameters specify the following behavior:</p> <p>If the number of bases left after adapter trimming is less than <code>--minimum-trimmed-read-length</code>, force the read length to be equal to <code>--minimum-trimmed-read-length</code> by masking adapter bases (replace with Ns) that fall below this length.</p> <p>If the number of ACGT bases left after this process falls below <code>--mask-short-adapter-reads</code>, mask all bases, resulting in a read with <code>--minimum-trimmed-read-length</code> number of Ns.</p> <p><u>Default: 22</u></p>

fastqが全てNと表記されてしまう

bcl2fastq2 Conversion Software Guide

`--minimum-trimmed-read-length`

Minimum read length after adapter trimming. bcl2fastq trims the adapter from the read down to the value of this parameter. If there is more adapter match below this value, then those bases are masked, not trimmed (replaced by N rather than removed).

Default: 35

`--mask-short-adapter-reads`

This option applies when a read is trimmed to below the length specified by the `--minimum-trimmed-read-length` option (default of 35). These parameters specify the following behavior:

If the number of bases left after adapter trimming is less than `--minimum-trimmed-`

アダプタートリムの設定で短いリードを読もうとしていませんか？

If the number of ACGT bases left after this process falls below `--mask-short-adapter-reads`, mask all bases, resulting in a read with `--minimum-trimmed-read-length` number of Ns.

Default: 22

本日の内容

- ▶ BasesSpaceの使用に関するお問い合わせ
 - fastq fileがダウンロードできない。
 - fastq fileがアップロードできない。
 - BaseSpace appsが使用できない。
 - 一般的によくあるお問い合わせ (tophat, cufflinksなど)
 - VariantStudioに関するお問い合わせ
- ▶ Bcl conversionに関するお問い合わせ
 - 全ての塩基でNとコールされてしまう
 - 解析が途中で停止し、fastqが生成されない

bcl conversionが途中でabortしてしまう

- ▶ bcl conversionを行ったが、conversionプロセスが途中で中断してしまう。
- ▶ 以下のようなメッセージがnohup.outなどのログに残る

Examples)

- Failed to read number of clusters from .../s_1_1101.bcl.gz
- Failed to read BCL file .../s_4_1103.bcl.gz
- Error: 2015-Jul-09 08:46:18: Invalid argument: /usr/.../BclReader.cpp(115)
: Throw in function unsigned int casava::alignment::BclReader::readClusterCount()'

途中でラン停止などのいつもと異なることはありませんでしたか？

シーケンサーとサーバー間の接続は中断されたりしませんでしたか？

以上の情報を添えてテクニカルサポートまでご一報ください。

bcl conversionが途中でabortしてしまう

- ▶ 状況を精査させていただき、対応させていただきたく存じます。まずは、下記のコマンドをランデータのルートディレクトリで走らせて頂き、ファイル構造をテキスト化していただだけませんか。

```
$ find | xargs ls -l | gzip -c > directory.structure.txt.gz
```

- ▶ このファイルが生成されたら、テクニカルサポートにお知らせください。こちらの情報を元に、各サイクルごとのbclファイルの生成状況を精査いたします。
- ▶ 恐れ入りますが、状況によってはfastqファイルの救出ができないことかもしれませんが、シーケンスの状況を鑑み、エラーの原因と今後についてご案内させていただければ幸いです。

参考資料

▶ イリミナサポートウェビナー

http://www.illumina.co.jp/events/webinar_japan/support_webinar.ilmn

[日本語でのBaseSpaceについての情報が充実しています]

▶ BaseSpace関連

https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/basespace/basespace-user-guide-15044182-e.pdf

[BaseSpaceの基本的な使い方はこちらに記載してあります]

https://support.illumina.com/downloads/basespace_core_apps_user_guides.html

[少し応用的ですが、BaseSpace appsの使用に際してはこちらが有用です]

▶ MiSeq Reporterやbcl2fastqなど、各種ソフトウェア

http://www.illumina.co.jp/support/sequencing/sequencing_software.ilmn

[こちらのページから全てのソフトウェアのユーザーガイドが入手できます]