

# RNA-Seqをもう一度： 情報解析をクラウドでより簡単・高速に！ DRAGEN™ RNA Pipelineのご紹介

2021年2月24日

テクニカルアプリケーションサイエンティスト 仁田原 翔太

M-JP-00001



# 本日の内容

1. RNA-Seq解析のワークフローについて
2. BaseSpace Sequence Hubの特徴
3. DRAGENの特徴
4. BaseSpace Sequence HubでのRNA-Seq解析
  - FASTQアップロードの方法について
  - BaseSpace DRAGEN Reference Builderの使い方
  - BaseSpace DRAGEN RNA Pipelineの使い方 / 結果の見方
  - BaseSpace DRAGEN Differential Expressionの使い方 / 結果の見方

# RNA-Seq解析のワークフローについて



# 日本語ウェビナー:これまでのRNA-Seqをはじめようシリーズ



【実験デザイン編 -これからRNA-Seqを始める方に-】

<https://jp.illumina.com/events/webinar/2018/webinar-180131-j.html>

【RNA-Seqをはじめよう ライブラリー調製編:絶対に失敗しないライブラリー調製】

<https://jp.illumina.com/events/webinar/2018/webinar-180228-j.html>

➡ RNA-Seqのライブラリー調製キット  
TruSeq Stranded mRNA/Total RNA Kitsをご紹介

**NEW!**



【RNA-Seqをもう一度:新しいライブラリー調製キットでより簡単・確実に】

<https://jp.illumina.com/events/webinar/2020/webinar-201125-j.html>

➡ RNA-Seqのライブラリー調製キット  
新製品Illumina RNA Prep Kitsをご紹介



【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】

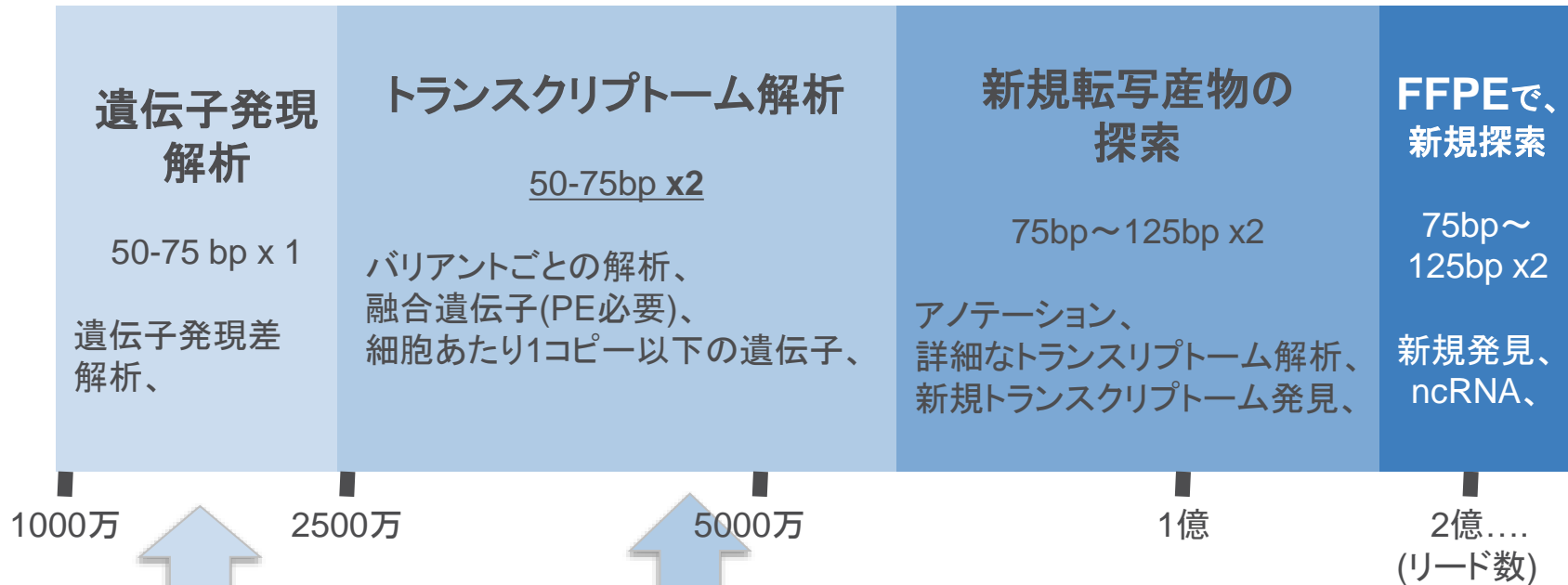
<https://jp.illumina.com/events/webinar/2018/webinar-180328-j.html>

➡ RNA-Seq Alignmentが開発終了になるため、後継のアプリ  
BaseSpace Sequence Hub, DRAGEN RNA Pipelineをご紹介  
(本日のウェビナーでご紹介)



# RNA-Seq の目的と実験デザイン (ヒト・マウスの例)

サンプルあたりのコスト:



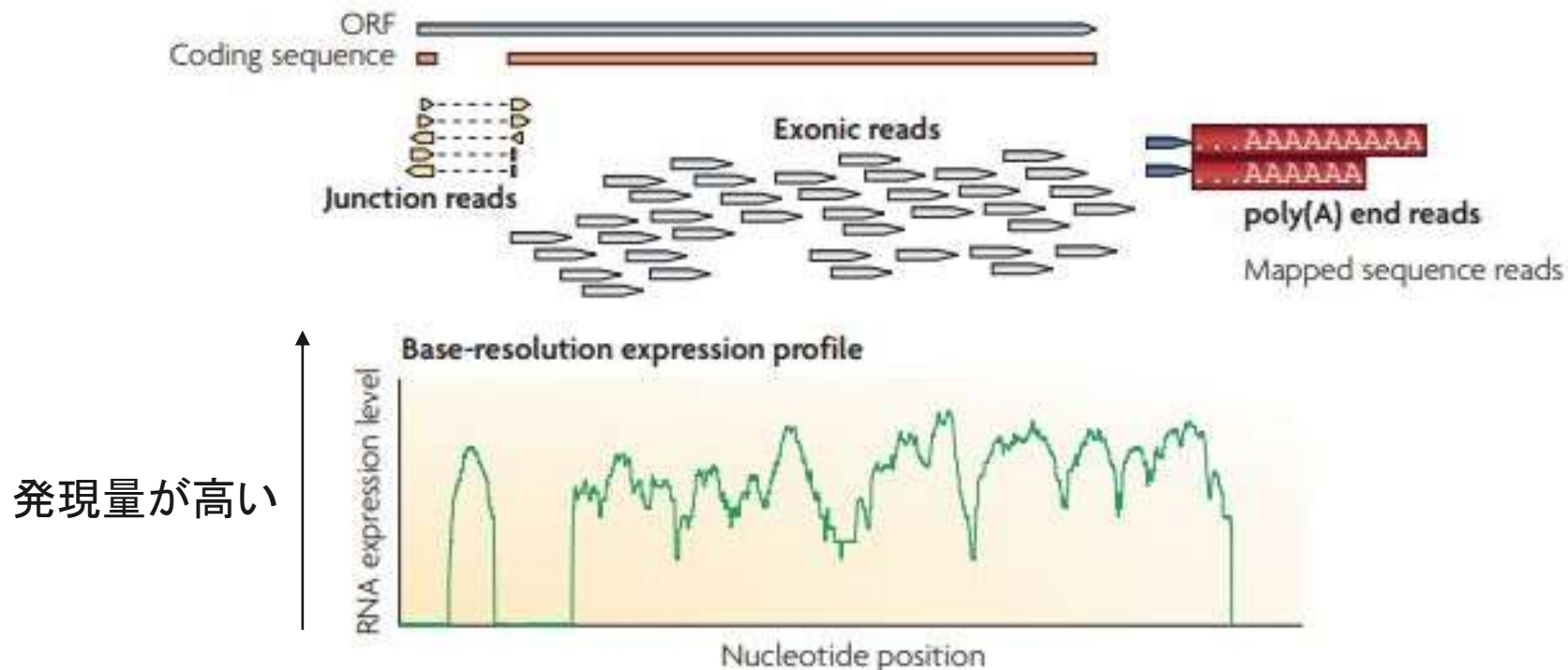
単純な遺伝子発現解析:  
mRNA-seq: 1,000万リード以上  
Total RNA-seq: 2500万リード以上

トランスクリプトごとの正確な解析:  
mRNA-seq: 2,500万リードx2以上  
Total RNA-seq: 5,000万リードx2以上

Considerations for RNA-Seq read length and coverage:

<https://jp.support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html>

# RNA-Seq 遺伝子発現頻度解析: マッピング、カウント



- 得られた配列を参照配列に**マッピング**
- ✓ **マッピング**: 得られたリードが参照配列と一致する部分を探す方法

マッピングされたリードの数を数えて、発現量を算出、比較することができる

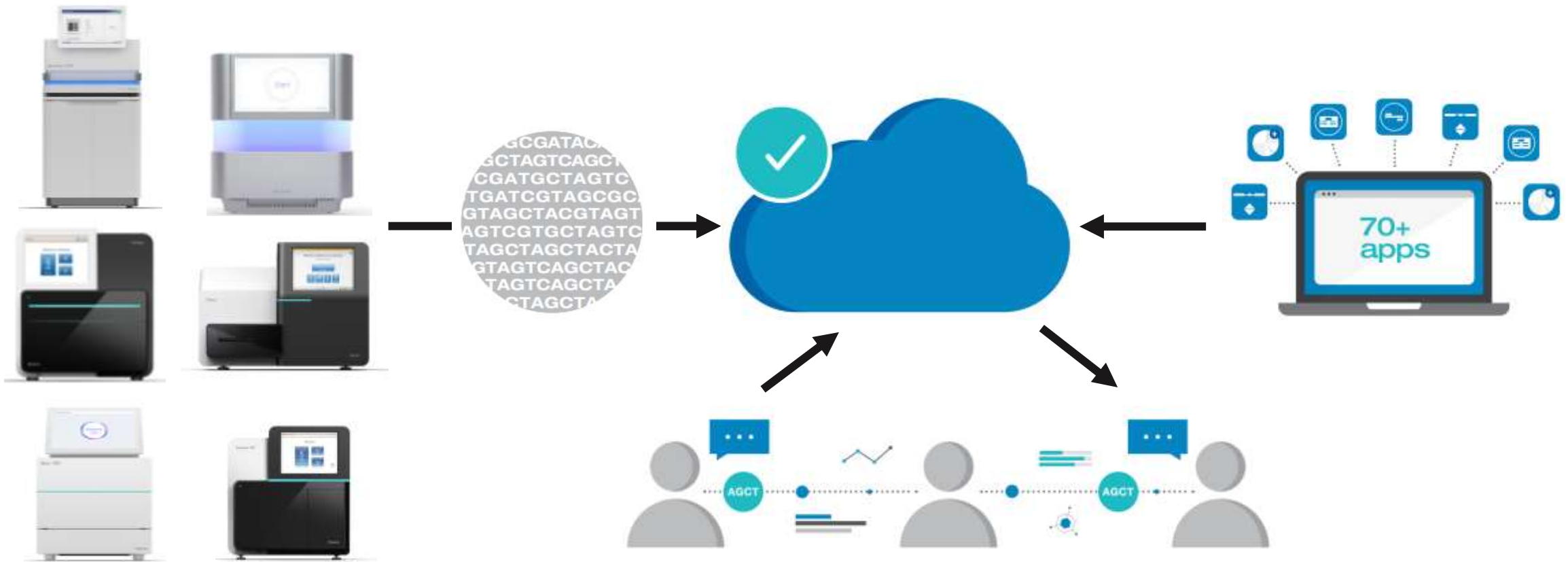
RNA-Seq: a revolutionary tool for transcriptomics

<https://doi.org/10.1038/nrg2484>

# BaseSpace Sequence Hubの特徴



# BaseSpace Sequence Hub, クラウドベースの解析プラットフォーム



- シーケンサーからランデータをアップロード、FASTQ生成を実行
- 解析のアプリケーションを実行
- 研究者同士でランや解析結果を共有



# BaseSpace Sequence Hubを使う利点



Amazon Web Service上に構築されたクラウドサービス  
- ご自身でサーバーなどインフラ構築する必要なし



必要に応じてスケールアップ可能なストレージ (有償)



HIPAA、ISO27001、ISO13485認証を受けたセキュリティ



DRAGENを含めクリックで実行可能なさまざまなアプリケーション  
- RNA-Seq, 変異解析、アセンブリ、16S metagenomicsなど  
- 複雑なコマンドラインを使う必要なし

# BaseSpace Sequence Hubのライセンス登録について

## BaseSpace Sequence Hub サービス一覧

	ベーシック	プロフェッショナル	エンタープライズ
年間ライセンス			
価格	無償	有償	有償
ストレージ利用			
ライセンス料に含まれる標準ストレージ	1TB まで	1TB まで	1TB まで
1TB を超える追加ストレージ (有償) <sup>*1</sup>	不可	可	可
アプリ利用			
初期および毎更新時の付与ポイント	対象外 <sup>*2</sup>	500 iCredit	500 iCredit
無償アプリ利用	可	可	可
有償アプリ利用 <sup>*1</sup>	不可	可	可

\*1. 月毎のご利用量 (追加ストレージ、アプリ) に応じて、翌月にご請求する事後従量精算制と、ご利用分に十分な iCredit を予め購入しアカウントにチャージいただく事前購入制のいずれかでお支払いいただけます。ご利用量の単位 1 iCredit は 125 円で換算します。(2020 年 4 月 13 日現在)

\*2. アカウント作成時には、30 日間有効のトライアルポイント (250 iCredit) が付与されます。


<https://basespace.illumina.com/home/index> から登録して使用可能です。

30日間、250 icreditの試用期間が付与されます。

試用期間終了後はプロフェッショナルライセンス (1年間、62,500円、500 iCredit付) に登録いただく必要があります。

(価格などの情報は2021年2月現在の情報となります)

# BaseSpace Sequence Hub iCredit 試算例



DRAGEN Germline  
Illumina, Inc.

[Bookmark this app](#) [Help](#)

Pricing


Compute cost ..... 4.80 iCredits  
per node hour

アプリケーション	コスト	詳細
ヒト全ゲノムシーケンス	7 iCredit	120 Gbase, x30 カバレッジ DRAGEN Germline
RNA-Seq トランスクリプトーム	2 iCredit	5000万リード DRAGEN RNA pipeline

アプリケーションごとに計算時間あたりの利用料金が設定されており  
計算時間に応じてiCreditが計算されます  
シーケンスデータ量により計算時間は前後しますので目安としてお考え下さい  
1 iCreditは125円となります (2021年2月現在)

<https://jp.illumina.com/products/by-type/informatics-products/basespace-sequence-hub.html>

# RNA-Seq Alignmentは2021年6月末で開発終了となります



RNA-Seq Alignment  
Illumina, Inc.

[Bookmark this app](#) [Help](#)

The RNA-Seq Alignment workflow performs the following main functions

- Read mapping using the STAR aligner
- Quantification of reference genes and transcripts using salmon

Pricing  
Compute cost ..... 3.00 iCredits per node hour

Version  
2.0.2

[LAUNCH APPLICATION](#)

## What's New

- **This app will be obsoleted by June 30, 2021. Illumina recommends usage of the DRAGEN RNA Pipeline app, whose outputs can be used with DRAGEN Differential Expression for differential expression analysis.**

Please visit these pages for more information:

- <https://basespace.illumina.com/apps/10500490/DRAGEN-RNA-Pipeline>
- <https://basespace.illumina.com/apps/10502492/DRAGEN-Differential-Expression>

後継のアプリケーションとしてDRAGEN RNA Pipeline, DRAGEN Differential Expressionを推奨しております。

本ウェビナーでは後継の両アプリケーションについてご紹介します。

# DRAGENの特徴





# DRAGENの特徴: NGS解析に特化したサーバー

## 1. 高速で出来る2次解析

- ✓ ヒト全ゲノム 30カバレッジのデータを25分で解析
- ✓ ヒトエキソーム 100カバレッジのデータを8分で解析

## 2. より正確な変異検出

- ✓ Germline / Somatic 変異を高い感度、特異度で検出
- ✓ ベンチマークテストで他のソフトウェアよりも高い正確性を発揮

## 3. 経済性とフレキシビリティ

- ✓ オンサイトサーバとクラウドを選択可能
- ✓ NextSeq 1000/2000には装置内にDRAGENを搭載



Illumina DRAGEN™ どらげん？どらじえん？ - やってきたNGS 高速解析の竜

<https://jp.illumina.com/events/webinar/2019/webinar-product-dragen-190625-j.html>

DRAGEN v3.7: Single Cell RNA, PrecisionFDA Accuracy Gains, and More

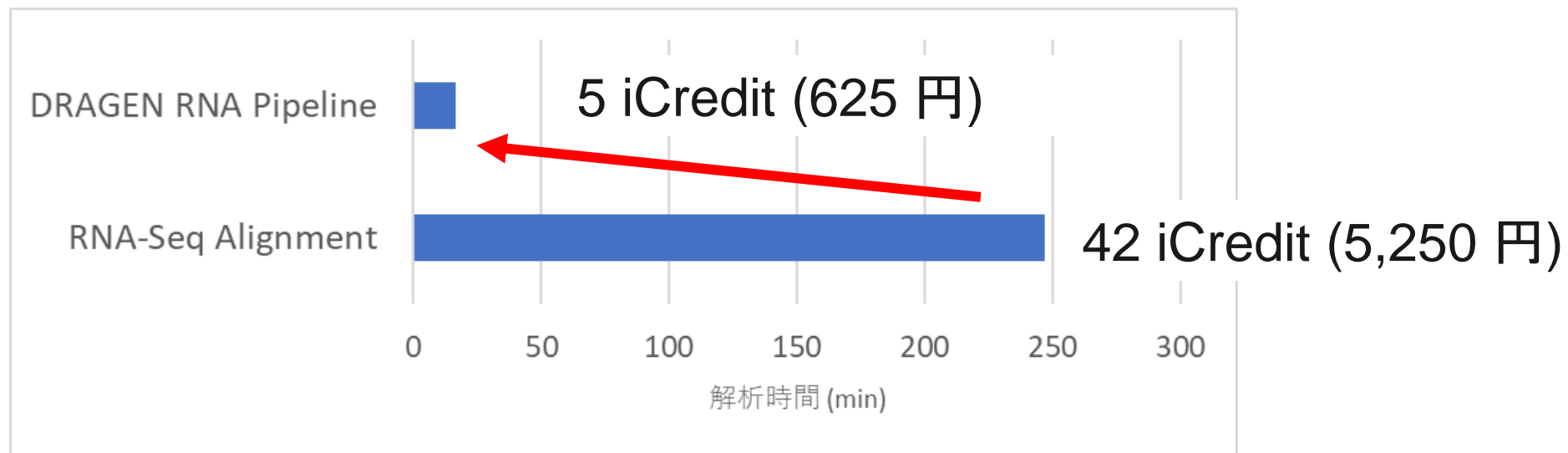
<https://blog.software.illumina.com/2020/11/10/dragen-v3-7-single-cell-rna-precisionfda-accuracy-gains-and-more/>

# DRAGEN RNA pipeline とRNA-Seq Alignment 解析時間の比較

サンプル条件: ヒトmRNAサンプル4検体

リード数: 9,500万リード、11,000万リード、12,000万リード、13,000万リード

リード長: 2 x 75 bp



DRAGENを使うことにより解析時間、解析コストを圧倒的に少なくすることが出来る。

# DRAGENの特徴: 2次解析ソフトウェア一覧

アプリケーション	BaseSpace Sequence Hub	オンサイトサーバー	NextSeq1000/2000
Demultiplexing BCL Convert	○	○	○
Map and align	○	○	○
RNA-Seq (gene fusion and quantification)	○	○	○
Single Cell RNA	×	○	○
Whole Genome / Whole Exome (Germline / Somatic)	○	○	Germline Only
Methylation	○	○	○
Joint Genotyping	○	○	○

2021年2月現在の情報となります。

# BaseSpace Sequence HubでのRNA-Seq解析



# BaseSpace Sequence HubでのRNA-Seq解析の流れ



FASTQをアップロード、  
もしくは  
ランをBaseSpaceに転送、  
クラウド上でFASTQ生成



DRAGEN Reference Builder  
Edico Genome Inc.

[Bookmark this app](#) [Help](#)

リファレンス配列を準備する(オプション)



DRAGEN RNA Pipeline  
Edico Genome Inc.

[Bookmark this app](#) [Help](#)

リファレンス配列にマッピングし、  
遺伝子発現頻度や融合遺伝子を解析する



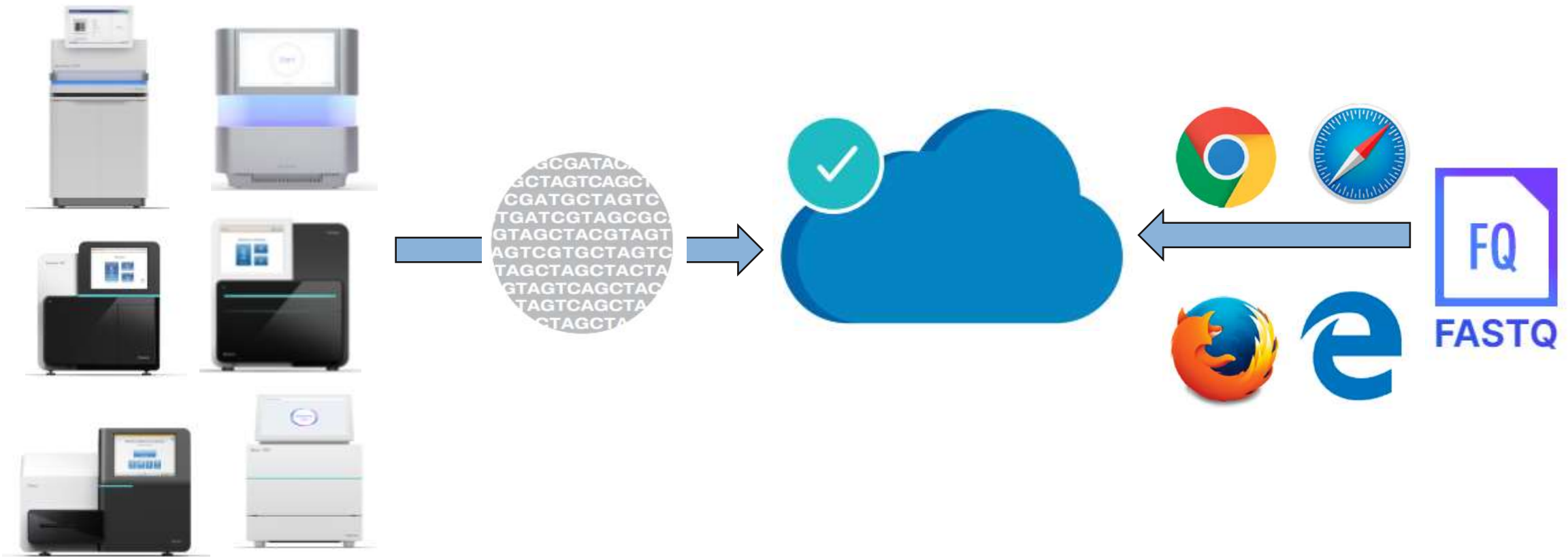
DRAGEN Differential Expression  
BaseSpace Labs

[Bookmark this app](#) [Help](#)

遺伝子発現頻度をグループごとに比較し、  
発現量の違う遺伝子リストを出す



# BaseSpace Sequence HubのFASTQアップロードについて

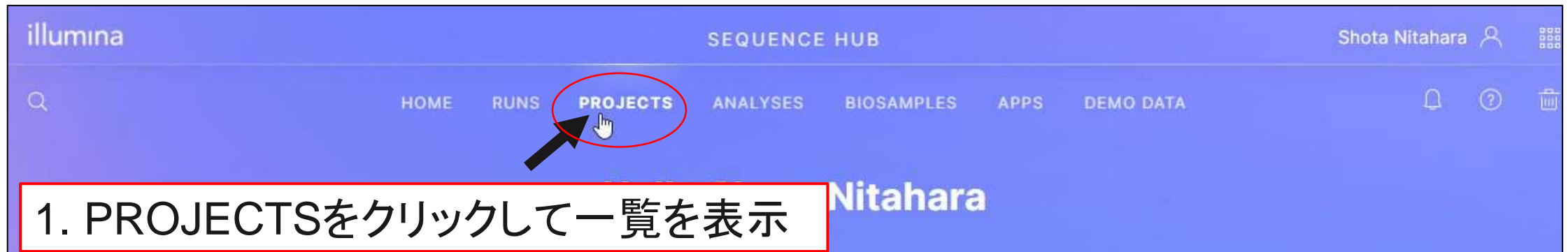


1. ラン開始時にランデータをBaseSpaceへアップロードする設定にする
2. FASTQファイルをブラウザ経由でアップロードする

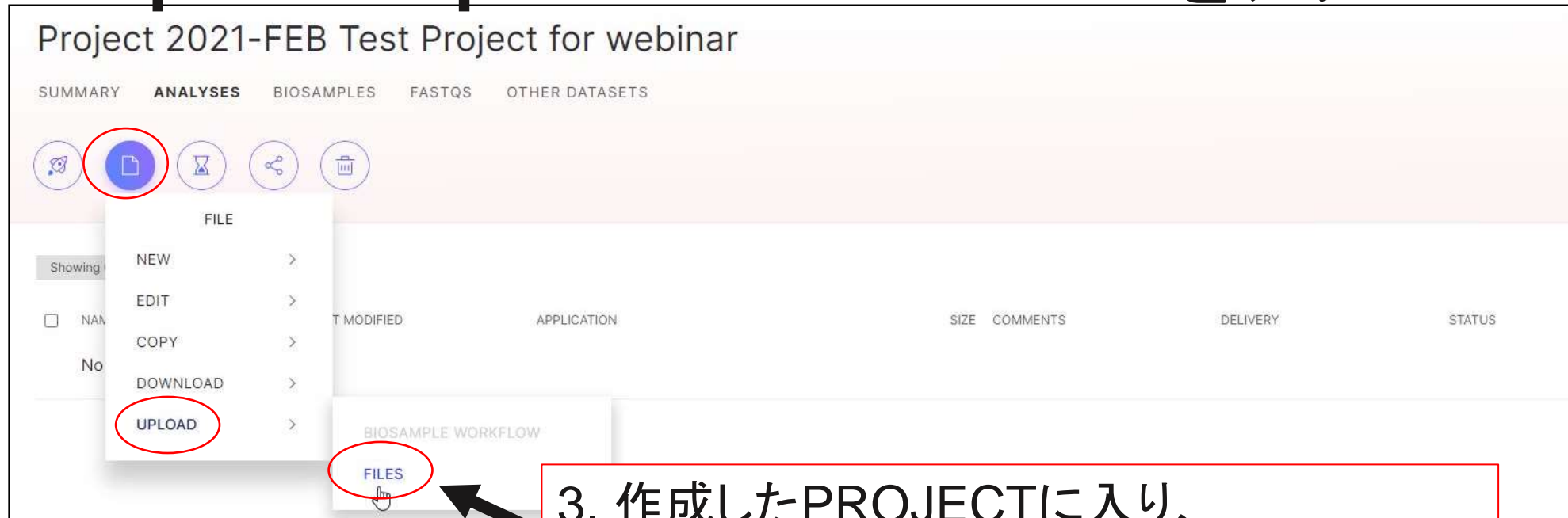
# FASTQアップロードの方法について



# BaseSpace Sequence HubにFASTQをアップロード (1/6)



# BaseSpace Sequence HubにFASTQをアップロード (2/6)

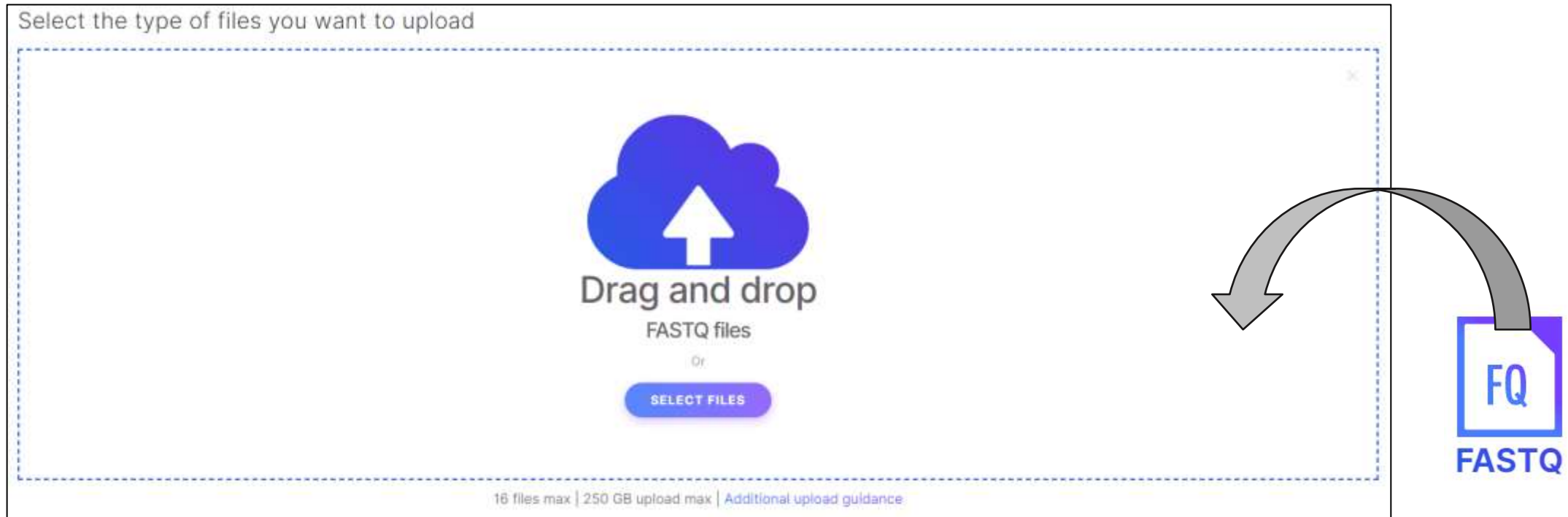


3. 作成したPROJECTに入り、  
アイコン > UPLOAD > FILESと選択する



4. アップロードするファイルの種類を選ぶ

# BaseSpace Sequence HubにFASTQをアップロード (3/6)



5. アップロードするFASTQファイルをドラッグ & ドロップする
  - ※ 1度に1サンプルのデータをアップロードしてください
  - ※ ペアエンドのデータの場合、R1/R2 両方のファイルをドラッグ & ドロップしてください



# BaseSpace Sequence HubにFASTQをアップロード (4/6)

Upload Files

Projects: 2021-FEB Test Project for webinar

### FASTQ Dataset

Default project  
2021-FEB Test Project for webinar  
(Not editable)

Save upload to\*  
**SELECT BIOSAMPLE**

Library name\*  
(No special characters, 60 max)

Library prep kit\*  
SELECT

FILE NAME	FILE TYPE	SIZE	UPLOAD STATUS
ForTest01_S1_L001_R1_001.fastq.gz	FASTQ	1.39 GB	
ForTest01_S1_L001_R2_001.fastq.gz	FASTQ	1.39 GB	

Drop more files or select files to add to this session

2 Files Total (16 max) | 3 GB Total (250 GB max) | [Additional upload guidance](#) | [Illumina FASTQ Guidance](#)

CANCEL FINISH UPLOAD

6. FASTQファイルのアップロード状況  
アップロードを完了するには、画面左側から  
サンプル情報を入れる必要がある

# BaseSpace Sequence HubにFASTQをアップロード (5/6)

Upload Files

Projects: 2021-FEB Test Project for webinar

---

FASTQ Dataset

Default project  
2021-FEB Test Project for  
(Not editable)

Save upload to \*  
**SELECT BIOSAMPLE**

Library name \*  
(No special characters, 60 max)

Library prep kit \*  
SELECT

7. BIOSAMPLE = Sample IDを入れる

Showing 0 of 0

BIOSAMPLE	DEFAULT PROJECT	UPDATED	OWNER	LAB STATUS	BIOSAMPLE STATUS
No matching records found					

NEW CANCEL SELECT

8. NEWから新しいBiosample名を入力する  
- Sample IDとなるため、ユニークな名前をつける  
9. Biosampleを選択する

BaseSpaceではサンプルをbiosample単位で管理している  
BiosampleにFASTQファイルを紐づけて管理する

Biosample

# BaseSpace Sequence HubにFASTQをアップロード (6/6)

Upload Files  
Projects: 2021-FEB Test Project for webinar

FASTQ Dataset

Default project  
2021-FEB Test Project for webinar  
(Not editable)

Save upload to \*  
SELECT BIOSAMPLE

Library name \*  
StrandedPrep\_2021Feb  
(No special characters, 60 max)

Library prep kit \*  
IDT-ILMN DNA-RNA UD INDEXE...

FILE NAME	FILE TYPE	SIZE	UPLOAD STATUS
ForTest01_S1_L001_R2_001.fastq.gz	FASTQ	1.39 GB	
ForTest01_S1_L001_R1_001.fastq.gz	FASTQ	1.39 GB	

2 Files Total (16 max) | 3 GB Total (250 GB max) | Addit

CANCEL FINISH UPLOAD

10. アップロードステータスに  
チェックが付くのを待つ

11. biosample を選択  
12. Library name の情報を入れる  
13. Library prep kitの情報を入れる  
Library name, Library prep kitの情報  
はメタデータであり、解析には影響しない

14. FINISH UPLOADをクリックしアップロードを完了させる

# FASTQファイルアップロード時のエラーについて



# よくあるFASTQアップロード時のエラー: その1

## File Upload Error

- Error: unable to extract metadata from reference file. Please ensure the FASTQ file adheres to the [Illumina FASTQ Guidance](#).

FASTQファイル名からアンダーバーで分けて切り出したサンプルの情報を取得できないことによるエラー

**解決策: SampleNameの部分にアンダーバーは避け、下記の形式にする**

**SampleName**\_SampleNumber\_Lane\_Read\_FlowCellIndex.fastq.gz

○ **Normal01**\_S1\_L001\_R1\_001.fastq.gz

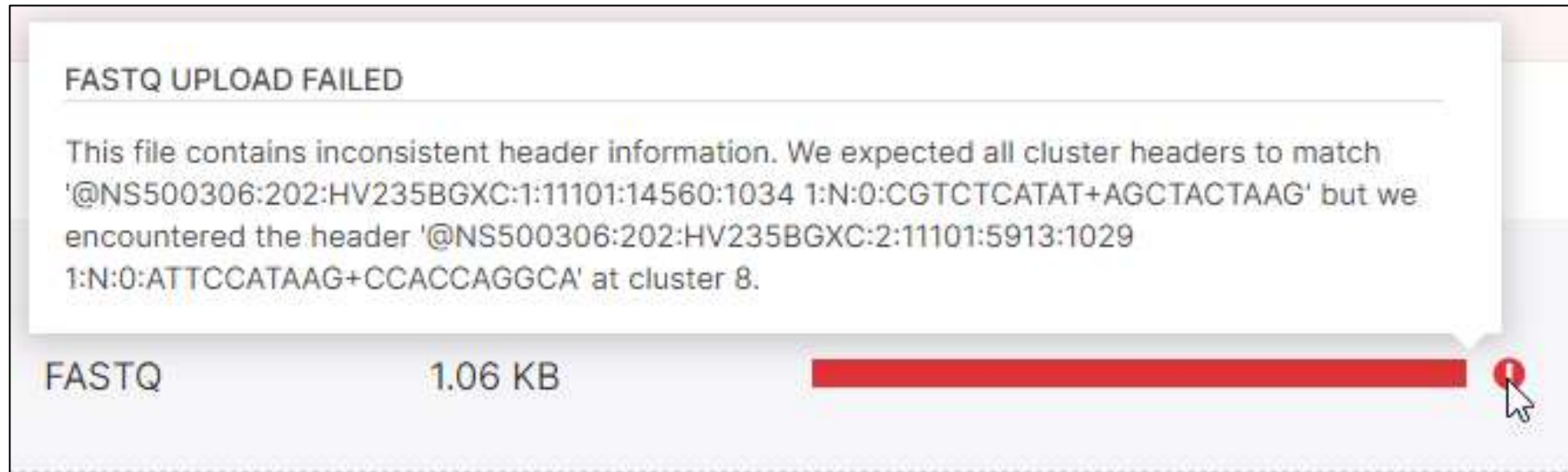
× **Normal\_01**\_S1\_L001\_R1\_001.fastq.gz

FASTQ File Upload Requirements

[https://support.illumina.com/help/BaseSpace\\_Sequence\\_Hub/Source/Informatics/BS/UploadFastqReq\\_swBS.htm](https://support.illumina.com/help/BaseSpace_Sequence_Hub/Source/Informatics/BS/UploadFastqReq_swBS.htm)



# よくあるFASTQアップロード時のエラー: その2



FASTQファイルに複数のラン、複数のレーンのデータが含まれていることによるエラー

## 解決策

1. bcl2fastqを使用する場合、`--no-lane-splitting`オプションを使わない
2. FASTQファイルを結合し、1つのファイルとしない

# よくあるFASTQアップロード時のエラー: その3



複数のサンプルのFASTQファイルを一度にアップロードしようとすることによるエラー

**解決策: 一度に1サンプルのFASTQファイルのみをアップロードする**

その他のエラーや対応策でも解決しない場合などお困りの際はテクニカルサポート (techsupport@illumina.com)までご連絡ください。

# BaseSpace DRAGEN Reference Builderの使い方



# BaseSpace Sequence HubでのRNA-Seq解析の流れ

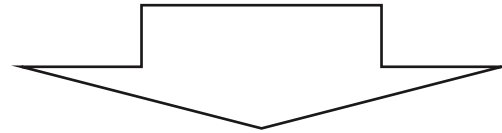


FASTQをアップロード、  
もしくは  
ランをBaseSpaceに転送、  
クラウド上でFASTQ生成



DRAGEN Reference Builder  
Edico Genome Inc.  
[Bookmark this app](#) [Help](#)

リファレンス配列を準備する(オプション)



DRAGEN RNA Pipeline  
Edico Genome Inc.  
[Bookmark this app](#) [Help](#)

リファレンス配列にマッピングし、  
遺伝子発現頻度や融合遺伝子を解析する



DRAGEN Differential Expression  
BaseSpace Labs  
[Bookmark this app](#) [Help](#)

遺伝子発現頻度をグループごとに比較し、  
発現量の違う遺伝子リストを出す

# DRAGENで使用するリファレンス配列について

- DRAGENではリファレンス配列をハッシュテーブルとして準備する必要がある
- Humanの主要なリファレンス配列のハッシュテーブルは予め用意がある
- 非ヒト生物などのほかのカスタムリファレンス配列は**DRAGEN Reference Builder**で作成することで使えるようになる

BasespaceDRAGEN RNA Pipelineで予め用意のあるHuman Reference

- Human (UCSC hg19 Alt-Aware)
- Human (UCSC hg19 No Alts, No decoys)
- Human (Ensembl GRCh37)
- Human (hg38 with HLAs)
- Human (hg38 without HLAs)
- Human (hg38 No Alts, with decoys)
- Human (Ensembl hs37d5: GRCh37+decoy)



# DRAGEN Reference Builder Appの使い方

Reference

HOME RUNS PROJECTS ANALYSES BIOSAMPLES **APPS** DEMO DATA

1. キーワードで検索でサーチ  
もしくはアプリケーション一覧から探す

Bowtie  
David Sachs  
Map reads to a reference using Bowtie.

BWA Aligner  
BaseSpace Labs  
Align samples with the BWA-MEM aligner to a reference genome, including custom references created from imported FASTA ...More

DRAGEN Reference Builder  
Edico Genome Inc.  
This app accepts FASTA files, and builds the proprietary reference used by the DRAGEN apps.

Categories: Resequencing

2. 目的のアプリケーションをクリック

DRAGEN Reference Builder  
Edico Genome Inc.  
Bookmark this app Help

FOR RESEARCH USE ONLY

Overview  
This app accepts a FASTA file and builds the proprietary DRAGEN Reference (also referred to as a Hash Table). This version of the app creates a V8 Hash Table. This reference can be used as input by any DRAGEN 3.5.x, 3.6.x, and 3.7.x apps which accept a Custom Reference File. The DRAGEN app typically provide built-in support for hg19, hg38 (HLA and non-HLA), GRCh37, and hg3745. The Reference Builder app is meant for users who need to generate a custom, non-human or non-standard reference. [READ MORE](#)

Version  
3.7.5

LAUNCH APPLICATION

3

3. ソフトウェアのバージョンを確認して  
LAUNCH APPLICATIONをクリック

# DRAGEN Reference Builder App 設定 1/2

Configuration

Analysis Name ⓘ  
DRAGEN Reference Builder 01/20/2021 3:01:22

Save Results To ⓘ  
[SELECT PROJECT](#)

FASTA Input file ⓘ  
[SELECT DATASET FILE\(S\)](#)

Output Filename Prefix ⓘ  
\_\_\_\_\_

SAM Liftover File ⓘ  
None ▼

Custom SAM Liftover File ⓘ  
[SELECT DATASET FILE\(S\)](#)



DRAGEN Reference Builder

Edico Genome Inc.

[Bookmark this app](#)

[Help](#)

4. 解析の名前を設定する

5. 保存するProjectを選択する

6. カスタムのリファレンス配列 FASTAを選択する

7. アウトプットの頭につける名前を設定する (Option)

8. Liftoverのファイルを指定する (Option)  
アセンブリ時に使用していなければ使用する必要はない

# DRAGEN Reference Builder App 設定 2/2

**Liftover Validation** ⓘ

Liftover Validation

Seed Length ⓘ

Check this, to build the RNA hash table (in addition to the DNA hash table)

**RNA** ⓘ

Include RNA Data in Reference

**CNV** ⓘ

Include CNV Data in Reference

**Methylation** ⓘ

Include Methylation Data in Reference

Advanced Settings (WARNING: for experienced users only)

Additional Arguments

**LAUNCH APPLICATION**



DRAGEN Reference Builder

Edico Genome Inc.

[Bookmark this app](#)

[Help](#)

9. RNAの解析用のデータを作成するオプション

10. CNV, Methylationの解析用のデータを作成するオプション

11. その他のAdvanced Option  
(変更の必要性は少ない)

12. 設定を確認して、Launch Applicationをクリックしてアプリケーションを開始  
解析が完了するとリファレンスが使える状態になる

# BaseSpace DRAGEN RNA Pipelineの使い方



# BaseSpace Sequence HubでのRNA-Seq解析の流れ

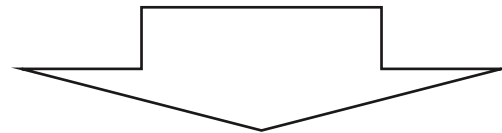


FASTQをアップロード、  
もしくは  
ランをBaseSpaceに転送、  
クラウド上でFASTQ生成



DRAGEN Reference Builder  
Edico Genome Inc.  
[Bookmark this app](#) [Help](#)

リファレンス配列を準備する(オプション)



DRAGEN RNA Pipeline  
Edico Genome Inc.  
[Bookmark this app](#) [Help](#)

リファレンス配列にマッピングし、  
遺伝子発現量や融合遺伝子探索をする

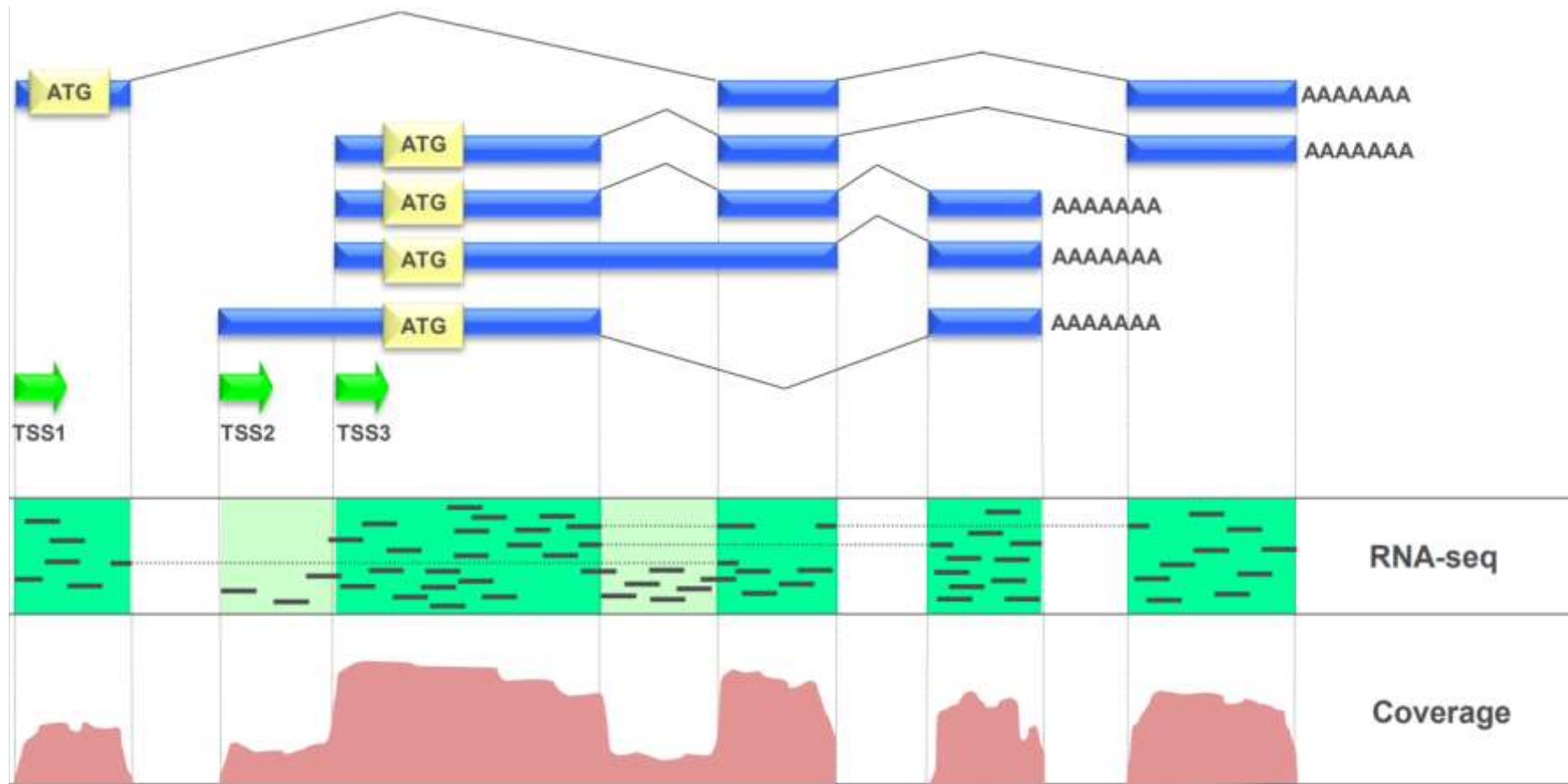


DRAGEN Differential Expression  
BaseSpace Labs  
[Bookmark this app](#) [Help](#)

遺伝子発現量をグループごとに比較し、  
発現量の違う遺伝子リストを出す

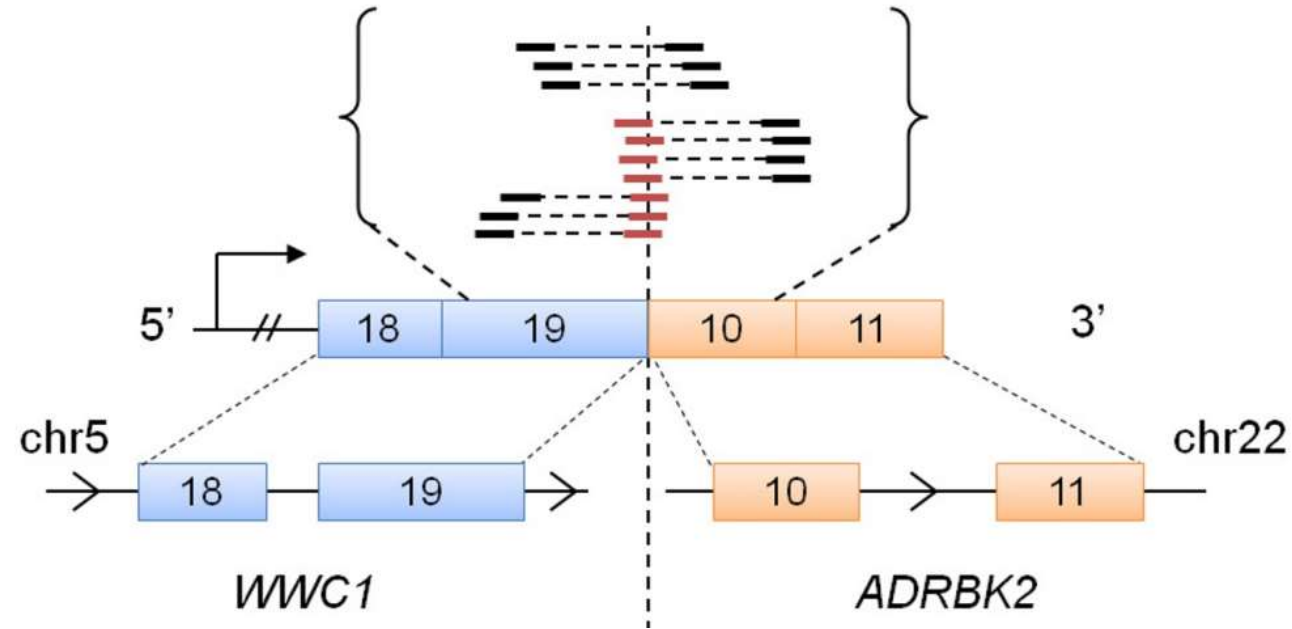


# RNA-Seq 遺伝子発現頻度解析: マッピング、カウント



RNA-Seqのマッピングでは、リードを分割し、Exonごとにマッピングする  
Exonへのマッピング状況からアイソフォームごとの発現量を算出できる

# RNA-Seq 融合遺伝子探索解析: マッピング、split-readsの検出



- ✓ 融合遺伝子: 染色体の再構成で異なる遺伝子同士が繋がった遺伝子

ペアエンドのリードがそれぞれ異なる遺伝子にマッピングされたリードやジャンクション部分にまたがってマッピングされたリード(split-reads)の情報を使い、融合遺伝子の検出を行なうことができる

Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines

<https://doi.org/10.1186/1755-8794-4-75>

# BaseSpace DRAGEN RNA Pipelineの使い方



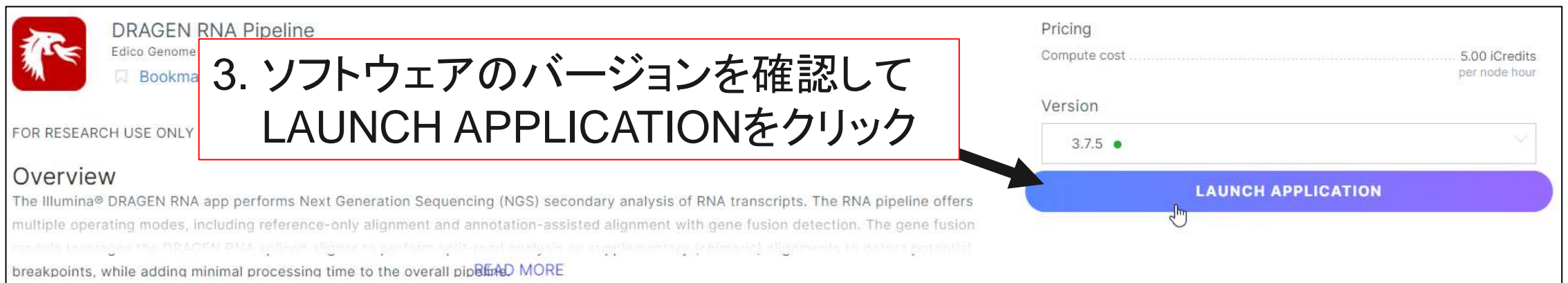
# BaseSpace DRAGEN RNA Pipelineの使い方



1. キーワードで検索でサーチ  
もしくはアプリケーション一覧から探す

2. 目的のアプリケーションをクリック

The screenshot shows a search interface with the text 'DRAGEN RNA' in the search bar. Below the search bar, two application cards are visible. The first card is for 'DRAGEN RNA Pathogen Detection' by Illumina Inc. The second card is for 'DRAGEN RNA Pipeline' by Edico Genome Inc. An arrow points from the first instruction to the search bar, and another arrow points from the second instruction to the 'DRAGEN RNA Pipeline' card.



3. ソフトウェアのバージョンを確認して  
LAUNCH APPLICATIONをクリック

The screenshot shows the application page for 'DRAGEN RNA Pipeline' by Edico Genome Inc. The page includes a 'FOR RESEARCH USE ONLY' warning, an 'Overview' section, and a 'Pricing' section. The 'Pricing' section shows a compute cost of 5.00 iCredits per node hour. The 'Version' section shows a dropdown menu with '3.7.5' selected. A large blue button labeled 'LAUNCH APPLICATION' is prominently displayed at the bottom right. An arrow points from the third instruction to the 'LAUNCH APPLICATION' button.

# BaseSpace DRAGEN RNA Pipeline オプション設定 1/4

Configuration

Analysis Name  
DRAGEN RNA Pipeline 01/20/2021 3:21:47

Save Results To  
**SELECT PROJECT**

2021-FEB Test Project for webinar

Biosample  
**SELECT BIOSAMPLE(S):**

UHRmRNA20131205: Unknown

BrainmRNA20131202: Unknown

1. 解析の名前を設定する

2. 保存するProjectを選択する

3. 解析対象のbiosampleを選択する



# BaseSpace DRAGEN RNA Pipeline オプション設定 2/4



RNA-seq library type:

- A - Auto-Detect, the default. For this value, DRAGEN examines the first reads/pairs in the dataset to automatically determine the correct library type.
- IU - Paired-end unstranded library.
- ISR - Paired-end stranded library in which read2 matches the transcript strand (eg, TruSeq RNA).
- ISF - Paired-end stranded library in which read1 matches the transcript strand.
- U - Single-end unstranded library.
- SR - Single-end stranded library in which reads are in reverse orientation to the transcript strand (eg, TruSeq RNA).
- SF - Single-end stranded library in which reads match the transcript strand.

```
--rna-quantification-library-type
```

RNA-Seq Alignmentなどのアプリでは、RNAの方向性を選択するオプションがある  
DRAGEN RNA Pipelineでも同じオプションがあるが、  
Auto-Detectのオプションがあるため入力する必要はない

# DRAGEN RNA Pipeline リファレンス-ヒトゲノムの場合 3-1/4

The screenshot shows the DRAGEN RNA Pipeline configuration interface. It includes sections for Reference, Custom Reference File, Gene Fusion, RNA Quantification, Annotation-Assisted Alignment, Custom Gene Annotation File, and Map/Align Output. Red boxes and arrows highlight specific configuration options, with corresponding numbered instructions in Japanese text boxes.

**Reference**  
Human (hg38 No Alts, with decoys)

**Custom Reference File**  
SELECT DATASET FILE(S)

**Gene Fusion**  
 Enable Gene Fusion Detection

**RNA Quantification**  
 Enable RNA Quantification

**Annotation-Assisted Alignment**  
 Enable Annotation-Assisted Alignment

**Custom Gene Annotation File**  
SELECT DATASET FILE(S)

**Map/Align Output**  
BAM

**Reference**  
Human (hg38 No Alts, with decoys)  
Human (UCSC hg19 Alt-Aware)  
Human (UCSC hg19 No Alts, No decoys)  
Human (Ensembl GRCh37)  
Human (hg38 Alt-Aware, with HLAs)  
Human (hg38 Alt-Aware, No HLAs)  
**Human (hg38 No Alts, with decoys)**  
Human (Ensembl hs37d5: GRCh37+decoy)  
Custom

**Map/Align Output**  
BAM  
**BAM**  
CRAM  
None

4. 使用するリファレンス配列を選ぶ

5. 融合遺伝子探索、発現量算出を行なうオプションを選ぶ  
いずれかを有効にするとAnnotation-Assisted Alignment  
を有効にする必要がある

6. Map/Alignの際のアウトプットを  
BAM/CRAM/出力しない から選ぶ

Use Only. Not for use in diagnostic procedures.



# DRAGEN RNA Pipeline リファレンス-カスタムゲノム 3-2/4

Reference ⓘ  
Custom

Custom Reference File ⓘ  
SELECT DATASET FILE(S)  
TAIR10-anchored.v8.tar

4-1. 使用するリファレンス配列で”Custom”を選ぶ

4-2.  
DRAGEN Reference Builderで作成したデータを選ぶ

Gene Fusion ⓘ  
 Enable Gene Fusion Detection

RNA Quantification ⓘ  
 Enable RNA Quantification

Annotation-Assisted Alignment ⓘ  
 Enable Annotation-Assisted Alignment

Custom Gene Annotation File ⓘ  
SELECT DATASET FILE(S)  
genes.gtf

5. 融合遺伝子探索、発現量算出を行なうオプションを選ぶ  
いずれかを有効にするとAnnotation-Assisted Alignmentを有効にする必要がある

Map/Align Output ⓘ  
BAM

Map/Align Output ⓘ  
BAM  
CRAM  
None

6. Map/Alignの際のアウトプットを  
BAM/CRAM/出力しない から選ぶ

Use Only. Not for use in diagnostic procedures.

# BaseSpace DRAGEN RNA Pipeline オプション設定 4/4



Advanced Settings (General)

Advanced Settings (Gene Fusion, Quantification)

Additional Arguments

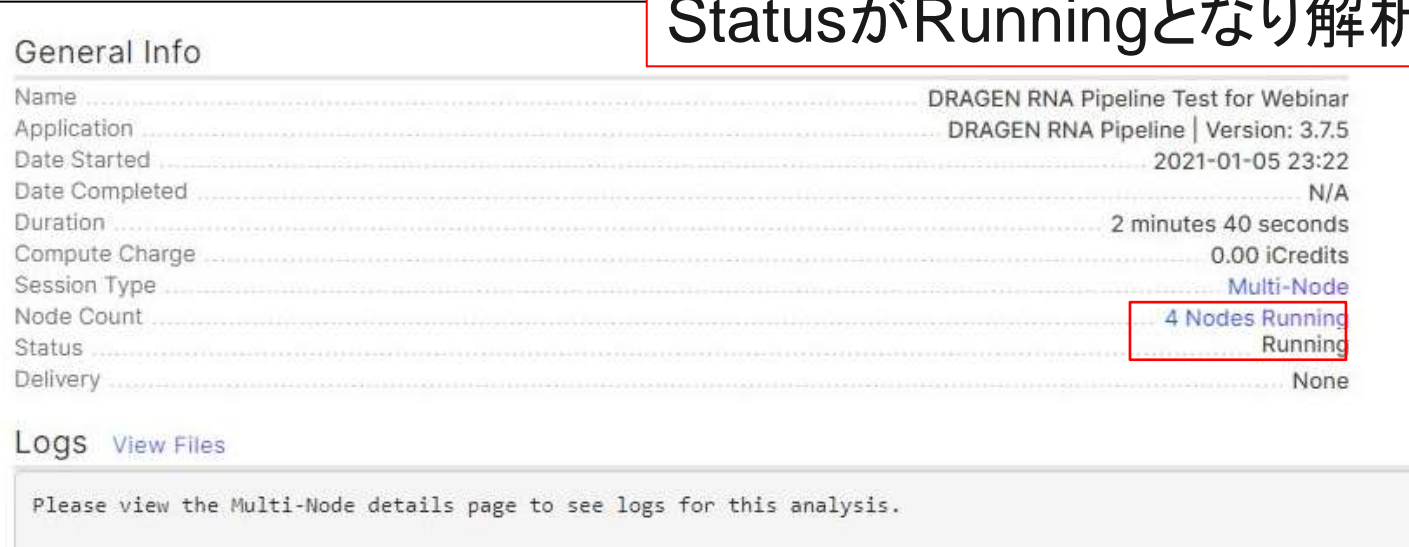
7. Advanced Option  
基本的に変更する必要はなし

8. Launch Applicationをクリックし、アプリをスタートさせる

LAUNCH APPLICATION

Detailed description: This block shows a screenshot of the 'Advanced Settings' menu in BaseSpace. Three menu items are listed: 'Advanced Settings (General)', 'Advanced Settings (Gene Fusion, Quantification)', and 'Additional Arguments'. A red box highlights the first two items, with an arrow pointing to them from a text box labeled '7. Advanced Option' which states '基本的に変更する必要はなし' (Basically, no need to change). Below the menu items is a blue 'LAUNCH APPLICATION' button, which is circled in red. An arrow points from a text box labeled '8. Launch Applicationをクリックし、アプリをスタートさせる' (Click Launch Application and start the app) to this button.

解析をスタートさせると画面が切り替わり、  
StatusがRunningとなり解析が始まる



General Info

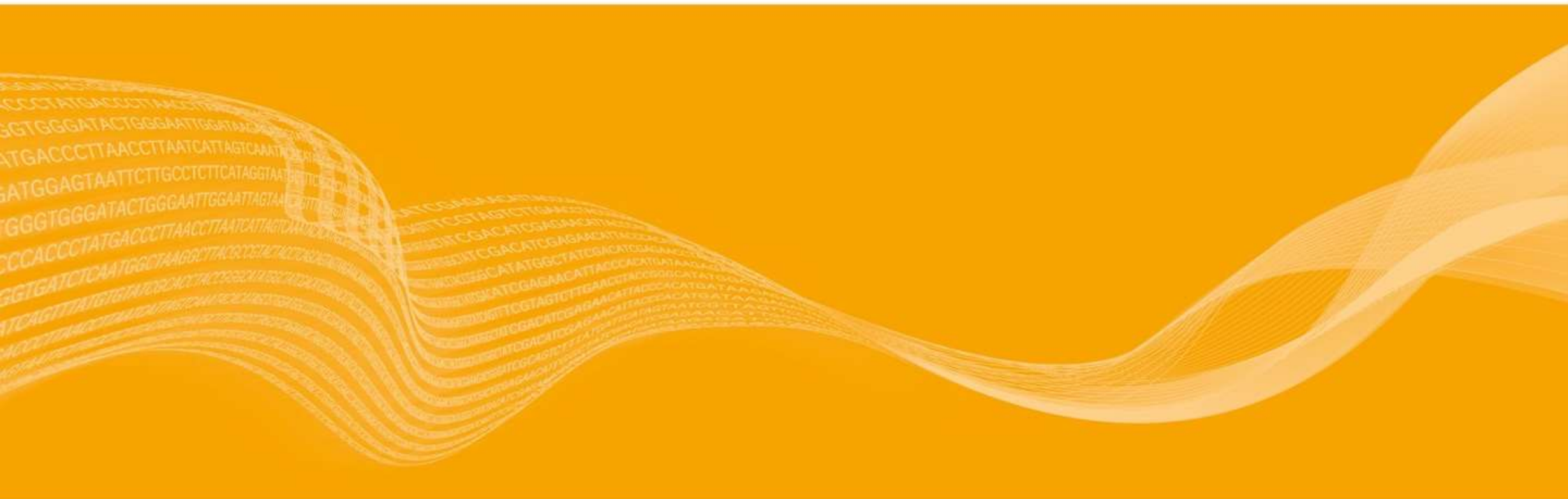
Name	DRAGEN RNA Pipeline Test for Webinar
Application	DRAGEN RNA Pipeline   Version: 3.7.5
Date Started	2021-01-05 23:22
Date Completed	N/A
Duration	2 minutes 40 seconds
Compute Charge	0.00 iCredits
Session Type	Multi-Node
Node Count	4 Nodes Running
Status	Running
Delivery	None

Logs [View Files](#)

Please view the Multi-Node details page to see logs for this analysis.

Detailed description: This block shows a screenshot of the 'General Info' section of the BaseSpace interface. It contains a table with the following data: Name: DRAGEN RNA Pipeline Test for Webinar; Application: DRAGEN RNA Pipeline | Version: 3.7.5; Date Started: 2021-01-05 23:22; Date Completed: N/A; Duration: 2 minutes 40 seconds; Compute Charge: 0.00 iCredits; Session Type: Multi-Node; Node Count: 4 Nodes Running; Status: Running; Delivery: None. The 'Node Count' and 'Status' rows are highlighted with a red box. Below the table, there is a 'Logs' section with a 'View Files' link and a message: 'Please view the Multi-Node details page to see logs for this analysis.'

# DRAGEN RNA Pipelineの結果の見方





# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

Analysis: DRAGEN RNA Pipeline Test for Webinar  
Project 2021-FEB Test Project for webinar

SUMMARY **REPORTS** INPUTS FILES

ここから結果を確認することができる

解析にかかった時間、解析に使用したiCreditなど  
解析の状況を確認することができる

General Info

Name	DRAGEN RNA Pipeline Test for Webinar
Application	DRAGEN RNA Pipeline   Version: 3.7.5
Date Started	2021-01-05 23:22
Date Completed	2021-01-05 23:39
Duration	16 minutes 45 seconds
Compute Charge	5.00 iCredits
Session Type	Multi-Node
Node Count	4 Nodes Complete
Size	31.78 GB
Status	Complete
Delivery	None

# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

The screenshot displays the DRAGEN RNA Pipeline web interface. At the top, there are navigation tabs: SUMMARY, REPORTS, INPUTS, and FILES. Below these are five circular icons representing different actions: a globe, a document, a timer, a share icon, and a trash can. The main content area is divided into two columns. The left column shows a 'Summary' section with a list of sample entries. The first entry, 'BrainmRNA20131201\_5...', is circled in red. Below it are other entries: 'BrainmRNA20131202\_5...', 'UHRmRNA20131205\_59...', and 'UHRmRNA20131206\_59...'. Each entry has links for 'Report' and 'Metrics'. The right column shows a 'Sample Report' for the selected sample: 'Sample: BrainmRNA20131201\_5900916 (ID=244607401)'. It includes version information: 'DRAGEN RNA Pipeline 3.7.5' and 'DRAGEN Host Software Version 05.021.595.3.7.5 and Bio-IT Processor Version 0x04261818'. Below this is a 'Multi-QC Report' section with a link to 'Open in new window' and a 'MultiQC v1.9.dev0' logo. A large red box with Japanese text is overlaid on the right side, pointing to the 'Metrics' link of the first sample entry.

個別の項目から  
各サンプルの数値情報などを確認できます

# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

## Sample Report

Sample: BrainmRNA20131201\_5900916 (ID=244607401)

DRAGEN RNA Pipeline 3.7.5

DRAGEN Host Software Version 05.021.595.3.7.5 and Bio-IT Processor Version 0x04261818

## Multi-QC Report

[Open in new window](#)

Open in new Windowをクリックすると、別のウィンドウとして開くことができます

MultiQC  
v1.9.dev0

General Stats

DRAGEN

Mapping metrics

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Toolbox

# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

**MultiQC**

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2021-01-05, 14:37 based on data in: `/data/output/appresults/220174962/BrainmRNA20131201_5900916`

📘 Welcome! Not sure where to start? [Watch a tutorial video \(6:06\)](#)

General Statistics

こちらのリンクよりYoutubeのチュートリアルを確認できます

# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

## General Statistics

Copy table    Configure Columns    Plot    Showing 1/1 rows and 7/24 columns.

Sample Name	M Input reads	Unmap	Dup	Prop pair
BrainmRNA20131201	128.3	2.0%	35.1%	97.3%

**Prop pair** (Properly paired reads): ペアエンドの両方のリードが、推定されるインサートサイズで適切にマッピングされたリードの割合

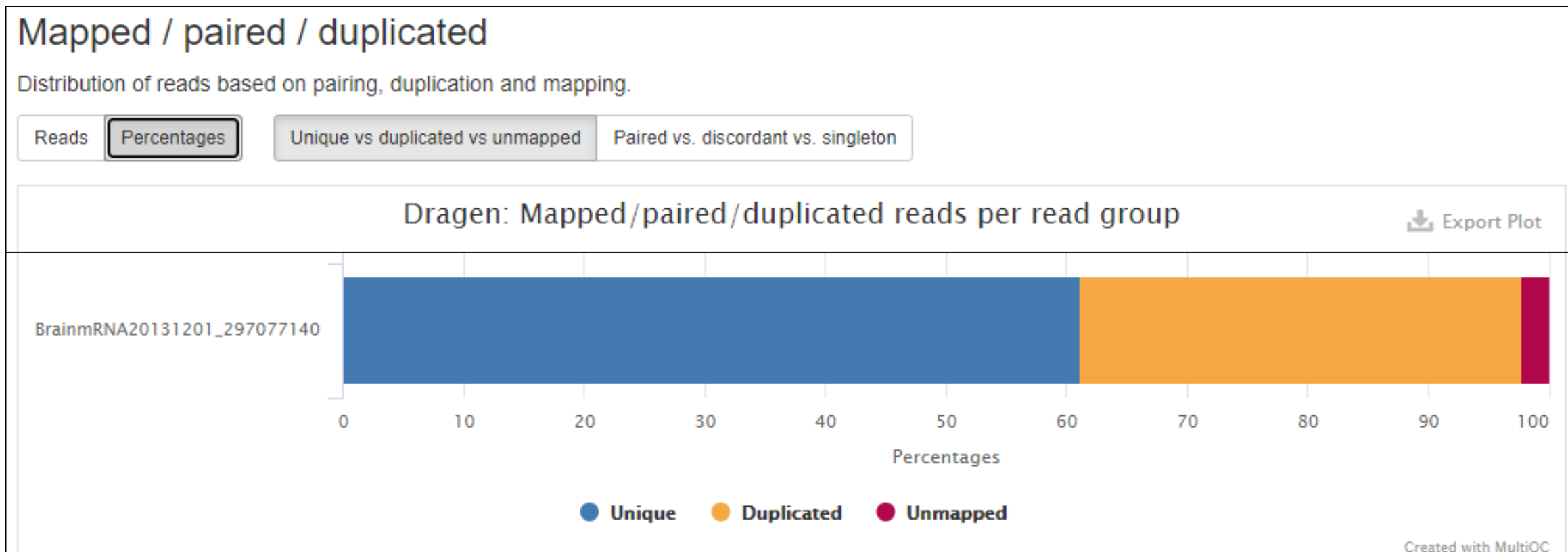
**Unmap** (Unmapped reads): リファレンス配列にマッピングされなかったリードの割合

**Dup** (Duplicate reads): PCR duplicate として判定されたリードの割合

適切にマッピングされたリードの割合は通常70-90 %程度になる  
低い場合、他生物のコンタミネーション、ライブラリー調製不良などが考えられる



# DRAGEN RNA Pipelineの結果の見方: 基本的な数値情報

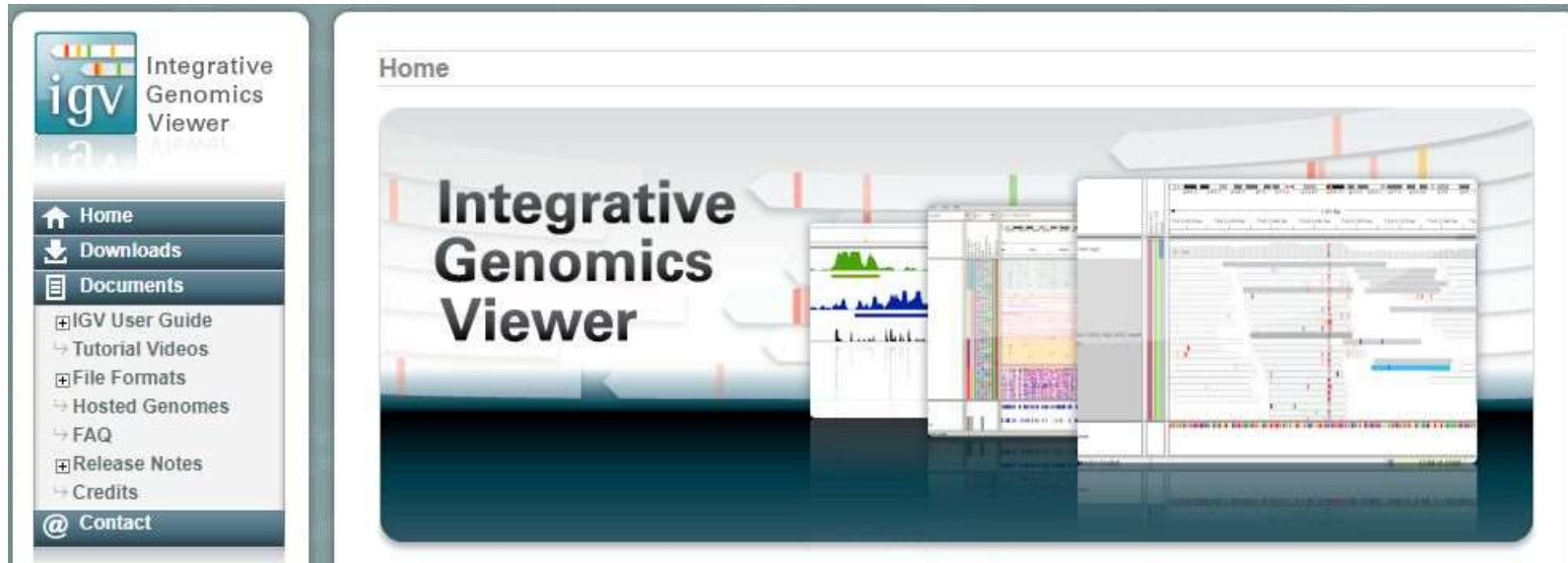


全ゲノムシーケンスなどではDuplicate Rateがライブラリ調製の良し悪しを示すが  
RNA-SeqではDuplicate Rateが高くとも異常ではない

# DRAGEN RNA Pipelineの主要アウトプットファイル

- **アライメントの結果ファイル (BAM / CRAM)**
- **Splice Junctionの位置情報ファイル (SJ.out.tab)**
- **遺伝子発現頻度解析結果**
  - 遺伝子発現解析 結果ファイル (\*.quant.gene.sf / \*.quant.sf)
- **融合遺伝子探索解析結果**
  - Split-Readsの情報を収録したファイル  
(\*.Chimeric.out.junction/\*.fusion\_candidates.features.csv)
  - Pre-filtered / post-filtered 融合遺伝子情報を収録したファイル  
(fusion\_candidates.preliminary / \*.final )

# アライメントの確認: Integrative Genomics Viewer



BAM/CRAM ファイルはバイナリ形式のファイルでテキストエディタなどで見ることはできない。  
Integrative Genomics Viewer (通称IGV)で見る事が可能になる。  
Broad Instituteで作成されたツールで下記URLから無料でダウンロード可能

<http://software.broadinstitute.org/software/igv/>



BaseSpace Appでも使用可能

※ BaseSpace Integrative Genomics ViewerではBAMファイルのみインプット可能

YouTubeでチュートリアルが公開されている

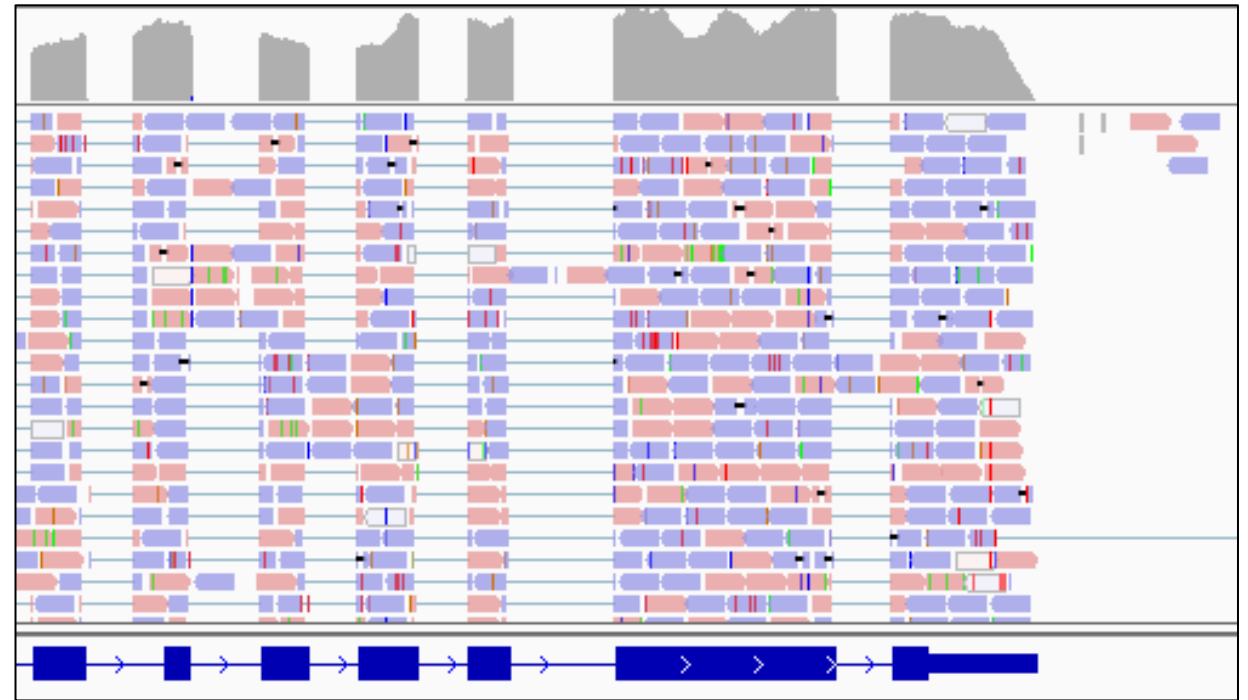
56 [https://www.youtube.com/channel/UCb5W5WqauDOwubZHb-IA\\_rA](https://www.youtube.com/channel/UCb5W5WqauDOwubZHb-IA_rA)

# アライメントの結果ファイル (BAM / CRAM)

マッピングされたリードのカバレッジグラフ

マッピングされたリードの様子  
表示設定でリードの向きなどで色分けできる

Exon (太い青枠部分)、Intronの表示



BAM / CRAMファイルとしてアライメント結果が収録される。  
Integrative Genomics Viewerで開くことでゲノムにマッピングされた  
状況を確認することができる。

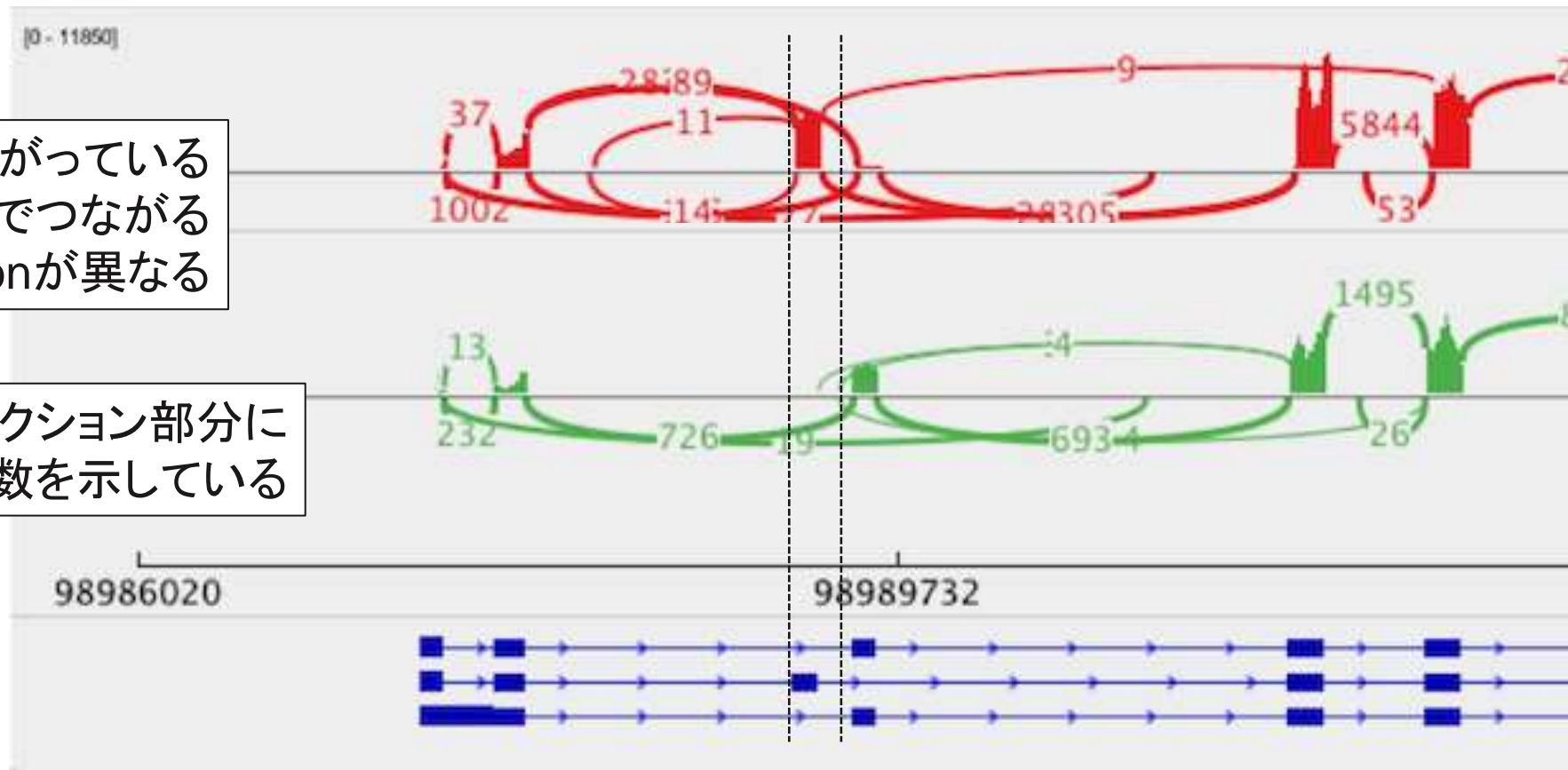
IGV User Guide

<http://software.broadinstitute.org/software/igv/UserGuide>

# アライメントの結果ファイル (BAM / CRAM), Sashimi Plot

Exon同士が線で繋がっている  
-> 赤と緑のサンプルでつながる  
Exonが異なる

それぞれの数字はジャンクション部分に  
またがったリードの数を示している



サンプルごとに異なるスプライスバリエントが発現している様子を確認することができる

What is sashimi\_plot?

<https://miso.readthedocs.io/en/fastmiso/sashimi.html>



# 遺伝子発現解析 結果ファイル (\*.quant.gene.sf / \*.quant.sf)

遺伝子の長さ

発現量数値, TPM (Transcript Per Million値)

Name	Length	EffectiveLength	TPM	NumReads
ENSG00000227232.5	1351	1173.52	12.312	490.3
ENSG00000238009.6	2261	2105.24	0.529	37.8
ENSG00000268903.1	755	551.54	12.448	232.98
ENSG00000269981.1	284	130.23	3.613	15.97
ENSG00000239906.1	323	166.89	0	0
ENSG00000241860.7	1049	895.74	12.282	373.33
ENSG00000222623.1	104	11.08	0	0
ENSG00000241599.1	457	298.53	0	0
ENSG00000279928.2	570	411.3	0	0

リード数

ファイルはタブ区切りテキストとなっている。エクセルなどで開くことができる。

遺伝子発現頻度の数値はリード数ではなく、  
遺伝子の長さで補正したTranscript Per Million値で算出される

# TPM 発現量の補正-1:長さでの補正

RNA-Seqでは遺伝子の長さと得られた総リード数で補正を行ない正規化する  
正規化して得られた値がTranscript Per Million: TPM という値  
この数値を用いて発現量の比較を行なう

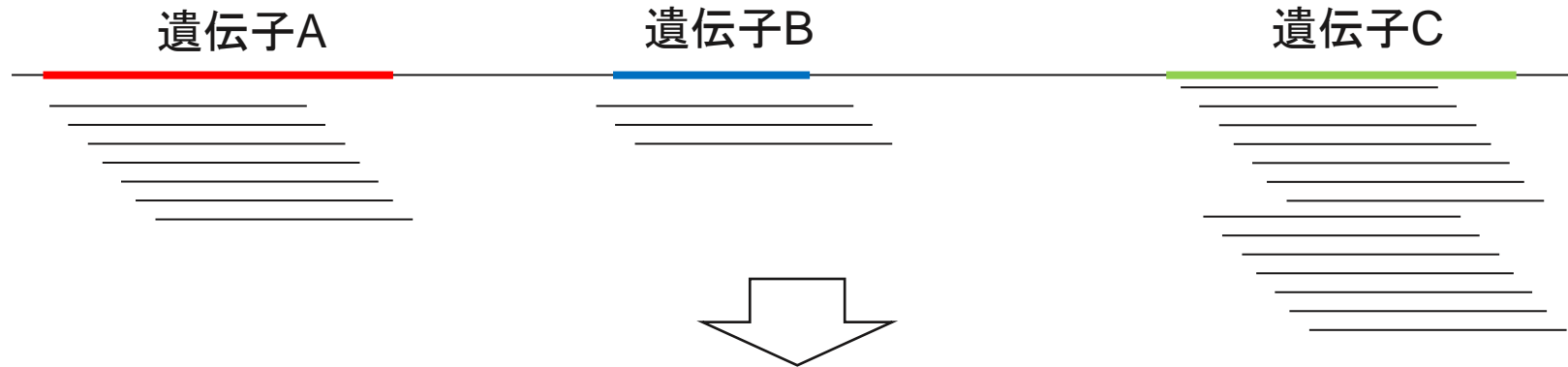
細胞内で同じ発現量の2つのmRNA



各遺伝子	遺伝子の長さ	リード数	遺伝子1000 bpあたりのリード数
遺伝子A	600 bp	6	10
遺伝子B	200 bp	2	10

遺伝子発現量が同じでも、遺伝子が長いほど、より多くのリードが得られる  
そこで遺伝子 1000 bpあたりのリード数に補正を行なう

# TPM発現量の補正-2: 総リード数での補正



各遺伝子のリード数	サンプル1	サンプル2
遺伝子A	10	40
遺伝子B	5	60
遺伝子C	15	20
総リード数	30	120



補正後	サンプル1	サンプル2
遺伝子A	333333.3	333333.3
遺伝子B	166666.7	500000
遺伝子C	500000	166666.7
合計	1000000	1000000

RNA-Seqでは、リード数をもとに遺伝子発現量を算出する  
 左表では遺伝子Aはサンプル2の方が4倍のリードがマッピングされているが総リード数も4倍のリードがある  
 そこで各サンプルから100万リード得られたと仮定し、補正を行なうことで正確な比較が可能になる

# Split-Readsの情報を収録したファイル (\*Chimeric.out.junction/\*.fusion\_candidates.features.csv)

A	B	C	D	E	F	G	H	I
#FusionGene	Score	LeftBreakpoint	RightBreakpoint	Filter	SplitScore	NumSplitReads	NumSoftClippedReads	NumSoftClipReadsGene1
BCAS4--BCAS3	0.999998	chr20:50795173:+	chr17:61368327:+	PASS	72	38	34	17
GAS6--RASA3	0.99989	chr13:113826996:-	chr13:113981858:-	PASS	38	33	5	3
TANC2--CA4	0.999398	chr17:63074014:+	chr17:60155314:+	PASS	9	9	0	0
ARFGEF2--SULF2	0.997803	chr20:48922010:+	chr20:47736942:-	PASS	18	14	4	2
BCR--ABL1	0.995384	chr22:23290413:+	chr9:130854064:+	PASS	11	4	7	2
ME1--PGM3	0.99474	chr6:83398367:-	chr6:83161242:-	PASS	2	1	1	0
RPS6KB1--VMP1	0.987878	chr17:59910611:+	chr17:59838295:+	PASS	11	7	4	2

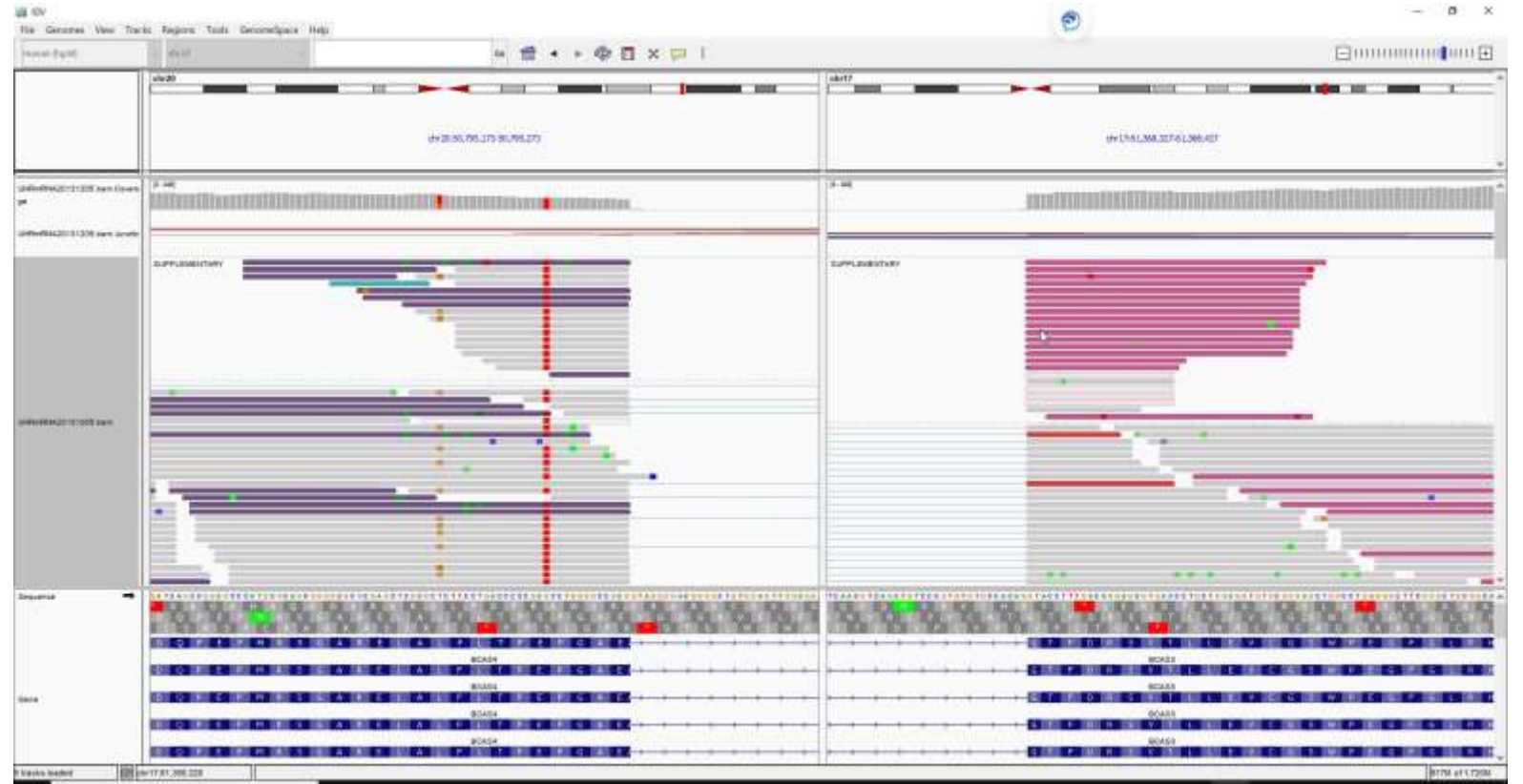
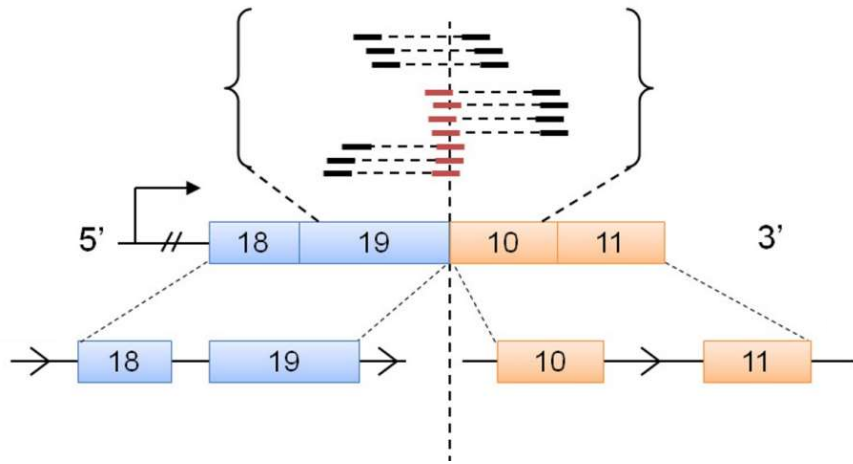
Fusion gene score  
(1に近いほど信頼性が高い)

Break pointの位置

Break pointをサポートするリードの数

Split-readの情報などから融合遺伝子を検出する  
その後フィルタリングを行ない最終的な結果として出力する

# Split-Readsの情報を収録したファイル (Chimeric.out.junction)



BAMファイルを確認することで  
融合遺伝子の様子を確かめることができる



# BaseSpace DRAGEN Differential Expressionの使い方



# BaseSpace Sequence HubでのRNA-Seq解析の流れ



FASTQをアップロード、  
もしくは  
ランをBaseSpaceに転送、  
クラウド上でFASTQ生成



DRAGEN Reference Builder  
Edico Genome Inc.

[Bookmark this app](#) [Help](#)

リファレンス配列を準備する(オプション)



DRAGEN RNA Pipeline  
Edico Genome Inc.

[Bookmark this app](#) [Help](#)

リファレンス配列にマッピングし、  
遺伝子発現量や融合遺伝子探索をする



DRAGEN Differential Expression  
BaseSpace Labs

[Bookmark this app](#) [Help](#)

遺伝子発現量をグループごとに比較し、  
発現量の違う遺伝子リストを出す

# DRAGEN Differential Expressionの使い方

1. キーワードで検索でサーチ  
もしくはアプリケーション一覧から探す

2. 目的のアプリケーションをクリック

3. ソフトウェアのバージョンを確認して  
LAUNCH APPLICATIONをクリック

DRAGEN Differential Expression  
BaseSpace Labs  
Bookmark this app Help

FOR RESEARCH USE ONLY

Overview  
The Illumina® DRAGEN RNA Differential Expression application performs Next Generation Sequencing (NGS) secondary analysis of RNA transcripts. It runs the DESeq2 algorithm on Salmon quantification files to output genes and transcripts that are differentially expressed between two sample groups.

Version  
3.6.3

5.00 iCredits per node hour

LAUNCH APPLICATION

READ MORE

# DRAGEN Differential Expression オプション設定 1/2

## Configuration

Analysis Name

DRAGEN Differential Expression 01/22/2021 5:28:42

1. 解析の名前を設定する

Save Results To

**SELECT PROJECT**

2. 保存するProjectを設定する

2021-FEB Test Project for webinar

3. 比較したい2群の名前を設定する

Control Group

Group Label

Brain

Comparison Group

Group Label

UHR

Control Salmon quantification files

**SELECT DATASET FILE(S)**

BrainmRNA20131201.quant.sf

BrainmRNA20131202.quant.sf

Comparison Salmon quantification files

**SELECT DATASET FILE(S)**

UHRmRNA20131206.quant.sf

UHRmRNA20131205.quant.sf

4. 解析データを設定する

# DRAGEN Differential Expression オプション設定 2/2

The screenshot shows the configuration interface for the DRAGEN Differential Expression application. It includes a dropdown menu for 'Reference Annotation' set to 'Human (hg38)', a section for 'Custom Gene Annotation File' with a 'SELECT DATASET FILE(S)' link, and a 'BaseSpace Labs Disclaimer' section with a checked checkbox. At the bottom, there is a blue 'LAUNCH APPLICATION' button. Three red-bordered boxes with arrows point to these elements, containing the following instructions:

1. RNA Pipelineで選択したリファレンス配列を選ぶ
2. BaseSpace Labs Appの規約に同意チェックする
3. アプリケーションをスタートする

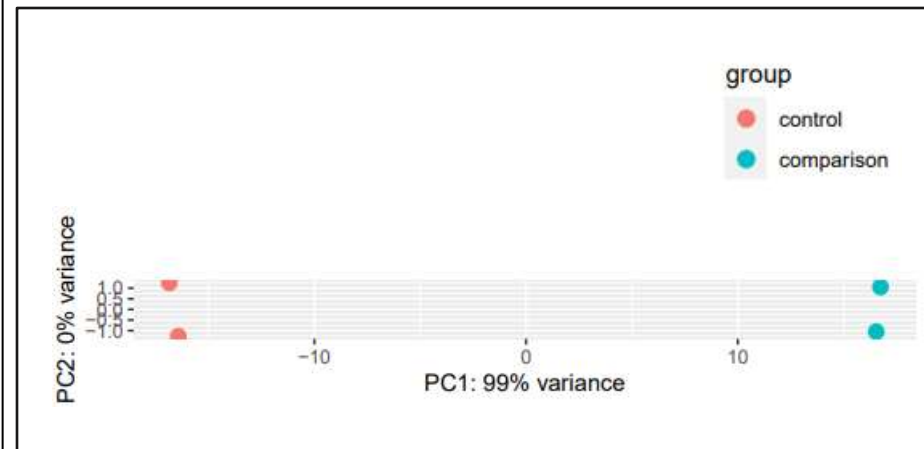
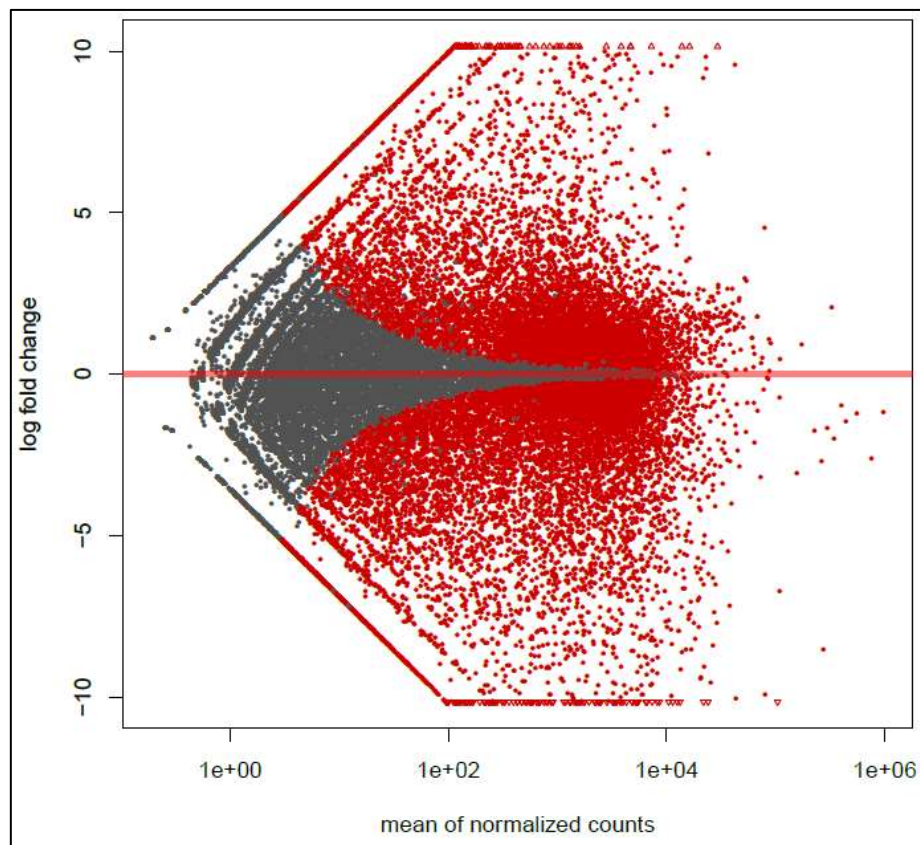
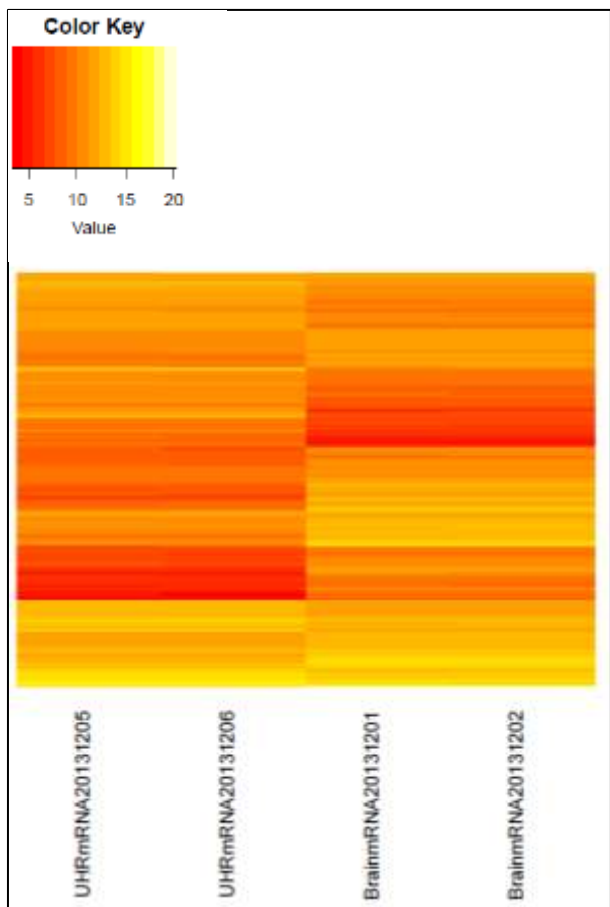


# DRAGEN Differential Expressionの結果の見方





# DRAGEN Differential Expressionの結果の見方



Heatmap、MA plot、PCAプロットでサンプル間の発現量比較や発現差のある遺伝子の様子が確認できる

# DRAGEN Differential Expressionの結果の見方

## Differential Expression Metrics

Annotation Gene Count	60609	<a href="#">Download</a>
Expressed Gene Count	22997	<a href="#">Download</a>
Differentially Expressed Gene Count	17674	<a href="#">Download</a>
Merged Gene Counts		<a href="#">Download</a>
Differentially Expressed Genes		<a href="#">Download</a>
Differentially Expressed Transcripts		<a href="#">Download</a>

トータルの遺伝子のカウント、  
発現している遺伝子のカウント、  
発現量の差があると判定された遺伝子の数

	BrainmRNA20131201	BrainmRNA20131202	UHRmRNA20131206	UHRmRNA20131205
ENSG000000000003.15	471	437	2354	1680
ENSG000000000005.6	4	2	68	57
ENSG00000000000419.12	1043	887	3280	2376
ENSG00000000000457.14	574	501	763	489

遺伝子発現量のカウントが横並びに並んでいる

# DRAGEN Differential Expressionの結果の見方

## Differential Expression Metrics

Annotation Gene Count	60609
Expressed Gene Count	22997
Differentially Expressed Gene Count	17674
Merged Gene Counts	<a href="#">Download</a>
Differentially Expressed Genes	<a href="#">Download</a>
Differentially Expressed Transcripts	<a href="#">Download</a>

トータルの遺伝子のカウント、  
発現している遺伝子のカウント、  
発現量の差があると判定された遺伝子の数

gene_id	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	status	control	comparison	gene_name
ENSG00000001461.17	5878.684075	-2.533805201	0.065508011	-38.67931833	0	0	OK	10026.04451	1731.323642	NIPAL3
ENSG00000003436.16	2176.064093	5.240982165	0.138811529	37.75610139	0	0	OK	112.1176779	4240.010508	TFPI
ENSG00000108797.12	7891.959122	-2.189539064	0.060232317	-36.35156655	2.48E-289	5.59E-288	OK	12945.8993	2838.018942	CNTNAP1
ENSG00000134057.15	3745.992871	4.153490773	0.114294221	36.3403393	3.73E-289	8.40E-288	OK	398.5921566	7093.393586	CCNB1
ENSG00000233237.8	377.5521598	-0.666707063	0.211237244	-3.156200343	0.00159839	0.002394826	OK	463.2704165	291.8339031	LINC00472
ENSG00000154736.6	357.541909	-0.645512258	0.204543695	-3.155864847	0.00160023	0.002397426	OK	436.2220164	278.8618015	ADAMTS5

発現差がどのくらいあるか  
+/- 1で2倍の発現差

2群間比較の(補正済み)p-value

遺伝子名

遺伝子発現量の差、有意差のある遺伝子、遺伝子名  
で絞り込みを行うことができる

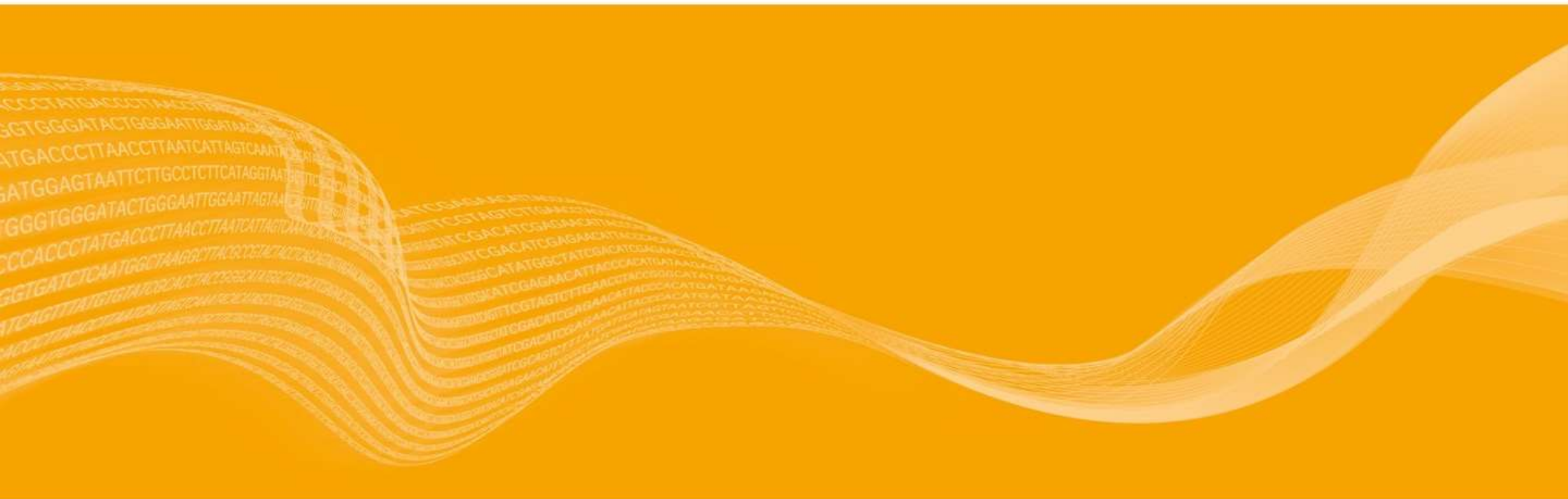
# まとめ



# まとめ

- BaseSpace Sequence Hubで複雑な環境構築や、コマンドラインを使うことなく解析が可能
- DRAGEN アプリを使うことでこれまでにない速さで解析が可能になる
- RNA-Seqの解析は2つのアプリケーションで行う
  - 遺伝子発現量の算出、融合遺伝子の探索にはDRAGEN RNA pipelineを使う
  - 2群間の遺伝子発現頻度の比較にはDRAGEN Differential Expressionを使う

ご清聴ありがとうございました。





# Appendix 資料



# 関連資料

BaseSpace Sequence Hub Help

[https://support.illumina.com/help/BaseSpace\\_Sequence\\_Hub/Source/HomePages/Home\\_Page\\_BaseSpace\\_Sequence\\_Hub.htm](https://support.illumina.com/help/BaseSpace_Sequence_Hub/Source/HomePages/Home_Page_BaseSpace_Sequence_Hub.htm)

Illumina DRAGEN Bio-IT Platform Online Help

[https://jp.support.illumina.com/content/dam/illumina-support/help/Illumina\\_DRAGEN\\_Bio\\_IT\\_Platform\\_v3\\_7\\_1000000141465/Content/SW/FrontPages/DRAGENBioITPlatform.htm](https://jp.support.illumina.com/content/dam/illumina-support/help/Illumina_DRAGEN_Bio_IT_Platform_v3_7_1000000141465/Content/SW/FrontPages/DRAGENBioITPlatform.htm)

# 関連ウェビナー

【実験デザイン編 -これからRNA-Seqを始める方に-】

<https://jp.illumina.com/events/webinar/2018/webinar-180131-j.html>

【RNA-Seqをはじめよう ライブラリー調製編：絶対に失敗しないライブラリー調製】

<https://jp.illumina.com/events/webinar/2018/webinar-180228-j.html>

【RNA-Seqをもう一度：新しいライブラリー調製キットでより簡単・確実に】

<https://jp.illumina.com/events/webinar/2020/webinar-201125-j.html>

# 関連ウェビナー

【NGS解析をはじめよう～よりよいデータ解析のためのFASTQファイルの前処理とデータ評価～】

<https://jp.illumina.com/events/webinar/2020/webinar-0527-j.html>

【NGS解析をはじめよう ～基礎からわかるバイオインフォマティクス入門編～】

<https://jp.illumina.com/events/webinar/2020/webinar-0325-j.html>

Illumina DRAGEN™ どらげん？どらじえん？ - やってきたNGS 高速解析の竜

<https://jp.illumina.com/events/webinar/2019/webinar-product-dragen-190625-j.html>