

2014年10月17日
イルミナサポートウェビナー

RNA Seq を始めよう！ BaseSpaceで行う かんたんNGSデータ解析 < RNA Express >

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)

© 2013 Illumina, Inc. All rights reserved.

Illumina, IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. The Genetic Energy 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®

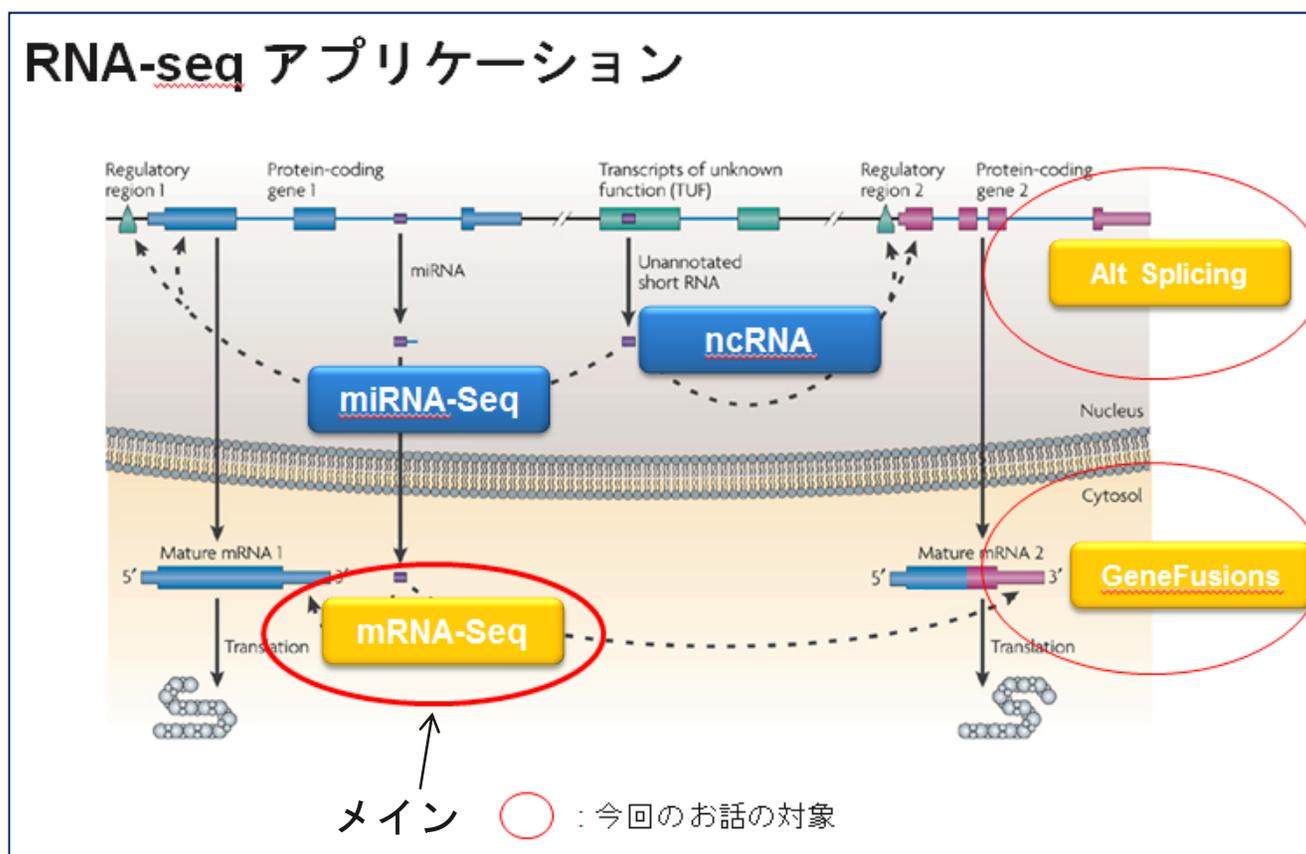
本日の内容

- ▶ RNA-Seq 概要
- ▶ BaseSpace のRNA Seqアプリ
- ▶ BaseSpace 実行例
 - BaseSpaceでアプリを起動
 - BaseSpaceでデモデータを取り込む
 - BaseSpaceでRNA Express Appを実行



スコープ

- ▶ 本回では、RNA Seqのうち、mRNA-Seq のことをRNA Seqと呼称します
- ▶ また既知のリファレンスがあることを前提としております



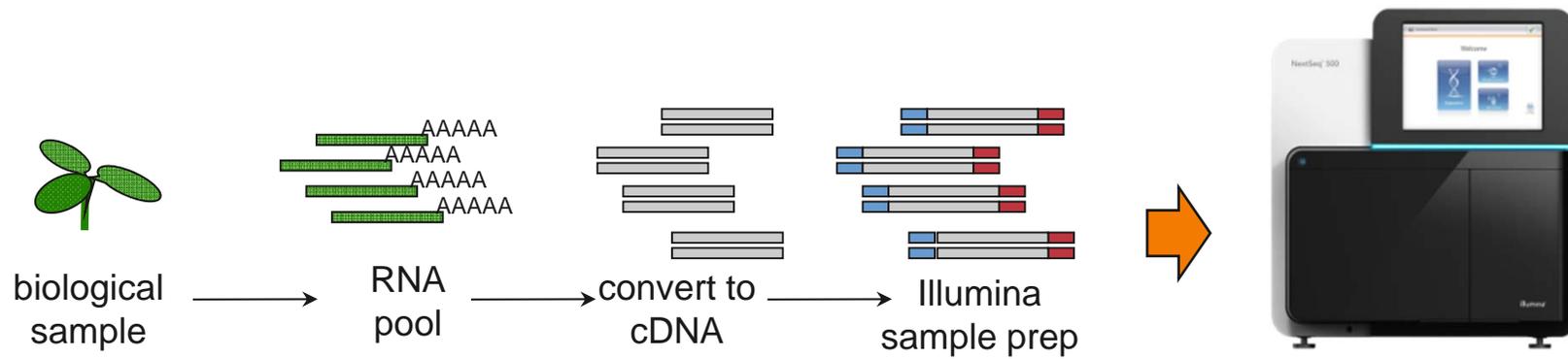
RNA-Seq による発現量の測定

- ▶ 特定の遺伝子領域にマップされたリードの数
= 遺伝子転写産物の存在量

ととらえ、解析をしています

RNA-Seq シーケンス前

<Brief review of RNA-Seq>



RNA-Seq シーケンス後

解析ソフトウェアで配列データを処理



RNA Seq アライメント工程 (例)

アライメント

サンプル毎のリードを
リファレンス配列に対してアライメント



取り除きたい余剰配列の処理



リファレンスゲノムへのアライメント



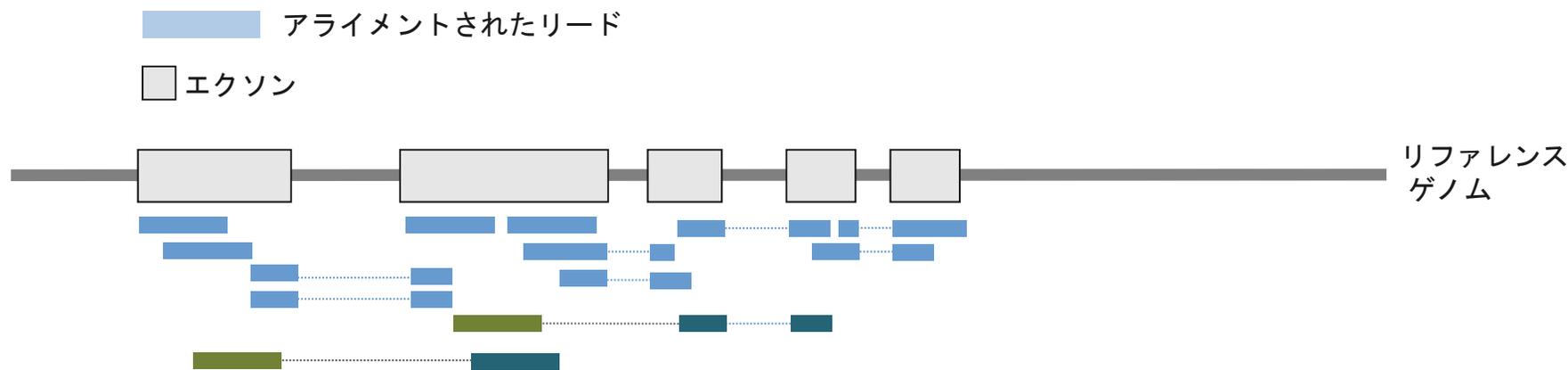
既知スプライスジャンクション
を考慮したアライメント

RNA Seq アライメント

アライメント

サンプル毎のリードを
リファレンス配列に対してアライメント

アライメントツールは、スプライスジャンクションを考慮したマッピングが必要
計算量が大きくなるので、ツールによって、それぞれ考慮の仕方の工夫を凝らしている



沢山のツールがあり、また計算量が多い。
自前でpipelineを構築する場合はしっかりした
計算機とツール選択やインストールなどの
メンテナンスが必要。



直ぐに無料で
お使い頂けます。

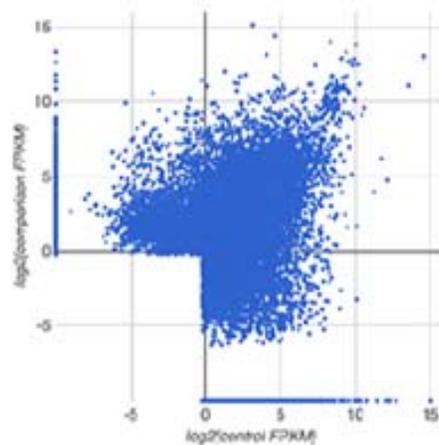
RNA-Seq 発現差解析



サンプル群 vs サンプル群で
発現の差異をみる

- ▶ リード数を数え、正規化した数値をもとに
統計検定を行い発現差異をみていく

サンプル群 B



サンプル群 A

※ 採用統計モデルは使用する
ツールにより様々であり、
開発が続けられている



RNA Seqでも複数アプリを搭載しているので
異なる統計モデルをお試し頂けます

もっと詳しく知りたい！ イルミナウェビナー RNA-Seq シリーズ

<p>2011/11/17 RNA-Seqをはじめよう！シリーズ 「Session 3 - データ解析のリテラシー」 東京大学大学院 農学生命科学研究科 門田 幸二 先生</p> <p>第3回のセッションは、東京大学の門田先生によるデータ解析です。</p> <p>(1) R...</p> <p>続きを開く</p>	<p style="border: 2px solid red; padding: 2px; display: inline-block;">正規化, DE 詳しい回</p>
<p>2011/10/13 RNA-Seqをはじめよう！シリーズ 「Session 2 - RNA-Seq実験ノート:リード長とリード数のデザインとウェット」 東京大学大学院 新領域創成科学研究科 鈴木 稯 先生</p> <p>第2回のセッションは、東京大学の鈴木先生をお迎えし、実際に実験を開</p> <p>続きを開く</p>	<p>2014/07/22 RNA-Seqシリーズ 「トランスクリプトームデータ解析戦略 2014」 東京大学・大学院農学生命科学研究科・アグリバイオインフォマティクス教育研究ユニット 門田 幸二 先生</p> <p>ILLUMINA HiSeqシリーズをまじめとしたトランスクリプトーム解析用機器の技術革新は...</p> <p>続きを開く</p>
<p>2011/09/08 RNA-Seqをはじめよう！シリーズ 「Session 1 - トランスクリプトーム解析の今昔:なぜマイクロアレイ?なぜRNAシーケンス?」 東京大学大学院 農学生命科学研究科 門田 幸二 先生</p> <p>遺伝子発現解析はこれまでマイクロアレイが使われてきました。しかし、次世代シーケンサー技術が急速に広がり、次第にシーケンサーを使った解析に手法が移行しています。第1回のセッションでは、これまでの手法と比べてRNAシーケンスはどのように違うのかについて、東京大学の門田先生にお話をいただきます。</p> <p>開じる</p>	<p>2014/06/24 RNA-Seqシリーズ 「進化するRNA-Seq:臨床検体からシングルセル解析まで - ウェット・ドライ解析の実験ノート」 東京大学大学院新領域創成科学研究科 情報生命科学専攻 鈴木 稯 教授</p> <p>2011年の秋に弊社ウェビナーでご講演いただきました東京大学大学院 鈴木 稯先生を再...</p> <p>続きを開く</p>

http://www.illumina.co.jp/events/webinar_japan.ilmn?ws=ws

本日の内容

- ▶ RNA-Seq 概要
- ▶ BaseSpace のRNA Seqアプリ
- ▶ BaseSpace 実行例
 - BaseSpaceでアプリを起動
 - BaseSpaceでデモデータを取り込む
 - BaseSpaceでRNA Express Appを実行



イルミナはBaseSpace 上にRNA Seq 用の 3つのフリーアプリをご提供しています

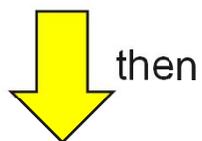
手法	BaseSpace アプリ	アイコン	内容
RNA-Seq	TopHat アライメント		<ul style="list-style-type: none"> • 業界標準のTopHat2を使ったRNA-Seqアライメントとカウンティング • 融合遺伝子のコール(オプション) • ISAAC Variant Callerを使ったcSNPコール • 結果はCufflinks Assembly & DE Appでさらに解析可能
	Cufflinks アセンブル & 遺伝子発現解析		<ul style="list-style-type: none"> • 詳細遺伝子発現差解析 • 選択的転写産物のアセンブルと新規転写産物予測
	RNAExpress		<ul style="list-style-type: none"> • 迅速な遺伝子発現プロファイルをSTARアライメントとDESeq2で実現 • 遺伝子レベルの遺伝子発現に特化

* 2014/10現在

アプリの違い：2通りの使い方



TopHat Alignment



Cufflinks Assembly/
Diff. Exp

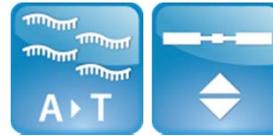
QC工程を
はさめる！



RNA Express

発現解析(Expression)
をノンストップ
特急 (Express)で！

アプリの違い：主要機能面



機能	TopHat/ Cufflinks	RNA Express
○ シーケンス量のフィルター	あり	あり
○ シーケンスアライメント	あり	あり
変異コール	あり	なし
融合遺伝子コール	あり	なし
転写産物アSEMBル	あり	なし
遺伝子量予測	あり	なし
転写産物量予測	あり	なし
○ 遺伝子発現差の解析	あり	あり

高機能、詳細解析の実行

かんたん、速い

アプリの違い : (内包するツールからもう少し詳しく)

1a. Abundance estimation (**genes and transcripts**), fusion detection, and variant calling workflow.



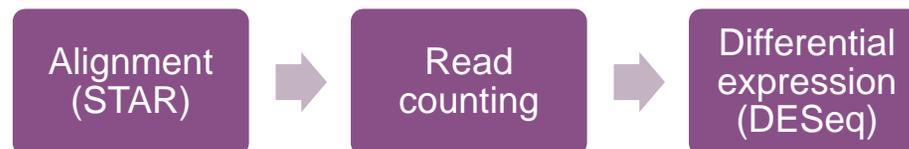
TopHat Alignment

1b. Creating and merging assemblies and differential expression calculation.



Cufflinks Assembly/
Diff. Exp

2. Fast Alignment and differential expression calculation of **genes**



RNA Express

本日の内容

- ▶ RNA-Seq 概要
- ▶ BaseSpace のRNA Seqアプリ
- ▶ BaseSpace 実行例
 - BaseSpaceでアプリを起動
 - BaseSpaceでデモデータを取り込む
 - BaseSpaceでRNA Express Appを実行



RNA-Seq App 解析実行・閲覧 デモ

BaseSpace デモデータと
RNA Express App



アライメント結果を可視化して確認したい

Analysis Info	
Inputs	
Output Files	
Analysis Reports Summary	

Output files	
Name	
BrainmRNA20131201.Log.final.out	
BrainmRNA20131201.Log.out	
BrainmRNA20131201.Log.progress.out	
BrainmRNA20131201.Log.std.out	
BrainmRNA20131201.SJ.out.tab	
BrainmRNA20131201.alignments.sorted.bam	}
BrainmRNA20131201.alignments.sorted.bam.bai	
BrainmRNA20131201.coverage.bedGraph.gz	}
BrainmRNA20131201.coverage.bedGraph.gz.tbi	
BrainmRNA20131201.parameters.txt	
renaming_log.txt	

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

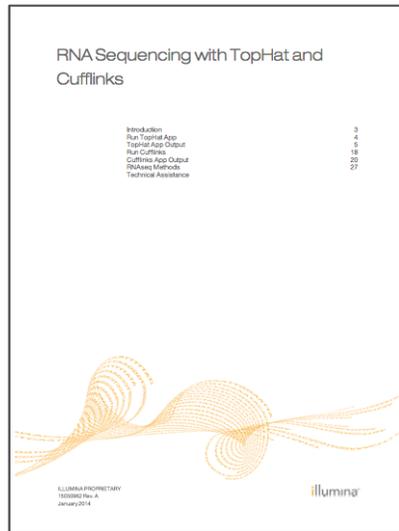


アライメントの結果ファイル
カバレッジの結果ファイル

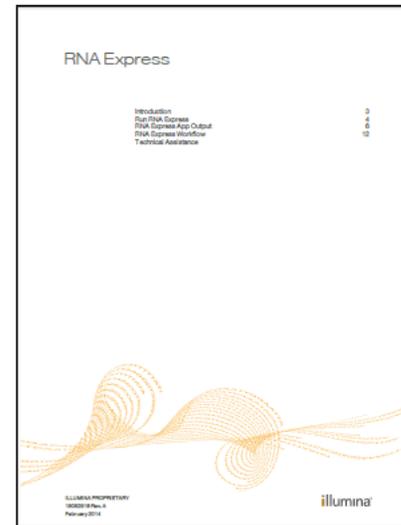
IGVなどで可視化可能

他にもDE解析中間ファイル等、解析途中データ、解析結果データ、Rスクリプト、描画画像などがOutput Filesの配下に置かれます。

詳細は User Guide をご確認ください



TopHat
Cufflinks



RNA
Express

本日ご紹介の
アプリ



ユーザーガイドのご提供場所

http://support.illumina.com/sequencing/sequencing_software/basespace/documentation.ilmn

The screenshot shows the Illumina support website interface. At the top left is the Illumina logo. A navigation bar contains links for APPLICATIONS, SYSTEMS, CLINICAL, SERVICES, SCIENCE, SUPPORT (circled), and COMPANY. A search box is on the right. Below the navigation bar is a breadcrumb trail: Support > Sequencing > Sequencing Software > BaseSpace > Documentation & Literature (circled). On the left, a 'BaseSpace Support' sidebar lists Overview, Computing Requirements, and Documentation & Literature (circled). The main content area is titled 'Documentation & Literature' and includes links for Documentation, Product Literature, and Tech Notes. Under a 'DESCRIPTION' section, there is a list of user guides: BaseSpace User Guide (expanded), BaseSpace Core App User Guides (collapsed), and several specific app guides: Instructions for using the BWA Whole-Genome Sequencing App (15050952 A), Instructions for using the BWA Enrichment App (15050958 A), Instructions for using the Tumor Normal App (15050950 A), and Instructions for using the TopHat and Cufflinks RNA-seq apps (15050962 A).

オンラインヘルプ (User Guideと基本的に同じ内容)

http://support.illumina.com/help/BaseSpace_App_enr_BWA_help/BWAEnrichmentHelp.htm
http://support.illumina.com/help/BaseSpace_App_enr_Isaac_help/IsaacEnrichmentHelp.htm

The screenshot shows the Illumina support website interface. The browser address bar displays the URL: support.illumina.com/help/BaseSpace_App_enr_BWA_help/BWAEnrichmentHelp.htm#Vault/Informatics/Sequencing. The page title is "Running BWA Enrichment".

Running BWA Enrichment

- 1 Navigate to the project or sample that you want to analyze.
- 2 Click the **Apps** button and select **BWA Enrichment** from the dropdown list.
- 3 Read the End User License Agreement and permissions, and click **Accept** if you agree.
- 4 Fill out the required fields in the BWA Enrichment input form:
 - a **Analysis Name:** Provide the analysis name. Default name is the app name with the date and time the analysis was started.
 - b **Save Results To:** Select the project that stores the app results.
 - c **Sample(s):** Browse to the sample you want to analyze, and select the check box. You can analyze multiple samples.
 - d **Reference Genome:** Select the reference genome. Currently, you can only use hg19.
 - e **Targeted Regions:** Select the targeted region of your enrichment.
 - f **Base Padding:** Select the padding you want. Padding defines the amount of sequence immediately upstream and downstream of the targeted regions that is also used in enrichment analysis.
 - g **Annotation:** Choose which gene and transcript annotation reference database to use.
- 5 If desired, fill out the advanced fields in the BWA Enrichment input form:
 - a **Depth Threshold:** The GATK variant caller filters variants if the coverage depth at that location is less than the specified threshold. Decreasing this value will increase variant calling sensitivity, but raise the risk of false positives. The variant caller reports the variants, but filters them in the VCF files by adding a LowDP flag. Default value for GATK: 5.
 - b **Trim Nextera Rapid Capture Adapters:** If selected, Nextera Rapid Capture adapters will be trimmed. Use this setting only if not already applied as a sample sheet setting.
 - c **Flag PCR Duplicates:** If selected, PCR duplicates are flagged in the BAM files and not used for variant calling. PCR duplicates are defined as two clusters from a paired-end run where both clusters have the exact same alignment positions for each read. Optical duplicates are already filtered out during RTA processing.
 - d **Generate Picard HS Metrics:** If selected, Picard HS metrics are generated. See [Picard Metrics](#) for more information.

Figure 2 BWA Enrichment Input Form

The input form shows the following fields:

- Analysis Name: BWA Enrichment 12/02/2013 1:49:23
- Save Results To: Select a Project
- Sample(s): Select Sample
- Reference Genome: Human (JCSC hg19)
- Targeted Regions: Nextera Rapid Capture Exom
- Base Padding: 150
- Annotation: RefSeq @ Ensembl

There is an "Advanced" section at the bottom of the form.

ご参考文献

<http://www.ncbi.nlm.nih.gov/pubmed>
からPMIDをご入力いただくとアクセスが簡便です

▶ イルミナシーケンシング

- Accurate whole human genome sequencing using reversible terminator chemistry.
Nature 456: 53-59 [PMID: 18987734]

▶ RNA-Seq, RPKM

- Mapping and quantifying mammalian transcriptomes by RNA-Seq
Nature Methods, Volume 5, 621 – 628 [PMID: 18516045]

▶ BaseSpace RNA-Seq アプリ

- TopHat: discovering splice junctions with RNA-Seq.
Bioinformatics. 2009 May 1;25(9):1105-11 [PMID: 19289445]
- Cufflinks/Cuffdiff
Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010 May;28(5):511-5 [PMID: 20436464]
- Cuffdiff2 Differential analysis of gene regulation at transcript resolution with RNA-seq
Nature Biotechnology 31, 46–53 (2013) [PMID: 23222703]
- TopHat-Fusion: an algorithm for discovery of novel fusion transcripts.
Genome Biol. 2011 Aug 11;12(8):R72. [PMID: 21835007]
- **STAR: ultrafast universal RNA-seq aligner.**
Bioinformatics, 2013 Jan 1;29(1):15-21. [PMID: 23104886]
- **DESeq**
Differential expression analysis for sequence count data.
Genome Biol. 2010;11(10):R106. [PMID: 20979621]
DESeq2: www.bioconductor.org/packages/2.13/bioc/html/DESeq2.html

ご参考サイト

イルミナ

<http://www.illumina.com/landing/basespace-core-apps-for-rna-sequencing/>
<http://res.illumina.com/documents/products/technotes/technote-basespace-rna-seq.pdf>
http://support.illumina.com/help/BaseSpace_App_RNAseq_help/RNAseq_Apps_Help.htm
<http://www.illumina.com/applications/sequencing/rna.ilmn>

業界フォーラム(英語)

<http://seqanswers.com/>
<https://www.biostars.org/>

日本語フォーラムサイト

<http://cell-innovation.nig.ac.jp/wiki/tiki-index.php>
<http://qa.lifesciencedb.jp/>

Next is Now

サポートウェビナーにご参加いただき
ありがとうございました！

本日のセッション終了後のご質問は、
techsupport@illumina.com
で承ります。



© 2014 Illumina, Inc. All rights reserved.

Illumina, IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

illumina®