

# RNA-seq解析



近年、様々な分野で次世代シーケンサーを用いた解析が行われています。コストもどんどん下がり試しにやってみようと思解析を行ってみたい研究者様も多いことでしょう。ですが実際にシーケンスしてみると、そのデータ量や解析の困難さに気づくはずで。また出てきたデータからどのような結果が導き出されるかということが分からない場合もあるかと思。そのためRNA-seq解析を例に挙げながら、解析手順など我々の解析サービスの実例について説明していきます。

## RNA-seq解析概要

次世代シーケンサーが登場する以前は、ノーザンハイブリダイゼーション、マイクロアレイ解析あるいはRealtime-PCRなどが、RNA発現解析の主流でした。しかし次世代シーケンサー登場後は、

RNA-seq解析により網羅的に発現を見て、目的とする遺伝子以外の発現も同時に確認でき、その関連因子や新規転写配列の予測も行うことができるようになったため、こちらも多く用いられるようになりました。またタンパク質をコードする

mRNA以外にも副次的に転写調節やエピジェネティックへの関与が認められている長鎖ノンコーディングRNA(lncRNA)の情報も得られるため、トランスクリプトーム解析の多様性解明に不可欠な手法となっています。

## 解析手法

まずお客様に次世代シーケンサーを使って、RNAサンプルをシーケンスをしていただきます。その時に出力されるデータを用いて、その後の解析を我々が行います。

データを頂きましたら、リードのクオリティチェックから行います。生データの中にはクオリティが低くそのまま解析に用いると、ノイズになってしまうリードが存在します。そのため、データをきれいにする必要があります。解析には様々なツールが存在しますが、お客さまからの要望がない場合は、主に、FastQC、FASTX-Toolkit、Cutadapt、Trimmomaticなどを使用しています。

次にリファレンスに対し、マッピングを行います。RNA-seqではスプライシング後のリードを用いているため、基本的にはエクソンにマッピングされるはずで。しかし、実際にはエクソンではない場所にマッピングされる場合があります(Fig1)。その場合、未知の遺伝子である可能性もあります。こういったデータが出るのも、RNA-seqの面白いところと言えます。

マッピングを終えたら、RPKM、FPKM値の算出を行います。これらは簡単に言えばノーマライズした発現量です。RPKMはRead単位でシングルリードで読んだとき、FPKMはペアエンドで読んだときの2つのReadを1組としたfragmentを単位としています。これらの値を元にリストやグラフを作成し、納品させて頂いております。ただし注意点として、RPKMやFPKMなどでノーマライズしたとしても、カウンティング方法によって、発現量の多い転写産物は実際よりもずっと多くカウントされてしまうなどのノイズ・バイアスがかかってくることもありますので、万能ではありません。有用遺伝子を特定したあとは、Realtime-PCRなどで確認してから、その後の実験に用いることをおすすめします。

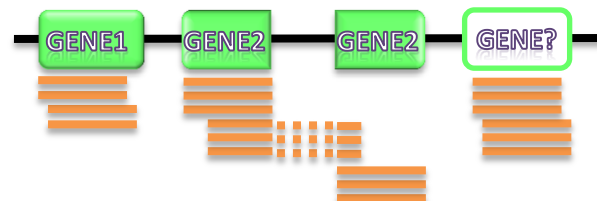


Fig 1. RNA-seq mappingの例

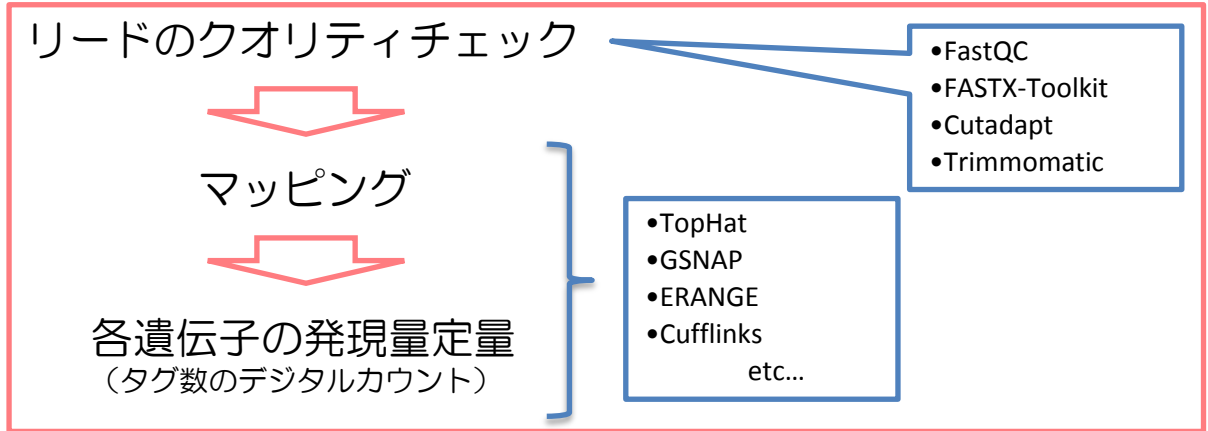
リファレンスの遺伝子に対し、RNAのリードがマップされている。右のGENE?はリファレンスでは示されていない箇所にリードがマップされたことを示している。



# 解析手順

サンプルの調整およびシーケンス

シーケンスのリードデータ



目星をつけた遺伝子の  
実際の発現を確認  
(例：Realtime-PCR)

例：特定した遺伝子をノックアウト  
あるいは過剰発現させて、  
Phenotypeを確認する

## 納品データ

Table.1 sample1とsample2における遺伝子間の発現量比較

Gene ID	Gene length	count tags		normalize		compare	
		sample1	sample2	sample1	sample2	sample1 / sample2	sample2 / sample1
gene1	4816	67	128	0	0	0.524541	1.714243
gene2	1234	223	322	3	4	0.660376	1.400939
gene3	1306	38	262	0	3	0.141460	6.008641
gene4	2800	129	208	0	1	0.594262	1.591410
gene5	1139	1278	1849	21	29	0.675245	1.351669
gene6	2484	753	704	4	4	1.035732	0.861264
gene7	2405	7600	8509	25	27	0.905136	1.055642
gene8	674	180	2427	7	96	0.068229	12.730983
gene9	6262	6779	10376	7	10	0.653873	1.425456
gene10	9535	580	194	0	0	2.762371	0.298133

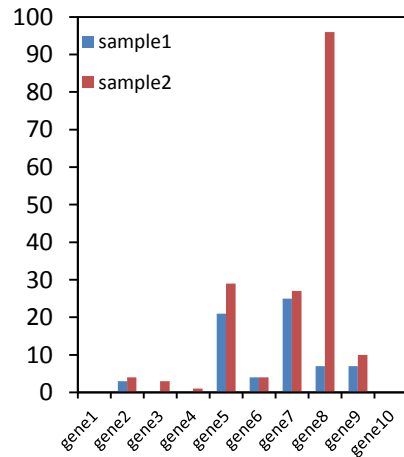


Fig.2 sample1およびsample2の  
遺伝子間発現量のグラフ



アクシオヘリックス株式会社

本社：沖縄県那覇市西2-16-3 屋島組本社ビル2-A号室

TEL：098-988-4235 FAX：098-988-4238

東京支社：東京都千代田区神田和泉町1番地12の17 久保田ビル8階

TEL：03-5823-4714 FAX：03-5823-4715

HP：http://www.pictbio.com E-mail：pictbio@axiohelix.com