

次世代シーケンサー(NGS)解析の 現状と失敗しないための心得

アクシオヘリックス株式会社

アクシオヘリックス株式会社について

本社

沖縄県那覇市

東京支社

東京都千代田区神田和泉町

代表取締役社長

シバスンタラン スハルナン

創立

2001年6月8日



事業ドメイン

ライフサイエンス

バイオインフォマティクス
に関連するサービスを提供

ITソリューション

システムの受託開発

ビジネスディベロップメント

新しい技術を取り入れた製
品を世界へ発信

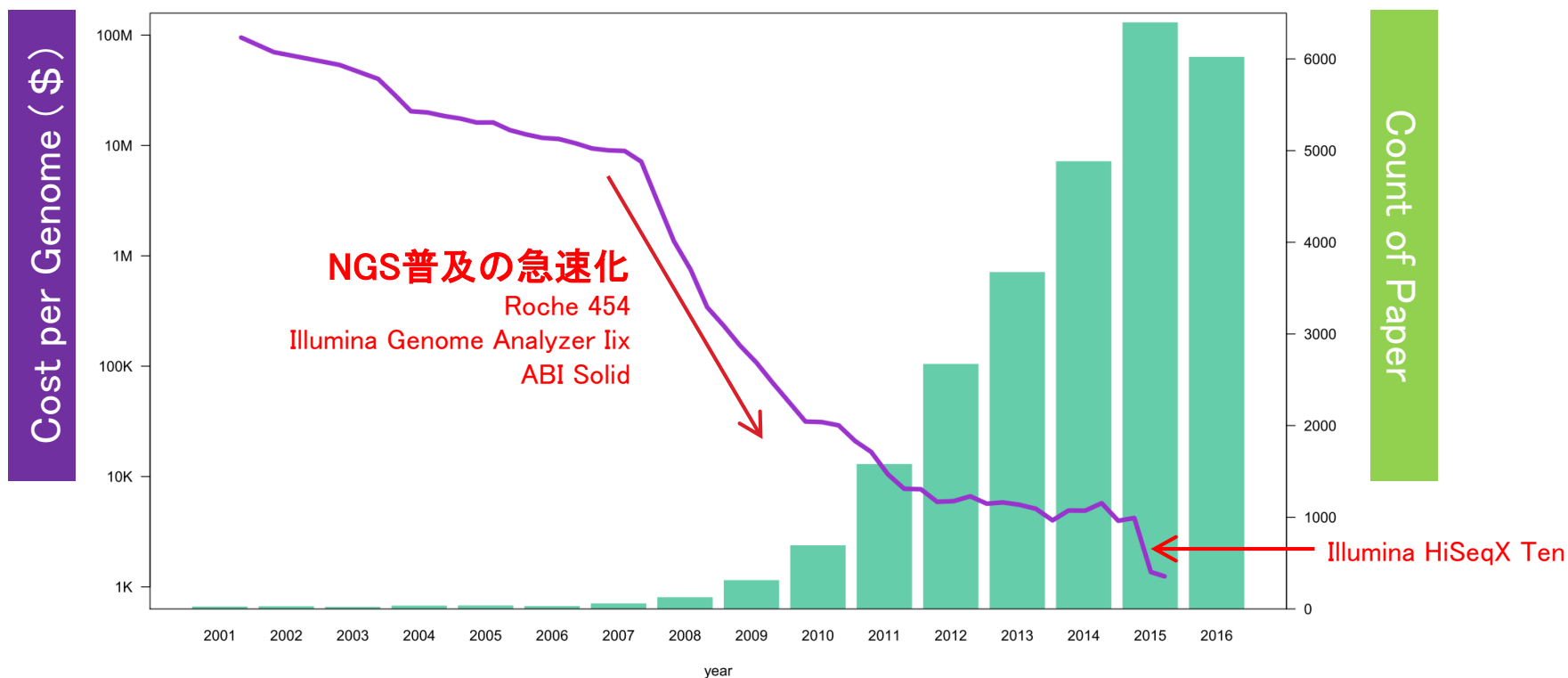
アジェンダ

- ▶ NGSを使った研究状況
- ▶ 解析例
- ▶ 失敗事例と解決法

NGSを使った研究状況

次世代シーケンサー(NGS)を使った研究状況

NGSのシーケンス価格と関連論文投稿数の推移

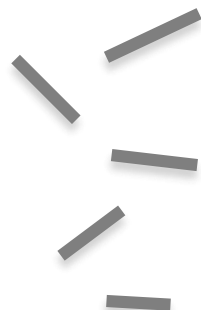
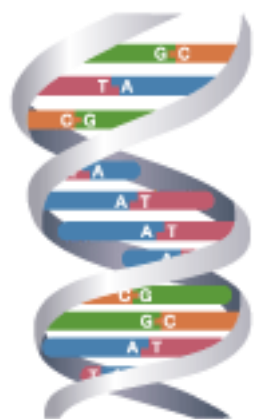


ゲノム1,000 \$時代到来

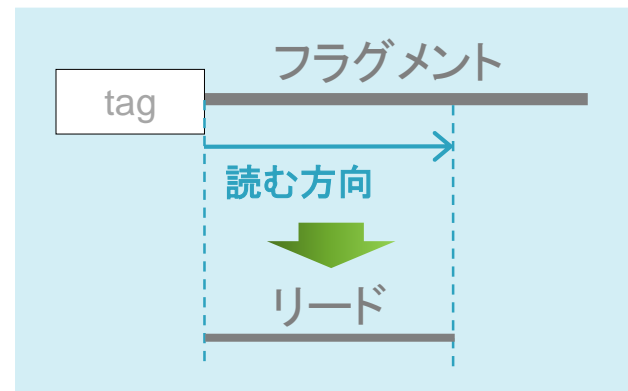
Cost per Genome
参考元: NIH
ゲノムはヒトゲノム基準

Count of Paper
参考元: PubMed
キーワード「Next Generation Sequencing」で検索

DNAの読み方



断片化



リード配列

read1 : ACTGANCTGCTTCTGGTGTGGTTGATATT...
read2 : GATGTNCCCATCTGAATGCAATGAAGAAAA...
read3 : CTACGNCCAGCAGCAGTGGGGAATTTTCCG...
read4 : CTACGNCTCGCAGCAGTGGGGAATCTTGGA...

⋮

NGSの特徴

▶ NGSとサンガー

	NGS ※ HiSeq 2000	サンガー ※ Sanger 3730xl
リード長	50 ~ 100 bp	400 ~ 900 bp
クオリティ	98%	99.999%
データ量	600 Gb	1.9~84 Kb

約 7×10^6 倍異なる

参考元 : Comparison of Next-Generation Sequencing Systems (Liu L et al. / 2012 / PubMed ID 22829749)

▶ 圧倒的な量のデータ取得できる

この量を生かして多くの研究が行われている

リードの種類

- ▶ シングルエンドリード

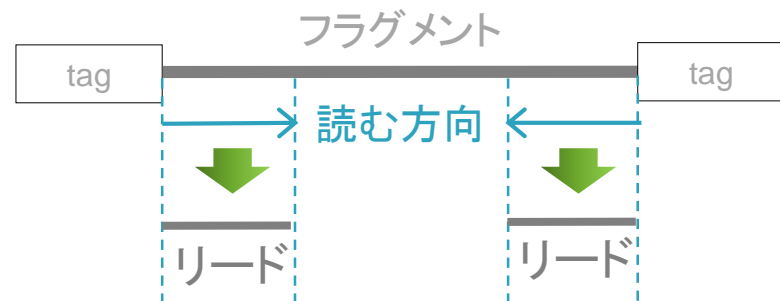
1フラグメントから1リードを得る。

- ▶ ペアエンドリード

1フラグメントから複数リードを得る。

一般的にフラグメントの両端から読む。

ゲノムへの張り付き方の計算時に位置関係を考慮することで**クオリティを上げる**。



解析例

NGSでよく行われている解析

DNA

- ▶ **変異解析**
- ▶ **メチル化解析 (バイサルファイト)**
- ▶ **ChIP-Seq解析**
- ▶ **新規ゲノム配列決定**
- ▶ **メタゲノム解析**

RNA

- ▶ **遺伝子発現解析**

NGS解析例(1) 変異解析

- ▶ 概要

既知の配列情報と比較し、異なる配列情報
(**変異**) を検出する。

- ▶ NGSの利点

ゲノム全体が見られる。

シーケンス手法によっては**構造変異**まで計算ができる。

- ▶ 利用される主要分野

遺伝子疾患の研究
育種

NGS解析例(1) 変異解析 ロジック



拡大

ゲノム

.....AGCAAAGCAGGGGAAATAAAAACAACCAAATGAAGGCAAAACACTAC.....

リード1 CAAAAGCAGGGGAA**C**ATA

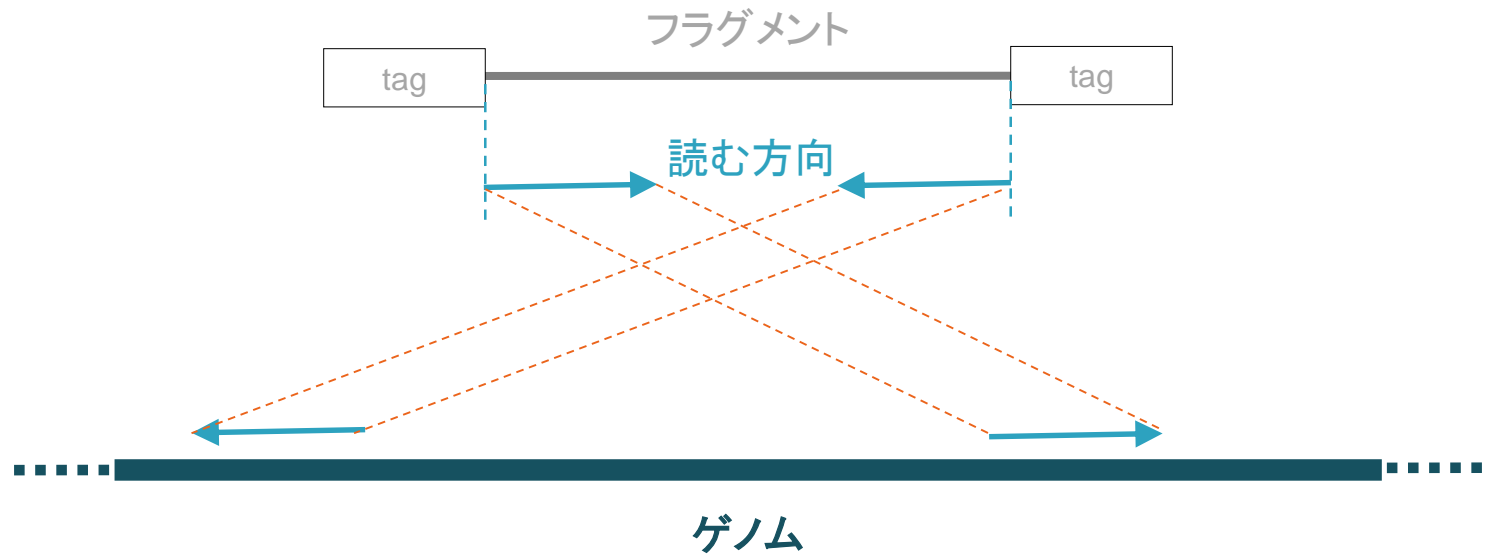
GGGAA**C**ATAAAAACAACCAAAT リード2

リード3 GCAGGGGAA**C**ATAAAAACAACCA

- ▶ 赤枠のゲノム位置では3つアライメントされたうち、3つのリードが『C』

サンプルでは『C』になっている

NGS解析例(1) 変異解析 ロジック

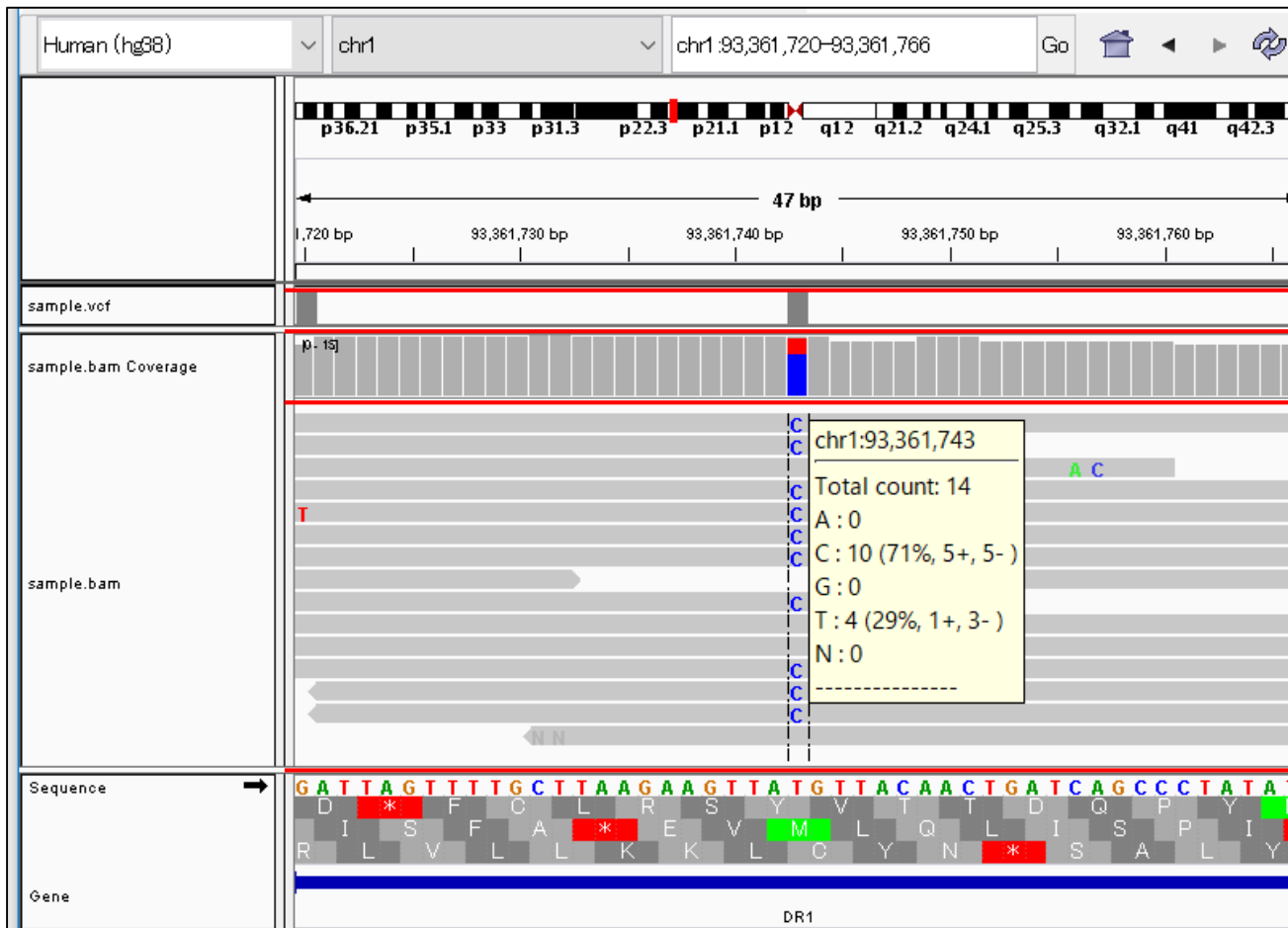


- ▶ ペアリードのマップした位置関係がおかしい

構造が変異している

NGS解析例(1) 変異解析

結果表示例



検出変異
塩基カウント

アライメント
リード

- ▶ 変異率
- ▶ 遺伝子のアミノ酸変異

NGS解析例(2) 発現解析

- ▶ 概要

既知の遺伝子位置情報を参照し、**遺伝子の発現量がわかる。**

- ▶ NGSの利点

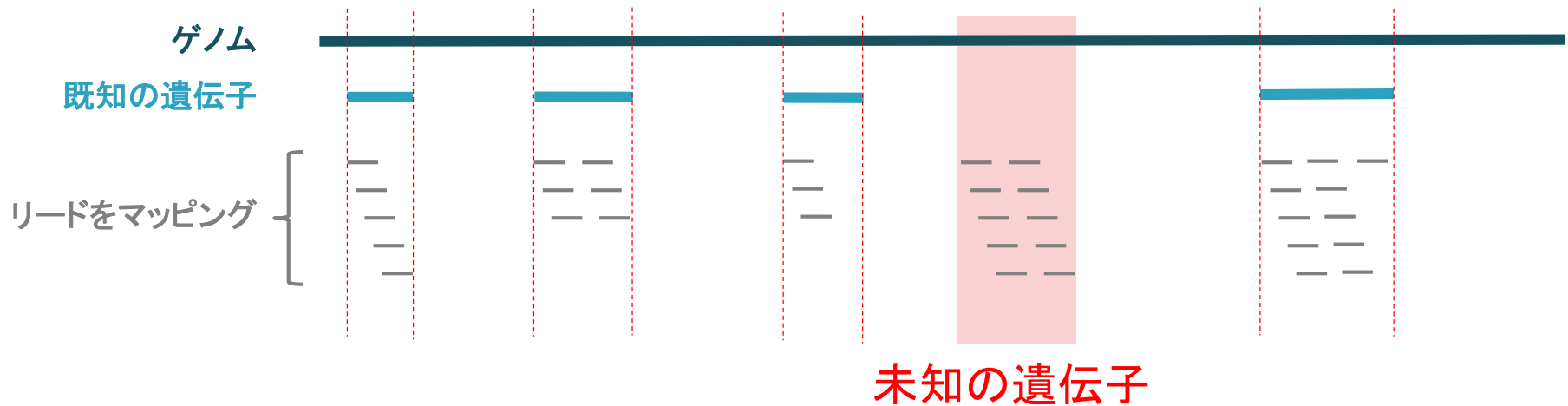
ゲノム全体が見られる。

既知の遺伝子だけでなく、**未知の遺伝子を予測することも可能。**

- ▶ 利用される主要分野

代謝など様々

NGS解析例(2) 発現解析 ロジック



- ▶ データは転写産物のリードのみ

未知の発現領域を同定

NGS解析例(2) 発現解析

結果表示例

ID	chr	tag d1	tag d2	tag d2	gene length	norm tag d1	norm tag d2	norm tag d2	n19 - n2	n19 - n2	n21 - n2	19 / 2	19 / 2	21 / 2	21 / 1
ENSG00000223971	29	224	144	46	1881	12.11250109	11.42530956	4.109532015	0.68719153	8.00296908	7.31577754	1.0601464	2.9474162	2.7801972	0.943265
ENSG00000223972	6	213	138	69	864	25.07497261	23.83744056	13.42019049	1.23753205	11.6547821	10.4172501	1.0519155	1.8684513	1.7762371	0.950646
ENSG00000223973	27	131	78	47	868	15.35062905	13.41124686	9.099163415	1.93938218	6.25146563	4.31208345	1.1446086	1.6870374	1.4738989	0.873661
ENSG00000223974	9	19	30	4	1731	1.116429078	2.586535635	0.388316867	-1.4701066	0.72811221	2.19821877	0.431631	2.8750466	6.6608892	2.316793
ENSG00000223975	29	1	2	0	2027	0.050178868	0.147255161	0	-0.0970763	0.05017887	0.14725516	0.3407614	#DIV/0!	#DIV/0!	2.934605
ENSG00000223976	9	132	41	23	632	21.24376357	9.681910366	6.115529842	11.5618532	15.1282337	3.56638052	2.1941707	3.4737405	1.5831679	0.455753
ENSG00000223977	27	150	134	142	3141	4.857333571	6.366945629	7.597028226	-1.5096121	-2.7396947	-1.2300826	0.7628985	0.6393728	0.8380837	1.310790
ENSG00000223978	X	379	185	119	949	40.62071878	29.0937562	21.07191865	11.5269626	19.5488001	8.02183755	1.3962006	1.9277181	1.3806885	0.716229
⋮															
ENSG00000223979	29	498	375	170	17712	2.859804503	3.159787985	1.612889631	-0.2999835	1.24691487	1.54689835	0.9050621	1.7730937	1.9590851	1.104896
ENSG00000223980	1	148	114	46	1788	8.419160859	9.515500055	4.323282841	-1.0963392	4.09587802	5.19221721	0.8847839	1.9474	2.2009895	1.130219
ENSG00000223981	6	386	369	88	3054	12.85561561	18.03232029	4.842135869	-5.1767047	8.01347974	13.1901844	0.7129208	2.6549473	3.7240426	1.402680
ENSG00000223982	27	60	35	11	936	6.520036216	5.580671703	1.97487753	0.93936451	4.54515869	3.60579417	1.1683246	3.3014889	2.8258318	0.855926
ENSG00000223983	27	1	0	0	1618	0.062863143	0	0	0.06286314	0.06286314	0	#DIV/0!	#DIV/0!	#DIV/0!	
ENSG00000223984	1	267	164	58	981	27.68323634	24.94991784	9.935330492	2.73331849	17.7479058	15.0145874	1.1095522	2.7863428	2.5112318	0.901264
ENSG00000223985	1	84	51	17	1044	8.183769595	7.290611506	2.736350684	0.89315809	5.44741891	4.55426082	1.122508	2.9907605	2.6643557	0.890862
ENSG00000223986	1	133	69	30	2141	6.318435844	4.809796508	2.354658445	1.50863934	3.9637774	2.45513806	1.3136597	2.6833768	2.0426727	0.761232

遺伝子上の
リード数
カウント

ノーマライズ結果

サンプル間比較

- ▶ 遺伝子間比較
- ▶ サンプル間比較

NGS解析例(3) メタゲノム解析

- ▶ 概要(NGSの利点)

環境内のDNAを丸ごとシーケンシングすることによって、その**環境にいるすべての生物のゲノムを取得**することができる。

- ▶ 利用される主要分野

水や土壌などの自然環境
生物の腸内や口腔や皮膚

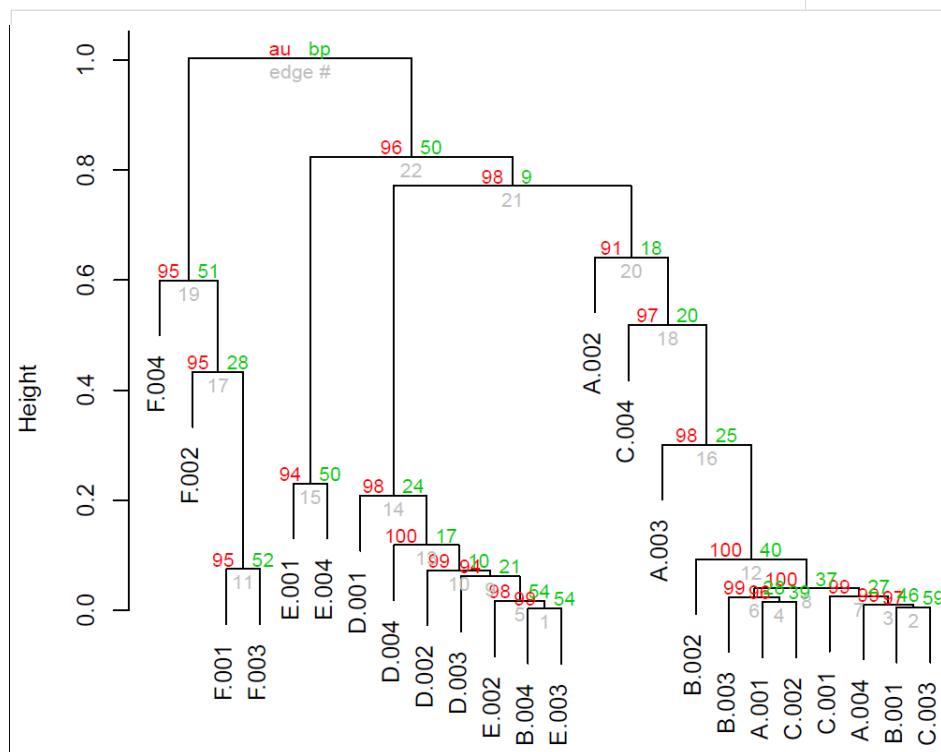
NGS解析例(3) メタゲノム解析

結果表示例

Species	A.001	A.002	A.003	A.004	B.001	B.002
A	0.01459	0.18231	0.02840	0.01524	0.00400	0.09411
B	0.08091	0.12643	0.11472	0.09238	0.16113	0.09282
C	0.02612	0.13254	0.16413	0.09667	0.09962	0.18589
D	0.14554	0.00878	0.04841	0.12232	0.10730	0.13183
F	0.10624	0.11113	0.07104	0.03973	0.01131	0.00022



- ▶ 環境内の生物の割合
- ▶ 環境間の類似度



環境のクラスタリングツリー

失敗事例と解決方法

NGS解析の流れ

サンプル調整

シーケンス

データ解析

つまづきポイント多

可視化

データ解析で必要なこと

専門知識

- ・ 解析ツールはあるが、多くはIT知識を要する。

⇒ 専門的知識

専用の計算機

- ・ データは容量は数TB以上になることも。
- ・ 大量計算は通常のPCでは終わらず、メモリが足りないことも。
- ・ 解析ツールが動く環境は多くがLinux。

⇒ 解析環境の構築

人材の確保

- ・ 作業に拘束されるため時間。
- ・ 知識を身に付ける時間や人員。

⇒ コスト

データ解析とは 例：変異解析

1. クオリティコントロール

クオリティチェックを行い、低クオリティのデータを除去する

2. マッピング作業 ➡ 低マップクオリティの結果が入っている

シーケンスリードをリファレンスゲノムにマッピングする

3. 重複リード除去

PCR duplicatesリードを取り除き、変異検出の際のバイアスを軽減させる

4. マッピング結果の補正

SNVやInDelの正確な位置を求めるために、マッピング結果の補正を行う

特につまづくPoint(1)

生物種などのデータの特異性に
左右される工程

データ解析とは 例：変異解析

5. 変異検出

SNVやInDelを検出する

6. 変異アノテーションの付加

検出された変異に遺伝子への影響や既知の変異の情報を付加する

7. サンプル間比較 ➡ 適した比較ツールがない

共通変異や、特異的変異、世代間の変化などを追う

8. 統計解析 ➡ 計算量が多くて実行できない

反復データを用いて優位さがあるかどうかを計算

特につまづくPoint(2)

一発で実行できる解析ツールがない

失敗しないためには

▶ データ解析の計画

- フローはあるか、途中の結果は大丈夫か
- 必要な手段はそろっているか

▶ 研究計画の再確認

▶ クオリティ、データ量の検討

相談先候補にNGSサービス業界

NGS解析サービスの種類と特徴

▶ シーケンス

NGSを使ったシーケンス解析を行う。

フラグメントの状態を送付することが多い。

生物種と解析の種類によって、機器や調整手法を選んでくれるところもある。

▶ データ解析

□ パッケージ

メニュー化された解析で低コスト。納品物は一定の形式に従う。フォローサービスは少ない。

□ オーダーメイド

研究目的に沿った解析フローから作成。または、依頼フローに従い解析。フォローサービスが充実。

PictBioのご紹介

- ▶ シーケンス、データ解析（パッケージ、オーダーメイド）すべてお取り扱い
- ▶ シーケンス前の計画段階からご相談可
- ▶ 研究者の方のご都合に合わせたご提案

ご清聴、有難うございました。

こちらに資料に関するお問い合わせは

pictbio@axiohelix.com

へお願いいたします。



ホームページの紹介

www.pictbio.com

PictBio
次世代シーケンサー 受託解析サービス

About 解析 ツール 実績 Public **解析メモ** お問い合わせ

New Release
RNA-Seq発現量比較データ作成サービス

PictBioとは
バイオインフォマティクスにおける
あらゆるお悩みを解決するサービスです

PictBio = Picture + Biology
膨大なバイオデータを可視化いたします

おすすめコンテンツ
解析メモ

タグ一覧

- ツール (14)
- 初心者向け (14)
- データベース (7)
- NGS (7)
- バイオインフォマティクス解析 (5)
- トラブルシューティング (4)
- サービス (3)
- トリミング (1)

4 / 4 < 1 2 3 4

ゲノム登録 (D-way) データベース

論文投稿する際に必ず必要になるNGS生データの登録。
タグ (アダプター) のトリミングがなっていない場合、メタデータがなくなったり、なかったり、意外に面倒です。

本家helpページ ⇒ [DDBJ Data Submission](#)
もし面倒でしたら私どもでも有償で代行しております。 ⇒ [お問い合わせ](#)

ご自身で行われる場合には、
マニュアルを熟読して理解したうえで登録作業を行ってください。

[続きを読む →](#)

開発・解析環境ツール

Twitter account : @PictBio

