

# 次世代シーケンサーの データ処理と解析について

アクシオヘリックス株式会社

# 次世代シーケンサ (NGS) の解析作業とは

## NGSから出力されたデータを研究に対して「意味のある」データに変換するための作業

### データ解析処理

- ・ 様々な解析アルゴリズムを用いた解析作業

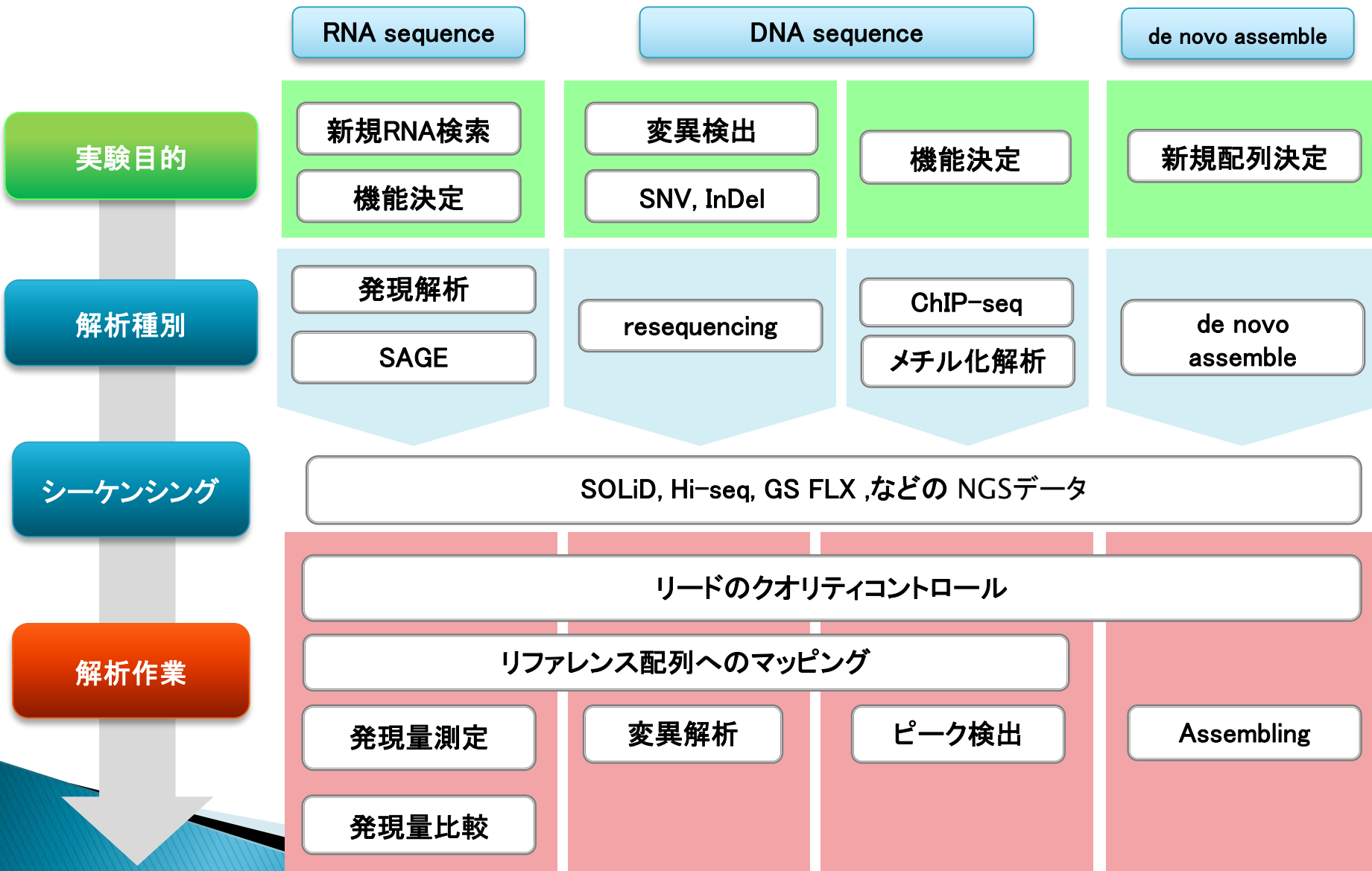
### 統計処理

- ・ データに適した統計処理によるデータマイニング

### グラフ化によるデータの視覚化

- ・ 論文に掲載可能なグラフの作図、分析

# 解析フロー



# 解析作業の流れ(変異解析の場合)

シーケンシングした配列と既知の配列情報を比較し、Insertion, Deletion, SNPsを検出することで、ゲノム上のどの位置で、どのような変異が起こったか知ることができます。

## 1. リードのクオリティコントロール

クオリティチェックを行い、低クオリティのリードを除去する

## 2. マッピング作業

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由来の重複したリードを取り除き、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

**5. 変異検出の実施**  
SNVやInDelを検出する

**6. 変異アノテーションの付加**  
検出された変異に対し、遺伝子のどの領域の変異なのか、もしくは既知の変異であるかの情報を付加する

**7. 変異情報の加工**  
要件に合わせてデータを整形する

**8. 納品物作成作業（データ、報告書、グラフデータ等）**  
ご依頼に応じて納品物として解析結果から納品物を作成する

**9. 納品とご報告**  
解析結果を納品物を元にご説明させていただき、次の解析の計画を検討します。

# NGSデータ解析作業で必要なこと

## 専門知識が必要

- ・解析ツールを利用するためにバイオ及びITの知識が必要

## 専用の環境が必要

- ・シーケンスデータは数千万～数億リード数、容量としては数TB以上のものもある。
- ・シーケンスデータは大量データの場合もあり、通常のPCでは解析が進められない場合も。
- ・専用の計算サーバーが必要となります。

## 解析に要する時間の確保

- ・解析作業に拘束されるため時間の確保ができない。
- ・知識を身に付ける時間や人員の確保が難しい。

結果として、第三者機関へ解析を依頼することが効率的

# NGSデータ解析サービスの種類と特徴

サービス形態	メリット	デメリット
シーケンシング ＋ データ解析	<ul style="list-style-type: none"> <li>シーケンシングから解析までを一括して依頼できる</li> <li>セット価格のため費用も抑えられる。</li> </ul>	<ul style="list-style-type: none"> <li>納品形式の指定ができない</li> <li>フォローが受けられない。</li> </ul>
データ解析	<ul style="list-style-type: none"> <li>シーケンサーを選択できる。</li> <li>メニュー化された解析なので、コストが抑えられる。</li> </ul>	<ul style="list-style-type: none"> <li>シーケンサーは別途依頼が必要</li> <li>納品形式を選択することができない。選択できた場合もコストが割高になってしまう。</li> <li>フォローが受けられない場合が多い</li> </ul>
データ解析 ＋ コンサルティング	<ul style="list-style-type: none"> <li>シーケンサーを選択できる。</li> <li>解析プランについて計画時に相談することができる。</li> <li>フォローが受けられる。</li> <li>解析途中で変更依頼ができる。</li> </ul>	<ul style="list-style-type: none"> <li>シーケンサーは別途依頼が必要</li> <li>コストが割高になってしまう。</li> </ul>

低

自由度・コスト

高

# データ解析＋コンサルティングサービス 「PictBio」のご紹介

## 特徴1 研究者様と研究目的に合わせた準備

- 解析方針についての打合せ、ヒアリング  
(Skypeなどのテレビ会議を推奨いたします。)
- 解析事例を交えて、解析提案

## 特徴2 オーダメイド解析

- 解析アルゴリズムの選択
- 新規解析アルゴリズム、パイプラインの開発

## 特徴3 作業内容の柔軟性

- 解析途中での結果報告、および作業方針再検討の受付
- 論文などへの掲載実績のない解析への対応

---

▶ Axiohelix の PictBio とは

- 膨大なNGSデータを可視化(Picture)するNGSデータ解析パッケージ



# シーケンシングデータ解析支援

## 変異解析

シーケンシングした配列と既知の配列情報を比較し、Insertion, Deletion, SNPsを検出することで、ゲノム上のどの位置で、どのような変異が起こったか知ることができます。

## 新規ゲノム配列決定

配列情報が未知の領域や、未知の生物種について、配列を決定するための手法です。

## トランスクリプトーム解析

RNAをシーケンシングし発現量を定量する手法で新規RNAの探索にも威力を発揮します。

## ChIP-seq解析

免疫沈降法を利用し、タンパク質と結合するDNA配列を解読することで、転写因子が結合する領域を明らかにすることができます。

## メチル化解析

遺伝子発現がどう制御されるか調べるためにゲノム配列のメチル化状態を解析する手法です。

## メタゲノム解析

複数生物種のゲノムをまとめて網羅的にシーケンシングすることで、異なる条件下での生物群の組成を調べたり、物質代謝を推測することが可能になります。

# その他の解析支援

## アノテーション 情報の付加

- 解析結果に対し公共データベースからのアノテーション付加や統計解析等を行ない、生物学的に意味のあるデータのみを抽出します。

## システム構築

- 大量データを管理するためのデータベースや、ラボ内のデータの統合アプリケーション等、ご要望に沿ったシステムを構築いたします。

## 人材教育

- バイオインフォマティクスの知識・技術を習得したい研究者や学生の方へ向け、プログラミングや解析に関するセミナーを行います。

# 解析サービスの流れ

## ▶ 1 解析案件につき 通常2～3ヶ月

### Step1

- ・シーケンシングが終了した後のデータについてヒアリング
- ・ご要望の解析、最終的な納品形態についてヒアリング
- ・ヒアリングに基づく解析プランのご提案

### Step2

- ・解析の実施

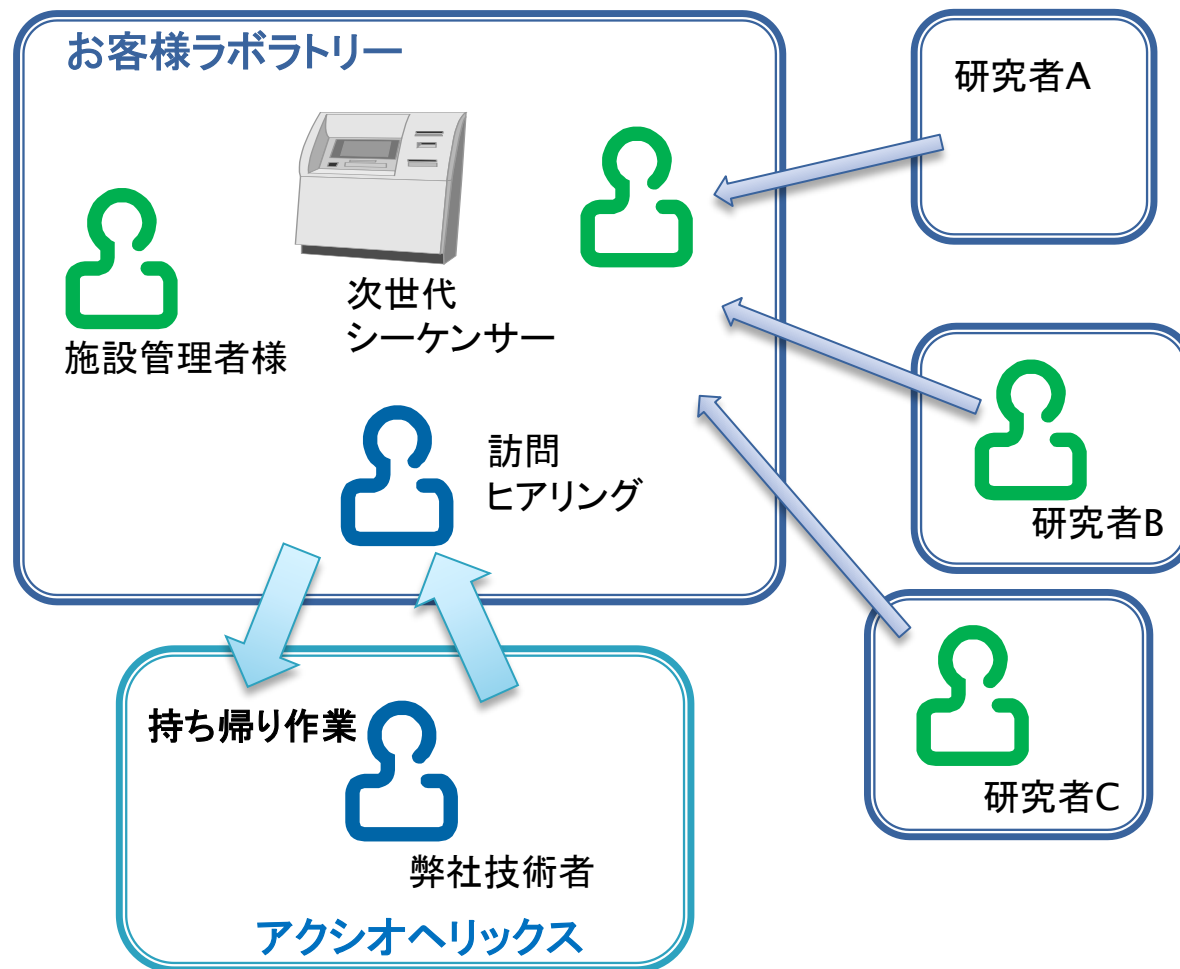
### Step3

- ・解析結果の提示、説明
- ・修正のご要望に応じて打ち合わせ

### Step4

- ・修正した解析結果の納品
- ・結果を踏まえて次の解析についてお打合せ

# 一括複数件でのご依頼の作業



複数案件に対して効率的なお打ち合わせを実施することで、コストを抑えることができます。

# 納品物について

- ▶ **ご要望、ご予算に応じた納品物を作成いたします。**  
例えば、Excel形式でソートを掛けた状態でのデータ作成や、指定の形式のグラフを使用した比較資料など論文で利用できる形式で作成いたします。
- ▶ **修正要望に対応いたします。**  
事前に納品物の内容をご確認いただき、修正などのご要望があれば、それを反映した最終納品物を作成いたします。
- ▶ **納品完了後のフォローアップに対応いたします。**  
納品完了後であっても、納品物についての説明などフォローアップを行っております。

# 納品物事例1：発現解析

目的: ある発生段階における遺伝子発現の経時的変化の比較

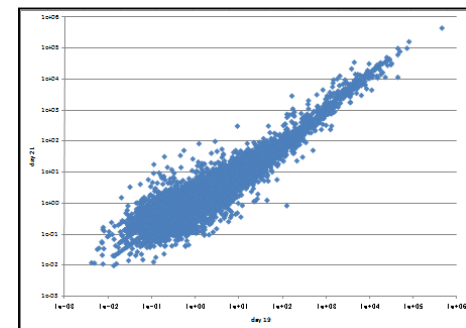
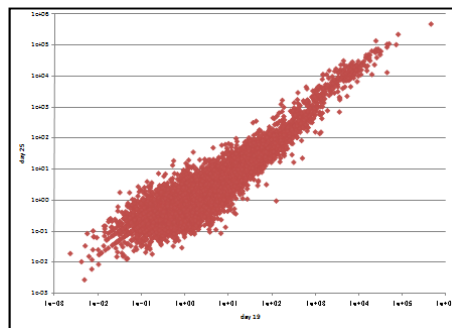
ID	chr	tag d1	tag d2	tag d2	gene length	norm tag d1	norm tag d2	norm tag d2	n19 - n2	n19 - n2	n21 - n2	19 / 2	19 / 2	21 / 2	21 / 1
ENSG000000000001	29	224	144	46	1881	12.11250109	11.42530956	4.109532015	0.68719153	8.00296908	7.31577754	1.0601464	2.9474162	2.7801972	0.943265
ENSG000000000002	6	213	138	69	864	25.07497261	23.83744056	13.42019049	1.23753205	11.6547821	10.4172501	1.0519155	1.8684513	1.7762371	0.950646
ENSG000000000004	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
ENSG000000000005	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
ENSG000000000006	29	1	2	0	2027	0.050178868	0.147255161	0	-0.0970763	0.05017887	0.14725516	0.3407614	#DIV/0!	#DIV/0!	2.934605
ENSG000000000007	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
ENSG000000000011	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
ENSG000000000012	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
ENSG000000000015	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
ENSG000000000016	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
ENSG000000000018	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
ENSG000000000020	27	60	35	11	936	6.520036216	5.580671703	1.97487753	0.93936451	4.54515969	3.60579417	1.1683246	3.3014889	2.8258318	0.855926
ENSG000000000024	27	1	0	0	1618	0.062863143	0	0	0.06286314	0.06286314	0	#DIV/0!	#DIV/0!	#DIV/0!	
ENSG000000000029	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
ENSG000000000030	1	84	51	17	1044	8.183769595	7.290611506	2.736350684	0.89315809	5.44471891	4.55426082	1.122508	2.9907605	2.6643557	0.890862
ENSG000000000031	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

カウント

比較

Excelの機能を使用して、選択されたデータに対して比較対象データが自動的にソートされる形式で資料を作成いたしました。

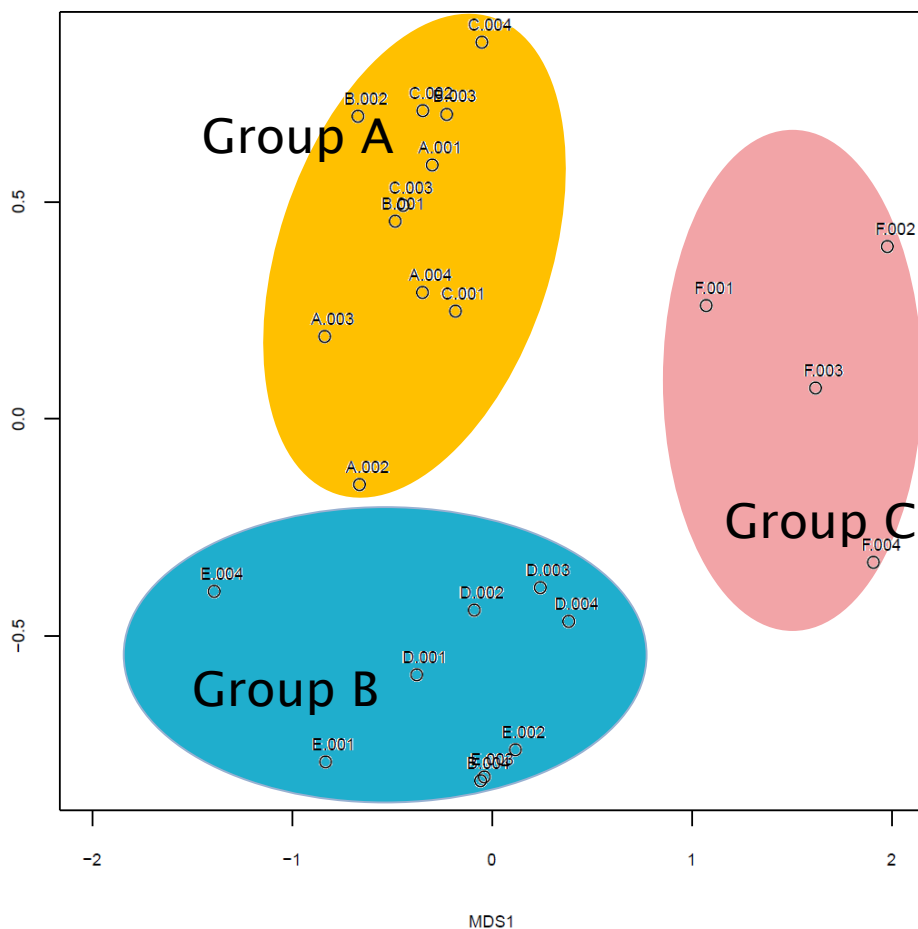
グラフ化



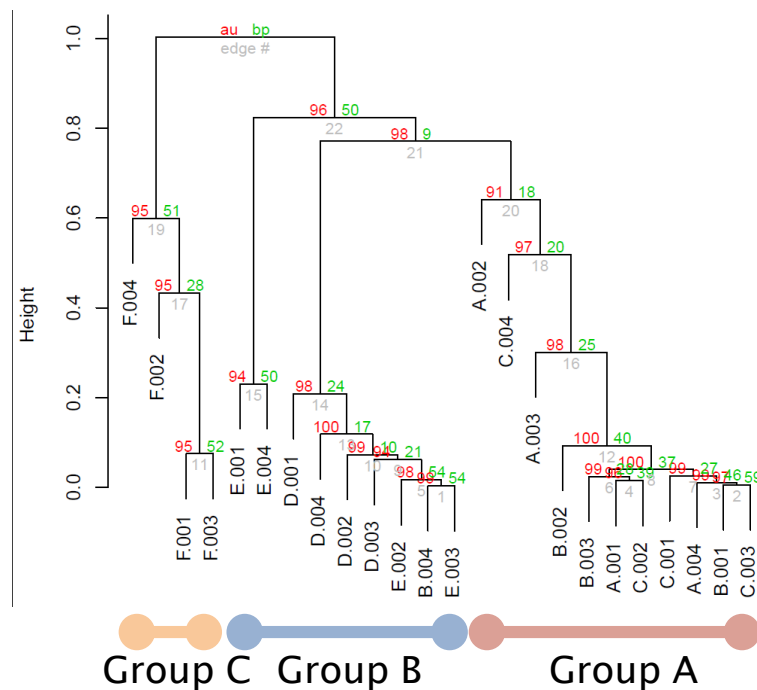
# 納品物事例2: 集団解析

## メタゲノム生物種クラスタリングのグラフ作成例

距離プロット図



クラスタリングツリー



# 失敗しないNGS解析のために

- ▶ NGSから得た大量の配列情報から解析を行い、研究に必要なデータを絞り込むことが重要

1. 研究の計画段階で以下のものを決定しておく。
  - (1) 注目すべきゲノム上領域をできるだけ選出しておく
  - (2) フィルタリングの為に用意するデータの選定  
コントロール、比較条件、同一条件 etc...

これらが決定されていないと・・・

- 結果範囲が広すぎて、変異を見つけられない。
- サンプル数が少なく、ターゲットの変異を絞り込めない
- コントロールサンプルが無い場合、目的ピーク以外のものが検出される



▶ 2. データの正確性はサンプルの状態に左右される

(1) クオリティの高いデータを取る

比較を意識していないサンプリング、コンタミ etc...は、データが偏る原因となりやすい

(2) 解析の種類、リード長によって、カバレッジを考慮し、必要量のデータを得る

→ミスアセンブルの発生

→サンプル間のデータ量に大きな差があるため、比較解析時に偏った値が出る

→コンタミにより、カバレッジ(データ量)が大幅に減ってしまい、条件が変化してしまう。

# NGS解析を失敗しないためには・・・

実験における最大の課題であり、

もっとも重要なのはサンプリング

見たい事象のデータを得られるよう、

慎重に計画立案をすることが必要

# 解析サーバー「BcMARx」のご紹介

# 解析サーバー「BcMARx」のご紹介

- ▶ 「BcMARx」は、ご自身で解析を行いたい方におすすめの解析サーバーです。
- コマンドラインを必要としない、インターフェースを実装  
難しい設定は不要です。最低限の入力のみで利用できます。
- 論文に引用しやすいオープンソースツールで構成  
論文に掲載実績のある信頼性の高いツールで構成されています。
- 頻度の高い解析手法を網羅  
弊社実績において、もっともニーズの多い5つの解析に対応しています。  
Bisulfite-seq、ChIP-seq、Mutation、Assemble、RNA-seq
- 充実のサポートサービス  
導入時に技術者による訪問設置サービスや、初回レクチャーサービス、  
解析結果の見方などを詳しく記載したマニュアルを準備しております。

# コマンドライン不要の親切UI

Linuxではコマンドラインによる操作が必須

【親切UI】

専用のUIを実装、Linux特有のコマンドラインを意識せずに利用することができます。

```

-p <FILE>: write profiling data to <FILE>
-w: wait for instrumentation to finish before returning
start profiling: am profile <PROCESS> start <FILE>
stop profiling: am profile <PROCESS> stop
start monitoring: am monitor [--gdb <port>]
--gdb: start gdbserver on the given port at crash/ANR
<INTENT> specifications include these flags:
[-a <ACTION>] [-d <DATA_URI>] [-t <MIME_TYPE>]
[-c <CATEGORY>] [-c <CATEGORY>] ...
[-el --es <EXTRA_KEY> <EXTRA_STRING_VALUE> ...]
[--esn <EXTRA_KEY> ...]
[--ez <EXTRA_KEY> <EXTRA_BOOLEAN_VALUE> ...]
[-el --ei <EXTRA_KEY> <EXTRA_INT_VALUE> ...]
[-n <COMPONENT>] [-f <FLAGS>]
[--grant-read-uri-permission] [--grant-write-uri-permission]

```

**B c M A R x**

解析メニュー 解析状況確認

### 変異解析パラメータ設定

以下の項目を設定してください。\* のついた項目は必須項目です。

設定ファイル ⓘ	<input type="button" value="ファイルを選択"/>	選択されていません
出力ディレクトリ名 ⓘ	<input type="text" value="/mutation"/>	
リードフォーマット ⓘ	<input type="radio"/> FASTA <input checked="" type="radio"/> FASTQ	
リファレンス配列 * ⓘ	<input type="text" value="Human(hg38)"/>	
既知の変異データ ⓘ	<input type="button" value="ファイルを選択"/>	選択されていません
ターゲットのデータ ⓘ	<input type="button" value="ファイルを選択"/>	選択されていません
リードタイプ ⓘ	<input checked="" type="radio"/> シングルリード <input type="radio"/> ペアリード	
リード ⓘ	<input type="button" value="ファイルを選択"/>	選択されていません

**B c M A R x**

解析メニュー 解析状況確認

解析メニュー

- Bisulfite** 【Epigenome解析】メチル化レベルの特定や、その度合いから形質発現の原因を探ります。
- ChIP-seq** 【Epigenome解析】クロマチンが遺伝子発現の原因の特定を行います。
- Mutation** 【DNA解析】遺伝子に対するゲノム上の変異を特定し、先天的形質発現の原因の特定を行います。
- Assemble** 【DNA解析】リードを繋ぎ合わせ、未知の配列の決定や興味のある部分の配列を決定します。
- RNA-seq** 【RNA解析】時系列や場所の違いによる発現量の変化の測定などから集団の特性の原因を探ります。
- マニュアル 【ユーザーマニュアル】困ったときや詳細な解説をご確認になりたいときにご参照ください。
- 弊社リンク 【PicBio】弊社サービスのPicBioについてのお問い合わせなどはこちらから。

5種類のパイプラインを準備

# ヘルプ機能や、予約機能などの便利機能を実装

リードフォーマット ⓘ FASTA  
FASTQ  
既に発見されている変異の情報ファイル (VCFフォーマット)を指定してください。正確なマッピングを行う際に必要となります。  
既知の変異データ ⓘ ⓘ ファイルを選択 選択されていません  
ターゲットのデータ ⓘ ⓘ ファイルを選択 選択されていません  
リードタイプ ● シングルリード

## 【ヘルプ機能】

設定項目ごとに吹き出し形式で設定情報について説明を表示します。

## 【進捗確認画面】

複数の解析を一度に設定して、予約することができます。

解析状況一覧

種別	解析名	予約日時	開始日時	終了日時	状況	中断	詳細確認
<input type="checkbox"/>	Mutation 変異解析_run20140401	2014/05/01 13:35:00	--	--	予約		<a href="#">詳細</a>
<input type="checkbox"/>	Mutation 変異解析_run20140301	2014/05/01 13:30:00	--	--	予約		<a href="#">詳細</a>
<input type="checkbox"/>	Mutation 変異解析_run20140201	2014/05/01 13:25:00	2014/05/01 13:26:00	--	実行	<a href="#">中断</a>	<a href="#">詳細</a>
<input type="checkbox"/>	Mutation 変異解析_run20140101	2014/05/01 13:20:00	2014/05/01 13:20:00	2014/05/01 13:26:00	終了		<a href="#">詳細</a>

[チェックした解析の子約を取り消す](#)

Copyright © AXIOHELIX

# SPEC

	ベーシック	アドバンス
解析パイプライン (選択) <sup>*1</sup>	<ul style="list-style-type: none"> <li>・ Bisulfite-seq (メチル化部位の特定)</li> <li>・ ChIP-seq (タンパク質結合部位の特定)</li> <li>・ Mutation (ゲノムの変異解析)</li> <li>・ Assemble (リードから元の長い塩基配列を再構築)</li> <li>・ RNA-seq (発現解析)</li> </ul> <p>上記より2つをご選択下さい</p>	
推奨ゲノム	微生物ゲノム (Assembleを除く <sup>*2</sup> )	ヒトゲノムなど (Assembleを除く <sup>*2</sup> )
リファレンスゲノム	ご指定のリファレンスゲノムを3種セットアップ致します。	
OS	Cent OS 6.5 (64bit)	
CPU	Intel Xeon E3-1275 v3 [3.5GHz/4Core]	Intel Xeon E5-2650 v2 [2.6GHz/8Core]
メモリ	32GB (8GB×4本)	64GB (16GB×4本)
ストレージ	HDD : 146GB×2台 1TB×2台	HDD : 146GB×2台 1TB×2台
RAID	ハードウェア方式 (RAID1)	
同梱物	21.5インチディスプレイ、マウス、キーボード、各種マニュアル	
保証期間	1年間 引き取り修理 (3年、5年の保証サービス有料延長可能)	
その他のサービス	訪問設置、初回レクチャーサービス、IGVインストール	

# お知らせ

NGS解析サービス PictBio紹介サイト

<http://www.pictbio.com>



お問い合わせシートダウンロードURL

[http://www.pictbio.com/download/Pictbio\\_order\\_sheet.xlsx](http://www.pictbio.com/download/Pictbio_order_sheet.xlsx)

- ▶ 「第37回日本分子生物学会」に出展いたします。
  - 日程 11月25日～27日
  - 場所 パシフィコ横浜 展示ホールA・B・C
  - ブース 163ブース(入り口 目の前の島)



ご清聴いただき有難うございました。