

次世代シーケンサー(NGS) 研究事例に基づく活用講座

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

メイン事業

ライフサイエンス (PictBio)

- ・ NGSデータ解析、統計解析
- ・ 解析アプリケーション開発
- ・ 講演会の実施など

その他事業

ITソリューション

ものづくり事業

本社

沖縄県那覇市

東京支社

秋葉原

代表取締役社長

シバスタランスハルナン
(スリランカ人)

創立

2001年6月8日



↑
ゲノム解読時代
麹菌ゲノム解読プロジェクトに参画

アジェンダ

1. NGSと現在の状況

1. NGS解析の仕組み

1. 注意点

1. 簡単に使うには

1. NGSと現在の状況

NGSとは？

Next Generation Sequencer

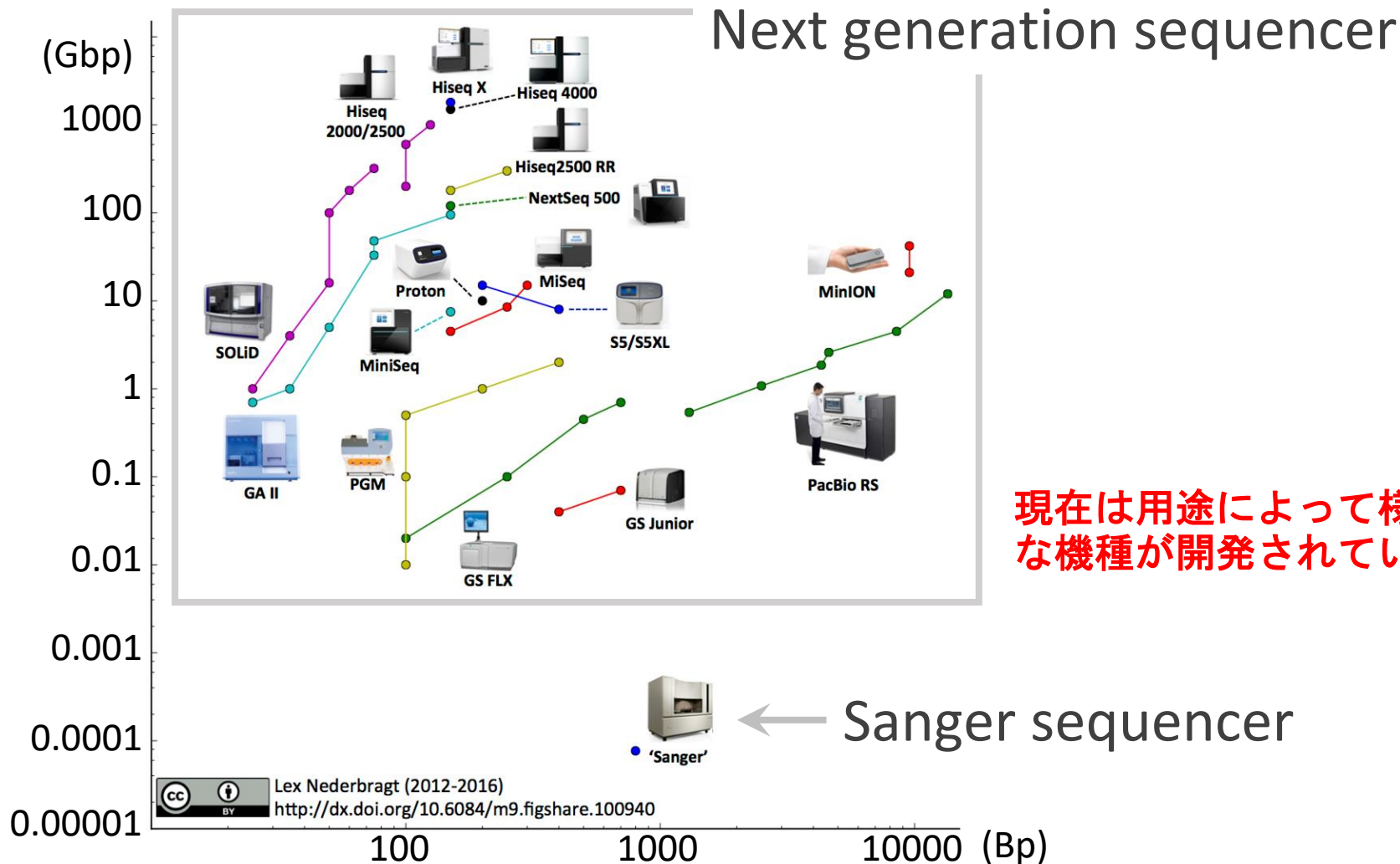
- ・ 高いコストパフォーマンス、短い期間で一度に大量のDNAの配列を決定できるDNAシーケンサー

| <i>HiSeq</i> vs. <i>Sanger</i> |  HiSeq ヒトゲノムの150倍以上 |  'Sanger' |
|--------------------------------------|--|--|
| データ量 | 最大500 Gbp | 1-100 Kbp |
| ヒトゲノム辺り | | |
| 費用 (/3Gbp) | 数十万 | 15億円以上 |
| 解析時間 | 1-2日 | 0.5-1日 |

網羅的な解析が可能になっている

機器ラインナップと性能

トータルデータ量



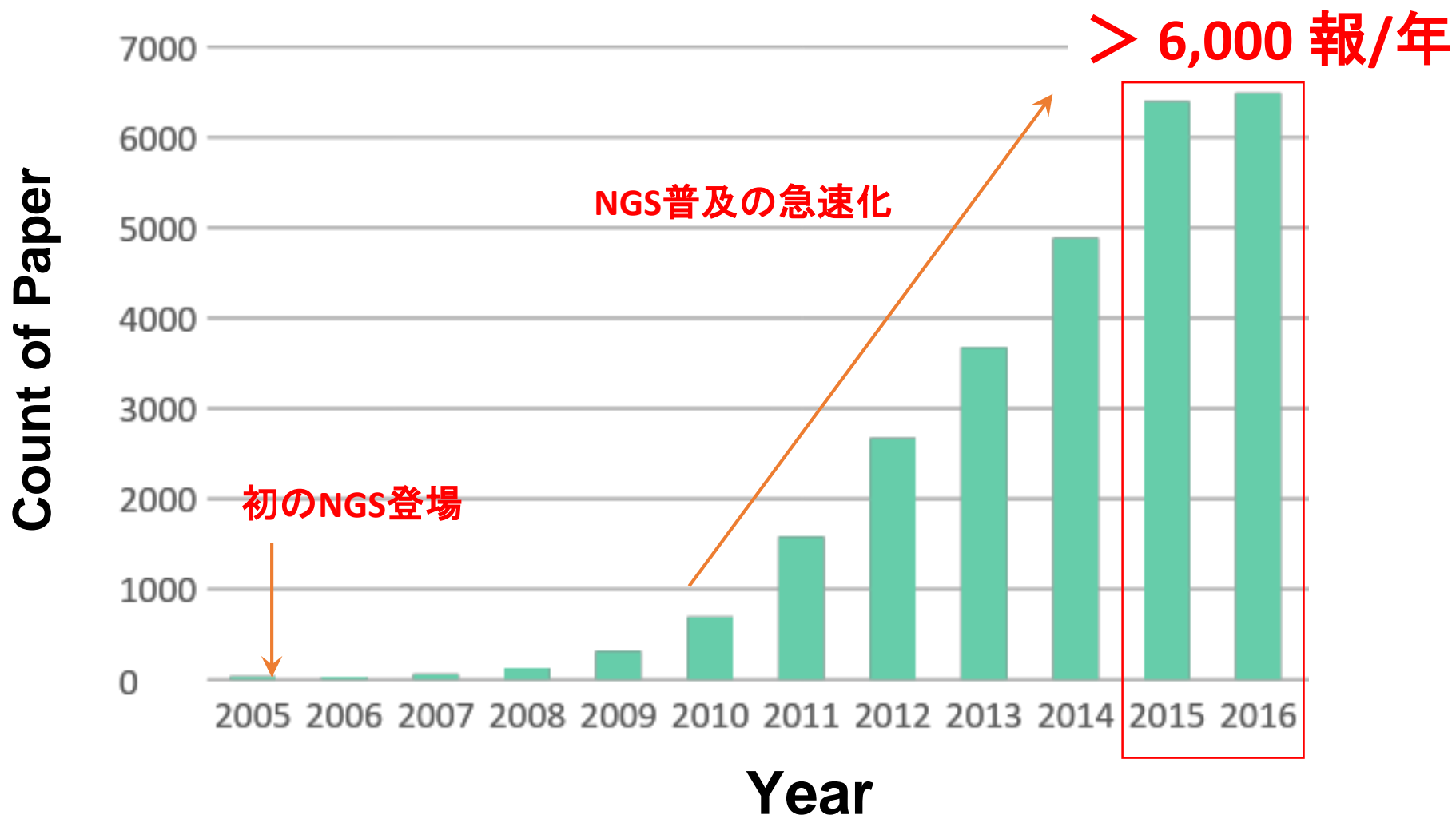
現在は用途によって様々な機種が開発されている

← Sanger sequencer

Lex Nederbragt (2012-2016)
<http://dx.doi.org/10.6084/m9.figshare.100940>

リード長

NGSを使った研究報告

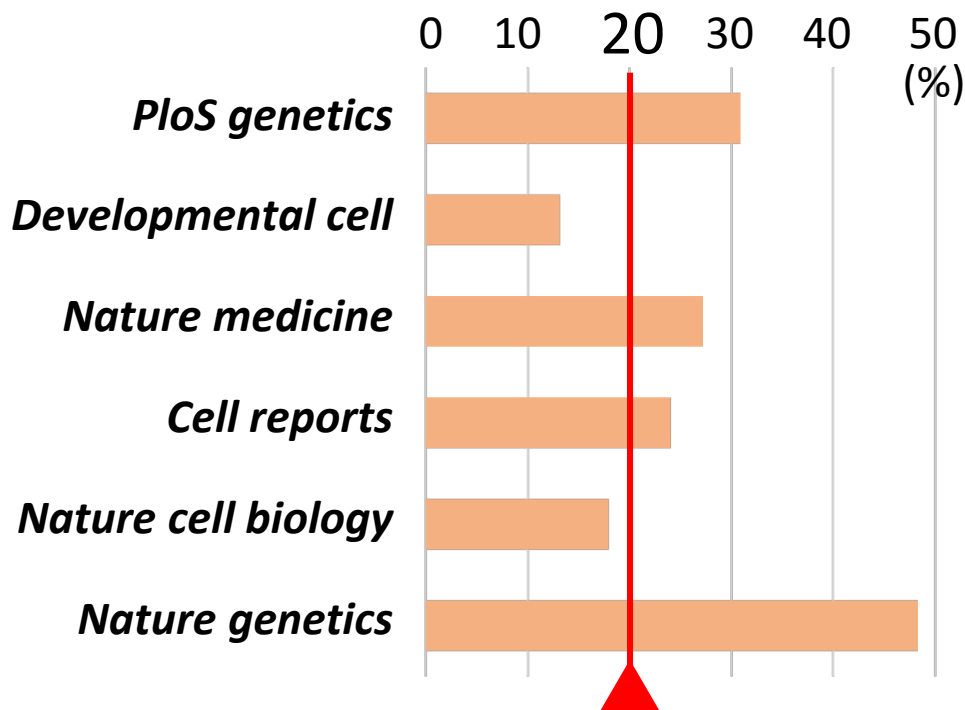


(PubMedで「Next Generation sequencer」を検索した際のヒット数 弊社調べ)

NGS関連論文数

各ジャーナルにおけるNGS関連論文の割合

(6誌をピックアップ：2017年のデータ 弊社調べ)



約5報に1報の割合でNGSが使用されている

まとめ

- NGSの出現により、網羅的解析が可能になった
- 目的によって様々なNGSが開発されている
- NGS関連論文数が増えてきている
- NGS関連の論文の割合が大きくなってきている

今後の研究はNGSを使用することが前提になる

2.NGS解析の仕組み

NGSでできること

代表的な手法

- ・ 変異の探索
- ・ 病気の原因遺伝子の特定
- ・ 個人の多様性を検出

Whole Genome-seq
Exome-seq

- ・ 網羅的転写解析
(Micro Arrayのよう
な)

- ・ 新規転写産物の検出

- ・ ヒストン修飾状態
- ・ タンパク質結合部位の同定

ChIP-seq

- ・ 網羅的CpGのメチル化状態の解析
Bisulfite-seq

- ・ ゲノムの立体構造解析
Hi-C

- ・ RNA上の翻訳領域の同定
Ribosome profiling

- ・ RNAの二次構造解析
SHAPE-seq

他にも数多くの網羅的手法が用意されており
NGSによって様々なことができる

NGS解析の内訳

NGS解析全体

1 次解析

NGSを使用した解析
DNA溶液からの配列決定

2 次解析

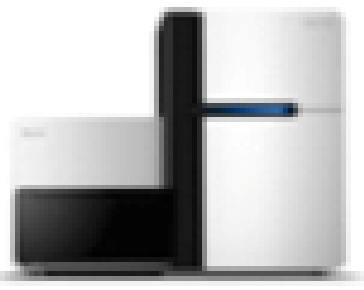
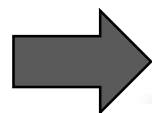
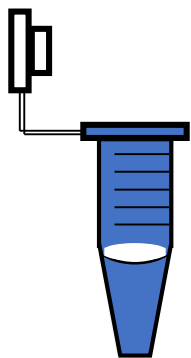
情報科学的な解析
コンピュータを使用して
大量な配列データを処理する。

一次解析

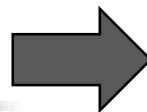
NGSはDNAの配列決定だけ行う

INPUT

DNA断片の溶液
(ライブラリ)



Hiseq



OUTPUT

数億本の配列ファイル

```
>fragment_1
  AACGGTGCCGG ...
>fragment_2
  CTGAGCTCAGG ...
>fragment_3
  CTTAGTAAGAA ...
>fragment_4
  CCATGGCACGT ...
  ⋮
```


DNA断片溶液に含まれていた断片の配列を
ランダムかつ大量に決定

2 次解析

NGSから得られるデータは**ただの配列データ**である

→ 考察できるデータに作り直す必要がある

```
Read_00001: ...TGC GAAGCGTGAGGCGGACGGATCTGGCGAGAATACTCCATTGGTTGCGATAGACCCGTGTCCAAGTGAACA
Read_00002: ...TATCACTAGTGGCTTTCTCCTAGGGCCTAATCGTCCGCGACATCGCTAGAGTCACTGTTGACGCGTGCCCG
Read_00003: ...GTGCCCGATGCCACCAGCACAGTCACCATAGCGAAAGCAGGCAAGTACCAGCTGATGTTAGGGAACAAATC
Read_00004: ...ATACTTCCCATCAGCAACTGTGTGTGGAACGAAGTCAGTAGGAAGCAGTCCCAGGTGATACTGTATGAAGCC
Read_00005: ...GTCATTACCCCGGACTGACCAAGGACAGACACTGTTGCTCTCCGTCAGGACGCTCATAGCCCTCCCGACTGCTCCG
```



| chr1 | position | Ref | Alt | Gene Symbol | AA change | Area | Refseq ID |
|-------|----------|-----|-----|-------------|--------------|-------------------------|----------------|
| chr17 | 73621432 | A | G | MYO15B | | intron_variant | NR_003587.2 |
| chr17 | 73625430 | G | A | MYO15B | | downstream_gene_variant | NR_003587.2 |
| chr17 | 73625430 | G | A | RECQL5 | p.His691 His | synonymous_variant | NM_004259.6 |
| chr17 | 73625430 | G | A | SMIM5 | | upstream_gene_variant | NM_001162995.2 |
| chr17 | 73625478 | T | C | MYO15B | | downstream_gene_variant | NR_003587.2 |
| chr17 | 73625478 | T | C | RECQL5 | p.Thr675Thr | synonymous_variant | NM_004259.6 |
| chr17 | 73625478 | T | C | SMIM5 | | upstream_gene_variant | NM_001162995.2 |
| chr17 | 73625629 | A | C | MYO15B | | downstream_gene_variant | NR_003587.2 |
| chr17 | 73625629 | A | C | RECQL5 | | intron_variant | NM_004259.6 |
| chr17 | 73625629 | A | C | SMIM5 | | upstream_gene_variant | NM_001162995.2 |
| chr17 | 73626208 | G | C | MYO15B | | downstream_gene_variant | NR_003587.2 |

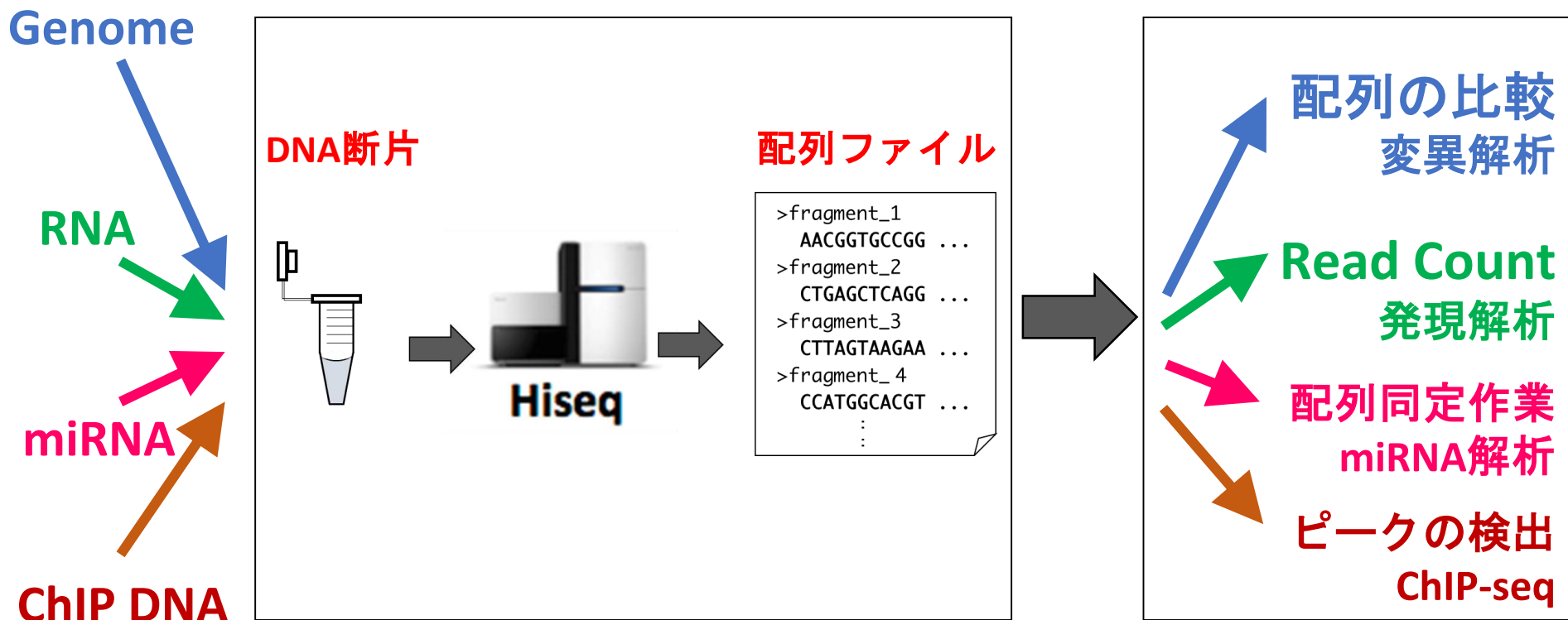
非常に膨大なデータ

→ コンピュータを使用して目的に沿った解析を行う

フローはすべての解析で共通

1 次解析

2 次解析



INPUTのサンプルと2次解析を変更することにより
手法の多様性が生まれる

NGSを使用した変異解析例



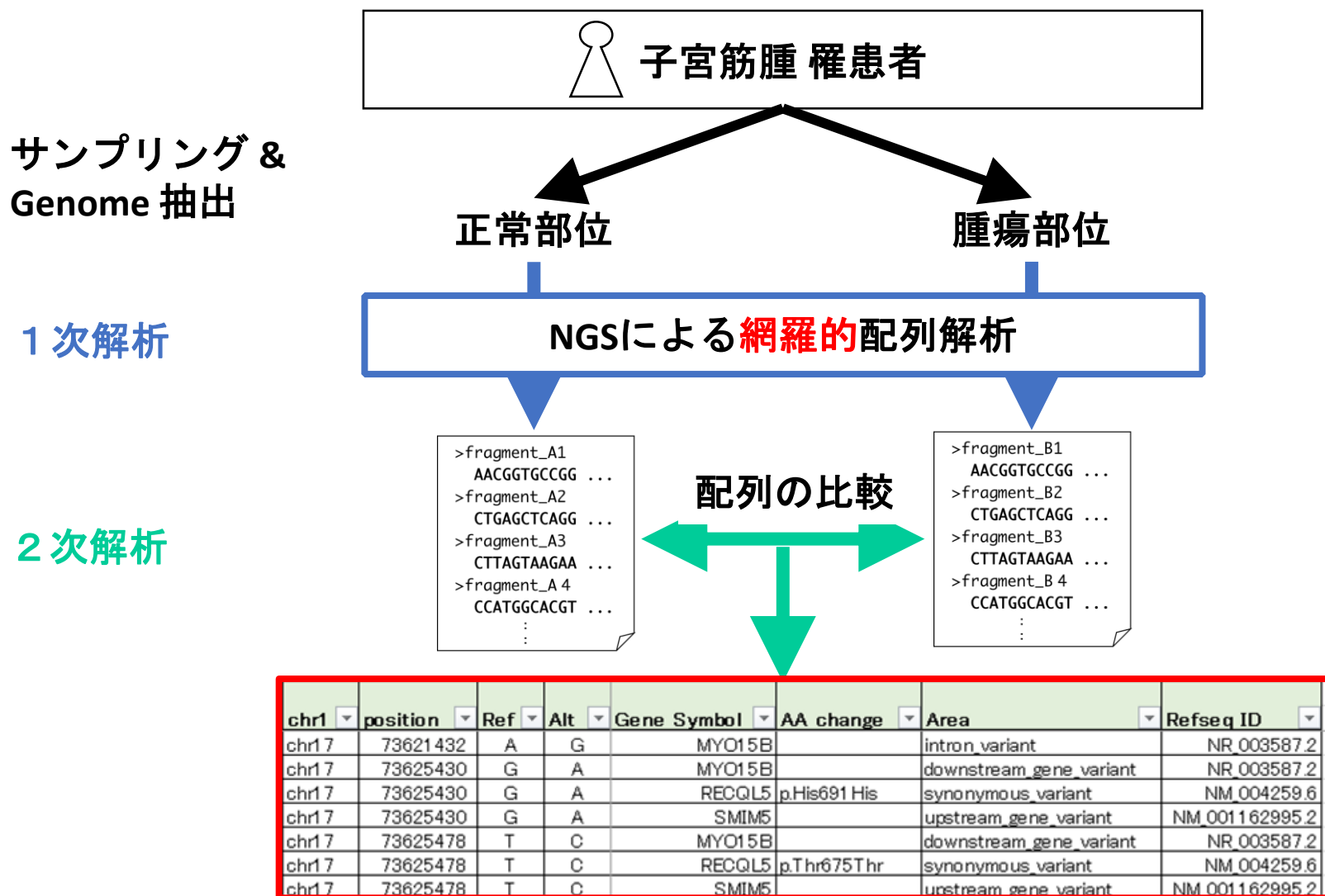
Published 25 August 2011 on Science Express

MED12, the Mediator Complex Subunit 12 Gene, Is Mutated at High Frequency in Uterine Leiomyomas

Netta Mäkinen,^{1*} Miika Mehine,^{1*} Jaana Tolvanen,¹ Eevi Kaasinen,¹ Yilong Li,¹ Heli J. Lehtonen,¹ Massimiliano Gentile,² Jian Yan,³ Martin Enge,³ Minna Taipale,^{1,3} Mervi Aavikko,¹ Riku Katainen,¹ Elina Virolainen,⁴ Tom Böhling,^{4,5} Taru A. Koski,¹ Virpi Launonen,¹ Jari Sjöberg,⁶ Jussi Taipale,^{1,3} Pia Vahteristo,¹ Lauri A. Aaltonen^{1†}

子宮平滑筋腫に関係のある遺伝子を見つけた論文

NGSを使用した変異解析例



腫瘍組織に存在する変異のリストを作成

NGSを使用した変異解析例

17人の罹患者で同様に解析



罹患者間で比較



17人の中で最も多かった変異を特定



**MED12遺伝子上で
腫瘍と最も関連がある変異を発見**

まとめ

NGS解析は 1 次解析と 2 次解析の段階に分かれる

- **1 次解析はNGSによる配列解析**
- **2 次解析はコンピュータを使用したデータ解析**

3. 注意点

NGSを使う前に覚えておくこと

▶ 1, 実験計画の重要性

2, 2次解析の大変さ

実験計画の重要性

本当にある相談

×「とりあえずNGSを使ってみた。
このデータ使って何かわからないかな??」

NGS解析には

- ・ 目的に適したシーケンサー
- ・ 目的に適した手法
- ・ 目的に必要なサンプルセット

が存在する

綿密な計画が大切

実験計画の重要性

失敗しないために必要なこと

1. NGSデータが得られたらどのように使うか？

具体的なデータおよび、解析方法を明確にイメージする。

2. NGSによる配列解析ではどうする？

シーケンサーの選択、手法の選択。

3. 予算計画

予算を有効に使うための戦略をたてる。

これらに注意することで、目的に沿った結果を得ることができる

NGSの専門家に相談する

小スケールなNGSを選択してみる

NGSを使う前に覚えておくこと

1, 実験計画の重要性

▶ 2, 2次解析の大変さ

2次解析の大変さ

よくある話

×「コンピュータ動かすだけなんだから
すぐに結果をだせるでしょ？」

2次解析は簡単ではない

- ・ コンピュータにも計算量の制限はある。
 - 処理の時間も必要
 - 処理が詰まっていれば、解析を待つ必要もある
- ・ データ解析処理を何度も行って最適解を模索する。
 - 実験目的に適した、ツールの使い方の模索が必要
 - 解析用にプログラムを書く場合が多い
 - 結果を一つ一つ見て確認する必要がある

想像以上の、期間と労力が必要

2次解析で必要なこと

期間が必要

- ・最低でも2-3ヶ月は必要
- ・長ければ半年以上の場合もある。

人手が必要

- ・情報科学と生物学が理解できる人材
- ・データ解析方法の試行錯誤する労力

あらかじめ**期間**と**人手**を確保しておく
必要がある

4. 簡単に使うには

NGSを使用する環境

外注環境が揃っている

| | 速さ※1 | 相談、計画 | NGS配列解析 | 2次解析 |
|------|------|-------|---------|------|
| 外部委託 | 早め | | | |
| 共同研究 | 遅め | | | |

※1 状況、人による ※2 少なくともアクシオではOK

現在は、解析の一部または全部を委託することが可能

NGS解析の委託先、相談先

具体的にNGSの環境を提供している機関

研究機関

- ・ 国立遺伝学研究所
- ・ 理化学研究所
- ・ 大学など教育機関
- ・ その他

企業

- ・ タカラバイオ
- ・ Eurofins Genomics
- ・ BGI Japan
- ・ コスモ・バイオ
- ・ 先端医療開発（沖縄綜研）
- ・ アクシオヘリックス

：

選択肢はたくさんある

外注で委託できる作業内容

例) アクシオヘリックスで行っているサービス

計画のお手伝

- ・ NGS研究に関する調査
- ・ 最新手法の調査、報告
- ・ 経験を元に適した手法をご提案

NGS配列解析

- ・ 適したシーケンサーのご提案
- ・ 現行の全てのシーケンサーによる配列解析を受託

情報解析

- ・ 目的のヒアリング、および情報解析手法の提案
- ・ データ解析方法の試行錯誤



一括受託

自身がNGSの専門家にならずとも解析ができるようになっている

必要な予算は？

**NGSの解析は多岐にわたるため、
解析によって大きくばらついてしまう。**

※NGS配列解析： データ量、サンプル数によって変化

※2次解析： 簡単な解析の場合 10万～

基本はどの外部委託先でも都度確認になる事が多い

スムーズな委託のために用意しておくべき情報

1. 解析予定の生物種は？
2. サンプル数は？
3. 想定している配列解析アプリケーションは？
4. 想定している情報解析方法は？
5. ゲノムまたは遺伝子データベースはあるか？

まとめ

これからNGSは研究の前提となる装置になっていくだろう

5報に1報がNGS関連の報告である

NGSによって、様々なことができる

様々な研究において取り入れるチャンスはある

計画はしっかりと

2次解析が必要不可欠、かつ重要

2次解析にも期間と人材を確保しておく必要がある

現在は環境が整備されている

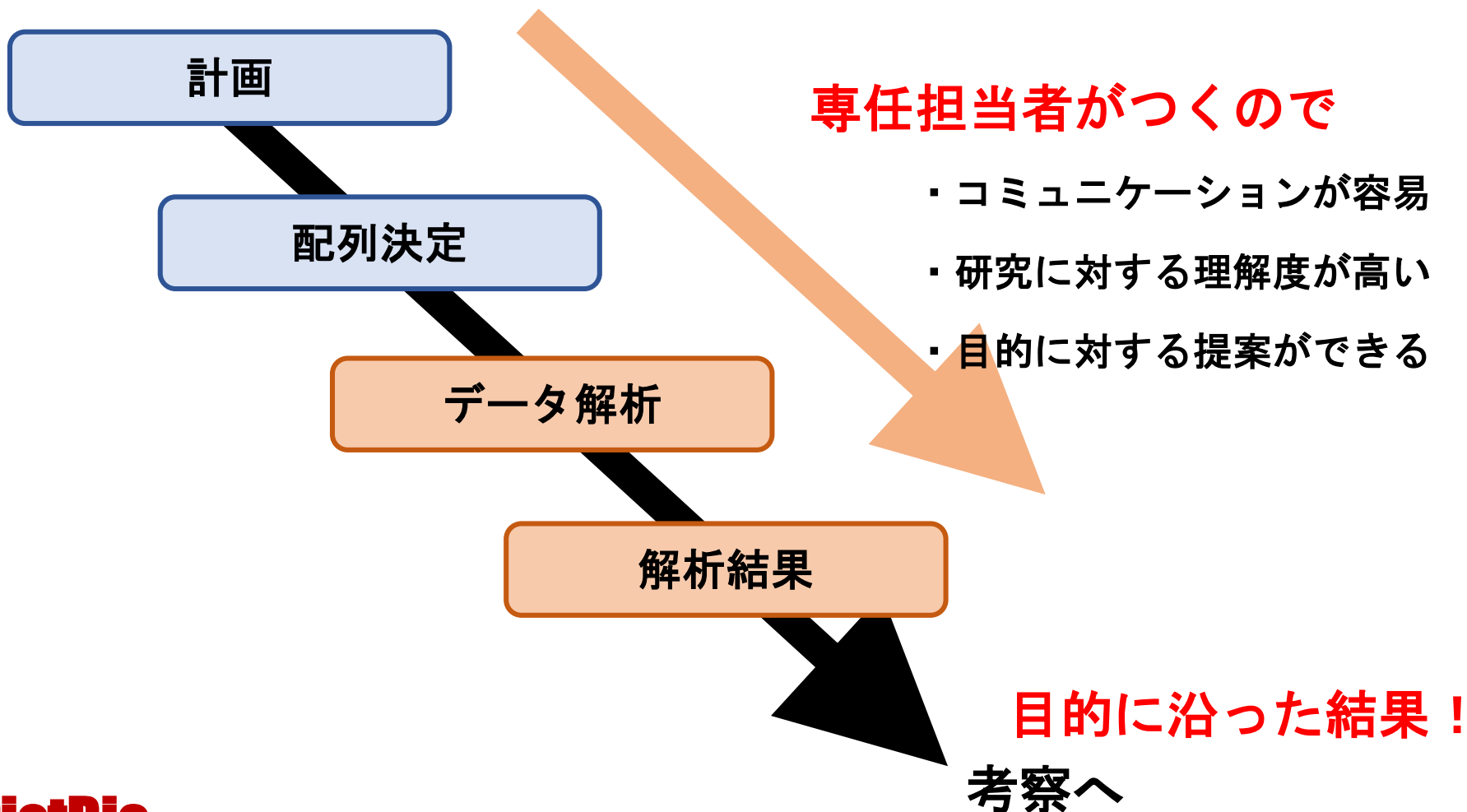
委託先の充実、および委託方法の充実
実験の計画段階から、結果出力までを委託することもできる

是非NGSを利用してみてください！

PictBioサービスの紹介

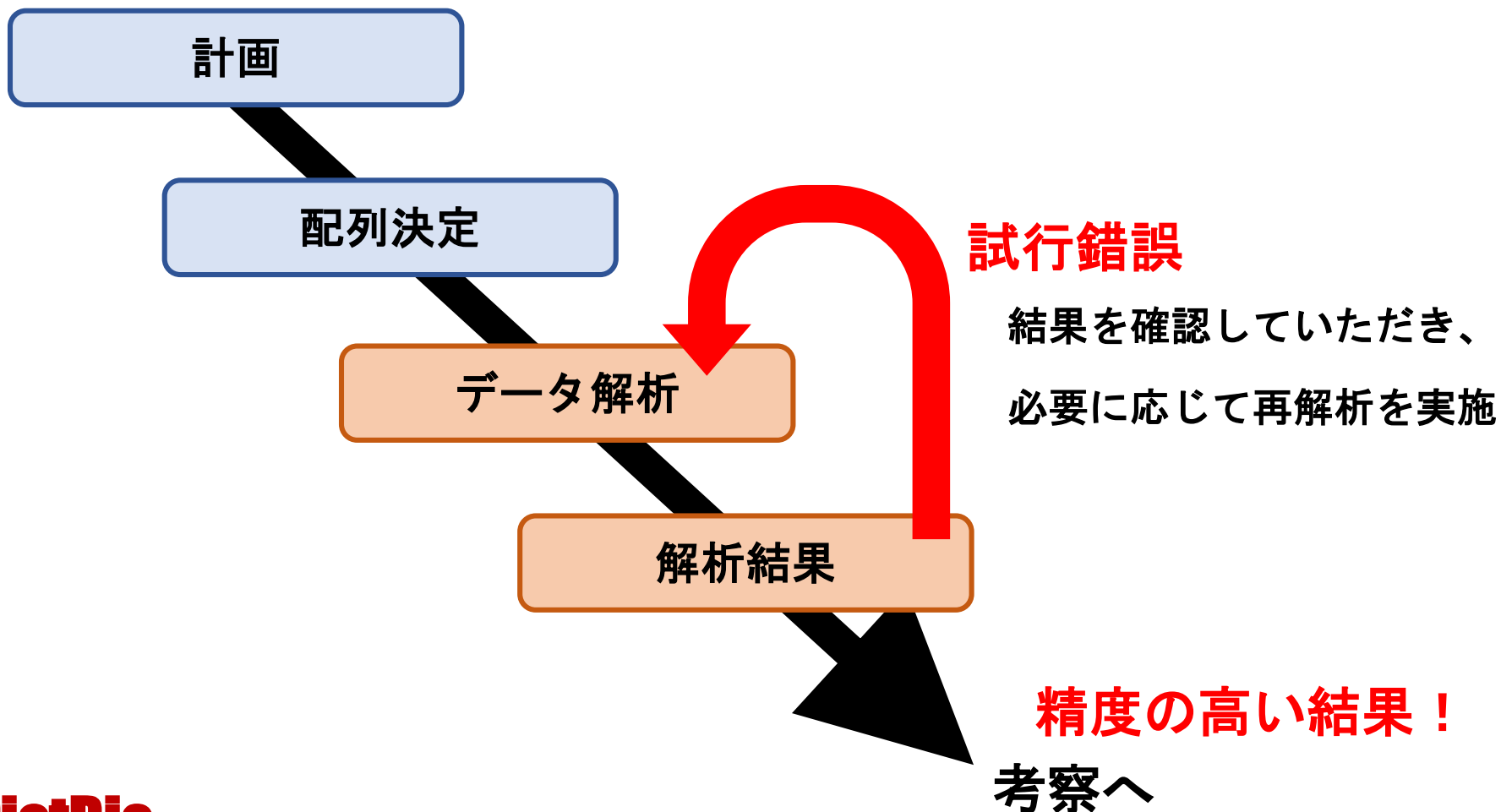
アクシオヘリックスの解析受託

専任担当者がつき、最初から最後までを責任持って解析



アクシオヘリックスの解析受託

2次解析における試行錯誤をしっかりと行います。



アクシオヘリックスの解析受託

2次解析終了後、考察フェーズからのお手伝いをします。

計画

配列決定

データ解析

解析結果

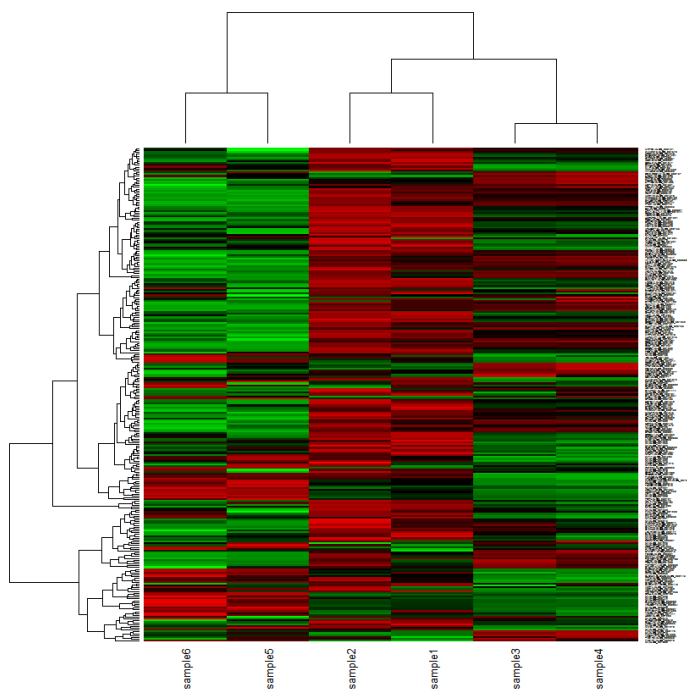
論文作成補助

- ・ 適した統計解析手法のご提案
 - ・ 統計解析処理
 - ・ データの見せ方のご提案
 - ・ 論文用の等の作成
- 考察へ

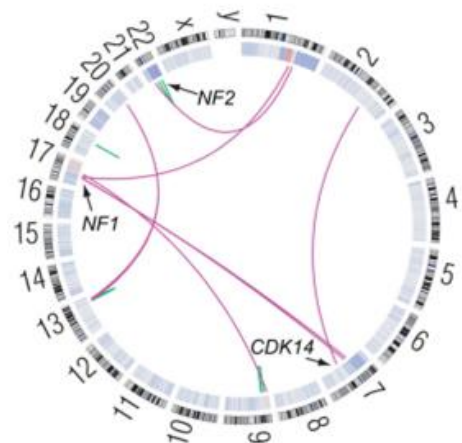
➡ 論文へ

論文Figureサービス

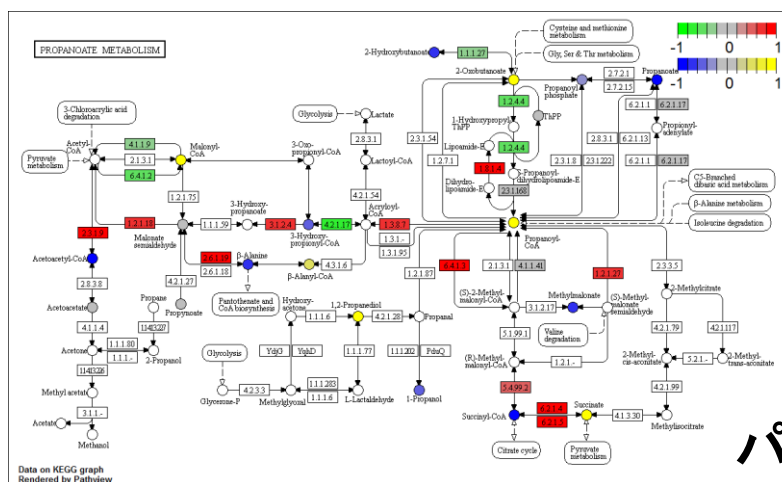
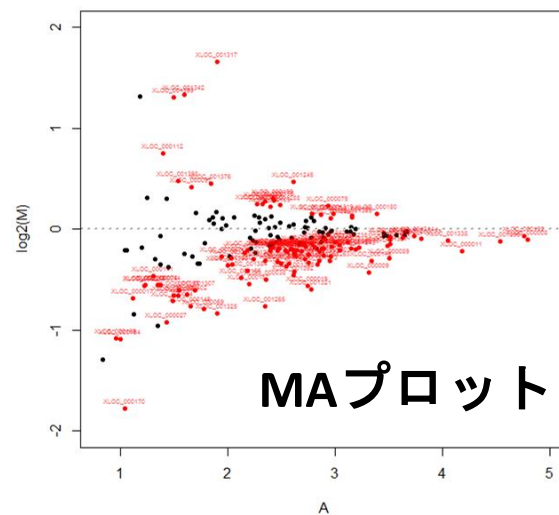
作図も任せてください！



ヒートマップ



サーコスプロット



パスウェイ

RNA-Seq発現量比較データ作成サービス

RNA-seq解析をお手軽に！

次世代シーケンサー
FASTQファイル

アップロード

[RNA] Expression comparison profiling analysis - Parameter settings

Select analysis conditions. Items marked with * are required.

Use the previous settings expComp --Select your previous setting-- Load your setting

Job title * expComp

Reference genome * data --Select your reference file--

Read type * Single-end Paired-end

Sample1 * forward -- Select your file -- reverse -- Select your file -- Insert size bps + -

Sample2 * forward -- Select your file -- reverse -- Select your file -- Insert size bps + -

Start Clear

ファイルと生物種を選択するだけ

発現量比較データシート

発現量比較データシートが作成できます！

| Gene ID | Gene symbol | locus | sample1_FPKM | sample2_FPKM | sample3_FPKM | Genbank ID | RefSeq ID | GO Term | Accession | GO Term Name | GO Term |
|------------------|-------------|----------------------|--------------|--------------|--------------|----------------------------|---------------------------|------------|------------|--|--|
| ENSG000000000419 | DFM1 | 205088918-50858555 | 3092.68 | 6502.68 | 4620.36 | AF007975,AK295569,AL034525 | NM_001317034.NM_001317035 | GO:0039195 | GO:0016000 | kinase activity | kinase activity |
| ENSG000000000457 | SOYL3 | 1169662006-169894267 | 22.7876 | 27.4915 | 100116 | AF540967,AK292507,AL031291 | NM_020423.NM_181063 | GO:0046304 | GO:0005515 | kinase activity | kinase activity |
| ENSG000000001036 | FUCA2 | 6143450806-143511690 | 192.405 | 171.891 | 71.6243 | AK075458,AK296485,AL031324 | NM_002020 | GO:0044560 | GO:0016787 | alpha-L-fucosidase activity | alpha-L-fucosidase activity |
| ENSG000000001167 | NFYA | 641026910-41099976 | 78.3511 | 62.2531 | 57.4387 | AL031778,BC039244,CH4710 | NM_002505.NM_021705 | GO:0032993 | GO:0016602 | protein-DNA complex | protein-DNA complex |
| ENSG000000001460 | STPG1 | 124307555-24472976 | 59.1732 | 38.2782 | 62.4779 | AL031431,BC017650,BC035054 | NM_001199012.NM_001199013 | GO:0006915 | GO:0006634 | apoptotic process | apoptotic process |
| ENSG000000001461 | NPAL3 | 124307555-24472976 | 163.143 | 41.4866 | 6.3365 | AK312433,AL031431,BC001261 | NM_001322854.NM_001322855 | GO:0005515 | GO:0015066 | protein binding | protein binding |
| ENSG000000001617 | SEMA3F | 348940006-50189075 | 97.454 | 14.8739 | 50.5654 | AC000063,AC104450,BC042929 | NM_001319798.NM_001319800 | GO:003215 | GO:0038191 | semaphorin receptor binding | semaphorin receptor binding |
| ENSG000000001630 | CYP51A1 | 792112150-92180725 | 13.7886 | 9.8132 | 0 | AC000020 | NM_000786.NM_001146152 | GO:0005776 | GO:0005776 | intracellular membrane-bounded organelle | intracellular membrane-bounded organelle |
| ENSG000000002030 | BAD | 1164269829-64289500 | 100.133 | 0 | 17.0897 | AB451254,AB451378,AF021761 | NM_004322.NM_002989 | GO:0008283 | GO:0006739 | cytosolic cytoskeleton | cytosolic cytoskeleton |
| ENSG000000002745 | WNT16 | 7121325366-121341104 | 94.3247 | 81.0882 | 76.7877 | AC006364,AF152584,AF169998 | NM_016067.NM_057168 | GO:0005776 | GO:2000078 | positive regulation of type I receptor signaling pathway | positive regulation of type I receptor signaling pathway |
| ENSG000000002746 | HECW1 | 742954134-43566001 | 0 | 0 | 17.9233 | AB002320,AB048365,AC004444 | NM_001281059.NM_015052 | GO:0005776 | GO:0005776 | intracellular membrane-bounded organelle | intracellular membrane-bounded organelle |
| ENSG000000002822 | MAD1L1 | 71815792-2233243 | 161.393 | 173.594 | 137.496 | AC005282,AC006433,AC006922 | NM_001013836.NM_001013837 | GO:0004842 | GO:0016740 | ubiquitin-protein transferase activity | ubiquitin-protein transferase activity |
| ENSG000000002834 | LASP1 | 1738869858-38921770 | 0 | 0 | 1.24342 | AC006441,AC110749,AK294714 | NM_001271608.NM_006148 | GO:0005776 | GO:0005643 | kinetochore nuclear core protein | kinetochore nuclear core protein |
| ENSG000000002919 | SNX11 | 1748103356-48123074 | 62.8594 | 14.2428 | 37.1196 | AC006468,AF121861,AK023914 | NM_013323.NM_152244 | GO:0005622 | GO:0006886 | intracellular membrane-bounded organelle | intracellular membrane-bounded organelle |

遺伝子名

遺伝子位置

発現量

アノテーション

*画像は開発中のものです。

- Total RNA-seqのデータから発現量の表を作成します。
- 定額なので何度でも使える
- WEBからマウス操作のみで使用可能

費用: 5万 / 1週間

今なら一週間無料です。

ご利用はこちらから → URL: <https://bcmarx.pictbio.com/>

ホームページの紹介

URL <https://www.pictbio.com>

PictBio
次世代シーケンサー 受託解析サービス

解析 ツール 実績 Public **解析メモ** お問い合わせ

PictBioとは
アクシオヘリックス株式会社が提供する、次世代シーケンサー（NGS）のデータ解析を筆頭にシーケンス解析、バイオインフォマティクス解析のあらゆるお悩みを解決する受託サービスです。

メインラインナップ

- バイオインフォマティクス**
NGSを中心とした
バイオデータ解析受託
- NGS受託解析 オーダーメイド**
データ条件に合わせて
フローをカスタマイズ作成
- RNA-Seq発現量比較解析サービス**
NGSデータ解析を
お手軽にクラウドで実行
- ツール開発**
新規アルゴリズム、ツール開発
公開データベース作成

バイオインフォマティクス受託解析について

バイオデータはその複雑性から膨大なデータとなり、巨大なデータベースが整備されるほどの量となっております。これらの研究・解析を行う上でLinux OSの計算サーバーを用いた大量計算や、統計を使用した解析が多く用いられており、インフォマティクスの技術が必要とされています。ですが、「Bio」「IT」の両知識を同時に取得することは、求めら

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) のトリミングしなくてはいけなかったり、メタデータ書がなくてはいけなかったり、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



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。

AX AXIOHELIX

アクセスはこちらのQRコードから



<http://www.axiohelix.com/>