

---

All Japan Annotator/Curator/System DB administrator

統合データベース講習会：AJACS 蝦夷@札幌医科大学、2018年11月9日

# ゲノムデータベースと 次世代シーケンスデータベース

---

東京大学大学院理学系研究科

河野 信

# 概要

---

- 本講習は、誰でも自由に使うことができる公共のデータベースやウェブツールを活用して、研究のさまざまな場面で利用することの多い（ヒトを含む）ゲノムデータベースや次世代シーケンスデータベースの使い方について学びます。
  - 次世代シーケンスの解析方法について知りたい方も多いかと思いますが、3時間という短い時間ではとてもカバーできませんので、概要と資料の紹介に留めます。
-

# 講習の流れ

---

1. 研究分野で頻繁に使われるDBやツールを知る：TogoTV
2. DNAシーケンス技術
3. 塩基配列データベース
4. 配列検索ツール
5. ゲノムデータベース
6. ヒト（ゲノム）データベース
7. ヒトに関する情報を検索するツール

## 講習に際しての注意とお願い

---

- みんなで同時にアクセスするとサイトにつながりにくくなる  
ことが予想されます。
    - 資料を見ながら自力で進められそうな方はどんどん先に、そうでない方は講師と一緒にすすめていきましょう。
    - サイトの反応が悪い時はタイミングをずらして実行してみてください。
    - 反応が無いからと言って何度もクリックするとますます繋がらなくなってしまいます。おおらかな気持ちで臨みましょう。
  - わからないことがあったら挙手にてスタッフにお知らせください。
    - 遠慮は無用です(そのための講習会です!)。おいてけぼりは楽しくありません。
-

---

# 1. 研究分野で頻繁に使われるDBやツールを知る：TogoTV

# 統合TVとは？

- 生命科学分野の有用なデータベースやツールの使い方を動画で紹介するウェブサイト <https://togotv.dbcls.jp/>

DBCLS Research Services Contact About

TOGO TV 生命科学系DB・ツール使い倒し系チャンネル

「統合TV」は、生命科学分野の有用なデータベースやツールの使い方を動画で紹介するウェブサイトです。

目的別に検索

- 講習会 実習資料 (AJACS)
- ゲノム・核酸 配列解析
- タンパク質 配列・構造解析
- 発現制御解析・可視化
- 文献・辞書・プログラミング
- 著名データベース
- その他講演・講習会
- 自由に使える画像を探す

関連するタグから検索

- ゲノム (327) 遺伝子 (492)
- タンパク質 (245)
- 配列解析 (278)
- 発現解析 (373) NGS (277)
- 文献検索 (302)
- 情報収集 (152)
- 環境設定 (145)
- DBCLS (193) English (235)
- ウェブツール (236)
- ソフトウェア (83)

Q 全番組のリストから、調べたいDBやウェブツールに関するキーワードで検索! (全 1500 件)

番組のタイトルや画像をクリックすると番組の再生ページへ移動します。番組リストやお問い合わせはこちらからどうぞ!

表示件数を選ぶ 検索窓にキーワードを入力すると、入力の数ごとに即座に候補の番組が絞り込まれます

GGGenome《ゲゲゲノム》を使って高速塩基配列検索をする 2018

GGGenomeは塩基配列を高速に検索することができるツールです。様々な生物種のゲノムや転写産物などを検索できます。検索できるデータベースは、ゲノム(350以上の生物種に対応)、転写産物(RefSeq, GENCODE, miRBaseなど)、DBDJの全塩基配列などがあり、6塩基以上の配列を検索可能です。今回の動画では、GGGenomeによる塩基配列検索の基本的な方法を説明します。さらにGGGenomeの機能を活用した例として、Google スプレッドシート(Spreadsheet)の機能を利用し、プライマー配列情報を半自動的かつ効率的に整理する方法についても紹介します。

Dataset2Tools でオミックスデータとその解析事例、計算ツールを検索し、再現性の高い再解析を行う

Datasets2Tools はマウント・シナイ医科大学の Ma'ayan 研究室で開発・運営されている、生物医学のデジタルオブジェクトの発見と評価のためのプラットフォームです。30,000 以上のバイオインフォマティクス解析、6,000 以上の生物学的データセット、4,000 以上の計算ツールを収録し、検索できるようにしています。Datasets2Toolsでは、Canned Analysis と呼ぶ永続的な URL を付与した解析データ一式を提供しており、これはバイオインフォマティクス解析の結果を FAIR原則に準拠して保存するために設計された新しいタイプのデジタルオブジェクトです。それらは解析のタイトルと内容の説明、もともとなるオリジナルのデータセット、適用している解析ツールとその出力などの情報が一つに整理されています。通常、論文に掲載する図は、著者の主張に応じて適切に編集されていますが、読者は最大限の情報を得るために未編集の形の図を閲覧したい場合があります。Canned Analysis という形式での提供は、そういった図を共有したり再解析することを可能にします。RNA-seq などのデータのクラスタリングとヒートマップ化をするツールである Clustergrammerや、多種多様なアノテーション情報を利用したエンリッチメント解析ツールEnrichrの簡単な使いかたも紹介しながら、Datasets2Toolsの使いかたについて紹介します。

文献情報関連サービス活用法 @ JBS2018

本日の統合TVは、2018年9月25日に開催された第91回日本生化学会大会(JBS2018)のシンポジウム「JBS-NBDC共催企画 統合データベース講習会:生命科学分野のデータベースとツールを使いこなす」から、ライフサイエンス統合データベースセンター (DBCLS) 山本 泰智 による「文献情報関連サービス活用法」をお送りします。約30分です。本講演では、主にPubMedの効率的な検索方法を紹介するとともに、DBCLSの提供する文献情報活用サービスであるAllieやinMeXes、Colliの使い方について紹介しています。

- 本講習の課題に対応するチュートリアル動画があります
    - ウェブサイトへのアクセスから結果の見方まで、操作の一挙手一投足がわかります
    - 講義・講習などの参考資料や後輩指導の教材として利用できます
    - その他、今回の講習に関連する内容の多くは、「ゲノム、核酸配列・構造解析」のカテゴリにあります
  - 過去の講習会の内容はそのほとんどが統合TVに収録されており、いつでもどこでも繰り返し復習できるようになっています。
    - お探しのDB・ツールが統合TV未掲載の場合には、統合TV番組リクエストフォー  
ムへどうぞ！
-

---

## 2. DNAシーケンス技術



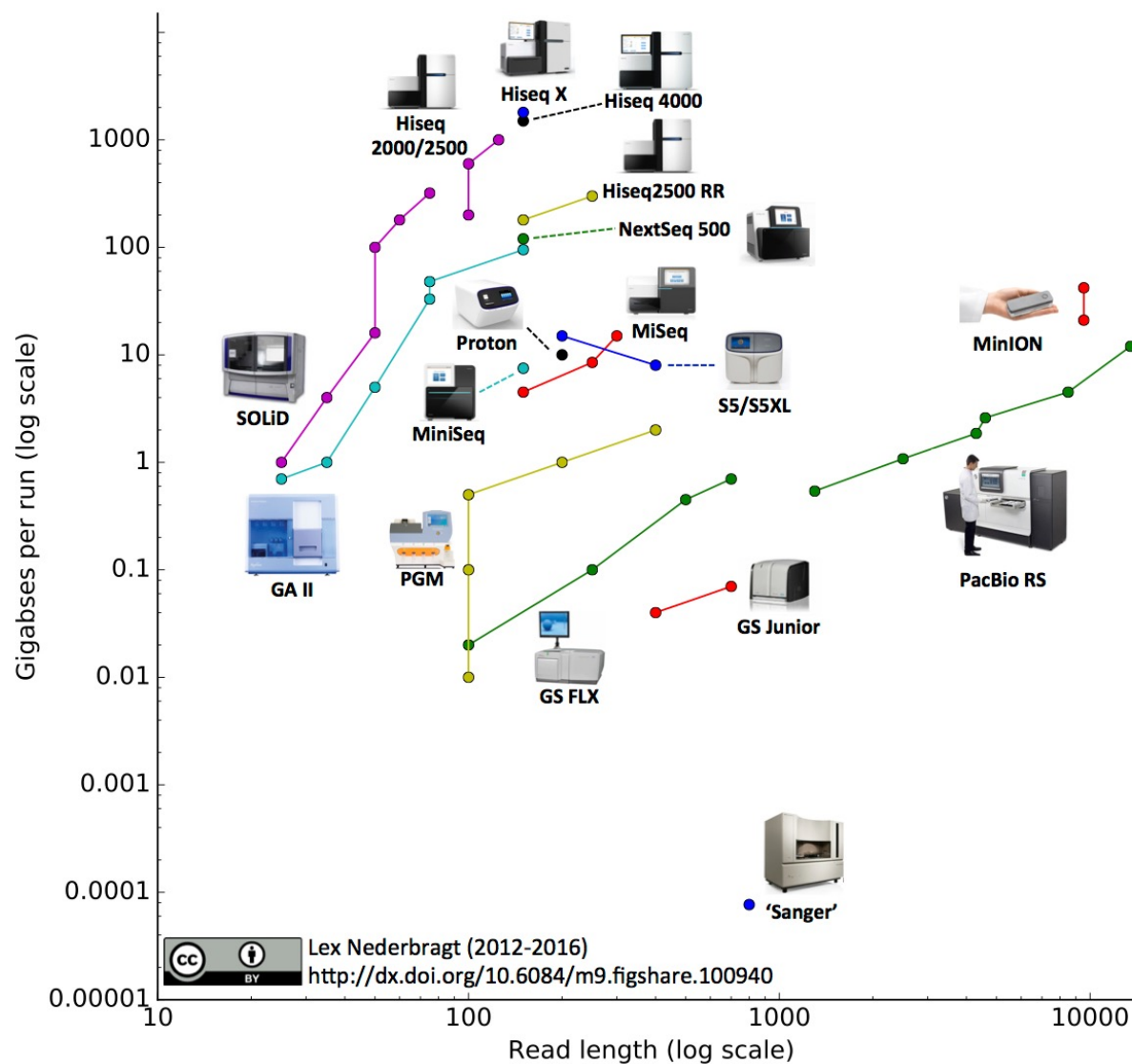
# サンガー法 v.s. 次世代シーケンス (NGS)

---

<https://www.nature.com/articles/nbt1486/figures/1>

---

# 次世代シーケンサー



---

### 3. 塩基配列データベース



# GenBank/ENA/DDBJ

---

# 【演習】 DDBJ 検索

---

# SRA/ERA/DRA

---

# 【演習】 DRA 検索

---



# GEO/ArrayExpress/GEA

---

# NGSデータ解析について

---

- NGSのデータ解析は

---

## 4. 配列検索ツール

# 配列解析入門

---

- 配列解析の基本である配列アラインメントについて、BLASTを例にその検索アルゴリズムを解説する
- ファイルフォーマット

ファイルフォーマット	ファイル拡張子	用途など
FASTA	.fa .fasta	塩基配列、アミノ酸配列
FASTQ	.fq .fastq	NGSからの塩基配列とそのquality
DDBJ(Genbank)	.dbj (.gbk)	メタデータを含んだ塩基配列やアミノ酸配列の記述
SRA	.sra	FASTQを圧縮したファイル形式
SAM/BAM	.sam .bam	リファレンスゲノム配列へのアラインメント
GFF(GTF)	.gff .gtf	ゲノムアノテーション
BED	.bed	ゲノムアノテーション
VCF	.vcf	バリアントの記述

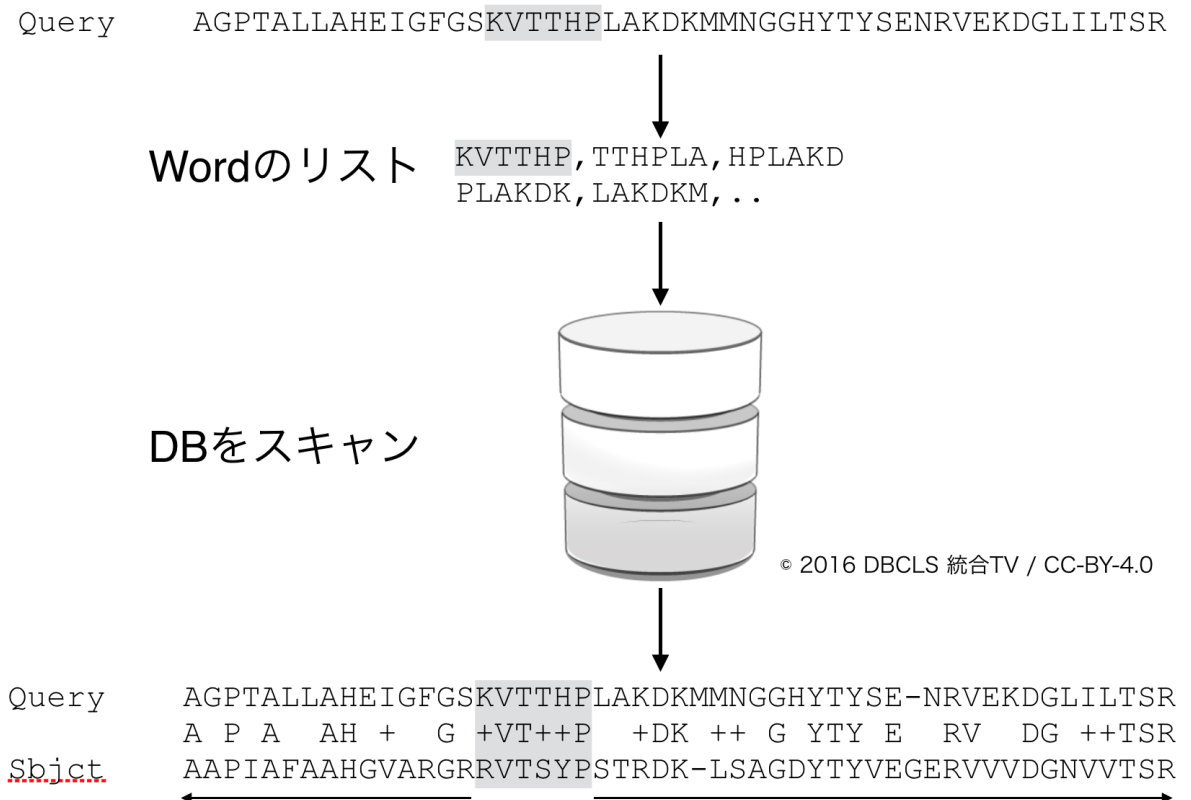
# BLAST

## ■ BLAST とは

- Basic Local Alignment Search Tool
- 配列類似性検索のデファクトスタンダード

## ■ BLAST の動作原理

- 質問配列 (Query)
- 検索対象DB (Sbjct)



# BLAST

---

- 質問配列と検索対象DBの組み合わせ

# 【演習】 BLAST 検索

---

# GGRNA/GGGenome

---



# 【実習】 配列の高速検索

---

# CRISPER direct

---

# 【実習】 CRISPER 配列を設計する

---