
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 次世代シークエンスデータベース

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ひ

- 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拳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/>

The screenshot shows the TOGO TV website interface. At the top, there's a navigation bar with links for DBCLS, Research, Services, Contact, and About. Below the navigation is a search bar with the placeholder "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)". To the left, there's a sidebar with a "目的別に検索" section containing links for various topics like "講習会 実習資料 (AJACS)", "ゲノム・核酸 配列解析", and "タンパク質 配列・構造解析". Below that is a "関連するタグから検索" section with tags such as "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)", and "ソフトウェア / R (23)". At the bottom of the sidebar, the URL "https://togotv.dbcls.jp/" is visible. The main content area displays search results for "GGGenome《ゲゲゲノム》を使って高速塩基配列検索をする 2018" and "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". Each result includes a thumbnail image of a video, its title, and a view count (e.g., 181025). The footer of the page also features a "TOGO TV" logo.

TogoTV

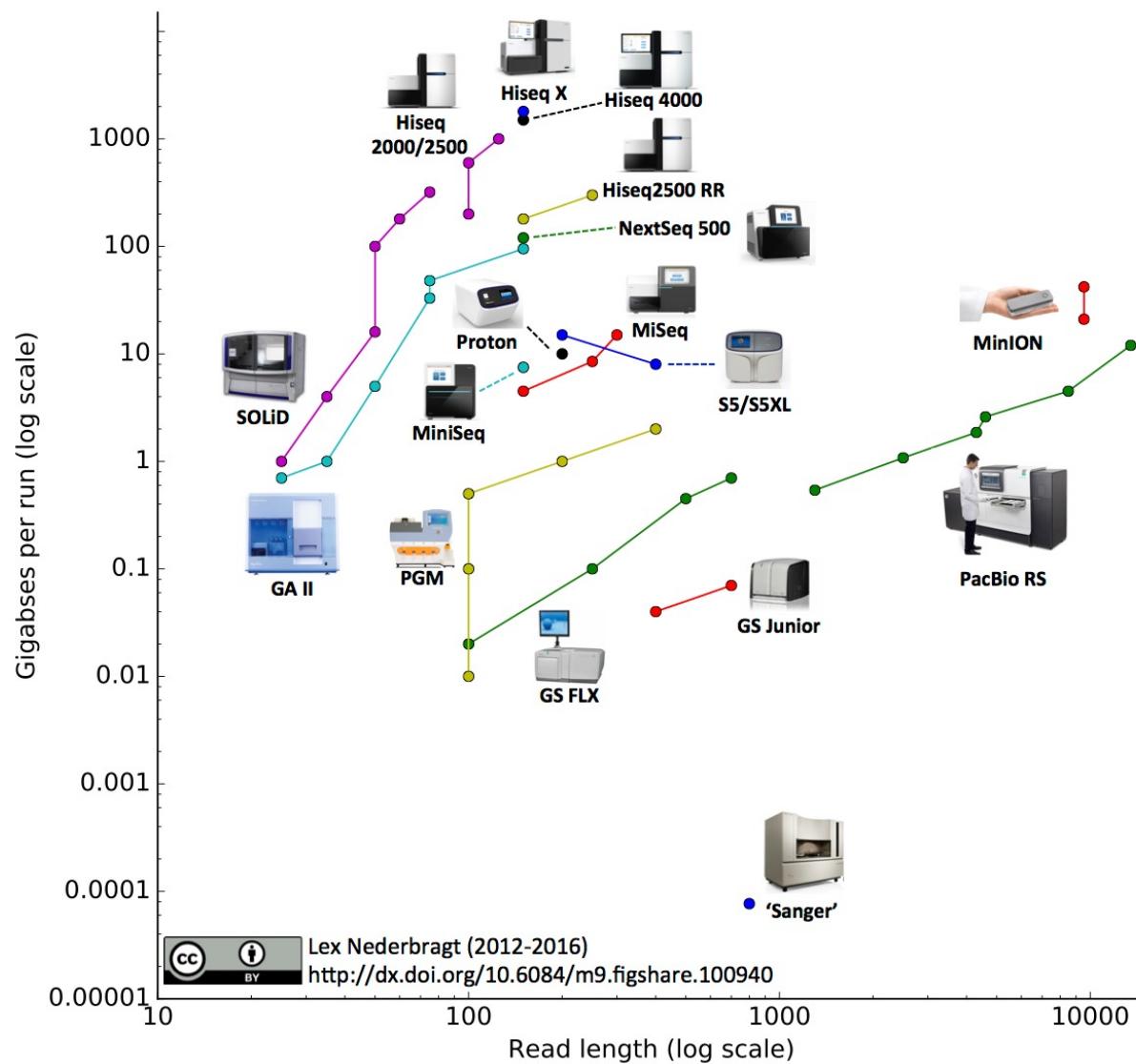
- 本講習の課題に対応するチュートリアル動画があります
 - ウェブサイトへのアクセスから結果の見方まで、操作の一挙手一投足がわかります
 - 講義・講習などの参考資料や後輩指導の教材として利用できます
 - その他、今回の講習に関連する内容の多くは、「ゲノム、核酸配列・構造解析」のカテゴリにあります
- 過去の講習会の内容はそのほとんどが統合TVに収録されており、いつでもどこでも繰り返し復習できるようになっています。
 - お探しのDB・ツールが統合TV未掲載の場合には、統合TV番組リクエストフォームへどうぞ！

2. DNAシークエンス技術

サンガーフラッシュ v.s. 次世代シーケンス (NGS)

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

次世代シークエンサー



NGS を用いたアプリケーション

- Sequencing Methods Poster (Illumina)

- <https://emea.illumina.com/techniques/sequencing/ngs-library-prep/library-prep-methods.html>
- DNA ([PDF](#))
- RNA ([PDF](#))
- Single Cell ([PDF](#))



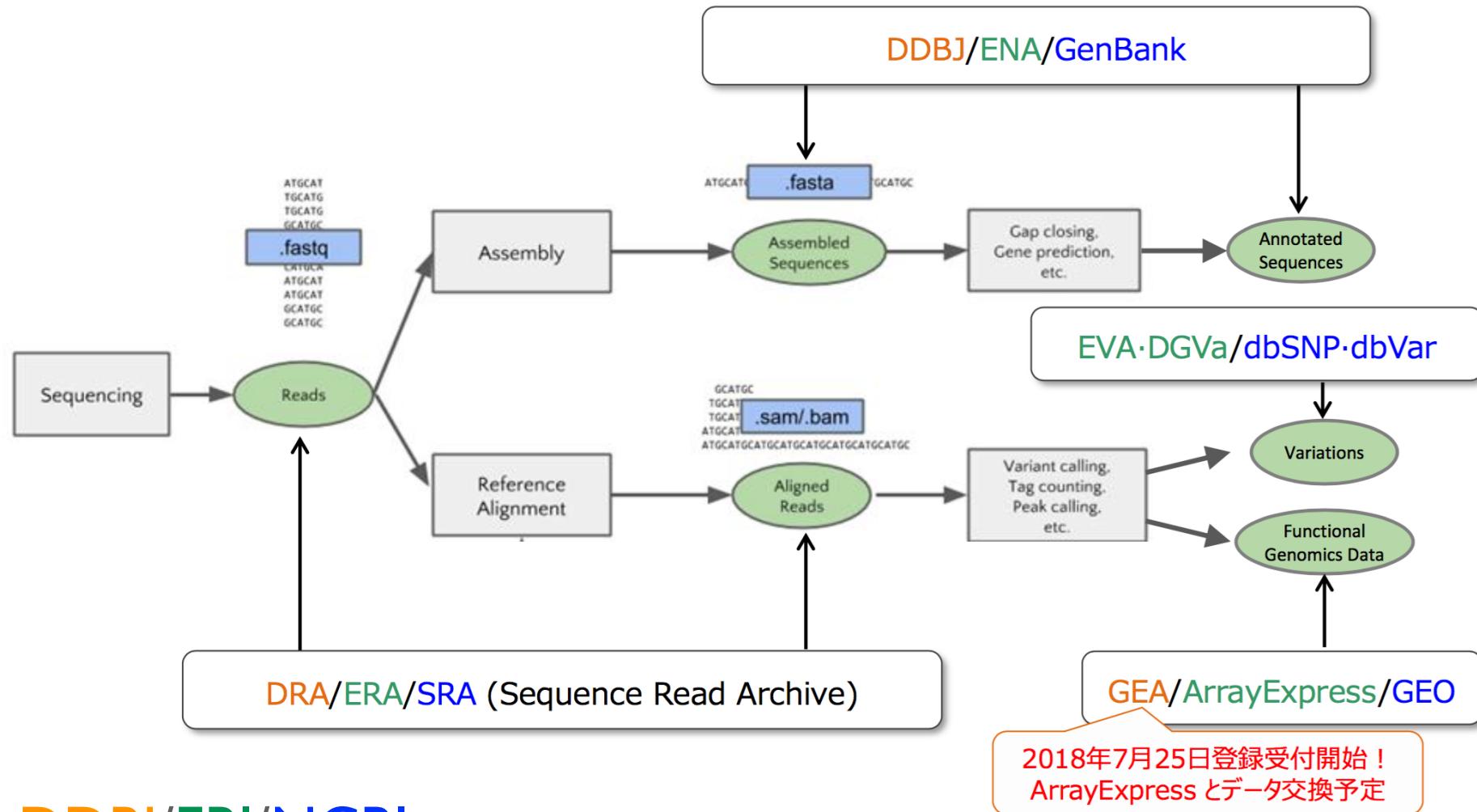
3. 塩基配列データベース

塩基配列データベースで使用され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/BCF	.vcf	バリアントの記述

参考：NGSハンズオン講習会資料

データの種類とデータベース



DDBJ/EBI/NCBI

一次データベースと二次データベース

	一次データベース Primary database	二次データベース Secondary database
別の呼び方	Archival database	Curated database; Knowledgebase
データソース	研究者（登録者）が実験で得たデータを直接登録	一次データベースのデータや文献を解析、解釈、キュレーションした結果
例	<ul style="list-style-type: none">• DDBJ/ENA/GenBank• GEA/ArrayExpress/GEO• DRA/ERA/SRA• EVA·DGVa/dbSNP·dbVar• PDB	<ul style="list-style-type: none">• RefSeq• Ensembl• Expression Atlas• ChIP-Atlas• UniProt

INSDC <http://www.insdc.org/>

- International Nucleotide Sequence Database Collaboration
- 日米欧 3 極での塩基配列データベースについての協力体制
 - それぞれのデータを交換しており、同じデータを保存・提供している
 - どこか1ヶ所にデータを登録すれば、自動的にデータが同期される
- 参加機関
 - 日本 : DNA Data Bank Japan ([DDBJ](#))
 - 欧州 : European Bioinformatics Institute ([EBI](#))
 - 米国 : National Center for Biotechnology Information ([NCBI](#))



DDBJ/ENA/GenBank

- DNA 塩基配列およびそのアノテーション情報を登録するためのデータベース
- 登録されたデータは日米欧で毎日交換
- DDBJ から登録すると、日本語で対応してもらえます
- データベース
 - DDBJ (DDBJ): <https://www.ddbj.nig.ac.jp/index.html>
 - ENA - European Nucleotide Archive (EBI): <https://www.ebi.ac.uk/ena>
 - GenBank (NCBI): <https://www.ncbi.nlm.nih.gov/genbank/>

【演習】DDBJ 検索

1. 「DDBJ」でググってトップページを開きます
2. 「検索・解析」の「ARSA」を開きます
 1. 「登録」からデータの登録が
 2. 「ダウンロード」からデータベースのまるごとダウンロードが
 3. 「スパコン」からスーパーコンピュータの利用申請（基本無料）ができます

The screenshot shows the DDBJ Center homepage with the following layout:

- Header:** DDBJ Services, Login & Submit, Contact, English.
- Top Bar:** DDBJ センター, DDBJ Center Web Sites, Google カスタム検索, Search icon.
- Alert Bar:** (11/15-22) DDBJ センターの登録、検索・解析サービス停止のお知らせ.
- Service Icons (Top Row):**
 - 検索・解析: Monitor icon.
 - 登録: Database icon.
 - ダウンロード: Download icon.
 - スパコン: Supercomputer icon.
- Service Icons (Bottom Row):**
 - 統計: Bar chart icon.
 - 活動: Pencil icon.
 - 講習会: Graduation cap icon.
 - センターについて: People icon.

【演習】DDBJ 検索

3. キーワード（「ALDH1A1 human」）を入力して検索します
4. ヒットした配列の Accession 番号をクリックします
 1. ユーザによる登録ごとにエントリが作られるのでたくさんヒットします

The screenshot shows the DDBJ ARSA search interface in English. It consists of two main panels: a left panel for search conditions and a right panel for search results.

Left Panel (Search Condition):

- Header: DDBJ DNA Data Bank of Japan
- Search bar: Google カスタム検索 (Search)
- Buttons: English, アプリを表示
- Menu: DDBJ の紹介, 利用の手引き, レポート・統計, FAQ
- Breadcrumb: HOME > 検索・解析 > ARSA
- Section: ARSA (Search Condition)
 - Quick Search: ALDH1A1 human (Search, AND)
 - Available Fields: ▶ Available Fields

Right Panel (Search Result):

- Header: DDBJ DNA Data Bank of Japan (English), Google カスタム検索 (Search)
- Buttons: English, アプリを表示
- Menu: DDBJ の紹介, 利用の手引き, レポート・統計, FAQ, お問い合わせ
- Breadcrumb: HOME > 検索・解析 > ARSA > ARSA (Search Result)
- Section: ARSA (Search Result) (ヘルプ)
- Section: Search Condition
- Section: Search Result
- Section: Facet
- Section: List of Entries
 - 1 - 30 entries / Number of founds: 121 (FlatFile, XML, Fasta, View selected, Download selected, Download All)
 - PrimaryAccessionNumber Definition SequenceLength MolecularType Organism
 - AY338497 Definition:Homo sapiens aldehyde dehydrogenase 1 family, member A1 (ALDH1A1) gene, complete cds. SequenceLength:55461 MolecularType:DNA Organism:Homo sapiens
 - KT585020 Definition:Homo sapiens clone HsUT00698041 ALDH1A1 (ALDH1A1) gene, 3' UTR. SequenceLength:471 MolecularType:DNA Organism:Homo sapiens
 - BV208366 Definition:ALDH1A1_2601 Rhesus macaque genomic DNA Macaca mulatta STS genomic clone MMA2601, sequence tagged site. SequenceLength:755 MolecularType:DNA Organism:Macaca mulatta

【演習】DDBJ 検索結果（一部）

これが DDBJ/GenBank フォーマットとよばれるものです

LOCUS AY338497 55461 bp DNA linear HUM 21-JUL-2003
DEFINITION Homo sapiens aldehyde dehydrogenase 1 family, member A1 (ALDH1A1 gene, complete cds.)
ACCESSION [AY338497](#)
VERSION AY338497.1
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 55461)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Wittek,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: <http://egp.gs.washington.edu>).
FEATURES Location/Qualifiers
[source](#) 1..55461
[/organism="Homo sapiens"](#)
[/mol_type="genomic DNA"](#)
[/db_xref="taxon:9606"](#)

CDS

```
join(1953..2018,14701..14805,23935..24075,25936..26065,  
27780..27841,29345..29473,30825..30938,36135..36237,  
37853..38037,42840..43004,45203..45360,48930..49004,  
53682..53754)  
/gene="ALDH1A1"  
/codon\_start=1  
/product="aldehyde dehydrogenase 1 family, member A1"  
/protein\_id="AAP88039.1"  
/translation="MSSSGTPDLPVLLDLKIQYTKIFINNEWHDSVSGKKFPVNPA  
TEELCQVEEGDKEDVDKAVKAARQAFQIGSPWRTMDASERGRLLYKLADLIERDRLL  
LATMESMNGGKLYSNAYLNDLAGCIKTLRYCAGWADKIQGRTIPIDGNFFTYTRHEPI  
GVCGQIIPWNFPLVMLIWKIGPALSCGNTVVVKPAEQTPLTALHVASLIKEAGFPPGV  
VNIVPGYGPAGAAISSHMDIKVAFTGSTEVGKLIKEAAGKSNLKRVTLEGGKSPC  
IVLADADLDNAVEFAHHGVFYHQGQCCIAASRIFVEESIYDEFVRRSVERAKKYILGN  
PLTPGVTQGPQIDKEQYDKILDIESGKKEGAKLECGGGPWGNKGYFVQPTVFSNVTD  
EMRIAKEIFGPVQQIMKFSLDDVIRANNTFYGLSAGVFTKDIDKAITISSALQAG  
TVWVNCYGVVSAQCPFGFKMSGNGRELGEYGFHEYTEVKTVKISQKNS"
```

BASE COUNT 17736 a 9840 c 10257 g 17628 t
ORIGIN

```
1 taagaagtga gatgacaagc caagtatgtt atgaaggctt gagcttcatt cgccctggaca  
61 tcaaaataaac cagtatttga atccacaaaac aactgtgact ctggggagaag aaagcagcaa  
121 aacacagctg tttgggcatg gccagagcgc cactctcaag ttatgttaaca gcttgtgt  
181 aattgtgcatt ttaactcaaa acacatttt gagaagaaaa acaatttagac tttccaaaaaa  
241 gaaagaaaaca gaaaaatgtt gaaaacac agaagttct atctggcat caaacataag  
301 acacggggga gtcaaaggca ttggaaatga tattgccta gaatctaggg agttggcccg  
361 tgcaccaaat ctggcctgcc ttctgtcttt ttaaataaag ttctatagga aaacaactac  
421 acccattaaat ttaccttatca tctacagttt cttttaataa gagtcaagta aattatatca  
481 gaaaccatcc agccccagg ctaaaaatat ttacagaaaa agtttgcaa attctgtct  
541 tagatgaata aaaagttatg tttaaatgcc tgtaagggtg actattcact gaaaaacca  
601 aagcaggta tgcttgaacc catgttaggat ttctcttgta gagaataggg tagaaatagt  
661 aaacagaaaat aagagtccag gttggggagt cagactggcc taaaatggaa ttctggct  
721 gcactaatag atctgtgacc ataagtaagc tactcacccct ctgtatgtct ctaaaatggg  
781 gatatgaaaat gagggtagaa agggaaacaa gtgctgcacc tgattggc aaacaattca
```

【演習】NCBI 検索

1. 「NCBI」でググってトップページを開きます
2. 先ほどとキーワード（「ALDH1A1 human」）を入力して検索します

The screenshot shows the NCBI homepage with a search bar containing 'ALDH1A1 human'. The search results are visible on the right side of the page.

NCBI Home

- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

November 14 Webinar: Variant Interpretation using NCBI Resources
06 Nov 2018

Next Wednesday, November 14, 2018, NCBI staff will show you how to use

【演習】NCBI 検索

- 遺伝子配列自体はたくさん登録されていますが、それらをまとめて整理した RefSeq は 1 エントリのみヒットしています

The screenshot shows the NCBI Gene search results for "ALDH1A1 human". The search bar at the top contains "ALDH1A1 human". The main result is for "ALDH1A1 – aldehyde dehydrogenase 1 family member A1" in "Homo sapiens (human)". Below the title, it says "Also known as: ALDC, ALDH-E1, ALDH1, ALDH11, HEL-9, HEL-S-53e, HEL12, PUMB1, RALDH1" and "GeneID: 216". There are three links highlighted with a red box: "RefSeq transcripts (1)", "RefSeqGene (1)", and "PubMed (294)". Below these are buttons for "Genome Browser", "BLAST", and "Download". At the bottom, there is a "RefSeq transcripts" section with a "+" button.

The screenshot shows the NCBI Databases search results for "ALDH1A1 human". It states "Results found in 26 databases for ALDH1A1 human (1 errors)". There are three main sections: "Literature", "Genes", and "Genetics". In the "Literature" section, "Bookshelf" has 13 results and "MeSH" has 1 result. In the "Genes" section, "Gene" has 242 results and "GEO DataSets" has 26 results. In the "Genetics" section, "ClinVar" has 31 results and "dbGaP" has 0 results. The "Gene" link in the "Genes" section is also highlighted with a red box.

【演習】NCBI 検索結果

- RefSeq は DDBJ/GenBank フォーマットで書かれていますが、これを基にした遺伝子のページが充実しています

ALDH1A1 aldehyde dehydrogenase 1 family member A1 [*Homo sapiens* (human)]

Gene ID: 216, updated on 4-Nov-2018

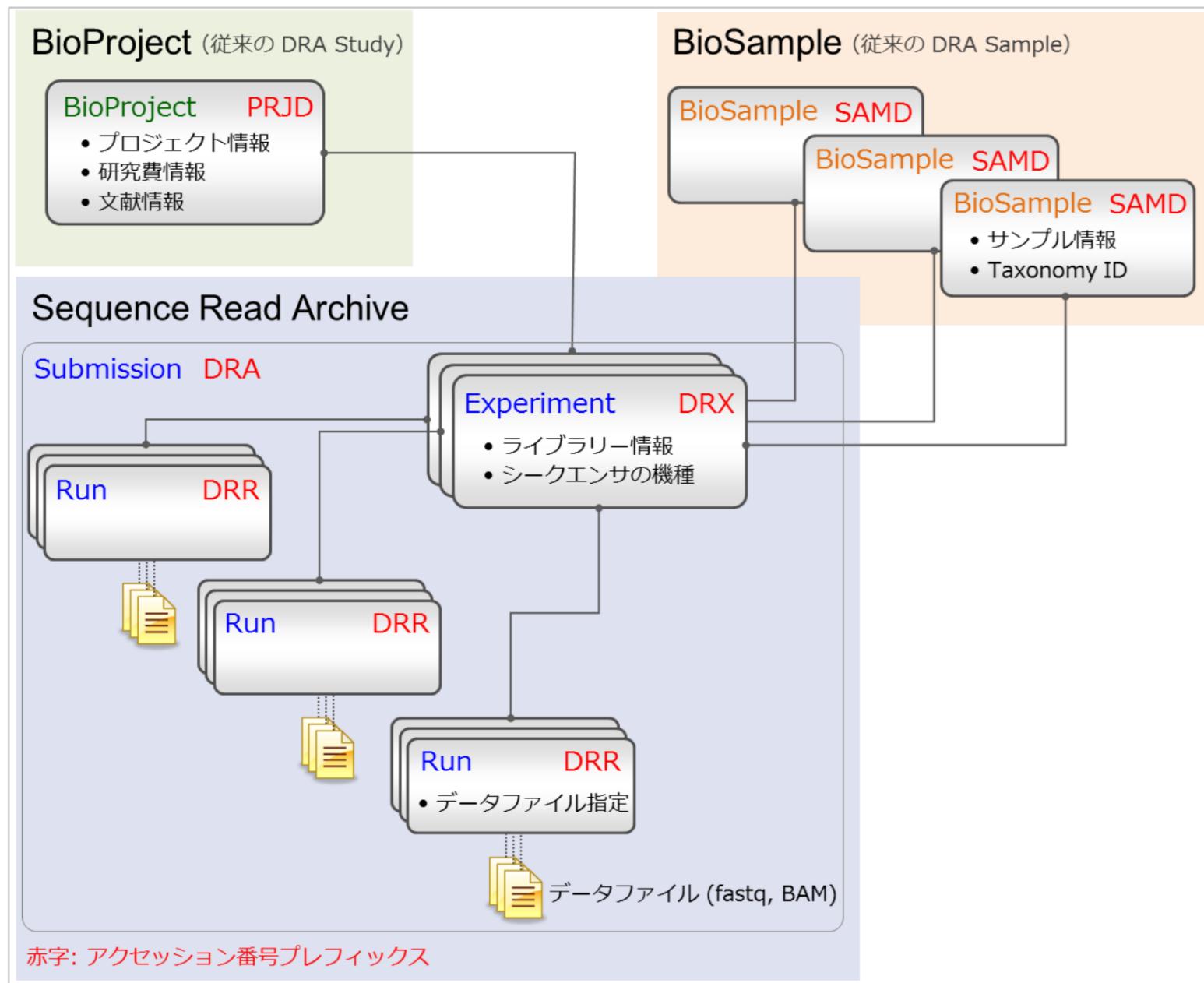


Official Symbol	ALDH1A1 provided by HGNC
Official Full Name	aldehyde dehydrogenase 1 family member A1 provided by HGNC
Primary source	HGNC:HGNC:402
See related	Ensembl:ENSG00000165092 MIM:100640 ; Vega:OTTHUMG00000020019
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	ALDC; ALDH1; HEL-9; HEL12; PUMB1; ALDH11; RALDH1; ALDH-E1; HEL-S-53e
Summary	The protein encoded by this gene belongs to the aldehyde dehydrogenase family. Aldehyde dehydrogenase is the next enzyme after alcohol dehydrogenase in the major pathway of alcohol metabolism. There are two major aldehyde dehydrogenase isozymes in the liver, cytosolic and mitochondrial, which are encoded by distinct genes, and can be distinguished by their electrophoretic mobility, kinetic properties, and subcellular localization. This gene encodes the cytosolic isozyme. Studies in mice show that through its role in retinol metabolism, this gene may also be involved in the regulation of the metabolic responses to high-fat diet. [provided by RefSeq, Mar 2011]
Expression	Broad expression in liver (RPKM 321.0), duodenum (RPKM 249.1) and 18 other tissues See more
Orthologs	all

DRA/ENA/SRA

- 次世代シーケンサから得られた生データもしくはマッピングデータを登録するデータベース
- データベース
 - DRA - DDBJ Sequence Read Archive (DDBJ): <https://www.ddbj.nig.ac.jp/dra/index.html>
 - ENA - European Nucleotide Archive (EBI): <https://www.ebi.ac.uk/ena>
 - SRA - Sequence Read Archive (NCBI): <https://www.ncbi.nlm.nih.gov/sra>
(一時期 Short Read Archive と呼ばれていた)

SRA データモデル



【演習】DRA 検索

1. 「DDBJ DRA」でググってトップページを開きます
2. 「検索」で検索ページを開きます
 1. 「ログイン・登録」からデータを登録できます

Screenshot of the DDBJ Sequence Read Archive (DRA) website:

The header includes: DDBJ Services, Login & Submit, Contact, English.

The main navigation bar includes: Home, Handbook, FAQ, Search, Downloads, About DRA.

A yellow banner at the top states: (11/15-22) DDBJ センターの登録、検索・解析サービス停止のお知らせ.

The central content area contains a paragraph about the DDBJ Sequence Read Archive (DRA), mentioning its purpose as a database for next-generation sequencing data from Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System, and others, and its role as a member of the International Nucleotide Sequence Database Collaboration (INSDC).

Below the text are three icons with labels: 検索 (Search), ハンドブック (Handbook), and ログイン / 登録 (Login / Register).

The right sidebar features a "Latest News" section with two entries:

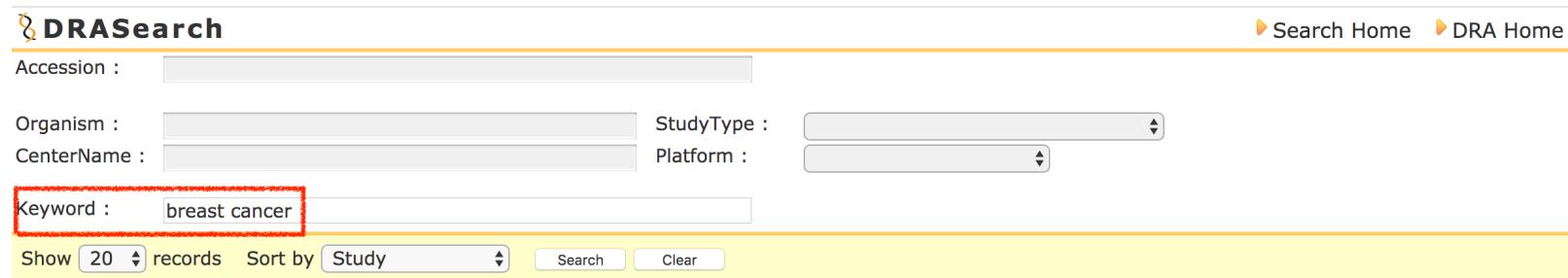
- 2018年10月31日 | メンテナンス
(11/15-22) DDBJ センターの登録、検索・解析サービス停止のお知らせ
- 2018年10月25日 | お知らせ
DDBJ センターの活動を報告した Nucleic Acids Research 論文が公開されました

Below the news is a brief summary: 遺伝研スーパー計算システムの定期メンテナンスに伴い、以下の日程で DDBJ センターの登録、検索・解析サービスが停止いたします。サービス名停止期間・DRA 検証処理・GEA 検証処理11月1

【演習】DRA 検索

3. キーワード（「breast cancer」とか）を入力して検索します

1. データベースに登録されている統計情報が表示されています

The screenshot shows the DRASearch interface. At the top, there are search filters for 'Accession', 'Organism', 'CenterName', 'StudyType', and 'Platform'. The 'Keyword' field contains 'breast cancer', which is highlighted with a red box. Below the search bar are buttons for 'Show 20 records' and 'Sort by Study'. On the right, there are links to 'Search Home' and 'DRA Home'. A timestamp 'Data Last Update 2018-11-08' is visible. The main content area is titled 'Statistics' and includes three tables: 'Released Entries' (listing Submission, Study, Experiment, Sample, and Run counts), 'Organism' (listing 10 entries from Homo sapiens to Zea mays), 'Study Type' (listing 10 entries from Other to Pooled Clone Sequencing), and 'Center Name' (listing 10 entries from BioProject to The Wellcome Trust Sanger Institute).
Released Entries

Type	Count
Submission	1037579
Study	166710
Experiment	5110728
Sample	4604529
Run	5815788

Organism

#	Organism Name	Study
1	Homo sapiens	15471
2	Mus musculus	12681
3	soil metagenome	4771
4	Populus trichocarpa	3062
5	Arabidopsis thaliana	2099
6	marine metagenome	1835
7	Panicum virgatum	1817
8	Drosophila melanogaster	1729
9	Oryza sativa	1717
10	Zea mays	1688

Study Type

#	Study Type	Study
1	Other	63982
2	Whole Genome Sequencing	53473
3	Metagenomics	24340
4	Transcriptome Analysis	22749
5	Population Genomics	796
6	Epigenetics	708
7	Exome Sequencing	264
8	Transcriptome Sequencing	170
9	Cancer Genomics	141
10	Pooled Clone Sequencing	35

Center Name

#	Center Name	Study
1	BioProject	98591
2	GEO	27777
3	DOE - JOINT GENOME INSTITUTE	2590
4	UMIGS	2557
5	JGI	2365
6	WUGSC	1399
7	JCVI	1148
8	BI	962
9	SC	902
10	The Wellcome Trust Sanger Institute	783

【演習】DRA 検索

4. 検索結果の Accession をクリックします

DRA Search

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword : Breast cancer

Show 20 records Sort by Study

Search Results (22001 records) << < 1 / 1101 > >>

Filtered by

document type:sample(10354) study(5334) experiment(2550) submission(2446) run(1314) analysis(3)
organism:Homo sapiens(13837) Mus musculus(2748) Salmonella enterica(808) Salmonella enterica subsp. enterica(560)
Campylobacter jejuni(501) Drosophila melanogaster(257)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA161444.submission.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION alias="bilateral breast cancer" submission	SRA161444	SRP041629	Homo sapiens Bilateral Breast Cancer Exome	Exome Sequencing	Homo sapiens	200G		BioProject
2	SRA480647.submission.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION lab_name="Breast Cancer Now" center_name="The	SRA480647	SRP090536	Whole genome sequencing analysis of APOBEC3B induced clones of HEK293 cell lines with inactivated TP53	Whole Genome Sequencing	Homo sapiens	0		BioProject
3	SRA654135.submission.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION lab_name="Breast Cancer Now" center_name="The	SRA654135	SRP131864	Optimized ARID1A immunohistochemistry is an accurate predictor of ARID1A mutational status in gynaecological cancers (IonProton)	Other	Homo sapiens	0		BioProject
4	SRA654548.submission.xml <?xml version="1.0" encoding="UTF-8"?> <SUBMISSION lab_name="Breast Cancer Now" center_name="The	SRA654548	SRP131907	Optimized ARID1A immunohistochemistry is an accurate predictor of ARID1A mutational status in gynaecological cancers (MiSea)	Other	Homo sapiens	0		BioProject

【演習】DRA 検索結果

5. Experiment や、サンプル、Run へのリンクから詳細な情報にアクセスできます
6. また、FASTQ 形式もしくは SRA 形式のファイルをダウンロードできます

DRA Search

SRA161444 FTP

Search Home DRA Home

Submission Detail	
Alias	bilateral breast cancer
Submission ID	
Submission Date	
Center Name	UCSD
Lab Name	

Navigation	
Study	SRP041629
Experiment	SRX532331
	SRX596150
	SRX596214
	SRX596259
	SRX596260
	SRX596261
Sample	SRS602507
	SRS636580
	SRS636587
	SRS636592
	SRS636593
	SRS636594
Run	SRR1374918
	SRR1411169
	SRR1411177
	SRR1411184
	SRR1411185
	SRR1411186
	SRR1411187

GEA/ArrayExpress/GEO

- 遺伝子発現のデータを格納するためのデータベース
- GEA が 2018 年 7 月から運用開始
- GEA ↔ ArrayExpress 間ではデータ交換予定
- ArrayExpress は GEO のデータを取り込んでいるが、
GEO は ArrayExpress のデータを取り込んでいない
- データベース
 - GEA - Genomic Expression Archive (DDBJ): <https://www.ddbj.nig.ac.jp/gea/index.html>
 - ArrayExpress (EBI): <https://www.ebi.ac.uk/arrayexpress/>
 - GEO - Gene Expression Omnibus (NCBI): <https://www.ncbi.nlm.nih.gov/geo/>

NGSデータ解析について

- NGSのデータ解析はコマンドラインで実行するものが多く、習得に時間がかかります（3時間ではとても無理です！）
- ここでは資料を提示するのにとどめます
- NGS ハンズオン講習会（資料、統合TVあり）
 - 平成 29 年 カリキュラム・講義資料・動画
 - 平成 28 年 カリキュラム・講義資料・動画
 - 平成 27 年 カリキュラム・講義資料・動画
 - 平成 26 年 カリキュラム・講義資料・動画
- 次世代シークエンサー DRY 解析教本
- ライフサイエンスQA <http://qa.lifesciencedb.jp/>
 - バイオ分野の質問サイトです



【参考】JGA-NGS データセット作成時の解析プログラム一覧

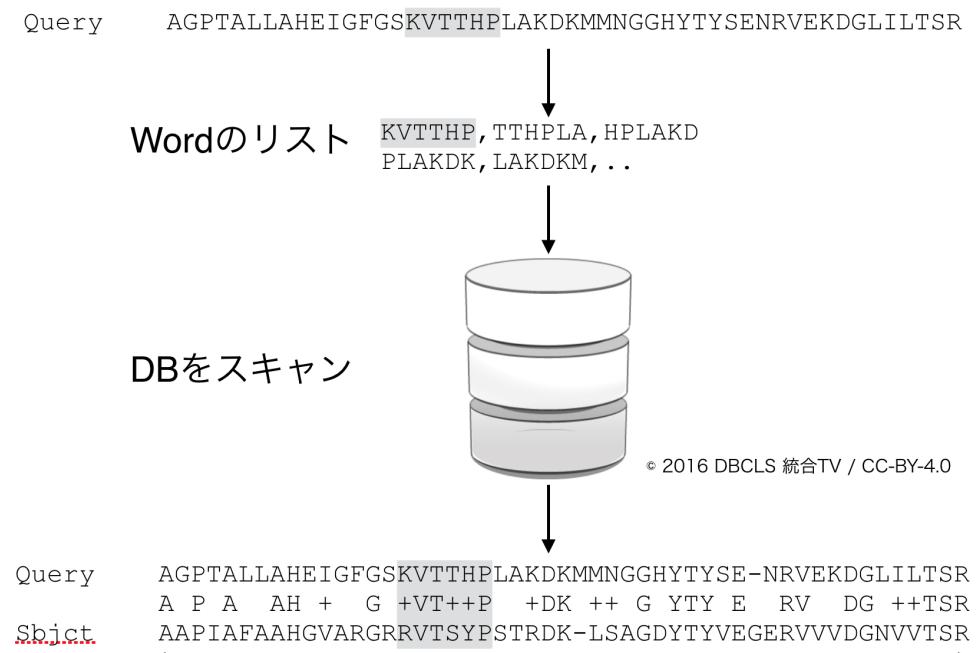
- TogoVar（後述）のデータ解析に使っているプログラム群です
- これ以外にも目的に応じてたくさんのプログラムが、多くの場合無償で提供されています

プログラム	バージョン	説明
FastQC	0.11.5	リード配列品質チェック
BWA	0.7.16a	リード配列をリファレンスゲノムにマッピング
SAM tools	1.6	SAM/BAMファイル等を操作するコマンド群
bedtools	2.13.3	BED関連ファイルの操作
BWA	0.7.16a	リード配列をリファレンスゲノムにマッピング
GATK	3.8.0-ge9d806836	BAMからバリアントを検出
SnpEff	v4.3_core	バリアントにアノテーションを付与
Picard	2.13.2	BAMやVCFファイルを操作するコマンド群

4. 配列検索ツール

BLAST

- 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
 - Basic Local Alignment Search Tool
 - 配列類似性検索のデファクトスタンダード
 - DNA 塩基配列とタンパク質アミノ酸配列の検索が可能
- BLAST の動作原理



BLAST

- 質問配列と検索対象DBの組み合わせ

- blastn だけが核酸配列レベルでの比較
- 残りはアミノ酸配列レベルの比較

DB→ query ↓	核酸 配列	アミノ酸 配列
核酸 配列	blastn tblastx	blastx
アミノ酸 配列	tblastn	blastp

【演習】BLAST 検索

1. 「NCBI BLAST」でググって、トップページを開く
2. Nucleotide BLAST を選ぶ

The screenshot shows the NCBI BLAST homepage. At the top, there's a navigation bar with the NIH logo, U.S. National Library of Medicine, NCBI National Center for Biotechnology Information, and a 'Sign in to NCBI' link. Below the navigation bar, the word 'BLAST®' is prominently displayed. To the right of 'BLAST®' are links for 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. A 'NEWS' sidebar on the left contains a news item about the BLAST widget. The main content area features three large boxes for 'Nucleotide BLAST', 'blastx', and 'tblastn', and another for 'Protein BLAST'.

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

Introducing the BLAST widget - integrating your BLAST results into NCBI's Genome Data Viewer!

Analyze your BLAST results in a genome browser and compare those results against other genome assembly annotations. Introducing the Genome Data Viewer (GDV) and the BLAST widget.

Tue, 19 Jun 2018 14:00:00 EST [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

【演習】BLAST 検索 - クエリ入力

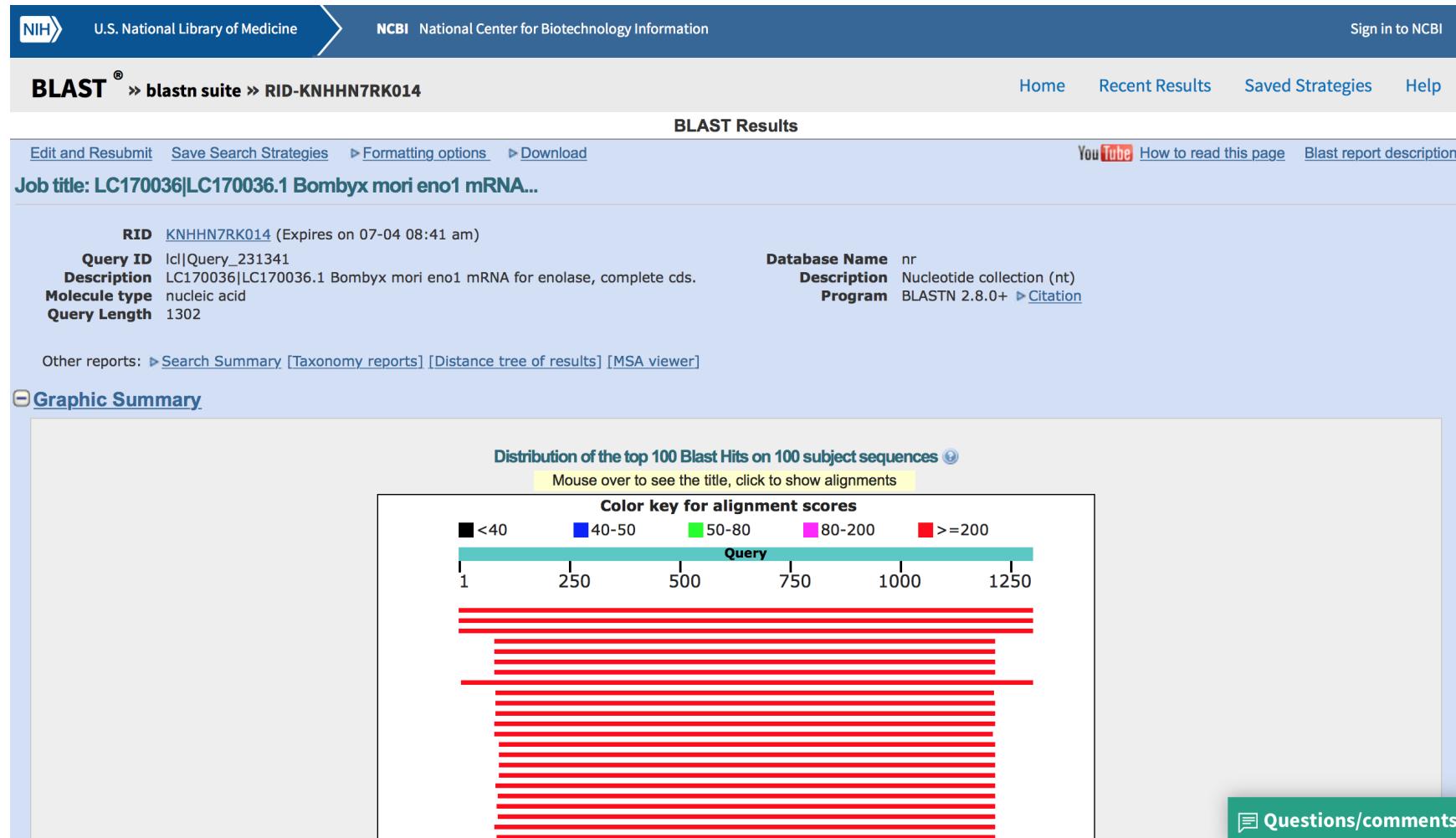
3. FASTA sequence (query) をペーストする。例1・例2

4. 検索対象DBとアルゴリズムを選ぶ。まずはデフォルトのまま
(nr/nt、 megablast) で

The screenshot shows the NCBI Standard Nucleotide BLAST search interface. The query sequence, >LC170036|LC170036.1 Bombyx mori eno1 mRNA for enolase, complete cds., is pasted into the 'Enter Query Sequence' field. In the 'Database' section, 'Nucleotide collection (nr/nt)' is selected. In the 'Program Selection' section, 'Highly similar sequences (megablast)' is selected. Both the 'Database' and 'Program Selection' sections are highlighted with red boxes.

【演習】BLAST 検索 - 結果1

5. 「BLAST」ボタンを押すと検索が始まり以下のような結果が得られる
 1. 検索の概要 + Graphic Summary



【演習】BLAST検索 - 結果2

5. 「BLAST」ボタンを押すと検索が始まり以下のような結果が得られる 2. ヒットした配列の概要

Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Bombyx mori eno1 mRNA for enolase, complete cds	2405	2405	100%	0.0	100%	LC170036.1
<input type="checkbox"/>	Bombyx mori enolase (LOC100037426), mRNA	2374	2374	100%	0.0	99%	NM_001098361.1
<input type="checkbox"/>	PREDICTED: Bombyx mori enolase (LOC100037426), transcript variant X1, mRNA	2300	2300	100%	0.0	99%	XM_021351854.1
<input type="checkbox"/>	Bombyx mori voucher Bmor1 putative enolase protein mRNA, partial cds	2069	2069	87%	0.0	99%	EU032900.1
<input type="checkbox"/>	Triuncina brunnea voucher Trbr putative enolase protein mRNA, partial cds	1297	1297	87%	0.0	87%	JQ789814.1
<input type="checkbox"/>	Racinoa sp. Ofic putative enolase protein mRNA, partial cds	1179	1179	87%	0.0	85%	JQ789671.1
<input type="checkbox"/>	Gastridiota adoxima voucher Gadx putative enolase protein mRNA, partial cds	1024	1024	87%	0.0	83%	JQ789534.1
<input type="checkbox"/>	PREDICTED: Helicoverpa armigera enolase (LOC110372787), mRNA	987	987	99%	0.0	80%	XM_021329748.1
<input type="checkbox"/>	Malacosoma californicum voucher Mcalifor putative enolase protein mRNA, partial cds	913	913	86%	0.0	81%	EU032918.1
<input type="checkbox"/>	Helicoverpa zea voucher Hzea putative enolase protein mRNA, partial cds	907	907	86%	0.0	81%	JQ789579.1
<input type="checkbox"/>	Helicoverpa zea voucher Hzea2 putative enolase protein mRNA, partial cds	905	905	86%	0.0	81%	JQ789580.1
<input type="checkbox"/>	Syntonarcha iriastis voucher Stna putative enolase protein mRNA, partial cds	872	872	87%	0.0	80%	JQ789781.1
<input type="checkbox"/>	Chelepteryx collesi voucher Ccol putative enolase protein mRNA, partial cds	869	869	86%	0.0	81%	EU032901.1
<input type="checkbox"/>	Oenosandra boisduvalii voucher Oeno putative enolase protein mRNA, partial cds	867	867	86%	0.0	81%	JQ789670.1
<input type="checkbox"/>	Habrosyne pyritooides voucher Hbpy putative enolase protein mRNA, partial cds	865	865	86%	0.0	81%	JQ789553.1
<input type="checkbox"/>	Pseudothyatira cymatophorooides putative enolase mRNA, partial cds	865	865	86%	0.0	80%	GQ283724.1
<input type="checkbox"/>	Pseudothyatira cymatophorooides voucher Pcly putative enolase protein mRNA, partial cds	865	865	86%	0.0	80%	EU032929.1
<input type="checkbox"/>	Noorda blitealis voucher Nblt putative enolase protein mRNA, partial cds	859	859	86%	0.0	80%	JQ789647.1
<input type="checkbox"/>	Acanthobrahmaea europaea voucher Aeuro putative enolase protein mRNA, partial cds	857	857	86%	0.0	80%	EU032894.1
<input type="checkbox"/>	Palthis asopialis voucher Paang putative enolase protein mRNA, partial cds	856	856	86%	0.0	80%	EU032894.1

Questions/comments

【演習】BLAST 検索 - 結果3

5. 「BLAST」ボタンを押すと検索が始まり以下のようないくつかの結果が得られる
 3. ヒットした配列のアライメント

Alignments

Download GenBank Graphics ▾ Next ▲ Previous Descriptions

Bombyx mori eno1 mRNA for enolase, complete cds
Sequence ID: LC170036.1 Length: 1302 Number of Matches: 1

Range 1: 1 to 1302 GenBank Graphics ▾ Next Match ▲ Previous Match

Related Information

Score	Expect	Identities	Gaps	Strand
2405 bits(1302)	0.0	1302/1302(100%)	0/1302(0%)	Plus/Plus

Query 1 ATG... 60
Sbjct 1 ATG... 60

Query 61 GAA... 120
Sbjct 61 GAA... 120

Query 121 ACT... 180
Sbjct 121 ACT... 180

Query 181 GGAG... 240
Sbjct 181 GGAG... 240

Query 241 AAC... 300
Sbjct 241 AAC... 300

Query 301 GAGA... 360
Sbjct 301 GAGA... 360

Query 361 GCTGGT... 420
Sbjct 361 GCTGGT... 420

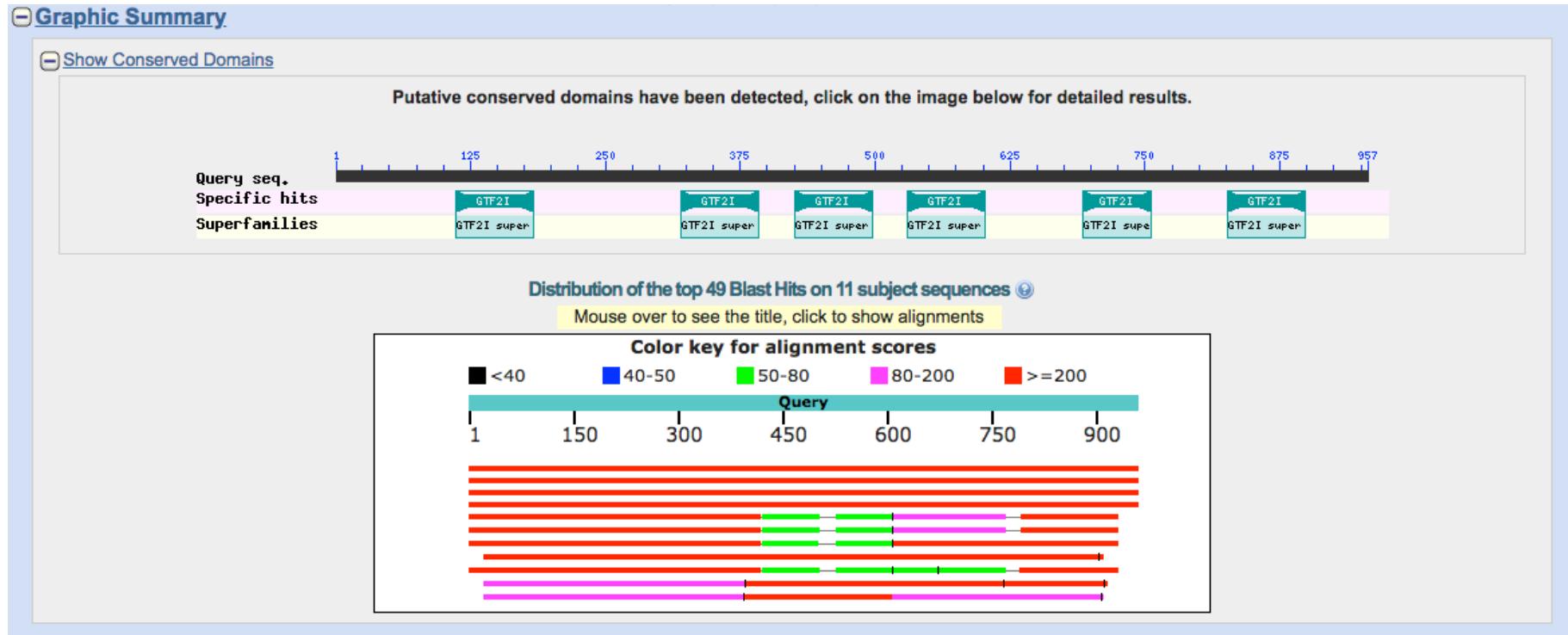
Query 421 AATG... 480
Sbjct 421 AATG... 480

Query 481 AATAAAC... 540
Sbjct 481 AATAAAC... 540

Query 541 GCCATG... 600
Sbjct 541 GCCATG...

【発展】 Protein BLAST、Local BLAST

- Protein BLAST も試してみましょう
 - GTFII-I のアミノ酸配列を UniProtKB/SwissProt データベースに対して検索した結果



- 大量に検索する場合、独自のデータベースに対して検索する場合は、自分の PC にインストールして使えます
 - 統合TV：Local BLASTの使い方～導入・準備編(MacOSX版)～2017
 - 統合TV：Local BLASTの使い方～検索実行・オプション(MacOSX版)～2017

GGGenome/GGRNA <https://gggenome.dbcls.jp/ja/> <https://ggrna.dbcls.jp/ja/>

■ 超絶高速ゲノム検索 GGGenome

- 検索対象：ゲノム配列
- 入力：塩基配列を入力して検索
- ミスマッチ/ギャップがあっても検索可能

■ 統合遺伝子検索 GGRNA

- 検索対象：RefSeq の転写産物
 - mRNA: NM、XM からはじまるエントリ
 - ncRNA: NR、XR からはじまるエントリ
- 入力：遺伝子名、キーワード、Accession 番号、塩基配列、アミノ酸配列、マイクロアレイのプローブ ID などから検索可能

【実習】GGGenomeによる塩基配列の高速検索

1. 「GGGenome」でググってトップページを開きます
2. 「塩基配列」と「検索対象ゲノム」を指定して検索します
(「検索例」に塩基配列があるのでコピペ or クリックします)

超絶高速ゲノム配列検索 Help | English

GGGenome

検索 Human genome, GRCh37/hg19 (Feb, 2009) 検索

許容するミスマッチ/ギャップの数: 1 (検索する塩基配列の長さの25%まで)

双方向を検索 +方向のみ検索 -方向のみ検索

検索例:

- [TTCATTGACAAACATT] 塩基配列を検索
- 許容可能な使い方
- 遺伝子や転写産物のキーワード検索は GGRNA 《ぐぐるな》へ
 - 例)ヒトの「nanog」を検索: <http://GGRNA.dbcls.jp/hs/nanog>

検索結果へのリンク:

- [http\[s\]://GGGenome.dbcls.jp/db/k/\[strand\]/sequence\[.format\]\[.download\]](http://GGGenome.dbcls.jp/db/k/[strand]/sequence[.format][.download])
 - db → hg19, mm10, dm3, ce10, TAIR10, pombe, refseq, etc. 省略時は hg19
 - 検索可能なデータベース一覧
 - k → 許容するミスマッチ/ギャップの数。あまり大きいとしぼうする。省略時は 0
 - strand → '+' ('plus') または '-' ('minus') で特定の方向のみ検索。省略時は両方向を検索
 - sequence → 塩基配列。大文字・小文字は区別しない
 - format → html, txt, csv, bed, gff, json。省略時は html
 - download → URLの最後に付加すると検索結果をファイルとしてダウンロードできる
- 例1:<http://GGGenome.dbcls.jp/TTCATTGACAAACATT>
 - ヒトゲノム hg19 (省略可) で
 - ミスマッチ/ギャップを許容せず (省略可)
 - TTCATTGACAAACATT を検索し
 - html 形式 (省略可) で結果を返す
- 例2:<http://GGGenome.dbcls.jp/mm10/2/+//TTCATTGACAAACATTGCGT.txt>
 - マウスゲノム mm10 で
 - 2 ミスマッチ/ギャップまで許容して
 - + 方向に限定して
 - TTCATTGACAAACATTGCGT を検索し
 - txt 形式 (タブ区切りテキスト) で結果を返す

【実習】GGGenomeによる塩基配列の高速検索結果

3. +方向に15配列、相補鎖に8配列ヒットしています

- 検索配列が短いとたくさんヒットします

超絶高速ゲノム配列検索 Help | English

GGGenome

TTCATTGACAACATT 検索 Human genome, GRCh37/hg19 (Feb, 2009)

許容するミスマッチ/ギャップの数: 0 (検索する塩基配列の長さの25%まで)

双方向を検索 +方向のみ検索 -方向のみ検索

2018-11-07 19:27:16, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

Summary:

- TTCATTGACAACATT (15)
- AATGTTGCAATGAA (8)
- **TOTAL (23)**

Results:

検索語に色がつきます(ミスマッチ・挿入欠失)。

chr1:83462476-83462490 ▼83462476
CTAATTAATGAATAGTAAAGTATATTTCTC **TTCATTGACAACATT** TTTCTCTTACTTTATTGTAAGAACATAAT

chr2:161223115-161223129 ▼161223115
ATTGCAAGAACAGCACATTATATTCTA **TTCATTGACAACATT** AGCATATAATGCAAATTACTTGGAACTT

chr3:15289790-15289804 ▼15289790
ATTGTAATAGCCAGAAATTGAAACATATA **TTCATTGACAACATT** TAAGATTATAATAGTCATATAATAGTCC

chr3:84619845-84619859 ▼84619845
TTGGTTTGATATATTAATAGTTGAAAC **TTCATTGACAACATT** AAGTTTGTCTAGGCTCTATTGGATAA

chr4:16134651-16134665 ▼16134651
AACGAAAGGCCTTGAGGAAAGCCTGATAATT **TTCATTGACAACATT** ACCATCAACCACACCTTACCACTTAGGA

chr4:63031140-63031154 ▼63031140
TCTTCAAATATTTACAGAGTTCAACTCTT **TTCATTGACAACATT** ATGACATTATGCTGATTGGATCCAGTGAGT

chr4:149833218-149833232 ▼149833218
CAGAATATATTATCGCTTAAACACAGATT **TTCATTGACAACATT** TTGTGGCACCCATTAAATATTAATATTAA

【実習】GGGenomeによるミスマッチ/ギャップ検索結果

4. 「許容するミスマッチ/ギャップの数」を指定してみます

- ヒット数が大幅に増えたのがわかります

超絶高速ゲノム配列検索 [Help](#) | [English](#)

GGGenome

TTCATTGACAACATT Human genome, GRCh37/hg19 (Feb, 2009)

許容するミスマッチ/ギャップの数: 1 検索する塩基配列の長さの25%まで

双方向を検索 十万方向のみ検索 一方向のみ検索

2018-11-07 19:25:39, GGGenome : Human genome, GRCh37/hg19 (Feb, 2009)

Summary:

- [TTCATTGACAACATT \(1007\)](#)
- [AATGTTGTCAATGAA \(943\)](#)
- TOTAL (1950)**

Results:

検索語に色がつきます(ミスマッチ・挿入欠失)。

[chr1:7709963-7709977](#) ▼7709963 TGCCAGGGAGTGGTCTGTCAAAGTTAGAATTTCATTG**CCAACATT**CAAATCTCCAGATTTCAGCCTCTTTAA

[chr1:14482580-14482593](#) ▼14482580 AAGAAATTTTTAATTTCCTTAATTTC**TTCATTGAC-ACATT**GGTCATTCAAGGACCATATTATTTAATTTC

[chr1:19318674-19318687](#) ▼19318674 GGTCACTTCTGTCTAGCTCCTGCATAAC**TTC-TTGACAACATT**CAACATAATTGCCATTCCCACCCCTTTGA

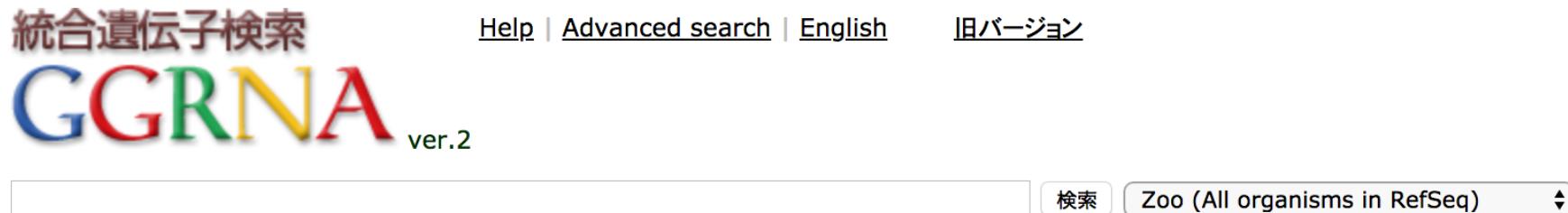
[chr1:24735451-24735464](#) ▼24735451 TCTATAAAAATTGATATTAAATAGAA**TT-ATTGACAACATT**CTTCAAGAAATTAAAGAGATTCAAAATAACA

[chr1:27076473-27076488](#) ▼27076473 TGCTAAAGAGGTGCTTGTAGATAA**TTCATTGAGCAACATT**TATCAAGCATTTGCGTTGTGTCTGG

[chr1:30202178-30202192](#) ▼30202178 TGCATTAACAGAAAATGCCCTGGATT**TTCATGGACAACATT**TATTGAAGAGTATAAAAGAAAATATGGAA

【参考】GGRNAによる遺伝子の高速検索

- GGRNAのトップページを開くと検索例のリンクがあるので、いろいろクリックしてみましょう



検索例：

- 「homeobox」「claudin」..... フリーワード検索
- 「"RNA interference"」..... ダブルクオートで囲ってフレーズ検索
- 「Argonaute "PAZ domain"」..... Argonaute かつ "PAZ domain" のAND検索
- 「NM_001518」「10579」..... RefSeq IDやGene IDなど各種IDから検索
- 「symbol:VIM」..... 遺伝子名(symbolまたはsynonym)から検索
- 「ref:Naito」..... 文献情報のなかからフリーワード検索
- 「1552311_a_at」..... マイクロアレイのプローブIDから塩基配列を検索
- 「aa:KDEL」..... アミノ酸配列を検索
- 「caagaagagattg」..... 塩基配列を検索
- 「comp:caagaagagattg」..... 相補鎖を検索
- 「iub:aggtcannrtgacct」..... N, R, Y 等のあいまいな塩基を含む塩基配列を検索
- 「詳細な使い方」
- ゲノム配列を検索したいときは **GGRNA**《ゲゲゲノム》へ
 - ヒトゲノムの検索例：<http://GGGenome.dbcls.jp/TTCATTGACAAACATT>

【統合TV】：[GGRNAで遺伝子をGoogleのように検索する](#)

CRISPRdirect <https://crispr.dbcls.jp/>

- ゲノム編集ツール CRISPR-Cas9 システムに使用するガイド
RNA 設計ツール
- PAM に隣接し、特異性の高い20塩基を選択
 - PAM (Protospacer adjacent motif) に隣接する配列の選択
 - オフターゲット配列検索 (GGGenome を利用)
- 多くの生物種に対応
 - 650 種以上の生物種に対応 (GGGenome と同じ)
 - Help ページに現在対応している生物種の一覧があります
 - CRISPRdirect/GGGenome 共、追加してほしい生物種（ゲノムが公開されているものに限る）がありましたら、ここからお知らせください！

【実習】CRISPER 配列を設計する

1. 「CRISPRdirect」でググってトップページを開く
2. 「Accession 番号」 or 「塩基配列」、「PAM」、「生物種」を入力

The screenshot shows the CRISPRdirect website interface. At the top, there is a logo with the text "CRISPR direct" and a subtitle "- Rational design of CRISPR/Cas target." Below the logo is a search bar with the placeholder text "Enter an accession number (e.g. NM_006299) or genome location (e.g. hg19:chr7:900000-901000)". To the right of the search bar is a "Help" button. Below the search bar is a "retrieve sequence" button. Further down, there is a section titled "or Paste a nucleotide sequence:" with a large text input area containing a sample sequence. Below this input area is another "retrieve sequence" button. At the bottom of the page, there are fields for "upload sequence file", "PAM sequence requirement" (set to "NGG"), and "Specificity check" (set to "Human (Homo sapiens) genome, GRCh37/hg19 (Feb, 2009)"). A "design" button is located at the bottom left.

【統合TV】 : [CRISPRdirectを使ってCRISPR/Cas法のガイドRNA配列を設計する](#)

【実習】CRISPERdirect 結果

Results: ?

Sequence name: sample sequence

PAM sequence: NGG

Specificity check: Human (Homo sapiens) genome, GRCh37/hg19 (Feb, 2009)

Time: 2018-11-07 11:19:48

- Highlighted target positions (e.g., 45 - 67) indicate sequences that are highly specific and have fewer off-target hits.
- Target sequences with '0' in '20mer+PAM' (in number of target sites column) are shown in gray. Such sequences may possibly span over exon-exon junctions, so avoid using these.
- Target sequences with TTTTs are also shown in gray. Avoid TTTTs in gRNA vectors with pol III promoter.

show **highly specific** target only

Show 20 entries

position start - end	target sequence 20mer+PAM (total 23mer)	sequence information					number of target sites ?		
		GC% of 20mer	Tm of 20mer	TTTT in 20mer	restriction sites	20mer +PAM	12mer +PAM	8mer +PAM	
4 - 26	- cccgcgtcgccccgaccagag [gRNA]	75.00 %	82.34 °C	-	Bme1580I	1 [detail]	3 [detail]	103 [detail]	
16 - 38	- cccgaccagagaaggcaagttcga [gRNA]	50.00 %	71.17 °C	-		1 [detail]	26 [detail]	4354 [detail]	
17 - 39	- ccgcaccagagaaggcaagttcgag [gRNA]	50.00 %	71.53 °C	-		1 [detail]	32 [detail]	5910 [detail]	
21 - 43	- ccaagagaaggcaagttcgagaacg [gRNA]	50.00 %	69.25 °C	-		1 [detail]	24 [detail]	8744 [detail]	
24 - 46	+ gagaaggcaagttcgagaacgg [gRNA]	50.00 %	69.25 °C	-		1 [detail]	2 [detail]	306 [detail]	
35 - 57	+ tcgagaacgaggaggaaaaattttt [gRNA]	40.00 %	66.49 °C	+		1 [detail]	25 [detail]	9901 [detail]	
66 - 88	- ccgcgagtgtagatataactaca [gRNA]	40.00 %	65.70 °C	-		0 [detail]	3 [detail]	1097 [detail]	
69 - 91	+ cgagtgtagatataactaca [gRNA]	40.00 %	65.70 °C	-		0 [detail]	26 [detail]	3790 [detail]	
70 - 92	+ gagtgtagatataactacac [gRNA]	40.00 %	65.27 °C	-		0 [detail]	3 [detail]	3604 [detail]	
77 - 99	+ agattaagtacacgggcttc [gRNA]	45.00 %	70.40 °C	-		1 [detail]	6 [detail]	1306 [detail]	
78 - 100	+ gattaagtacacgggcttc [gRNA]	45.00 %	70.01 °C	-		1 [detail]	9 [detail]	6616 [detail]	
83 - 105	+ agtacacgggcttcaggagac [gRNA]	60.00 %	78.32 °C	-		1 [detail]	63 [detail]	6397 [detail]	
93 - 115	+ ctccagggaccggcccccacg [gRNA]	75.00 %	84.21 °C	-		1 [detail]	7 [detail]	2568 [detail]	
102 - 124	+ ccggccccacgaggAACGCCAG [gRNA]	80.00 %	86.67 °C	-		1 [detail]	1 [detail]	1244 [detail]	
102 - 124	- ccgcgccccacgaggAACGCCAG [gRNA]	75.00 %	84.31 °C	-		1 [detail]	19 [detail]	3777 [detail]	
106 - 128	- cccgcacgaggAACGCCAGCAGC [gRNA]	70.00 %	80.55 °C	-		1 [detail]	1 [detail]	1098 [detail]	
107 - 129	- cccgcacgaggAACGCCAGCAGC [gRNA]	70.00 %	82.51 °C	-		1 [detail]	2 [detail]	1116 [detail]	
108 - 130	- ccaagggaggAAACGCCAGGACCGCT [gRNA]	70.00 %	82.52 °C	-		1 [detail]	2 [detail]	1011 [detail]	
120 - 142	- ccaaggcagcgttccaggAAACGCCCT [gRNA]	65.00 %	80.61 °C	-		1 [detail]	6 [detail]	1510 [detail]	
130 - 152	+ ttccaggaaacgcgttccaggAAACGCCGA [gRNA]	65.00 %	80.58 °C	-		1 [detail]	4 [detail]	289 [detail]	

Showing 1 to 20 of 77 entries

First Previous 1 2 3 4 Next Last

標的サイトの位置と塩基配列

配列に関する情報

ゲノム中に何か所あるか
(オフターゲット)

オススメ

オススメ

5. ゲノムデータベース



ゲノムデータベースとは？

- ゲノム配列をはじめとした（遺伝）情報を生物種ごとにまとめたデータベース
- 狹義にはゲノム配列のデータベースをいう
- さまざまなゲノムデータベース
 - NCBI の Genome <https://www.ncbi.nlm.nih.gov/genome/>
 - 生物種ごと (Browse by Organism) <https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/>
 - PlantGDB <http://www.plantgdb.org>
 - Plant Genome Database Japan (PGDBj) <http://pgdbj.jp/>
 - MicrobeDB.jp <http://microbedb.jp/MDB/>

コミュニティによるゲノムデータベース

- Mouse Genome Informatics (MGI) - マウス
 - <http://www.informatics.jax.org/>
- Rat Genome Database (RGD) - ラット
 - <https://rgd.mcw.edu/>
- WormBase - 線虫
 - <https://www.wormbase.org/>
- FlyBase - ショウジョウバエ
 - <http://flybase.org/>
- The Arabidopsis Information Resource (TAIR) - シロイヌナズナ
 - <https://www.arabidopsis.org/>
- Saccharomyces Genome Database (SGD) - 酵母
 - <https://www.yeastgenome.org/>
- CyanoBase - シアノバクテリア (光合成細菌)
 - <http://genome.microbedb.jp/cyanobase/>

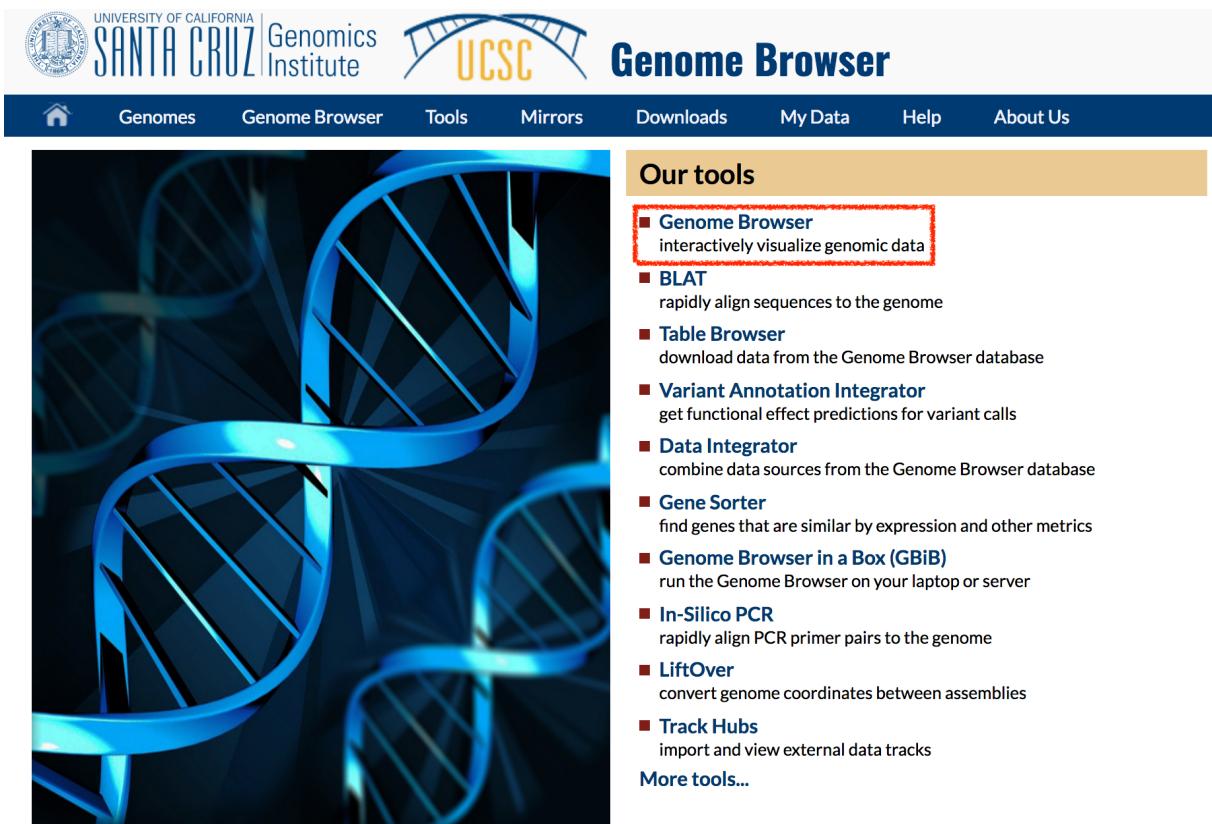


ゲノムブラウザとは？

- 塩基配列解読したゲノム配列とそこに付与（アノテーション）された情報を見るための仕組み
- オンライン型とローカル型
 - オンライン型：ウェブブラウザ上でサーバにあるゲノムデータベースから必要な情報を取り出してこれる
 - UCSC Genome Browser: <https://genome.ucsc.edu/>
 - Ensembl Genome Browser: <https://www.ensembl.org/>
 - NCBI Genome Data Viewer: <https://www.ncbi.nlm.nih.gov/genome/gdv/>
 - TOGO GENOME: <http://togogenome.org/>
 - ローカル型：手元のコンピュータにインストールして使用
 - Integrative Genomics Viewer (IGV): <https://software.broadinstitute.org/software/igv/>

【実習】UCSC ゲノムブラウザを使ってみる

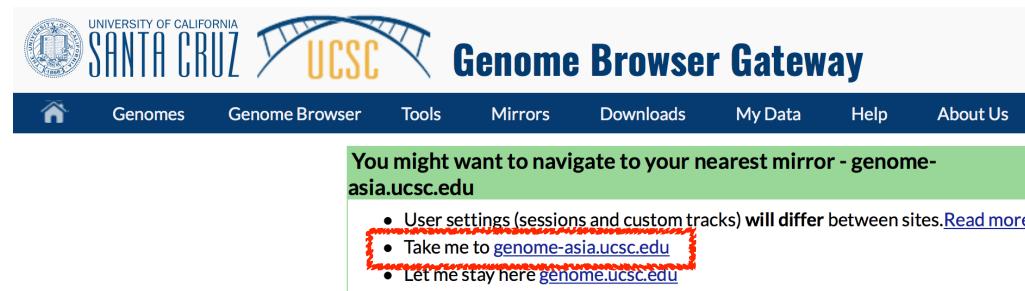
1. 「UCSC Genome Browser」でググってトップページを開く
2. トップページにはツール名がリストされている。一番上にある「Genome Browser」をクリックする。



【統合TV】：UCSC Genome Browser を使って様々な組織、細胞における遺伝子発現データをゲノムブラウザで表示する

【実習】UCSC ゲノムブラウザを使ってみる - 検索項目の入力

3. 最寄りのミラーサイトに接続する



4. Genome Browserのページが開くので、生物種(Human)とアッセンブリ(Feb.2009/(GRC37/hg19))を選んで、検索語を入力する (ここではFAM32A)

Human Assembly
Feb. 2009 (GRCh37/hg19)

Position/Search Term
FAM32A

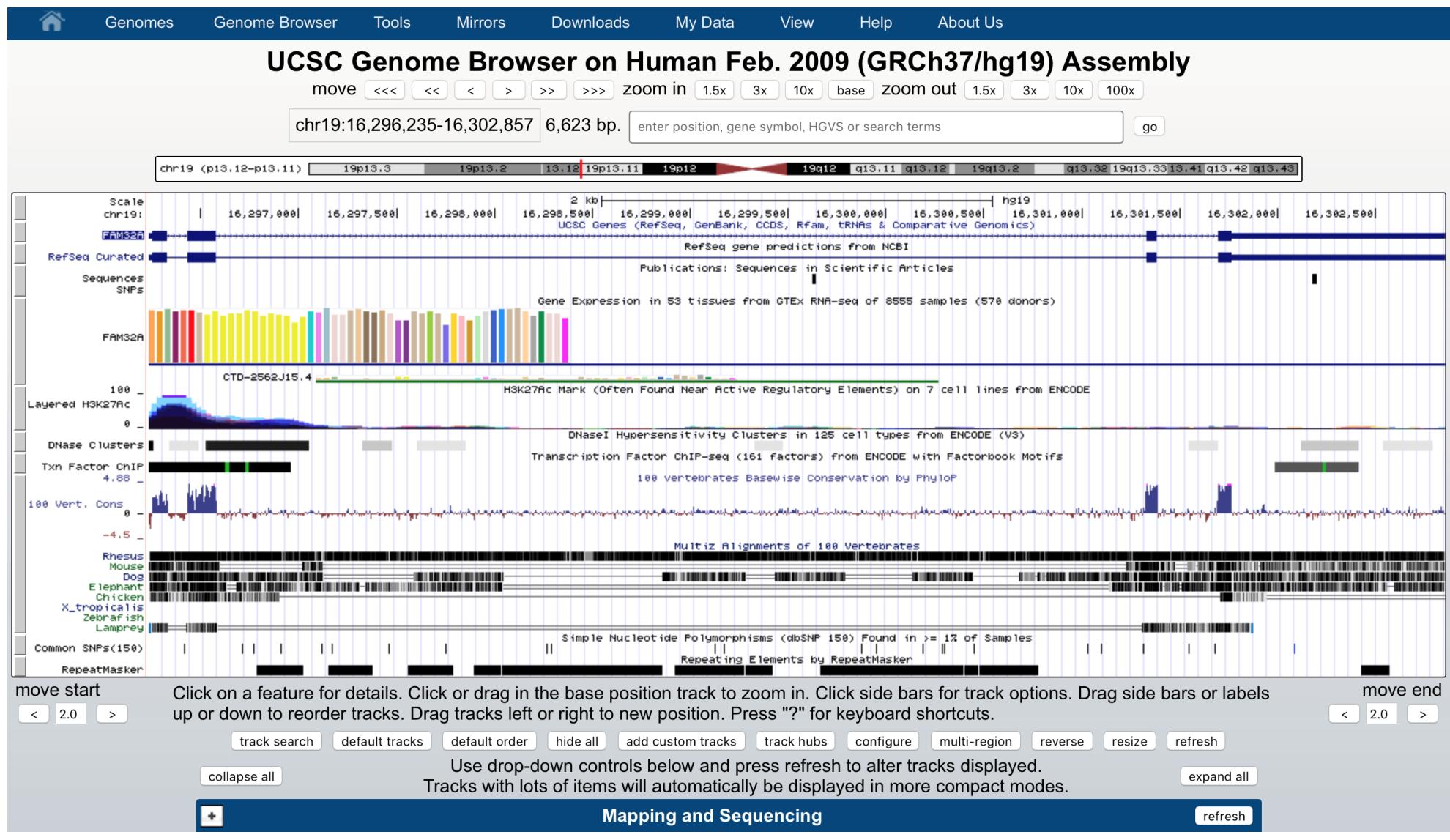
FAM32A (Homo sapiens family with sequence similarity 32, member A (FAM32A), mRNA.)

Human Genome Browser - hg19 assembly

view sequences

【実習】UCSC ゲノムブラウザを使ってみる - ゲノム領域の表示

5. FAM32A遺伝子のゲノム領域が表示される



【実習】UCSC ゲノムブラウザを使ってみる - 表示項目の追加

6. 「Regulation」の「ENC TF Binding...」を「hide」から
「show」に変更して、「refresh」ボタンを押す

The screenshot shows the UCSC Genome Browser interface with the 'Regulation' track selected. In the 'Regulation' section, the 'ENC TF Binding...' item is highlighted with a red box. Its dropdown menu is open, showing 'hide' and 'show' options. The 'show' option is selected, indicated by a blue arrow pointing to it. A red box also highlights the 'refresh' button in the same section.

Other ESTs Other mRNAs ⑯ Poly(A) PolyA-Seq SIB Alt-Slicing Spliced ESTs

UniGene hide

GTEx Gene GTEx Transcript Affy Exon Array Affy GNF1H ⑯ Affy RNA Loc Affy U95

Affy U133 hide Affy U133Plus2 Allen Brain Burge RNA-seq ⑯ CSHL Small RNA-seq ⑯ ENC Exon Array...

ENC ProtGeno... hide ⑯ ENC RNA-seq... ⑯ GIS RNA PET GNF Atlas 2 GWIPS-viz Riboseq ⑯ Illumina WG-6

PeptideAtlas qPCR Primers ⑯ RIKEN CAGE Loc ⑯ Sestan Brain

Expression refresh

Regulation refresh

ENC Encode Regulation... New GeneHancer GTEx Combined eQTL GTEx Tissue eQTL ⑯ CD34 Dnase1 CpG Islands...

ENC Chromatin... hide ⑯ ENC DNA Methyl... ⑯ ENC Histone... ⑯ ENC RNA Binding... ⑯ ENC TF Binding...

FSU Repli-chip hide ⑯ Genome Segments ⑯ NKI Nuc Lamina... ORegAnno Stan Nucleosome SwitchGear

⑯ SwitchGear TSS hide ⑯ TFBS Conserved TS miRNA sites UCSF Brain Methyl UMMS Brain Hist ⑯ UW Repli-seq

Vista Enhancers hide

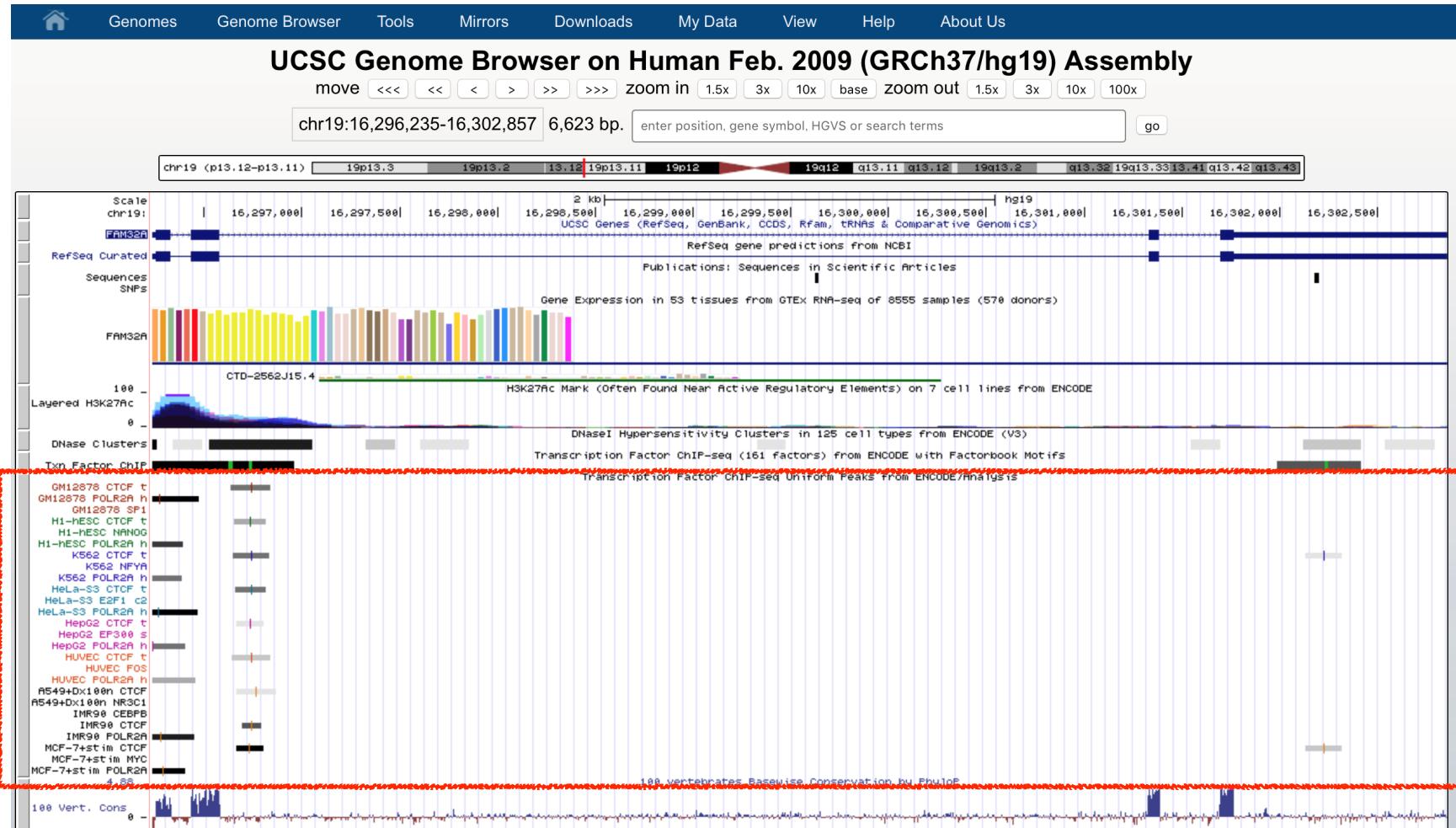
Comparative Genomics refresh

Conservation Cons 46-Way ⑯ Cons Indels MmCf ⑯ Evo Cpg GERP phastBias gBGC

full hide ⑯ hide hide hide hide

【実習】UCSC ゲノムブラウザを使ってみる - 表示項目の追加

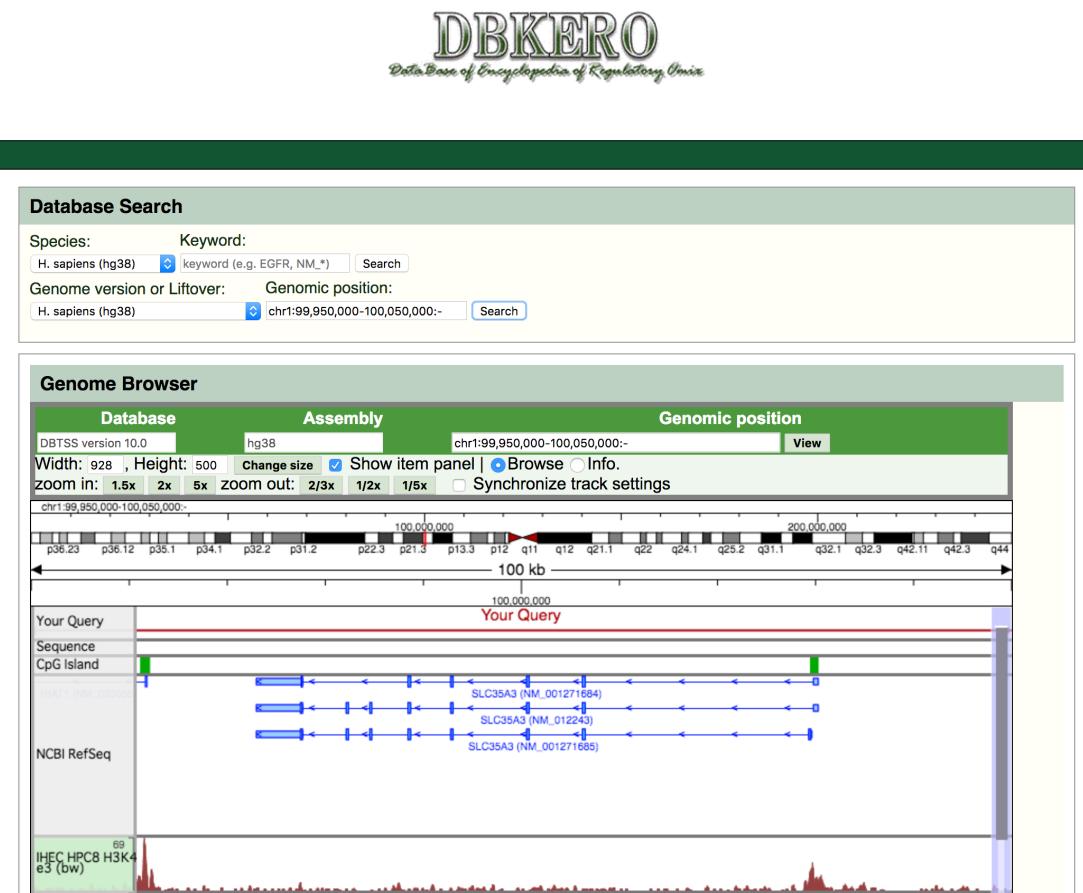
7. 転写因子結合サイトの情報が追加される



8. いろいろ変更して表示してみましょう。わからなくなったら、図の下に並んでいるボタンの「default tracks」を押すと最初の状態に戻せます。

DBTSS/DBKERO <https://dbtss.hgc.jp/> <http://kero.hgc.jp/>

- DataBase of Transcriptional Start Site/DataBase of Kashiwa Encyclopedia for human genome mutation in Regulatory region and their Omics contexts
- ヒト細胞の各種オミクスデータを集積したデータベース
 - 全ゲノム (WGS)
 - トランスクリプトーム (RNA-seq)
 - エピゲノム (BS-seq, ChIP-seq)
 - 転写開始地点 (TSS-seq)
 - シングルセルデータ
 - ロングリードデータ

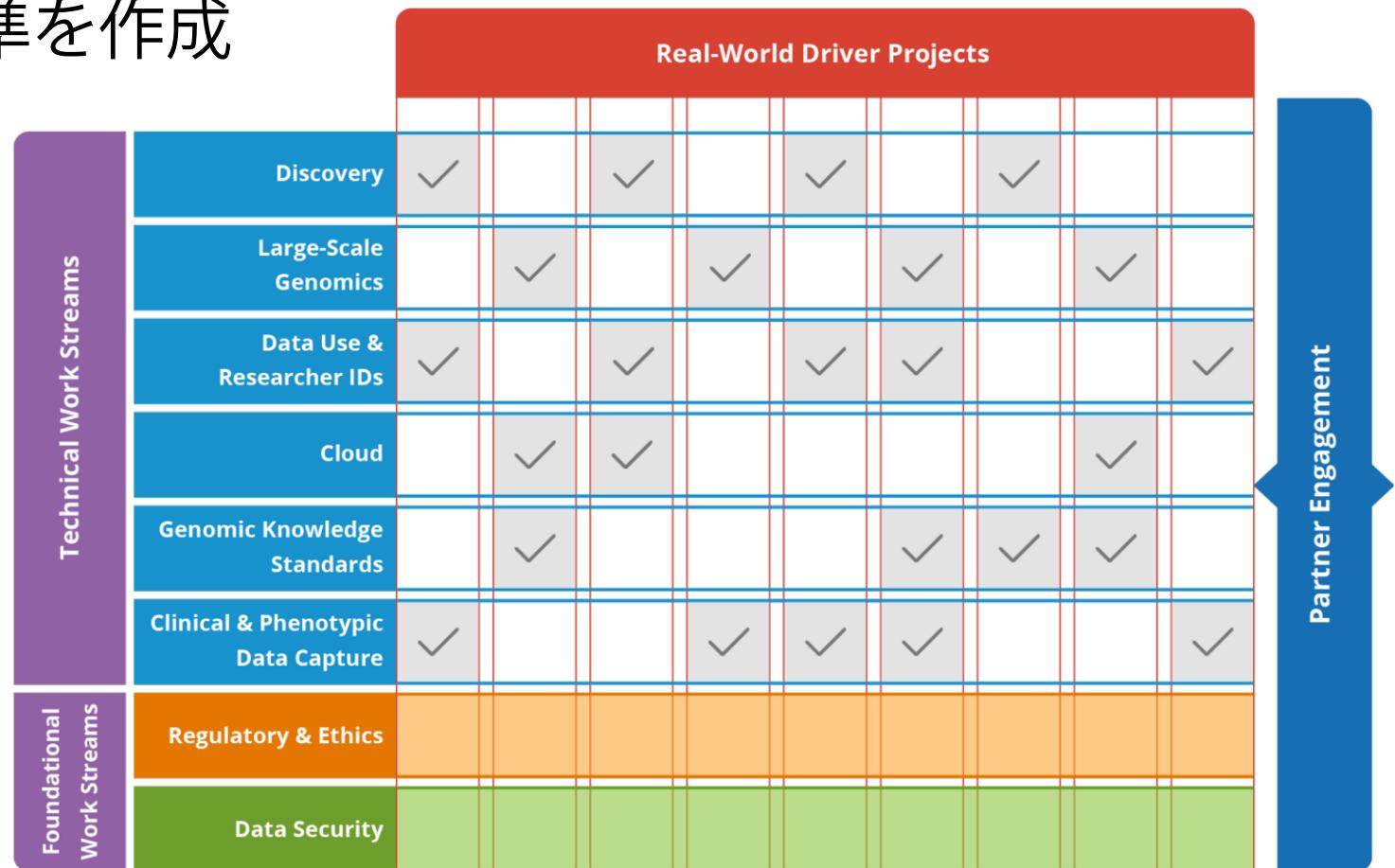


6. ヒト（ゲノム）データベース



GA4GH <https://www.ga4gh.org/>

- GA4GH: Global Alliance for Genomics and Health
- ヒトのゲノムデータ・医療情報を国際的に共有するためのルール・標準を作成



GA4GH <https://www.ga4gh.org/>

■ Work Streams

- Data Security
- Regulatory & Ethics
- Cloud
- Data Use & Researcher Identities (DURI)
- ...

■ Driver Project

- MatchMaker Exchange
- ELIXIR Beacon
- Clinical Genome Resource (ClinGen)
- ENA/EVA/EGA
- (AMED Umbrella Project)
- ...



Matchmaker
Exchange



All of Us
RESEARCH PROGRAM



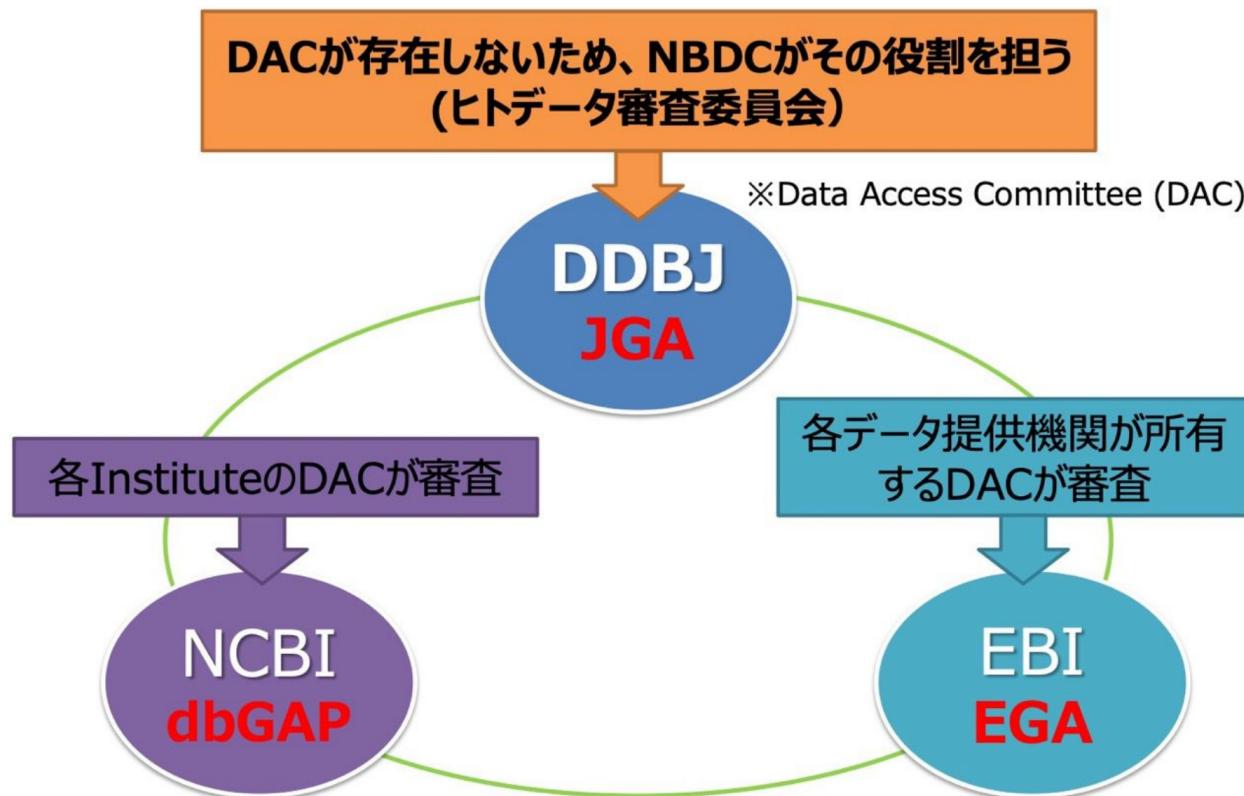
monarch
INITIATIVE

NIH NATIONAL CANCER INSTITUTE
Genomic Data Commons



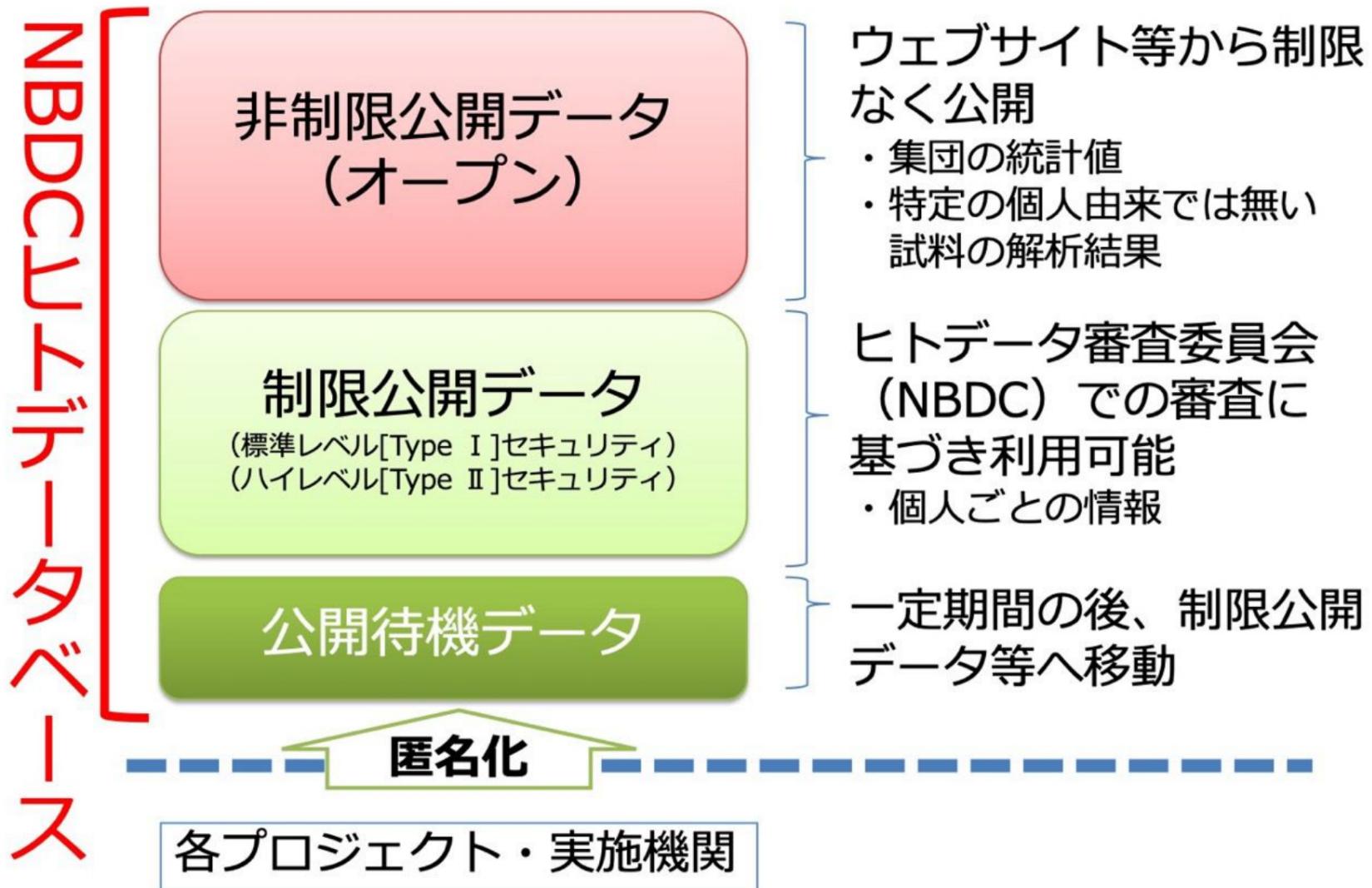
制限公開データベース

- Controlled Access (制限公開) が必要なヒトに関するデータを収集・公開する
 - NCBI dbGaP (The database of Genotypes and Phenotypes)
 - EBI EGA (European Genome-phenome Archive)
 - DDBJ JGA (Japanese Genotype-phenotype Archive)



NBDCヒトデータベース <https://humandbs.biosciencedbc.jp/>

■ データの種類



NBDCヒトデータベース - JGA のデータ提供申請・審査

■ データ提供の必要性

- 論文投稿時に公的 DB へのデータの登録とアクセスション番号の記載が必須
- 予算申請時に AMED データマネージメントプラン提出の義務化

■ データ提供に必要な手続き

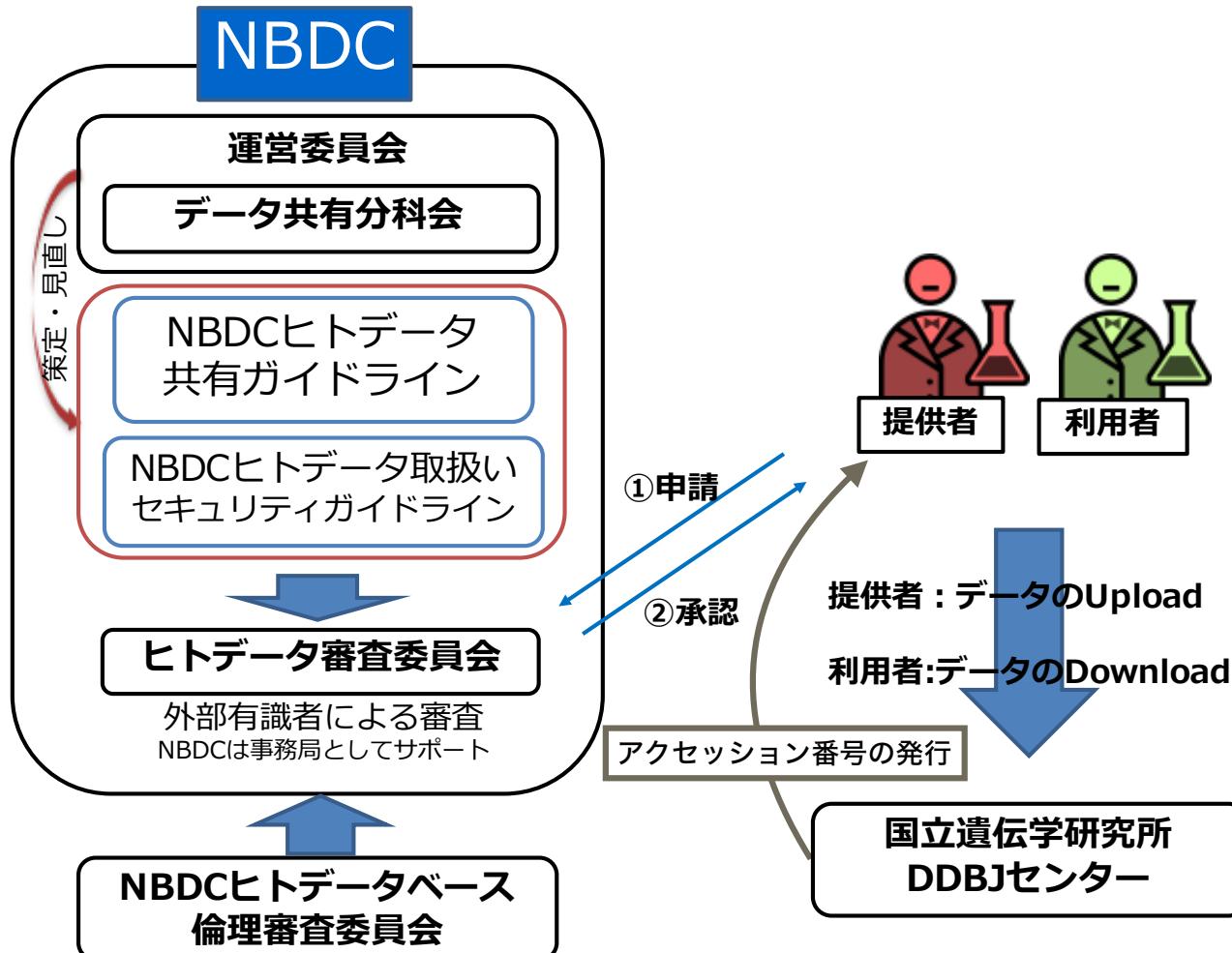
- 研究計画書、同意文書（IC）および説明文書、倫理審査の承認通知書等
- 特にデータ提供に関する同意を取得しておくことが重要

The screenshot shows the NBDC Hit Database homepage. The navigation bar includes links for Home, Data Utilization, Data Submission (highlighted with a red box), Guidelines, External Server, NBDC Hit Database Review Committee, Publication, Contact, and FAQ. A search bar is also present. A red box highlights the 'Data Submission' link. A warning message at the top states: '遺伝研サーバ停止に伴い、『JGA Submission/Download Tool』および『JGA Meta Viewer』のJGA登録・検索サービスが一時停止します。停止期間:2018/11/16 13:00 - 2018/11/21 24:00(予定)' (Due to the temporary shutdown of the Genetic Research Institute server, the registration and search services for the JGA Submission/Download Tool and JGA Meta Viewer will be temporarily unavailable. Stop period: November 16, 2018, 13:00 - November 21, 2018, 24:00 (estimated)). Below this, the 'Data Submission' section contains information about the submission process, including a note about the new Personal Information Protection Act and the Ethics Guideline, and a list of required documents such as the NBDC Hit Database Shared Guideline, research plan, and informed consent forms.

NBDCヒトデータベース <https://humandbs.biosciencedbc.jp/>

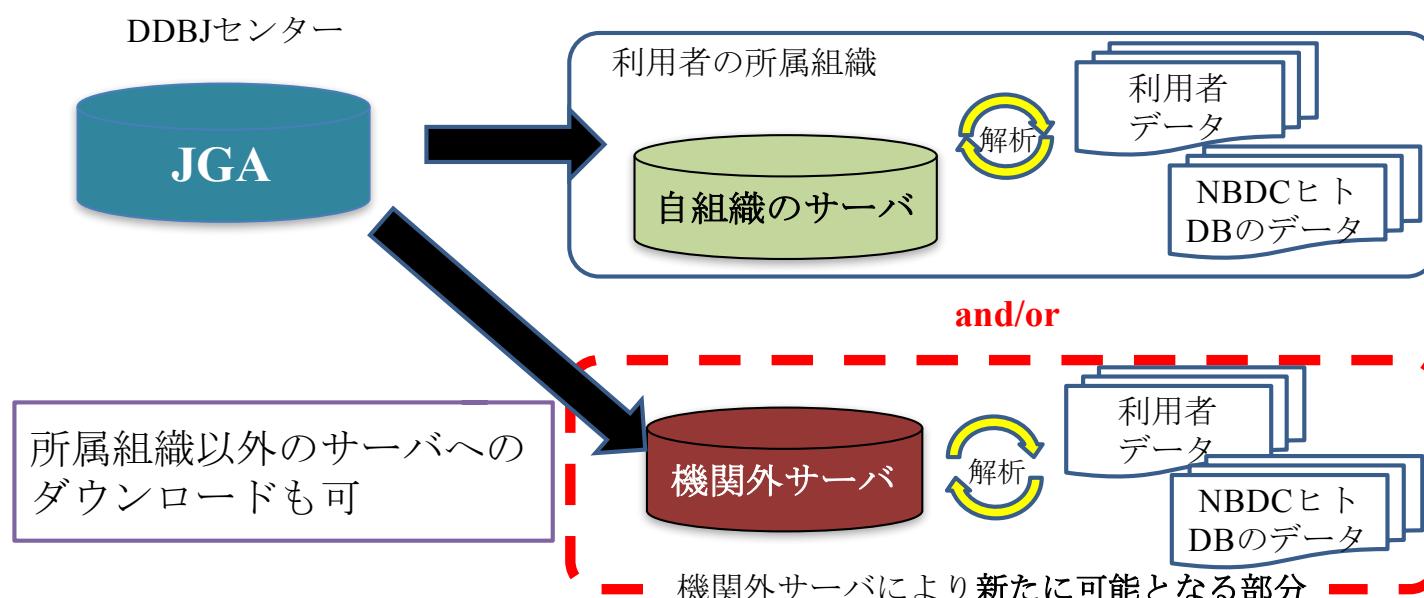
■ データ共有の仕組み

NBDCヒトデータベース運営の概要



NBDCヒトデータベース 制限公開データの利用

- データ利用申請・審査が必要
 - 研究計画、倫理審査の承認通知書等
 - 年1回利用報告義務あり
- 利用者の所属組織以外のサーバでも、認定された機関外サーバであれば、ダウンロードして利用可能



NBDCヒトデータベースとTogoVarの関係

	一次データベース Primary database	二次データベース Secondary database
別の呼び方	Archival database	Curated database; Knowledgebase
データソース	研究者（登録者）が実験で得たデータを直接登録	一次データベースのデータや文献を解析、解釈、キュレーションした結果
例	<ul style="list-style-type: none">DDBJ/ENA/GenBankGEA/ArrayExpress/GEODRA/ERA/SRAEVA·DGVa/dbSNP·dbVarPDB	<ul style="list-style-type: none">RefSeqEnsemblExpression AtlasChIP-AtlasUniProt

転載元：次世代シーケンスデータベースの紹介 (DDBJ児玉博士 作成)
(https://github.com/AJACS-training/AJACS71/blob/master/05_kodama/AJACS71_05_kodama.pdf)



NBDCヒトデータベース

個人別の情報（個人情報）

データ提供者（登録者）
データ利用者
の両方が存在する



個人別の情報でない（集計情報）

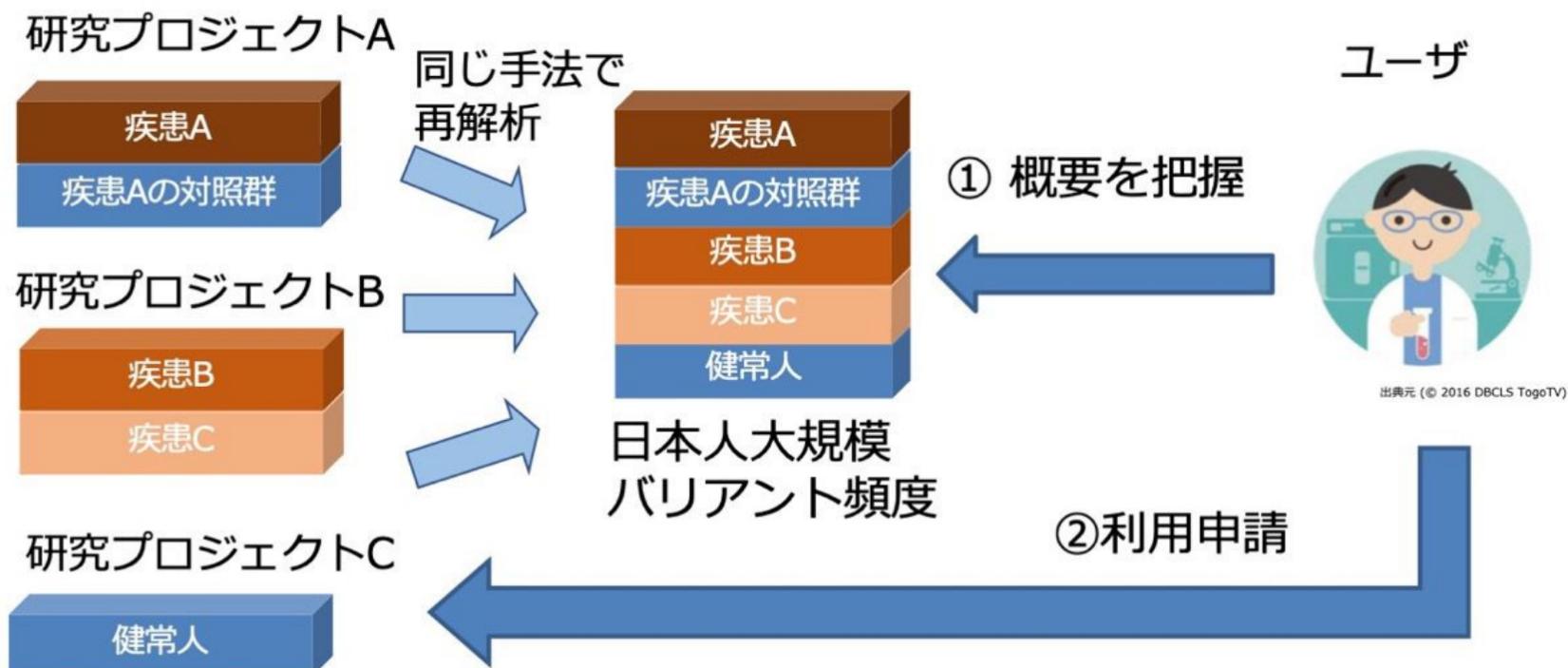
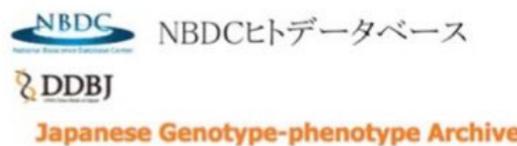
データ利用者のみ

TogoVar <https://togovar.biosciencedbc.jp/>

- 日本人ゲノム多様性統合データベース
 - NBDCヒトデータベースを基に個人特定されない加工データ（頻度情報）を提供
 - 日本や海外で公開されている頻度情報、ゲノム多様性と疾患との関連情報を統合、ワンストップで検索可能に
- 検索対象データベース
 - NBDC ヒトデータベース（125人全エクソーム/183884人マイクロアレイ）
 - iJGVD: Integrated Japanese Genome Variation Database（3554人全ゲノム）
 - 東北メディカルメガバンク
 - HGVD: Human Genetic Variation Database（1208人全ゲノム）
 - 京都大学（長浜コホート）
 - ExAC: Exome Aggregation Consortium（60706人全エクソーム）
 - Broad Institute
 - ClinVar（バリアントの疾患関連知識ベース）
 - PubTator (NCBI)/Colil (DBCLS)（文献情報）

TogoVar - NBDC ヒトデータベースデータの再解析

- NBDC ヒトデータベースに登録された日本人のゲノムデータの一部から集計した大規模なバリアントの頻度情報のデータセットを公開
 - 今後1026人分の全ゲノムデータを追加予定



【演習】TogoVar を使ってみる

1. 「TogoVar」でググってトップページを開きます
2. 遺伝子名、疾患名、rs 番号、ゲノム位置や範囲での検索が可能です
 1. ここでは「Breast-ovarian cancer, familial 2」で検索してみます

The screenshot shows the TogoVar web application interface. On the left, there are several filter panels: 'Data set' (checkboxes for All, JGA NGS, JGA SNP, 3.5KJPN, HGVD, ExAC, ClinVar), 'Alternative allele frequency' (dropdown set to JGA NGS, ex: 0.1), 'Add filter' (button), 'Update' (button), 'Variant type' (checkboxes for All, SNV, Insertion, Deletion, Indel, Substitution), and 'Clinical significance' (checkboxes). The main search area has a search bar containing 'breast' with a dropdown menu listing various diseases. A warning message 'Please select one from the list below to search gene or disease.' is displayed. Below the search bar, the query 'breast-ovarian cancer, familial 2' is highlighted in blue. The results table lists four variants:

ID	Position	Ref	Alt	Symbol	Frequency ⓘ	Consequence	SIFT ⓘ	Polyphen ⓘ	ClinVar conditions
tgv21 rs55998931	1:10492	C	T			intergenic_variant			
tgv27 rs189107123	1:10611	C	G			intergenic_variant			
tgv29	1:10623	T	C			intergenic_variant			
tgv41 rs62635286	1:13116	T	G	DDX11L1		intron_variant non_coding_transcript_variant			

【統合TV】 : TogoVar でヒトゲノムに存在するバリアントに関する情報を調べる

TogoVar の画面構成

TOGO VAR Home Datasets Downloads Terms Contact About Help

+ Add filter

Update

Variant type

All

S SNV (2,787)
Ins Insertion (482)
Del Deletion (1,203)
Ind Indel (65)
Sub Substitution (8)

Clinical significance

All

Not in ClinVar (0)
Pathogenic (2,242)
Likely pathogenic (37)
Uncertain significance (1,106)
Likely benign (106)
Benign (672)
Conflicting interpretations of pathogenicity (450)
Drug response (0)
Association (0)
Risk factor (0)
Protective (0)
Affects (0)

Breast-ovarian cancer, familial 2

Example - Disease: Breast-ovarian cancer, familial 2, Gene: ALDH2, refSNP: rs114202595, TogoVar: tgv421843, Position(GRCh37/hg19): 16:48258198, Region(GRCh37/hg19): 10:73270743-73376976

Search Clear

Display 25 variants per page

Showing 1 to 25 of 4,545 variants related to disease "Breast-ovarian cancer, familial 2". (filtered from 19,537,677 variants)

1 2 3 4 5

ID	Position	Ref	Alt	Symbol	Frequency	Consequence	SIFT	Polyphen	ClinVar conditions
S tgv48208253	13:32889080	A	G	ZAR1L		intron_variant			Breast-ovarian cancer, familial 2
S tgv48208258	13:32889363	G	A	ZAR1L		intron_variant			Breast-ovarian cancer, familial 2
S tgv48208267	13:32889669	C	T	BRCA2		5_prime_UTR_variant			Fanconi anemia Hereditary breast and ovarian cancer Breast-ovarian cancer, familial 2
S tgv48208269	13:32889682	G	A	BRCA2		5_prime_UTR_variant			Breast-ovarian cancer, familial 2
Del tgv48208271	13:32889783	AAG		BRCA2		5_prime_UTR_variant			Fanconi anemia Hereditary breast and ovarian cancer Breast-ovarian cancer, familial 2

各種フィルター

TogoVar ID dbSNP ID

変異の種類

ゲノム上のポジション リファレンスと変異

各データセットでの 頻度情報

遺伝子中の位置

立体構造・機能 への影響スコア

ClinVar による アノテーション

【演習】TogoVar を使ってみる

3. 左側にリストされている項目で検索結果の絞り込みができます

1. データセット、頻度、変異の種類、疾患の関連度など
2. まず「SNV」のみを絞り込んでみます

TOGO VAR Home Datasets Downloads Terms Contact About Help

Data set

- All
- JGA NGS (110)
- JGA SNP (55)
- Tm 3.5KJPN (173)
- HGVD (49)
- ExAC (803)
- ClinVar (2,787)

Alternative allele frequency

JGA NGS ex: 0.1

+ Add filter

Update

Variant type

All

S SNV (2,787)

Ins Insertion

Del Deletion

Ind Indel

Sub Substitution

Clinical significance

All

Not in ClinVar (0)

Breast-ovarian cancer, familial 2

Example - Disease:Breast-ovarian cancer, familial 2, Gene:ALDH2, refSNP:rs114202595, TogoVar:tgv421843, Position(GRCh37/hg19):16:48258198, Region(GRCh37/hg19):10:73270743-73376976

Search Clear

Display 25 variants per page

Showing 1 to 25 of 2,787 variants related to disease "Breast-ovarian cancer, familial 2". (filtered from 19,537,677 variants)

ID	Position	Ref	Alt	Symbol	Frequency	Consequence	SIFT	Polyphen	ClinVar conditions
S tgv48208253	13:32889080	A	G	ZAR1L		intron_variant			B Breast-ovarian cancer, familial 2
S tgv48208258	13:32889363	G	A	ZAR1L		intron_variant			B Breast-ovarian cancer, familial 2
S tgv48208267	13:32889669	C	T	BRCA2		5_prime_UTR_variant			Fanconi anemia Hereditary breast and ovarian cancer : Breast-ovarian cancer, familial 2
S tgv48208269	13:32889682	G	A	BRCA2		5_prime_UTR_variant			B Breast-ovarian cancer, familial 2
S tgv48208273	13:32889792	A	G	BRCA2		5_prime_UTR_variant			Fanconi anemia Hereditary breast and ovarian cancer : Breast-ovarian cancer, familial 2

【演習】TogoVar を使ってみる

3. 左側にリストされている項目で検索結果の絞り込みができます

3. 次に JGA NGS データセットでのアリル頻度が 0.1 (10%) 以下の変異で絞り込んでみます

【演習】TogoVar を使ってみる

4. ID 欄に表示されている tgv ID をクリックすると変異の詳細が表示されます

1. ここでは上から13番目にある「tgv48209632」（全データセットで検出されている変異）をクリックしてみます

TOGO VAR Home Datasets Downloads Terms Contact About Help

Variant Information Frequency Genomic context Transcripts Publications

Variant report

ID: tgv48209632

Variant	Variant type	ClinVar
Chromosome 13	S SNV	B Benign
Position 32906480		
Ref allele A		
Alt allele C		

JGA-NGS	JGA-SNP	3.5KJPN	HGVD	ExAC
25 / 250 0.100	38198 / 296304 0.129	897 / 7100 0.126	321 / 2410 0.133	6233 / 120366 0.052

TogoVar ID: tgv48209632
refSNP ID: rs766173

【演習】TogoVar を使ってみる

5. 下にスクロールしていくと詳細情報が表示されます

1. 各データセットでの詳細な頻度情報（データがあれば Genotype 情報も）

TOGOVAR Home Datasets Downloads Terms Contact About Help

Variant Information Frequency Genomic context Transcripts Publications

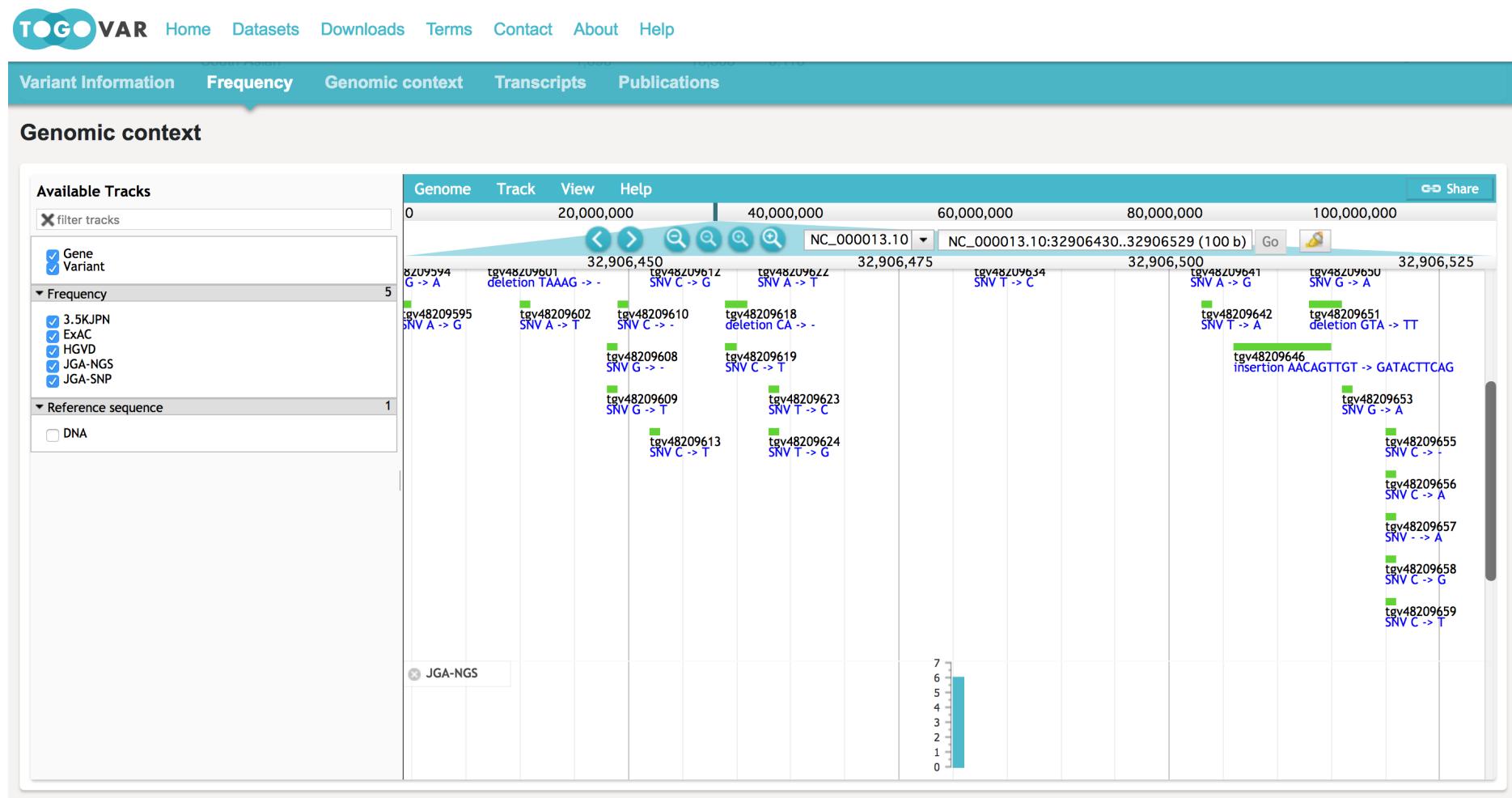
Frequency

Dataset	Population	Allele count		Alt allele freq	Genotype count			Filter status
		Alt	Total		Alt / Alt	Alt / Ref	Ref / Ref	
JGA-NGS	Japanese	25	250	0.1				PASS
JGA-SNP	Japanese	38,198	296,304	0.129	2,551	33,096	112,505	N/A
3.5 KJPN	Japanese	897	7,100	0.126				PASS
HGVD	Japanese	321	2,410	0.133				PASS
ExAC	Total	6,233	120,366	0.052				PASS
	African	243	10,144	0.024				-
	East Asian	866	8,640	0.1				-
	European (Finnish)	94	6,608	0.014				-
	European (Non-Finnish)	2,317	66,232	0.035				-
	Latino	769	11,528	0.067				-
	Other	54	906	0.06				-
	South Asian	1,890	16,308	0.116				-

【演習】TogoVar を使ってみる

5. 下にスクロールしていくと詳細情報が表示されます

2. 周辺の変異情報（ゲノムブラウザ）



7. ヒトに関する情報を検索するツール



NBDC Beacon <https://humandbs.biosciencedbc.jp/beacon/api>

- Beaconとは興味のある変異がデータセット中にあるかどうかを知るためのサービスです
 - 例：1番染色体の12345番目のTがAであるデータセットはDB中にあるか？
- NBDC ヒトデータベースのオープンデータを使って試験公開しています
- 将来的には制限アクセスデータに対して検索できるようになる予定？です

【実習】NBDC Beacon を使ってみる

1. 「NBDC ヒトデータベース」でググって、トップページを開く
2. Example にあるリンクをクリックする
 1. GRCh37で12番染色体の112241766番目の塩基が 'A' のデータがデータセット中にあるか？

The screenshot shows the NBDC Human Database Beacon homepage. At the top, there is a navigation bar with links for Home, Data Utilization, Data Provision, Guidelines, External Server, NBDC Human Database Review Committee, Research成果, Contact, and FAQ. There is also an English language selection and a search bar with a site search button.

A red warning banner at the top states: "遺伝研スパコン停止に伴い、『JGA Submission/Download Tool』および『JGA Meta Viewer』のJGA登録・検索サービスが一時停止します。停止期間:2018/11/16 13:00 - 2018/11/21 24:00(予定)" (Due to the temporary shutdown of the GIGA supercomputer, the registration and search services for "JGA Submission/Download Tool" and "JGA Meta Viewer" will be temporarily unavailable. Stop date: November 16, 2018, 13:00 - November 21, 2018, 24:00 (estimated)).

The main content area has two boxes: "NBDCヒトデータベースについて" (Information about NBDC Human Database) and "新着情報" (New Information). The "NBDCヒトデータベースについて" box contains text about the database's purpose and how it supports research. The "新着情報" box lists recent updates from November 5, 2018, and November 29, 2018, from the Osaka University Frontier Research Center for Immunology and the Faculty of Medicine, Toyonaka Campus.

At the bottom, there is a search bar with the placeholder "Search NBDC Human Database Beacon for Alternative Alleles [API help]" and a note that the database is a member of the GA4GH Beacon Network. A specific search query "GRCh37 e.g. 12:112241766 A" is shown, with the result "ALDH2 Variant (GRCh37, '12:112241766 A')".

【実習】NBDC Beacon を使ってみる

3. hum0013, hum0015, hum0029 には変異を含むデータが存在し、
hum0014 には存在しないことがわかる

NBDC Human Database

Home Data Use Data Submission Guidelines Data Access Committee

Search NBDC Human Database Beacon for Alternative Alleles [API help]

A NBDC Human Database Beacon is a member of [GA4GH Beacon Network](#).

GRCh37 12:112241766 A Search Example: ALDH2 Variant (GRCh37, '12:112241766 A')

Beacon Query

- Reference: GRCh37
- Chromosome: 12
- Position: 112241766
- Allele: A

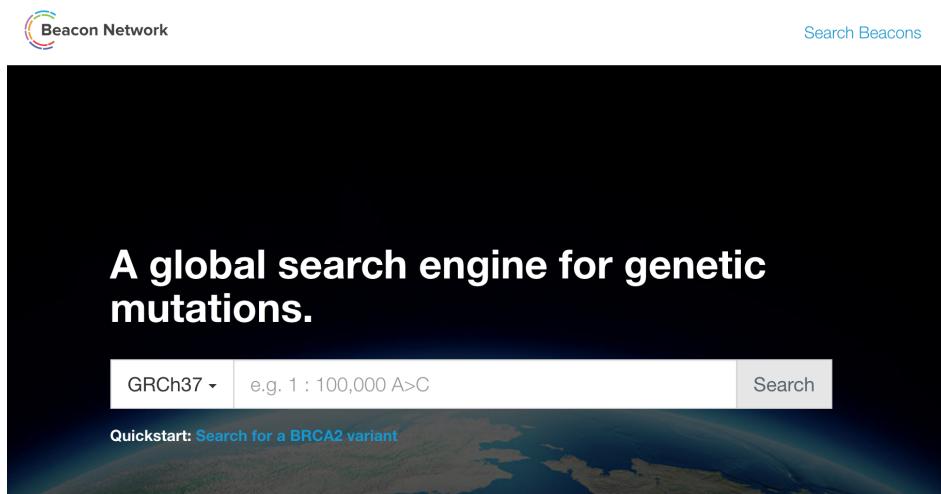
Response by Dataset

Dataset	Found	Link to Download Site	Error Messages
hum0013.freq.v1	Found	http://humandbs.biosciencedbc.jp/en/hum0013-v1	
hum0014.v3.T2DM.1.v1	Not Found	http://humandbs.biosciencedbc.jp/en/hum0014-v3	
hum0015.freq.v1	Found	http://humandbs.biosciencedbc.jp/en/hum0015-v1	
hum0029.freq.v1	Found	http://humandbs.biosciencedbc.jp/en/hum0029-v1	

Site Policies | Sitemap | Contact Us | Supported Browsers

【参考】Beacon Network <http://beacon-network.org/>

- Beacon Network は世界中の Beacon サーバーを横断的に検索するシステムです
 - 現在 38 機関、66 サーバが検索対象



The screenshot shows the search results page for the query "GRCh37 ▾ 13 : 32936732 G > C". The search bar at the top right has the same query. The results are listed in a table format:

Response	All	None
<input checked="" type="checkbox"/> Found	7	
<input checked="" type="checkbox"/> Not Found	37	
<input type="checkbox"/> Not Applicable	22	

Below the table, there is a section titled "Organization" with a "All" button and a "None" button. Under "All", several organizations are listed with checkboxes:

- AMPLab, UC Ber...
- Australian Genomi...
- Belgian Medical G...
- BGI
- Bioinformatics Ar...
- BioReference Lab...
- Brazilian Initiative o...
- BRCA Exchange

The results table lists six entries:

	AMPLab - 1000 Genomes Project	Not Found
	BioReference	Not Found
	BRCA Exchange	Show Metadata <input checked="" type="button"/> Found
	Cafe CardioKit	Not Found

PubCaseFinder <https://pubcasefinder.dbcls.jp/>

- 希少疾患・症例を検索できる希少疾患診断支援システム
- 患者の症状をキーワードとして、疾患名および症例報告を関連性の高い順にランキング提示する
 - 希少疾患DBのOrphanet (<https://www.orpha.net>) : 4000件の疾患
 - 遺伝性疾患DBのOMIM (<https://www.omim.org>) : 7000件の疾患
 - PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) : 30万件の症例報告
- 特徴
 - Human Phenotype Ontology (HPO) による症状の標準化
 - テキストマイニングによる症状と疾患の自動抽出
 - 日本語の症例報告にも対応予定

【実習】PubCaseFinder を使ってみる

1. 「PubCaseFinder」でググって、トップページを開く
2. 入力ボックスの下にある例をクリックする
 1. 英語でも日本語でも入力できます
 2. 一部の文字を入力すると候補が出るのでそこから選択できます
 3. 入力した症状をクリックすると、症状の詳細な説明、上位概念・下位概念の症状が表示されるので、より適当な症状を選択することができます



【実習】PubCaseFinder を使ってみる

3. 「疾患を検索」をクリックして検索します

1. 関連度順に疾患名がリストされます
2. 原因遺伝子の候補がある場合さらに絞り込むことができます
3. 疾患に関する画像（Google検索）や症例報告を見ることができます

PubCaseFinder

患者の 徴候・症状 を入力 + Upload File (HPO ID):

HP:0001009 毛細血管拡張 ✕ HP:0001249 知的障害 ✕ HP:0001250 発作 ✕ HP:0002072 舞踏病 ✕ HP:0002315 頭痛 ✕

疾患を絞り込む + Upload File (Entrez Gene ID):

結果の要約をダウンロード 疾患を検索 クリア

希少疾患 (Orphanet) 4,066 件 遺伝性疾患 (OMIM) 6,969 件

合計: 4,066 件 1 2 3 … 407 > 10 (表示件数)

結果をダウンロード

順位 (類似度)	疾患名 (疾患ID)	対応する徴候・症状	疾患原因遺伝子
1 (100.0%)	Moyamoya disease (ORDO:2573)	偏頭痛 毛細血管拡張 発作 知的障害 舞踏病	ACTA2 RNF213

Moyamoya disease (MMD) is a rare intracranial arteriopathy involving progressive stenosis of the cerebral vasculature located at the base of the brain causing transient ischemic attacks or strokes.
>> 翻訳 (Google)

画像検索 (Google) 症例報告検索

【おまけ】DDBJのスパコンについて <https://sc.ddbj.nig.ac.jp/ja>

- 日本のアカデミア機関に所属していれば基本的に無料で利用できます
 - 年度末の利用報告義務あり
- 大量のデータを扱う場合（研究グループで30TBを超える場合）は有料
- 個人ゲノム解析環境を利用する場合も有料
- 基本的にコマンドラインからログインして利用する
 - 一部、ウェブから利用できるものもある（MiGAP、DDBJ Read Annotation Pipeline）
- 無料なので結構混んでいるらしいです
- 夏場はたまに停電で停止することがあります