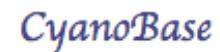

5. ゲノムデータベース

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/>

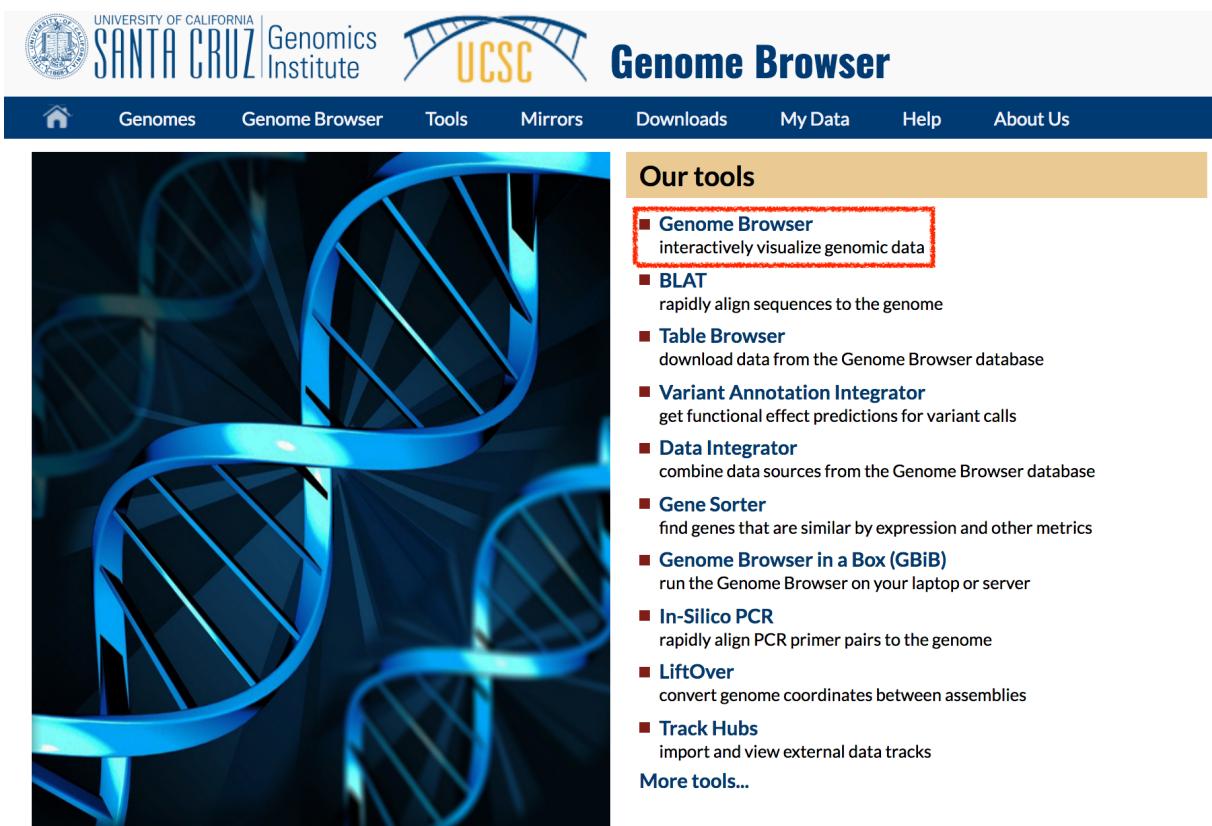


ゲノムブラウザとは？

- 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。



【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

You might want to navigate to your nearest mirror - genome-asia.ucsc.edu

- User settings (sessions and custom tracks) will differ between sites. [Read more.](#)
- Take me to genome-asia.ucsc.edu
- Let me stay here genome.ucsc.edu

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）

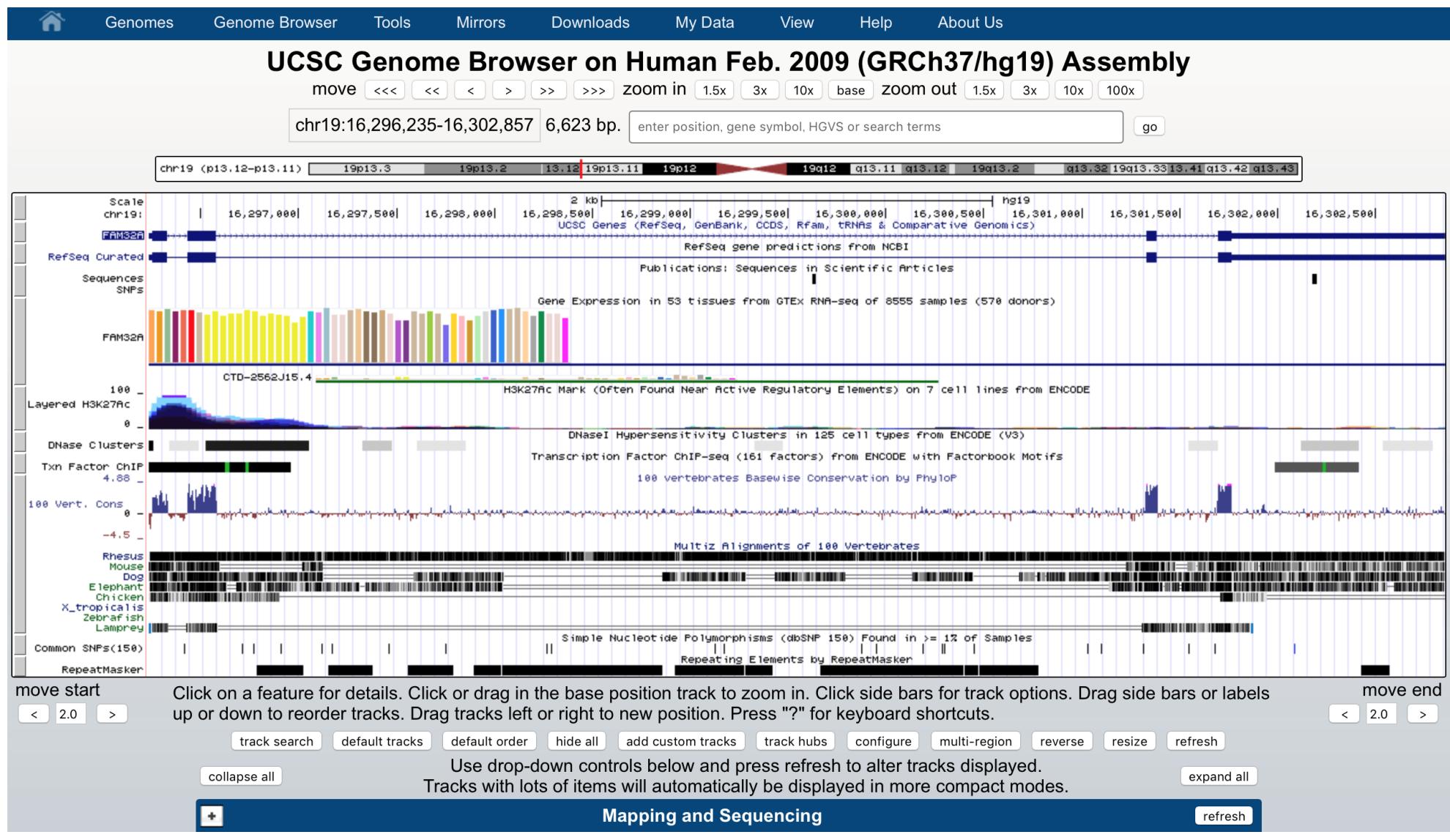
Human Assembly
Feb. 2009 (GRCh37/hg19)

Position/Search Term
FAM32A

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

【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 'Regulation' track settings in the UCSC Genome Browser. The 'Regulation' tab is highlighted with a red border. A dropdown menu for 'ENC TF Binding...' is open, showing 'hide' (selected) and 'show'. A red box highlights this dropdown. A 'refresh' button is located to the right of the dropdown.

Expression

- GTEx Gene
- GTEx Transcript
- Affy Exon Array
- Affy GNF1H
- Affy RNA Loc
- Affy U95
- Affy U133
- Affy U133Plus2
- Allen Brain
- Burge RNA-seq
- CSHL Small RNA-seq
- ENC Exon Array...
- ENC ProtGeno...
- ENC RNA-seq...
- GIS RNA PET
- GNF Atlas 2
- GWIPS-viz Riboseq
- Illumina WG-6
- PeptideAtlas
- gPCR Primers
- RIKEN CAGE Loc
- Sestan Brain

Regulation

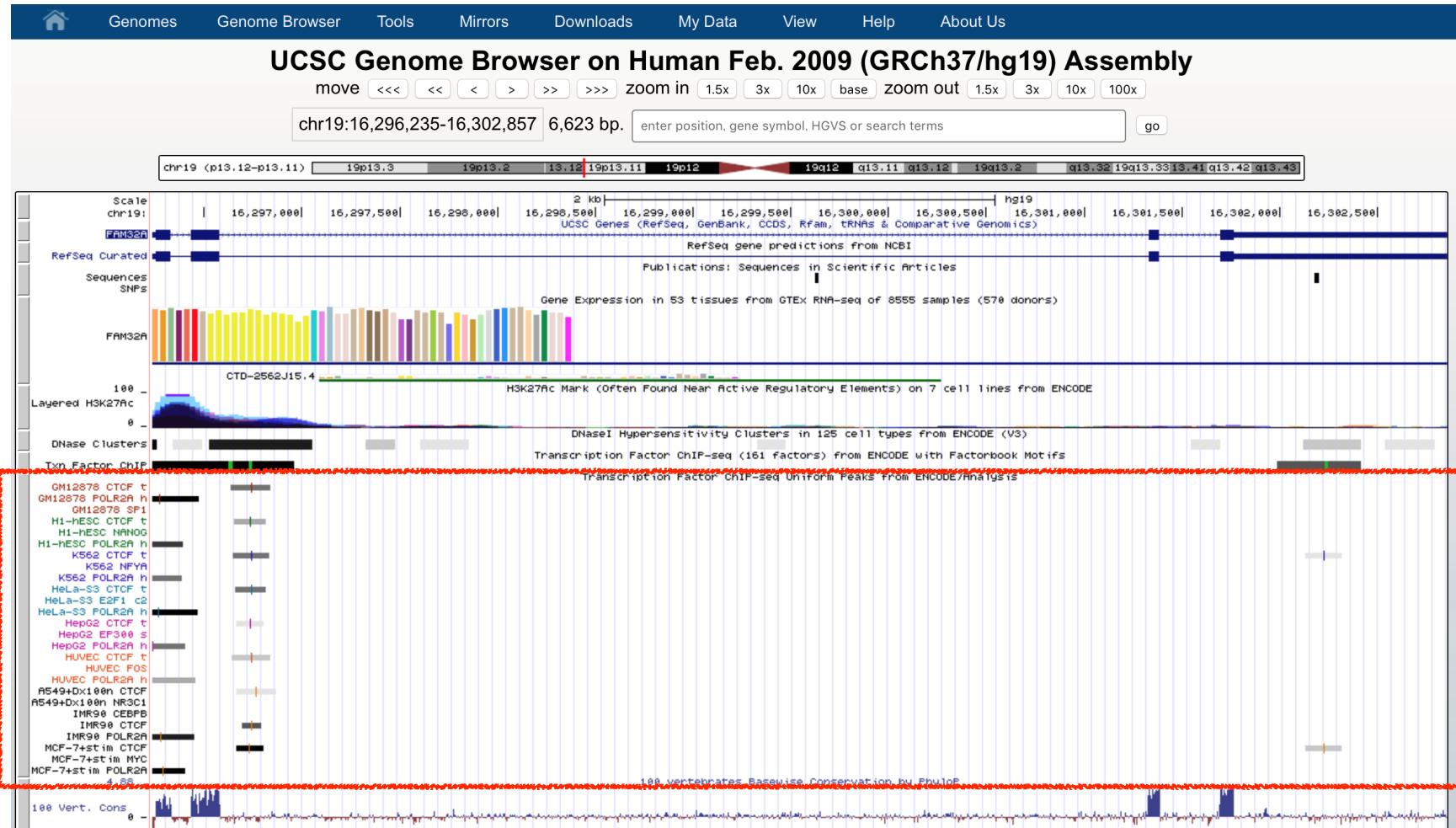
- ENCODE Regulation...
- New GeneHancer
- GTEx Combined eQTL
- GTEx Tissue eQTL
- CD34 DnaseI
- CpG Islands...
- ENC Chromatin...
- ENC DNA Methyl...
- ENC DNase/FAIRE...
- ENC Histone...
- ENC RNA Binding...
- ENC TF Binding...
- FSU Repli-chip
- Genome Segments
- NKI Nuc Lamina...
- ORegAnno
- Stanf Nucleosome
- SwitchGear
- SwitchGear TSS
- TFBS Conserved
- TS miRNA sites
- UCSF Brain Methyl
- UMMS Brain Hist
- UW Repli-seq
- Vista Enhancers

Comparative Genomics

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

【実習】UCSC ゲノムブラウザを使ってみる - 表示項目の追加

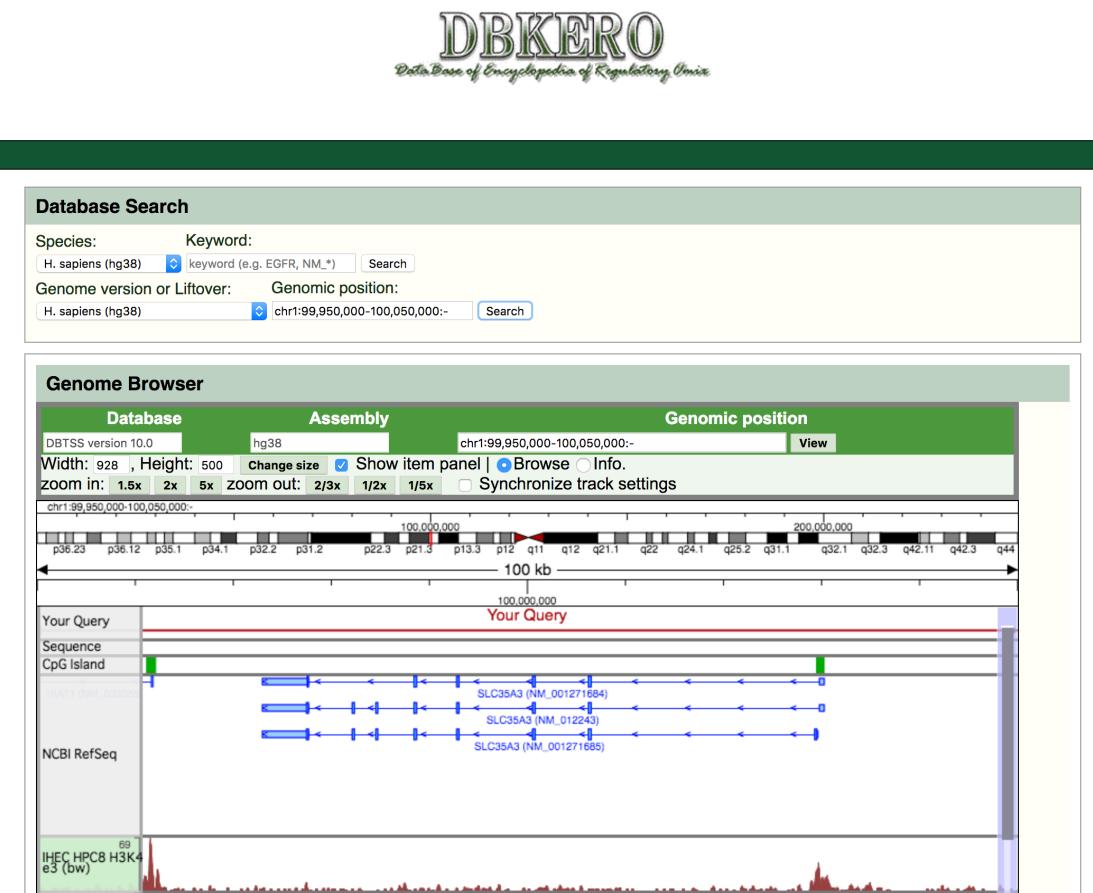
7. 転写因子結合サイトの情報が追加される



17.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/>

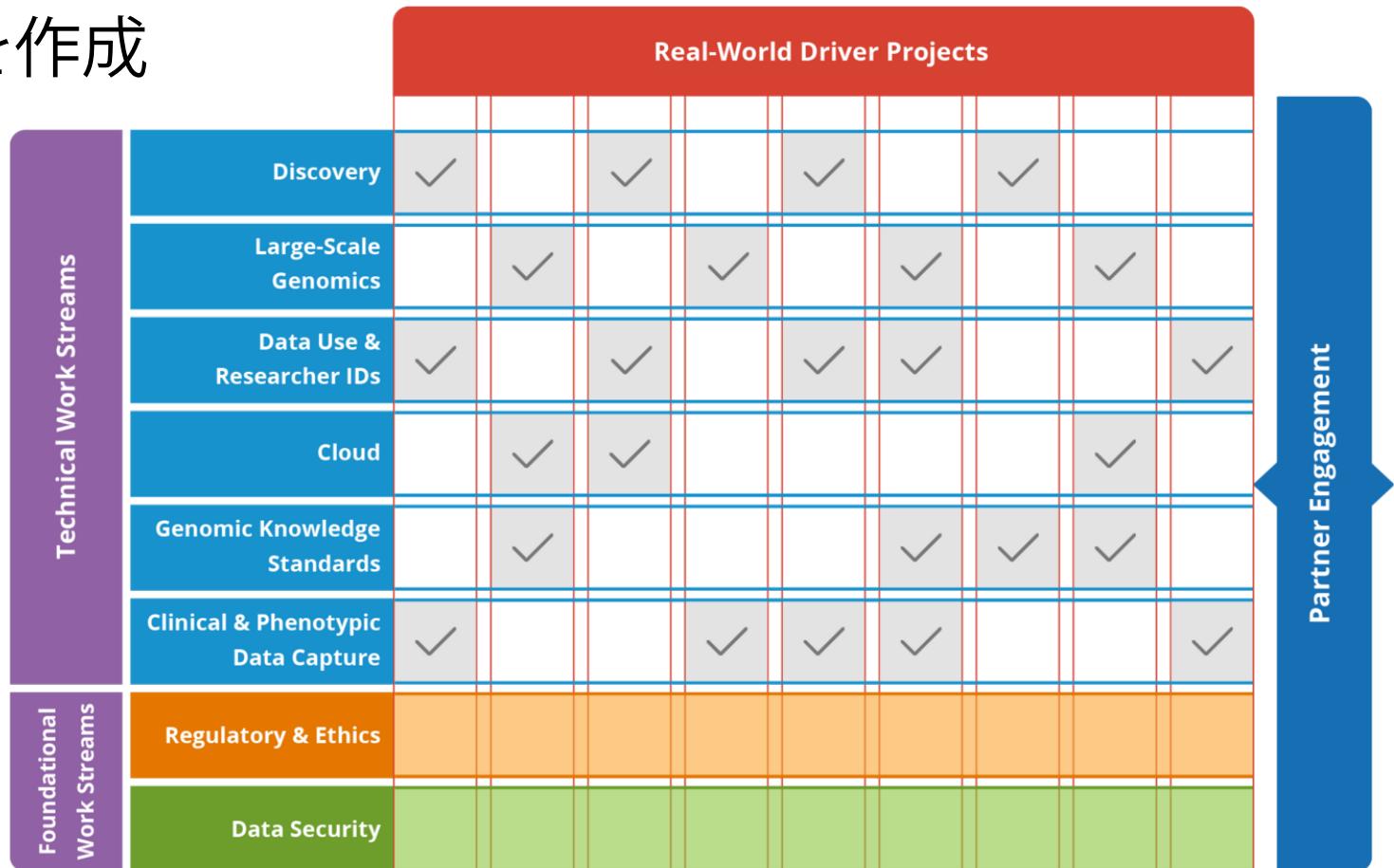
- DataBase of Transcriptional Start Site/DataBase of Kashiwa Encyclopedia for human genome mutation in Regulatory region and their Omics contexts
- ヒト細胞の各種オミクスデータを集積したデータベース
 - 全ゲノム (WGS)
 - トランスクリプトーム (RNA-seq)
 - エピゲノム (BS-seq, ChIP-seq)
 - 転写開始地点 (TSS-seq)
 - シングルセルデータ
 - ロングリードデータ



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

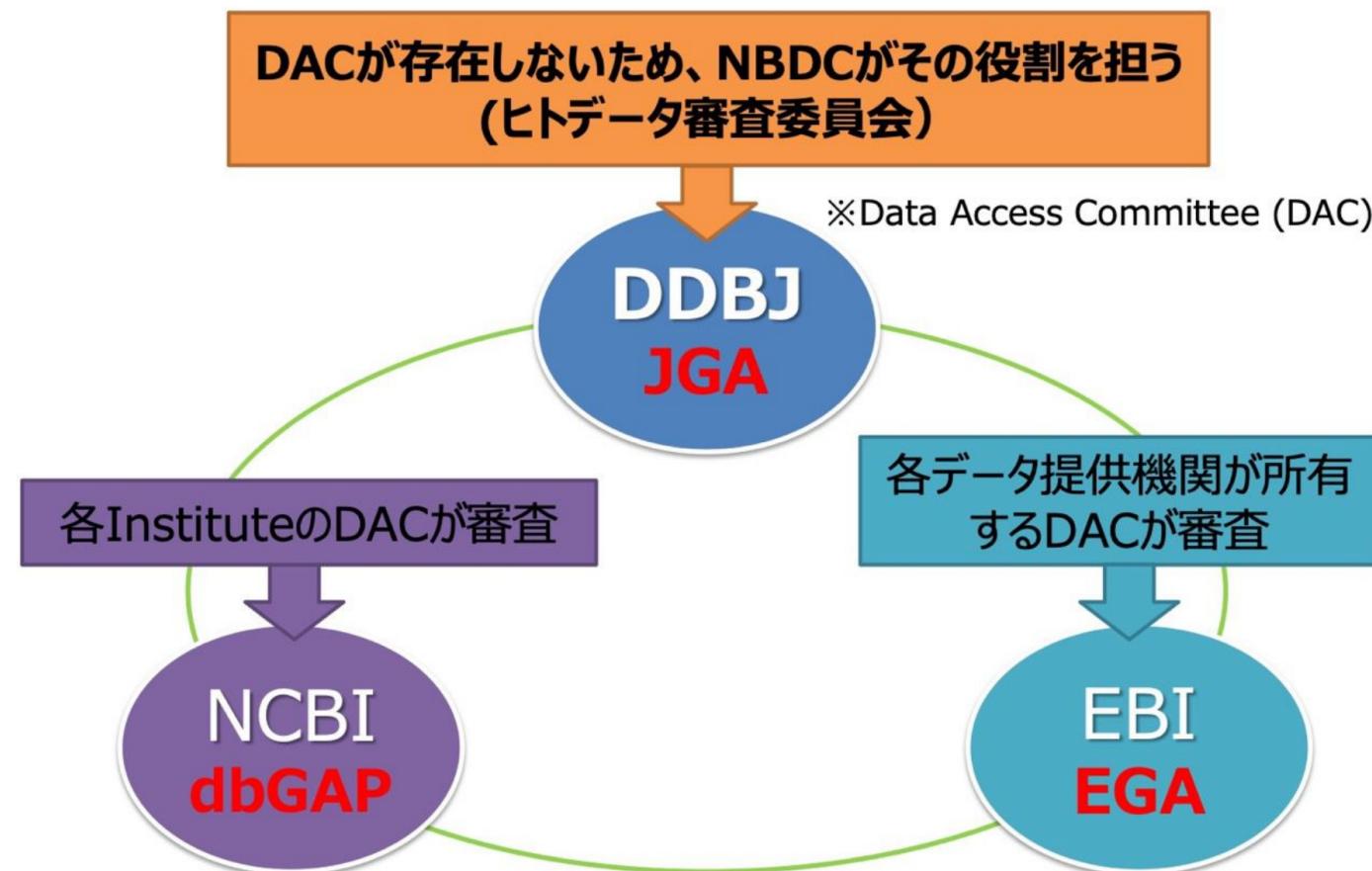


NIH NATIONAL CANCER INSTITUTE
Genomic Data Commons

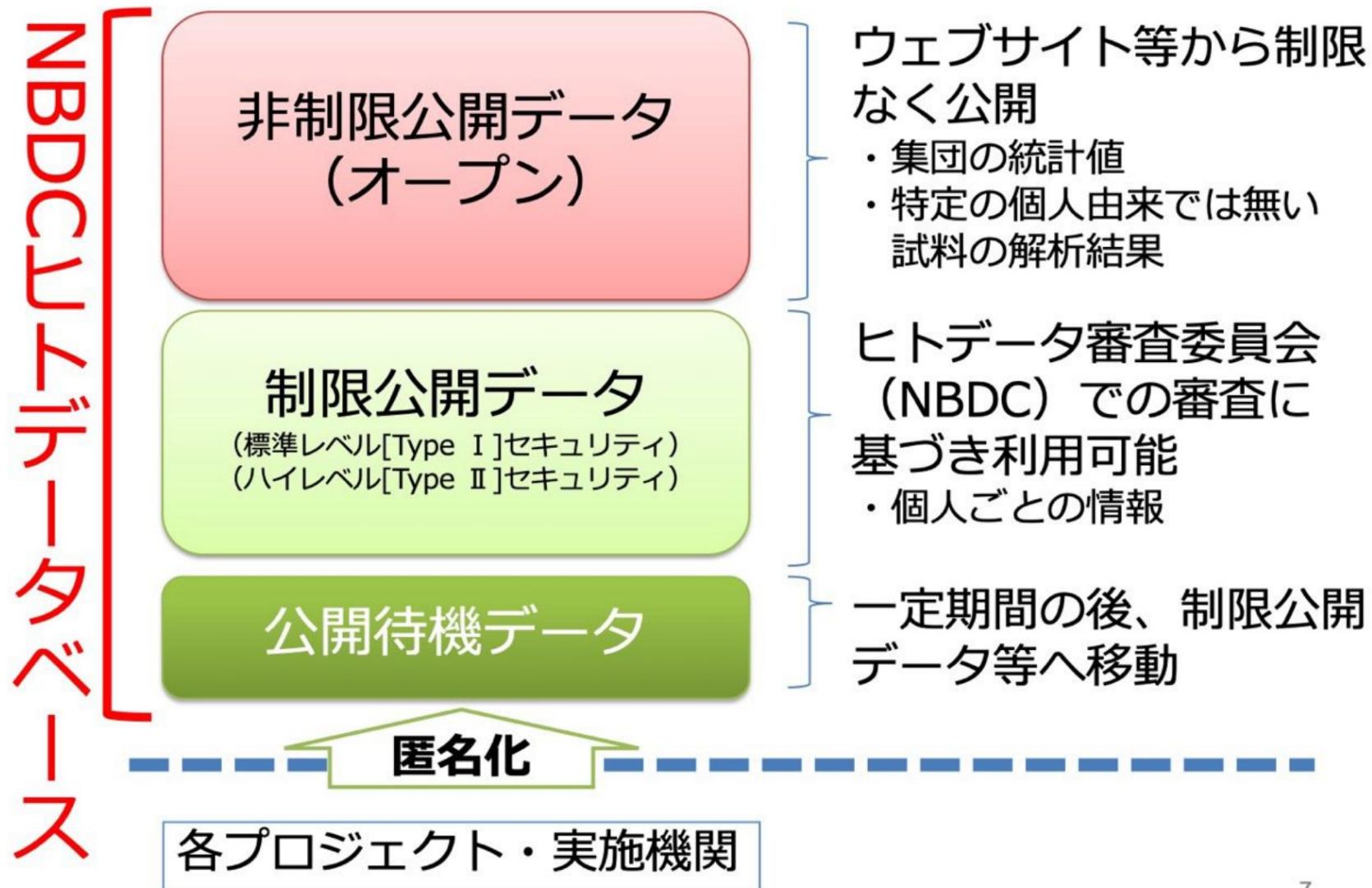


制限公開データベース

- 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)



■ 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. At the top, there is a navigation bar with links for Home, Data Utilization, Data Submission (highlighted with a red box), Guidelines, External Server, NBDC Hit Database Review Committee, Results Publication, Contact, and FAQ. A search bar is also present. A red warning box at the top of the main content area 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 Genetic Research Supercomputer, the registration and search services for "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 is shown, containing information about the submission process and a list of required documents.

データの提供

2017年5月30日より改正個人情報保護法および改正倫理指針が施行されます。データ提供を行う際には、新指針に沿った手続きを実施する必要がありますのでご留意ください。詳細はFAQをご参照ください。

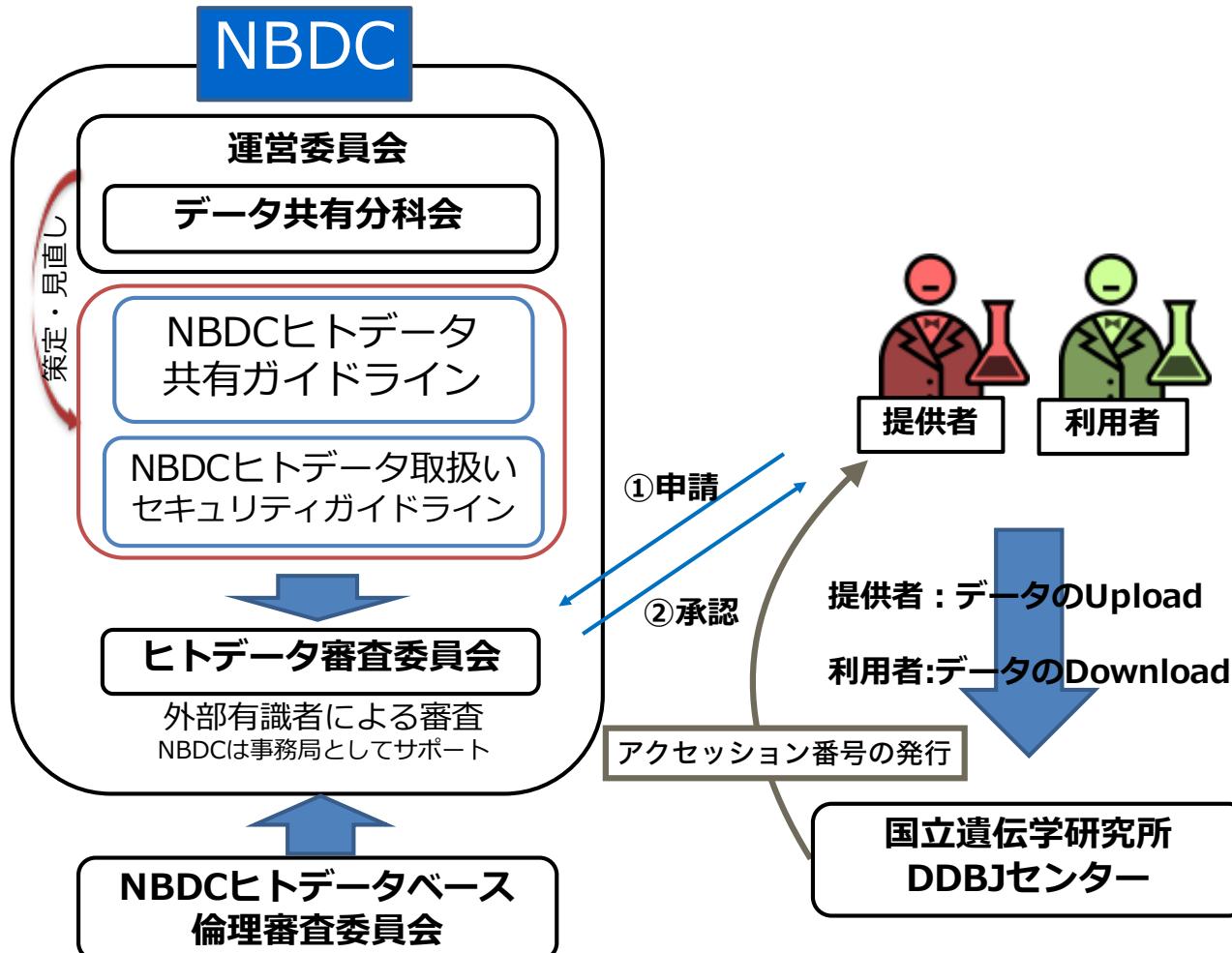
NBDCヒトデータベースへのヒトに関するデータ(遺伝情報、臨床情報、画像情報など)の提供手続きは以下の通りです。

1. NBDCヒトデータ共有ガイドラインを熟読して下さい。
2. 以下の情報を準備して下さい。
 - 提供するデータについての情報:研究内容の概要(目的、方法、対象、発表論文など)、データの種類(名称・量、アクセス制限レベルの分類に関する情報、公開可能日など)
 - 研究代表者の氏名、所属情報、連絡先
 - 所属機関の長の氏名、職名、メールアドレス *所属機関の長とは、倫理審査委員会によって承認された研究計画の実施を許可する者を指します。
 - 研究計画書(倫理審査申請書)(PDF形式 5MB以内)
 - 同意文書及び説明文書のフォーム(PDF形式 5MB以内) *実際に署名されたものではなくフォーム(書式)をご提出ください。
 - 研究内容についての倫理審査の承認通知書(PDF形式 5MB以内)

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

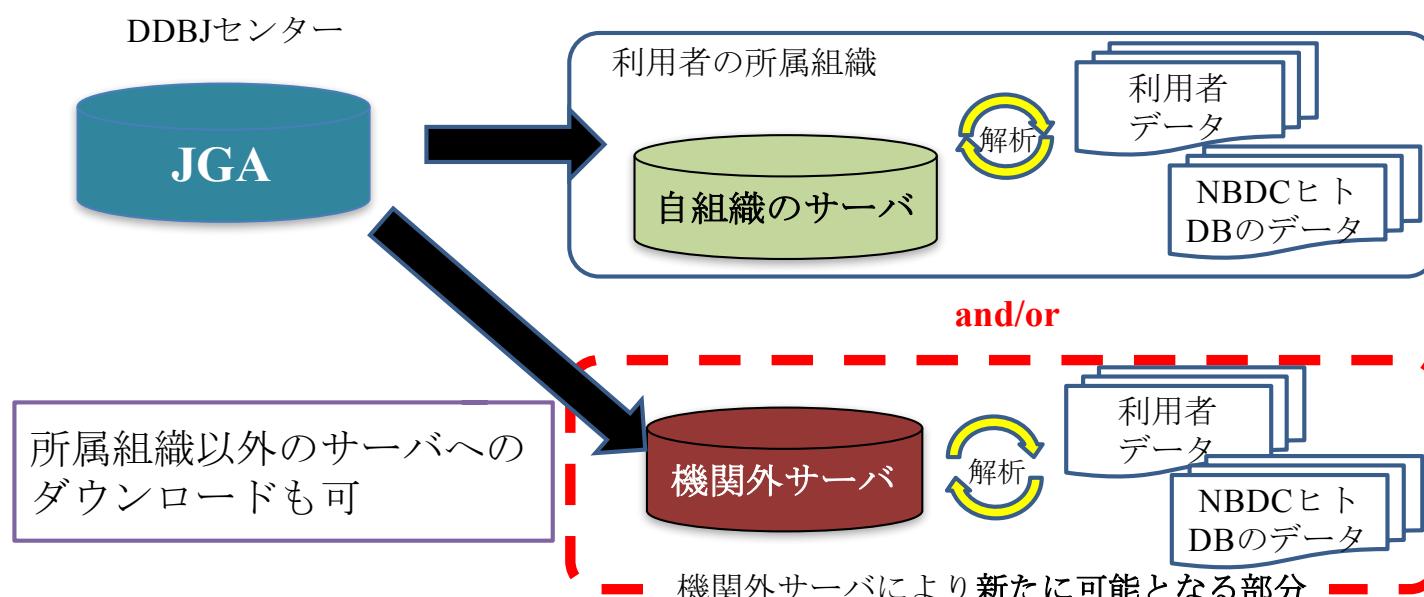
■ データ共有の仕組み

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）

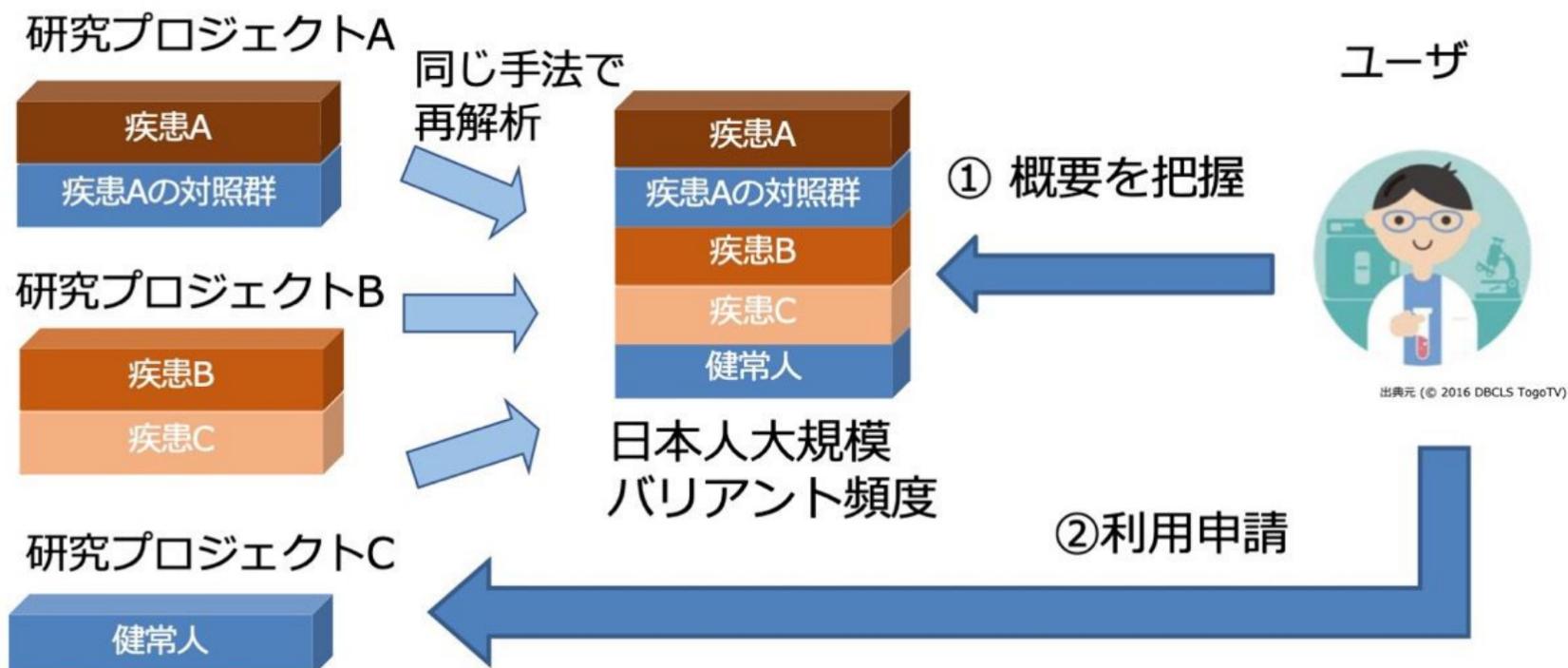
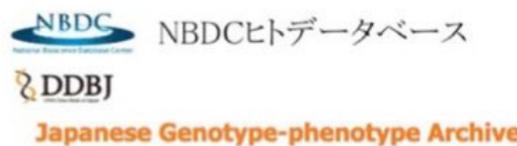
データ利用者のみ

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

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's a navigation bar with links for Home, Data Utilization, Data Submission, Guidelines, External Servers, NBDC Human Database Review Committee, Research Outputs, Contact, and FAQ. There's also an English language switch and a search bar.

A red box highlights the search bar at the bottom, which contains the text "GRCh37" in a dropdown menu, "e.g. 12:112241766 A" in the input field, and a "Search" button. To the right of the input field, it says "Example: ALDH2 Variant (GRCh37, '12:112241766 A)".

The main content area has two sections: "NBDCヒトデータベースについて" (Information about the NBDC Human Database) and "新着情報" (Recent News). The "Information" section contains text about the database's purpose and how to use it. The "Recent News" section lists two entries:

- 2018/11/05 大阪大学免疫学フロンティア研究センター 実験免疫学 からの制限公開データ(Type I)を公開しました(hum0125)
- 2018/10/29 福井大学医学部 耳鼻咽喉科・頭頸部外科学 からの制限公開データ(Type I)を公開しました(hum0128)

【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

The screenshot shows the NBDC Human Database Beacon search interface. At the top, there is a logo for 'NBDC National Bioscience Database Center' and a navigation bar with links to Home, Data Use, Data Submission, Guidelines, and Data Access Committee. Below the navigation bar, a search bar displays the query 'GRCh37 12:112241766 A'. To the right of the search bar, it says 'Example: ALDH2 Variant (GRCh37, '12:112241766 A')'. The main content area is titled 'Beacon Query' and lists the following parameters:

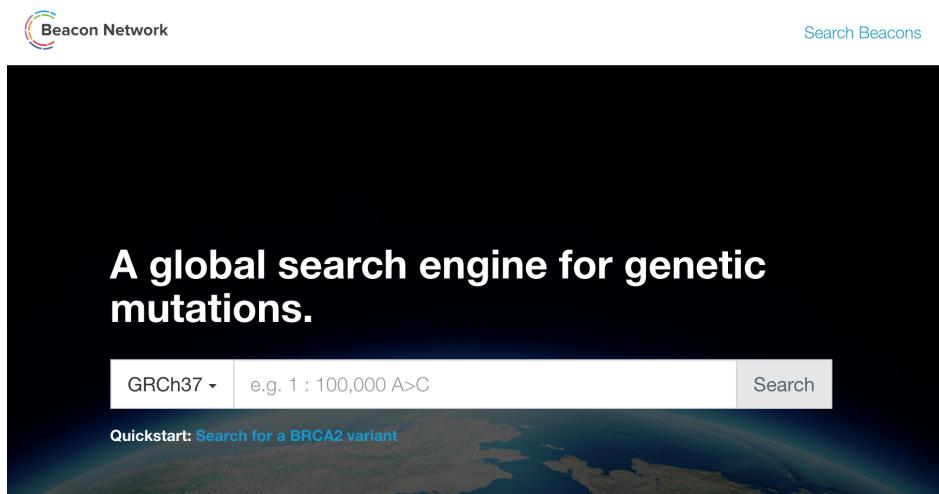
- Reference: GRCh37
- Chromosome: 12
- Position: 112241766
- Allele: A

Below this, a section titled 'Response by Dataset' shows a table of results:

Dataset	Found	Link to Download Site	Error Messages
hum0013.v1.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.v1.freq.v1	Found	http://humandbs.biosciencedbc.jp/en/hum0015-v1	
hum0029.v1.freq.v1	Found	http://humandbs.biosciencedbc.jp/en/hum0029-v1	

【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 of the Beacon Network. At the top right is the "Search Beacons" button. Below it is a search bar with "GRCh37" and "13 : 32936732 G > C" entered. To the right of the search bar is a "Search" button. The results are listed in a table format. Each row includes a logo, the organization name, and a status indicator. The first row is for "AMPLab - 1000 Genomes Project" with a "Not Found" status. The second row is for "BioReference LABORATORIES" with a "Not Found" status. The third row is for "BRCA Exchange" with a "Found" status. The fourth row is for "Cafe CardioKit" with a "Not Found" status.

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

Organization	All	None
<input checked="" type="checkbox"/> AMPLab, UC Ber...		
<input checked="" type="checkbox"/> Australian Genomi...		
<input checked="" type="checkbox"/> Belgian Medical G...		
<input checked="" type="checkbox"/> BGI		
<input checked="" type="checkbox"/> Bioinformatics Ar...		
<input checked="" type="checkbox"/> BioReference Lab...		
<input checked="" type="checkbox"/> Brazilian Initiative o...		
<input checked="" type="checkbox"/> BRCA Exchange		

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)

順位 疾患名 (疾患ID)

1 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