

多型データベースの紹介 (TogoVarなど)

令和元年6月5日@AJACS徳島

国立研究開発法人 科学技術振興機構 (JST)
バイオサイエンスデータベースセンター (NBDC)
豊岡 理人

多型/バリアントとは

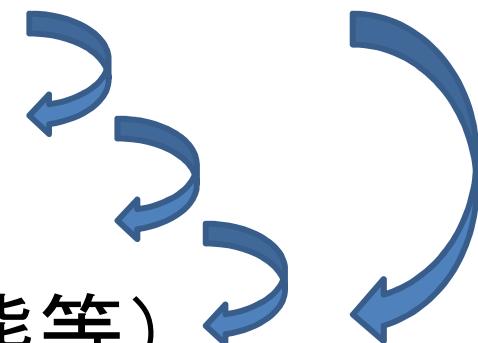
- ・ 多型(polymorphism) / バリアント (variant)
 - ヒトゲノムにおいて観察されるreferenceと異なる塩基配列のこと
- ・ ゲノムと形質との関連を明らかにするうえで重要な基礎データ

ゲノム、エピゲノム

トランスクリプトーム

プロテオーム

形質(疾患、薬剤副反応、形態等)

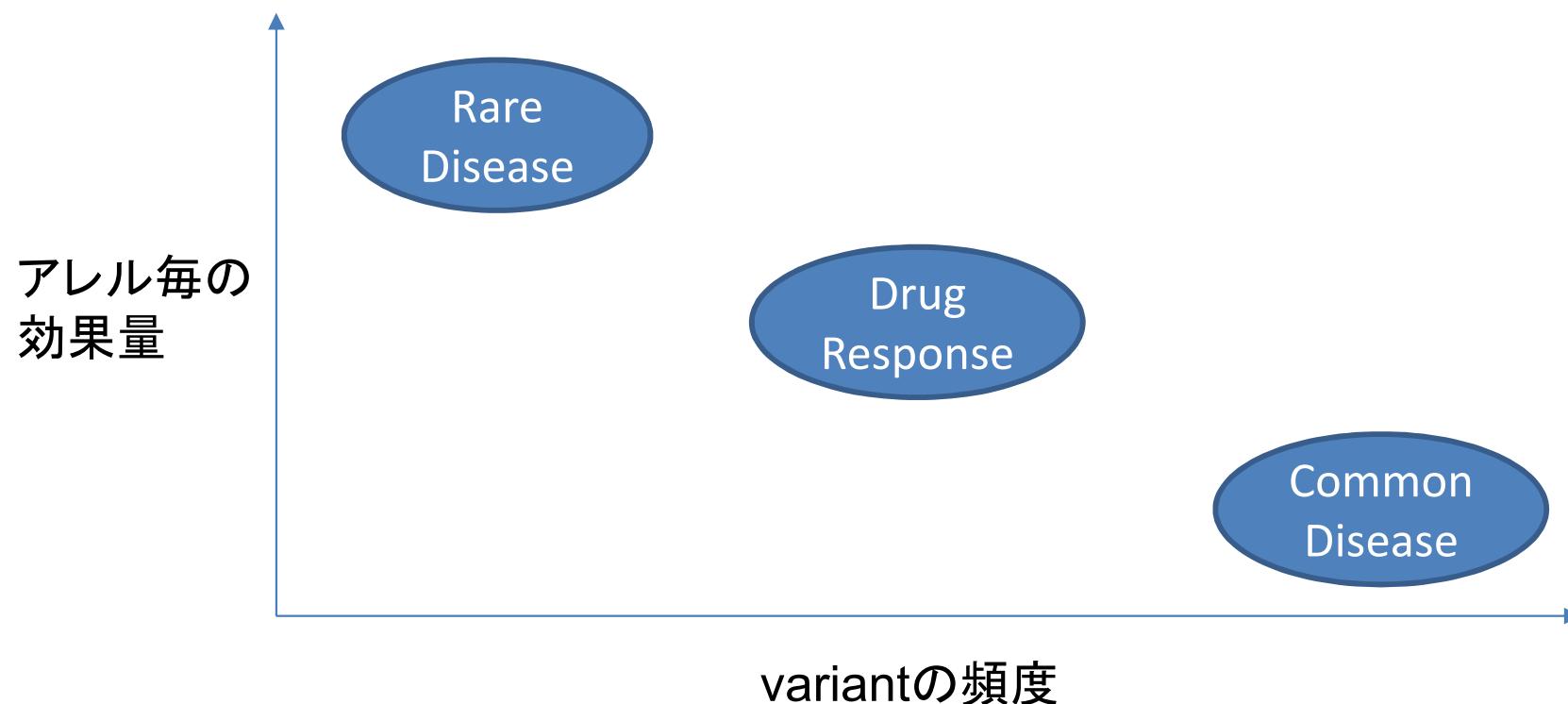


多型/バリアントの種類

- ・ 一塩基多型 (SNP: Single Nucleotide Polymorphism)
- ・ コピー数多型 (CNV: Copy Number Variation)
- ・ タンデムリピート、レトロトランスポゾン
- ・ 構造多型 (SV: Structural Variant)

バリエントの頻度と形質の関係

- Common Disease Common Variant仮説
- Rare Disease Rare Variant仮説



疾患と関連する遺伝子同定の研究概要

- ありふれた疾患
RareのVariantはPrimaryの探索から外されることが多い。
 - 稀少疾患
CommonのVariantはPrimaryの探索から外されることが多い。
 - Trio解析
 - 発症者 + 未発症の父母の3人で、発症者におけるde novoな variantを抽出
- 一般集団のvariant頻度情報は基礎的情報として有用

進化するプラットフォーム

- ・サンガーシークエンサー
 - 配列情報
- ・SNP Chip
 - 60-90万箇所のgenotype情報
- ・次世代シークエンサー
 - 全exome領域や全genome領域等のgenotype情報、haplotype情報
- ・PacBio
 - 特定の領域のgenotype情報、haplotype情報



多型データベースの紹介 1:dbSNP

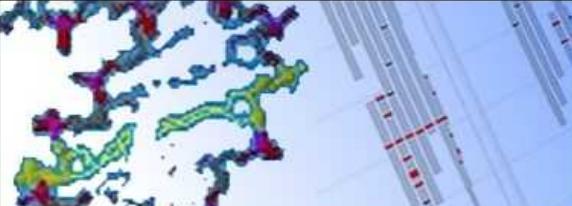
dbSNP

- 運営元: NCBI
- URL: <https://www.ncbi.nlm.nih.gov/snp/>
- 特徴
 - 様々の研究において検知された多型/バリアントの情報を受け付け、ss番号を付与。その後、ss番号をまとめてrs番号を付与
 - ss番号 → rs番号
 - 50bp以下の多型を対象に情報を収集
 - 50bpよりも大きい多型 dbVarに収載
 - 多型の有無の確認

dbSNPの画面説明、操作

NCBI Resources How To Sign in to NCBI

dbSNP SNP Advanced Search Help

 dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

Getting Started

- [Overview of dbSNP](#)
- [About Reference SNP \(rs\)](#)
- [Factsheet](#)
- [Entrez Updates](#)

Submission

- [Clinically Associated Human Variations](#)
- [All Other Variations](#)
- [Hold Until Published \(HUP\) Policies](#)
- [Submission Search](#)

Access Data

- [Variation Services API](#)
- [FTP Download](#)
- [Tutorials on GitHub](#)

April 2, 2019: Starting next week dbSNP will release the new Entrez updates that will include the latest dbSNP build 152 data for user testing. The updates are listed [here](#).

Notifications:

- As previously announced on (April 19, 2018), dbSNP Entrez currently only houses human data. In addition, the Entrez and eUtils report formats RS docsum, XML, ASN.1, FASTA, and FLAT format will no longer be available. dbSNP Entrez eUtils will transition to a new single compact eSummary format, the spec and sample data are [here](#). Please contact snp-admin@ncbi.nlm.nih.gov if you have any comments or concerns.

dbSNP News and Announcements

[RSS Feed](#)  dbSNP News and Announcements(RSS) Feed

[Email List](#)

Related Sites

[Variation Portal](#)
[Variation Tools](#)

Variation Databases

[dbVar](#)
[dbGaP](#)
[ClinVar](#)
[GTR](#)

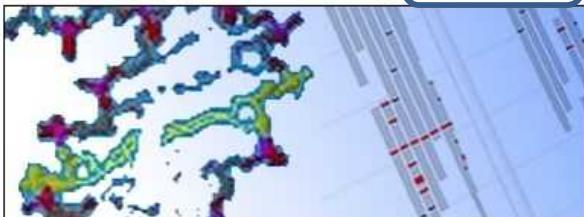
検索対象語を入力だけ→何を対象に調べているのか？

9

dbSNPにおける拡張検索機能

NCBI Resources How To Sign in to NCBI

dbSNP SNP Advanced Search Help



dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

Getting Started	Submission	Access Data
Overview of dbSNP	Clinically Associated Human Variations	Variation Services API
About Reference SNP (rs)	All Other Variations	FTP Download
Factsheet	Hold Until Published (HUP) Policies	Tutorials on GitHub
Entrez Updates	Submission Search	

April 2, 2019: Starting next week dbSNP will release the new Entrez updates that will include the latest dbSNP build 152 data for user testing. The updates are listed [here](#).

Notifications:

- As previously announced on (April 19, 2018), dbSNP Entrez currently only houses human data. In addition, the Entrez and eUtils report formats RS docsum, XML, ASN.1, FASTA, and FLAT format will no longer be available. dbSNP Entrez eUtils will transition to a new single compact eSummary format, the spec and sample data are [here](#). Please contact snp-admin@ncbi.nlm.nih.gov if you have any comments or concerns.

dbSNP News and Announcements

[RSS Feed](#)  dbSNP News and Announcements(RSS) Feed

[Email List](#)

Related Sites

[Variation Portal](#)

[Variation Tools](#)

Variation Databases

[dbVar](#)

[dbGaP](#)

[ClinVar](#)

[GTR](#)

dbSNPの拡張検索の組み立て

NCBI Resources How To Sign in

SNP Home Help

SNP Advanced Search Builder

Use the builder below to create your search

Edit Clear

Builder

All Fields AND Search All Fields

Accession Allele Base Position Chromosome

Clinical Significance Contig Position Create Build ID Filter Function Class Gene Description

rs11568821 Sort by: SNP_ID 2 11:55:11

Sort by: SNP_ID 3020 11:47:47

03962 Sort by: SNP_ID 5 11:47:33

Sort by: SNP_ID 5 11:47:23

Download history Clear history

You are here: NCBI > Variation > SNPs (dbSNP)

GETTING STARTED

- NCBI Education
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials
- Submit Data

Gene Name Genotype Global Minor Allele Frequency Local SNP ID LocusLink ID Method Class Modification Date Organism

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

FEATURED Genetic Testing Registry GenBank Reference Sequences Gene Expression Omnibus Genome Data Viewer Human Genome Mouse Genome Influenza Virus Primer-BLAST Sequence Read Archive

NCBI INFORMATION

- About NCBI
- Research at NCBI
- NCBI News & Blog
- NCBI FTP Site
- NCBI on Facebook
- NCBI on Twitter
- NCBI on YouTube
- Privacy Policy

Advanced Searchを使用しないと、検索語をAll Fieldsに対して検索

dbSNPの検索結果

NCBI Resources How To Sign in to NCBI

dbSNP SNP PDCD1 rs11568821 Search Help

Annotation Cited in PubMed PubMed nucleotide Global MAF Custom range... Validation Status by-cluster by-frequency Clear all Show additional filters

Display Settings: Summary, Sorted by SNP_ID Send to: Filters: Manage Filters

Find related data Database: Select

検索式の表示

検索候補が表示

過去の検索内容の表示

Search results Items: 2

rs11568821 [Homo sapiens]

- Variant type: SNV
Alleles: C>G,T
Chromosome: NT_187527.1:63858
Gene: LOC105373977 (GeneView), PDCD1 (GeneView)
Functional Consequence: downstream_transcript_variant,genic_downstream_transcript_variant,upstream_transcript_variant,intron_variant
Clinical significance: risk-factor
Validated: by frequency,by cluster
T=0.0409/205 (1000Genomes)
T=0.0711/2196 (GnomAD)
T=0.0746/9373 (TOPMED)
T=0.1157/446 (ALSPAC)
T=0.1162/431 (TWINSUK)
HGVS: NC_000002.12:g.241851760C>G, NC_000002.12:g.241851760C>T,
NC_000002.11:g.242793912C>G, NC_000002.11:g.242793912C>T,
NG_012110.1:g.12147G>C, NG_012110.1:g.12147G>A, NT_187527.1:g.63858C>G,
NT_187527.1:g.63858C>T
[PubMed](#) [Varview](#)
- rs606231173 has merged into rs11568821 [Homo sapiens]

Search details PDCD1 [All Fields] AND rs11568821 [All Fields]

Recent activity Turn Off Clear

- PDCD1 rs11568821 (2) SNP
- PDCD1 (3020) SNP
- 12 : 111803962 (5) SNP
- rs671 (5) SNP

dbSNPの検索結果（詳細画面）

旧表示形式での表示も可能

参照配列毎の位置
ClinVarの情報
頻度情報
Submission情報
rs番号の変遷
文献情報を
Tabで切り替えられます

Reference SNP (rs) Report

[Switch to classic site](#)

rs11568821

Organism: *Homo sapiens*

Position: chr2:241851760 (GRCh38.p12)

Alleles: C>G / C>T

Variation Type: SNV Single Nucleotide Variation

Frequency: T=0.07464 (9373/125568, TOPMED)
T=0.0711 (2196/30874, GnomAD)
T=0.041 (205/5008, 1000G) (+ 2 more)

Clinical Significance: Reported in ClinVar

Gene : Consequence: PDCD1: Intron Variant
LOC105373977: 2KB Upstream Variant

Publications: 32 citations

Genomic View: See rs on genome

Variant Details

Genomic Placements

Sequence name	Change
GRCh37.p13 chr 2	NC_000002.11:g.242793912C>T
GRCh37.p13 chr 2	NC_000002.11:g.242793912C>G
GRCh38.p12 chr 2	NC_000002.12:g.241851760C>T
GRCh38.p12 chr 2	NC_000002.12:g.241851760C>G
GRCh38.p12 chr 2 alt locus HSCHR2_3_CTG15	NT_187527.1:g.63858C>T
GRCh38.p12 chr 2 alt locus HSCHR2_3_CTG15	NT_187527.1:g.63858C>G
PDCD1 RefSeqGene	NG_012110.1:g.12147G>A
PDCD1 RefSeqGene	NG_012110.1:g.12147G>C

dbSNPである遺伝子を対象に検索

NCBI Resources How To

dbSNP SNP PDCD1 Create alert Advanced

Annotation Cited in PubMed PubMed nucleotide protein structure

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Search results

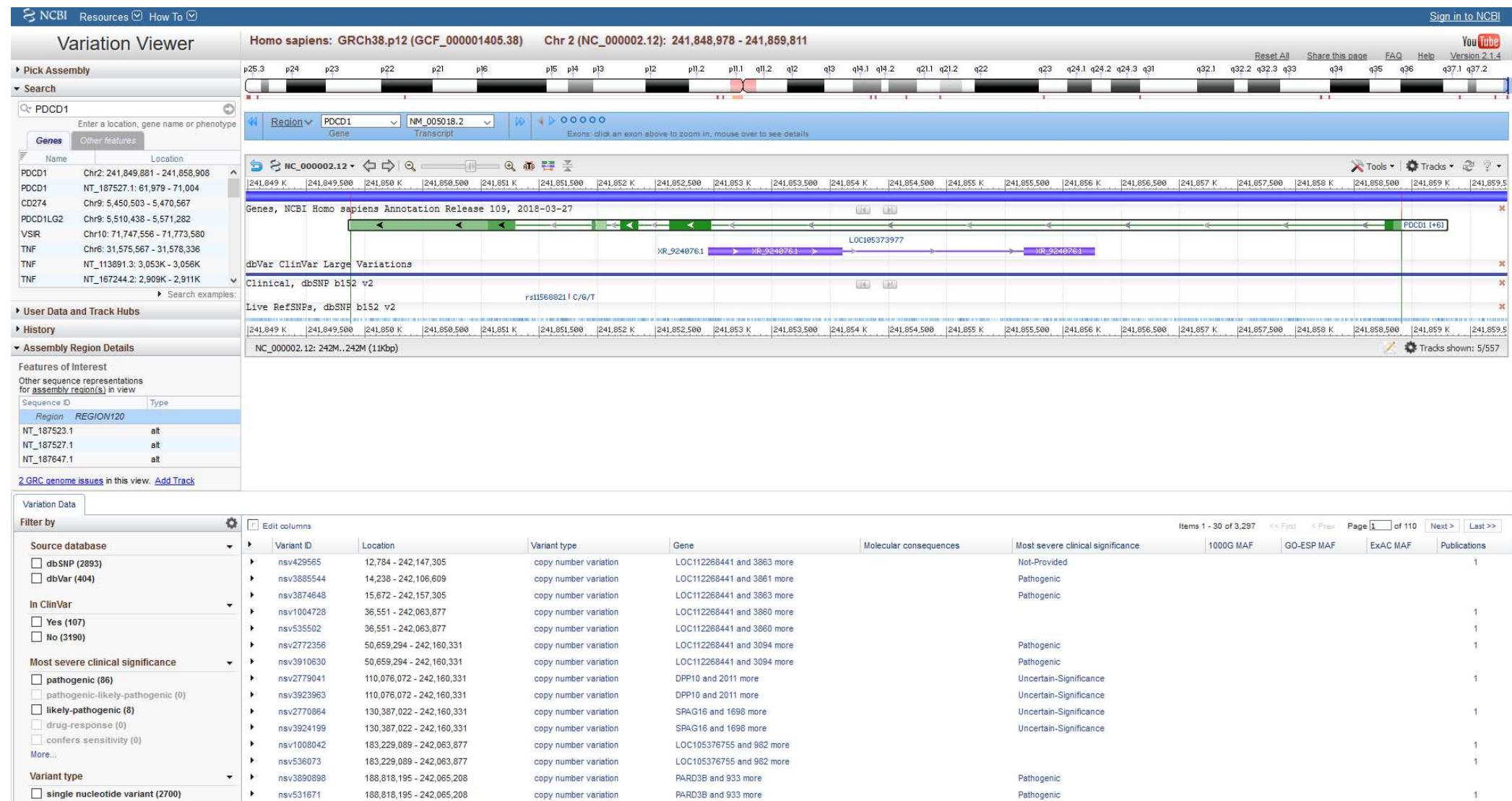
Items: 1 to 20 of 3020 <<First <Prev Page 1 of 151 Next

rs11568821 [*Homo sapiens*]

1.

Variant type:	SNV
Alleles:	C>G,T
Chromosome:	NT_187527.1:63858
Gene:	LOC105373977 (GeneView), PDCD1 (GeneView)
Functional Consequence:	downstream_transcript_variant,genic_downstream_transcript_variant,upstream_transcript_variant,intron_variant,splice_donor_variant,splice_acceptor_variant
Clinical significance:	risk-factor
Validated:	by frequency,by cluster
Global MAF:	T=0.0409/205 (1000Genomes) T=0.0711/2196 (GnomAD) T=0.0746/9373 (TOPMED) T=0.1157/446 (ALSPAC) T=0.1162/431 (TWINSUK)
HGVS:	NC_000002.12:g.241851760C>G, NC_000002.12:g.241851760C>T, NC_000002.11:g.242793912C>G, NC_000002.11:g.242793912C>T,

NBDC dbSNPの遺伝子ビュー(Variant Viewer)



dbSNPおよびdbVar, ClinVarの情報を genome browser上で確認可能
自分の持つ情報をUploadして併せて表示も可能。

dbSNP以外からの利用方法

NCBI Resources How To Sign in to NCBI

All Databases Search

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

Have you tried BLAST+ (2.9.0) and version 5 BLAST databases (dbV5)?
24 May 2019

We recently updated the version 5 BLAST protein databases (dbV5) on our
[New BLAST results to become the default view August 1, 2019](#)
20 May 2019

We have been offering the new BLAST results page ([Figure 1](#)) for you to try out
[March-April 2019 RefSeq eukaryotic](#)

NCBIのWeb site (<https://www.ncbi.nlm.nih.gov/>) から
All Databasesを対象に遺伝子名(例: CYP2D6)で検索すると、、、

[ハンズオン1] dbSNP

[basic]

- ・「rs75961395」と関連が報告された形質は？
- ・「rs75961395」のアリル頻度はどのようなプロジェクトから報告されている？

[advanced]

- ・「rs75961395」はどのプラットフォームで検知可能かを調べましょう

多型データベースの紹介 2: ExAC/gnomAD

ExAC

- 運営元: 米国ブロード研究所
- URL: <http://exac.broadinstitute.org/>
- 様々のプロジェクトからの60,706名の全エクソームシーケンス(WES)データを併せて解析
 - 日本人は76名分のみ
- 次世代シーケンサーにおけるCoverageの情報トラックあり
- 7のpopulationにおけるアリル頻度の違いを表示

ExACの画面説明、操作

ExAC Browser Beta About Downloads Terms Contact Jobs FAQ

Interested in working on the development of this resource? [Apply here.](#)

ExAC Browser (Beta) | Exome Aggregation Consortium

Examples - Gene: PCSK9, Transcript: ENST00000407236, Variant: 22-46615880-T-C, Multi-allelic variant: rs1800234, Region: 22:46615715-46615880

About ExAC

The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientific community.

The data set provided on this website spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. The ExAC Principal Investigators and groups that have contributed data to the current release are listed [here](#).

All data here are released under a [Fort Lauderdale Agreement](#) for the benefit of the wider biomedical community - see the terms of use [here](#).

Sign up for our mailing list for future release announcements [here](#).

Recent News

August 8, 2016
- CNV calls are now available on the ExAC browser

March 14, 2016
- Version 0.3.1 ExAC data and browser (beta) is released! ([Release notes](#))

January 13, 2015
- Version 0.3 ExAC data and browser (beta) is released! ([Release notes](#))

October 29, 2014

gnomAD

- 運営元： ブロード研究所
- URL: <https://gnomad.broadinstitute.org/>
- ExACの後継プロジェクト
 - 125,748名の全エクソームシークエンスデータ
 - 15,708名の全ゲノムシークエンスデータ
- ExACよりも、より細かい条件での頻度情報を取得できる
- 同じ解析手法により解析している為、dbSNPより高精度

gnomADの画面説明、操作

gnomAD browser About Downloads Terms Contact Jobs FAQ

This is a new version of the gnomAD browser. The old version is available at <http://gnomad-old.broadinstitute.org>

gnomAD

genome aggregation database

Search by gene, region, or variant

Examples - Gene: PCSK9, Variant: 1-55516888-G-GA

遺伝子名またはバリアントを指定
バリアントは、rs番号、HGVS、染色体番号-位置-参照アリル-代替アリルの様式で指定

The Genome Aggregation Database (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.

The data set provided on this website spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease specific and population

gnomADの検索結果(遺伝子)



[ハンズオン2] gnomAD

[basic]

- 2番染色体の位置242794191はどのようなアリルが報告されていますか？
- 上記の多型が存在する遺伝子の名前を調べましょう。

[advanced]

- rs1801133は、日本人集団において男女のアリル頻度に差があるか、また、非がんサンプルと非神経疾患サンプルでアリル頻度に差があるか調べましょう。

多型データベースの紹介 3: jMorp、HGVD

jMorpとは

- 運営元: 東北メディカル・メガバンク機構
- URL:
<https://jmorp.megabank.tohoku.ac.jp/201905/>
- 日本人の約3,500名の住民コホートからの
全ゲノムシークエンスデータ
- 以前は、iJGVDにて公開。

jMorpの画面説明、操作

The screenshot shows the homepage of the Japanese Multi Omics Reference Panel (jMorp). The top navigation bar includes links for Sequence, Variation, Proteome, Metabolome, Repository, Downloads, Help, and a link to the Light Toyo-oka project. The main banner features a green background with two hands raised, and the text "Welcome to Japanese Multi Omics Reference Panel". On the left, there is a vertical list of omics datasets: Phenome (To be provided), Metabolome, Proteome, Transcriptome (Iwate Medical Megabank Organization; iMethyl), Methylome (Iwate Medical Megabank Organization; iMethyl), Genome Variation (highlighted with a blue border), and Genome Sequence. The right side contains two sections: "jMorp release 201905" (May 8th, 2019, 3.5KJPv2 Genotype Frequency dataset released) and "jMorp Publication" (two academic papers by Tadaka et al. and Tadaka et al. about the 3.5KJPv2 dataset).

jMorp release 201905

May 8th, 2019
3.5KJPv2 Genotype Frequency dataset
We released 3.5KJPv2 Genotype Frequency dataset calculated from 3.5K Japanese individuals. You can download it from [Downloads](#) page.

More

jMorp Publication

Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, Koshiba S, Yamamoto M, Kinoshita K. "jMorp: Japanese Multi Omics Reference Panel" Nucleic Acids Research. 2016 Jan 4;46(D1):D551-D557. [\[PubMed\]](#)

Tadaka S, Katsuoka F, et al., "3.5KJPv2, An allele frequency panel of 3,552 Japanese Individuals" bioRxiv 529529; doi: <https://doi.org/10.1101/529529> [\[bioRxiv\]](#)

More

Phenome
To be provided

Metabolome

Proteome

Transcriptome
Iwate Medical Megabank Organization; iMethyl

Methylome
Iwate Medical Megabank Organization; iMethyl

Genome Variation

Genome Sequence

jMorp release 201905

May 8th, 2019
3.5KJPv2 Genotype Frequency dataset
We released 3.5KJPv2 Genotype Frequency dataset calculated from 3.5K Japanese individuals. You can download it from [Downloads](#) page.

More

jMorp Publication

Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, Koshiba S, Yamamoto M, Kinoshita K. "jMorp: Japanese Multi Omics Reference Panel" Nucleic Acids Research. 2016 Jan 4;46(D1):D551-D557. [\[PubMed\]](#)

Tadaka S, Katsuoka F, et al., "3.5KJPv2, An allele frequency panel of 3,552 Japanese Individuals" bioRxiv 529529; doi: <https://doi.org/10.1101/529529> [\[bioRxiv\]](#)

More

jMorpの検索方法

The screenshot shows the jMorp search interface. At the top, there are links for Sequence, Variation, Proteome, Metabolome, Repository, Downloads, Help, and Licht Toyo-oka. Below that is a green header with the jMorp logo and the text "jMorp -Japanese Multi Omics Reference Panel-". The main section is titled "Genomic Variants". It has three search options: "Search by gene name", "Search by rs#", and "Search by region (GRCh37/hg19)". A large input field for "Gene name" is highlighted with a blue border. Below it, examples like ALDH2, NFE2L2, GATA1 are listed. A green "Search" button is to the right. A large blue arrow points upwards from the "Gene name" input field towards the search button. The footer contains the text "jMorp release 201404 / Last update: May 26, 2019 / Conditions of use / Keio Medical Information Organization, Keio University C"

遺伝子名、rs番号、領域を指定し、検索可能

jMorpの検索結果

gnomADにおける頻度

ToMMoにおける頻度

Type	Position	RefAlt	rs#	Annotation	Gene	MeanDepth (162PE)	JP4	ToMMo 3.5KJPv2 Allele Frequency Panel	gnomAD AFR	gnomAD AMR	gnomAD ASJ	gnomAD EA%	gnomAD NFE	
SNV	12:112199811	A/T		upstream_gene_variant	ALDH2	22.0/22.0		0.0001						
SNV	12:112199828	G/T	rs1022625657	upstream_gene_variant	ALDH2	21.9/21.9			0.0010					
SNV	12:112199876	C/T	rs944812066	upstream_gene_variant	ALDH2	22.5/22.5			0.0001					0.0001
SNV	12:112199884	T/C	rs539004647	upstream_gene_variant	ALDH2	22.4/22.4			0.0006					
SNV	12:112199912	G/A	rs914760172	upstream_gene_variant	ALDH2	22.3/22.3			0.0006					

jMorpの検索結果 (タンパク質立体構造上へのマッピング)

Type	Position	RefAlt	rs#	Annotation	Gene	MeanDepth (162PE)	JPA	ToMMo 3.5KJPNv2 Allele Frequency Panel	gnomAD AFR	gnomAD AMR	gnomAD ASJ	gnomAD EA%	gnomAD NFE
SNV	12:112241766	G/A	rs671157	missense_variant (p.Glu504Lys)	ALDH2	18.4/18.4	V18V2		0.1959	0.0002	0.0012		0.2671

↓

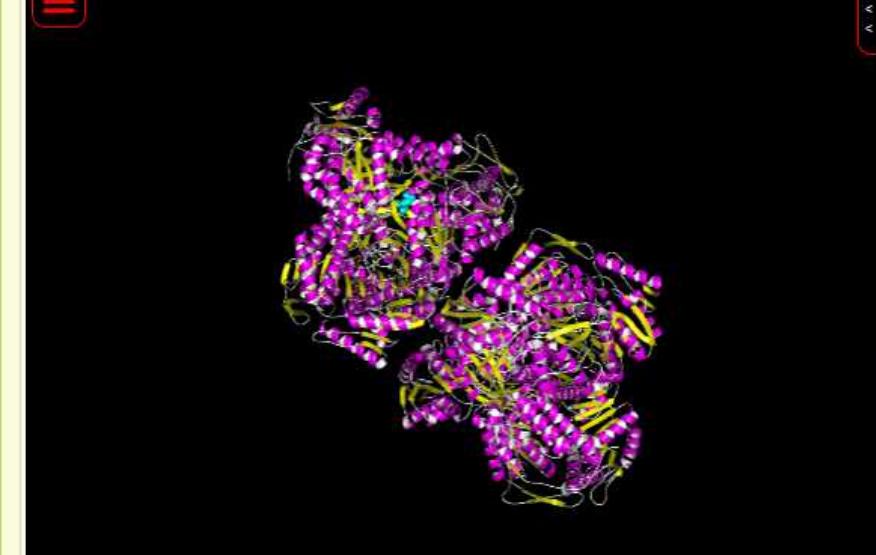
**多型情報を
タンパク質立体構造上に
マッピングされた情報への
リンク**

(タンパク質立体構造が解
析済みかつ多型がエクソ
ン領域にある場合に、リンク
が表示)

Blast result

mRNA info: Homo sapiens aldehyde dehydrogenase 2 family (mitochondrial) (ALDH2), transcript variant 1, mRNA.
mRNA change: NM_000690.3:c.1510G>A_NP_000681.2:p.504E>K
Query: ALDH2 (GeneID: 217) | NM_000690.3 | NP_000681.2
Subject: 5I13C
e-value: 0.0000
Sequence Identity: 100.00

5I13_entity_1 (1/1)



Chain:A (1/1)

Single
• Secondary Structure: E
• Relative ASA: 0.374

BioUnit:1
• delta Relative ASA: 0.295

BioUnit:2
• delta Relative ASA: NA

Mapped Position
• auth_asym_id: A, auth_seq_id: 487
• label_asym_id: A, label_seq_id: 504

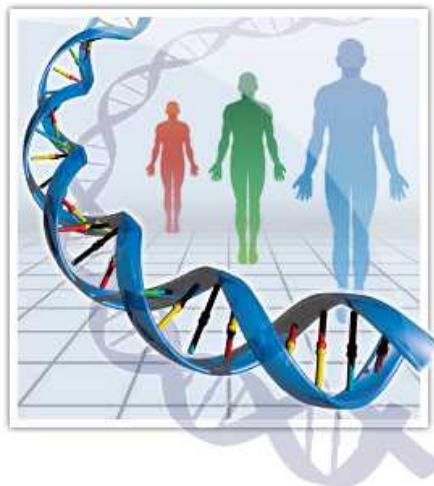
Human Genome Variation Database: HGVD

- 運営元： 京都大学
- URL：
<http://www.hgvd.genome.med.kyoto-u.ac.jp/>
- 1208名の全エクソームシークエンスデータからの多型情報 および 3,248名のSNP chipからの多型情報
- 長浜コホートからのサンプルが中心

HGVDの画面説明、操作

Human Genetic Variation Database

Home About Statistics Link Download Repository Contact How to Use Login



Welcome to Human Genetic Variation Database

Search database

Gene name/ID

dbSNP rsID

Pathogenic Variation

Chromosome

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

NGS Bioinformatics
Kyoto Course



Page views: 3,289,095

Total downloads: 6,259

What's New?

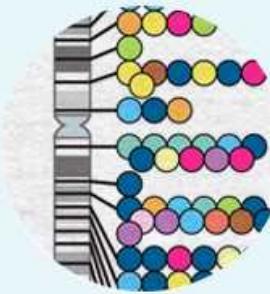
- ▶ 08/02/2017 HGVD version 2.3 is now downloadable.
- ▶ 04/17/2017 New HLA typing software 'HLA-HD' is released. [link](#)

多型データベースの紹介 4: GWAS-catalog

GWASの結果を集めた多型データベース GWAS-catalog

- 運営元：米国国立ヒトゲノム研究所(National Human Genome Research Institute) および 欧州バイオインフォマティクス研究所(EMBL-EBI)
- URL: <https://www.ebi.ac.uk/gwas/>
- 今までGWASによって同定された形質(疾患、形質、薬剤副反応等)と関連が報告されたSNPをマニュアルキュレーション

多型データベースの紹介: GWAS-catalog



GWAS Catalog
The NHGRI-EBI Catalog of published genome-wide association studies

Search the catalog Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000 Search icon

feedback

疾患名、rs番号、著者名、遺伝子名、領域で検索可能

Take a tour of the NEW search interface...

Search

Search the Catalog in a number of ways, including by trait, SNP identifier, publication, gene and genomic location.

Diagram

Explore an interactive visualisation of all SNP-trait associations with genome-wide significance ($p \leq 5 \times 10^{-8}$).

Download

Download a full copy of the GWAS Catalog in spreadsheet format as well as current and older versions of the GWAS diagram in SVG format.

多型データベースの紹介: GWAS-catalog

Show SNPs for	
Digestive system disease	333
Cardiovascular disease	1042
Metabolic disease	403
Immune system disease	1879
Nervous system disease	1269
Liver enzyme measurement	159
Lipid or lipoprotein measurement	993
Inflammatory marker measurement	448
Hematological measurement	4955
Body weights and measures	1726
Cardiovascular measurement	955
Other measurement	11855
Response to drug	325
Biological process	2155
Cancer	1415
Other disease	1633
Other trait	1281



多型データベースの紹介: GWAS-catalog

The screenshot shows the GWAS Catalog homepage. On the left, there is a decorative graphic of colored circles and lines. The main title "GWAS Catalog" is displayed prominently. Below it, the subtitle "The NHGRI-EBI Catalog of published genome-wide association studies" is shown. A search bar contains the query "breast carcinoma". Below the search bar, a placeholder text says "Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000". The URL in the browser bar is "GWAS / Search / breast carcinoma".

フェノタイプのサブグループが表示される

Search results for *breast carcinoma*

T *breast carcinoma* EFO_0000305

A carcinoma arising from the breast, most commonly the terminal ductal-lobular unit. It is the most common malignant tumor in females. Risk factors include country of birth, family history, menstrua... Show more >

Associations 1640 Studies 112

T *HER2 positive breast carcinoma* EFO_1000294

A biologic subset of breast carcinoma defined by high expression of HER2, GRB7, and TRAP100, and by lack of expression of estrogen receptor (ER).

Associations 0 Studies 1

Refine search results

Traits 10
Genes 2

Catalog stats

- Last data release on 2019-05-03
- 3989 publications
- 90031 SNPs
- 138312 associations
- Genome assembly GRCh38.p12
- dbSNP Build 151
- Ensembl Build 96

[ハンズオン3] GWAS-catalog

[basic]

- 疾患「estrogen-receptor negative breast cancer」と関連が報告されたSNPと論文数を調べましょう
- SNP「CYP2C19」と関連が報告された形質およびその集団を確認しましょう

多型データベースの紹介 5: TogoVar

日本人多様性データベース: TogoVar

- 運営元: NBDC, JST
- URL:
<https://togovar.biosciencedbc.jp/>
- 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、データの概要を把握可能に
- 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
- 2018年6月7日公開
- 現状germline variantのみ、somaticなし

NBDCヒトデータベースの必要性 研究データ公開・再利用の世界的潮流から

- オープンサイエンス
学術論文のオープン化 + 研究データの公開

2014年米国NIHのGenomics Data Sharing Policy

ヒトゲノム研究データのデータリポジトリへの登録を要求

2018年3月 AMEDデータマネージメントプラン提出の義務化

- FAIR原則
Findable(見つけられる)、Accessible(アクセスできる)、
Interoperable(相互運用できる)、Reusable (再利用できる)
- Global Alliance for Genomics and Health(GA4GH)
国際的な研究情報共有や統一プロトコルでのゲノム解析の仕組みや、
研究者認証等についての枠組みを討議するアライアンス

NBDCヒトデータベースの必要性 制限公開データベースの必要性（プライバシーの担保）

- 研究データのオープン化の懸念
 - プライバシー保全とのバランス
 - データ公開・利用についての適切な審査
 - →制限公開データベース
- 制限公開データベースの発足
 - 2007年米国NIH
 - Database of Genotypes and Phenotypes (dbGaP)
 - 2008年欧州EMBL-EBI
 - European Genome-phenome Archive (EGA)
 - 2013年10月 NBDCおよび国立遺伝学研究所DDBJセンター
 - NBDCヒトデータベース/Japanese Genotype-phenotype Archive(JGA)

非制限公開データ
Open data

- ・頻度や統計量
 - ・特定の個人由来では無い試料の解析結果
- Webからダウンロード可能

制限公開データ
Controlled-Access data

- ・個人ごとのゲノムデータ
- 利用者を限定
(国内外のアカデミアや企業内研究者)

二重匿名化

各プロジェクト・実施機関

NBDCヒトデータベースのデータ例 バイオバンクジャパン (BBJ) : hum0014

心筋梗塞：1666症例、対照：3198名

健常者：934名 (JSNP)

35疾患：各約190症例 (JSNP)

食道癌：182症例 (JSNP)

ALS：92症例 (JSNP)

T2DM：9817症例、対照：6763名

T2DM：5646症例、対照：19,420名

AD：1472症例、対照：7966名

AF：8180症例、対照：28,612名

BMI：158,284名

Genotypeデータ：182,505名

POAG：3980症例、対照：18,815名

58臨床検査値：162,255名

初潮年齢データを有する女性：67,029名

閉経年齢データを有する女性：43,861名

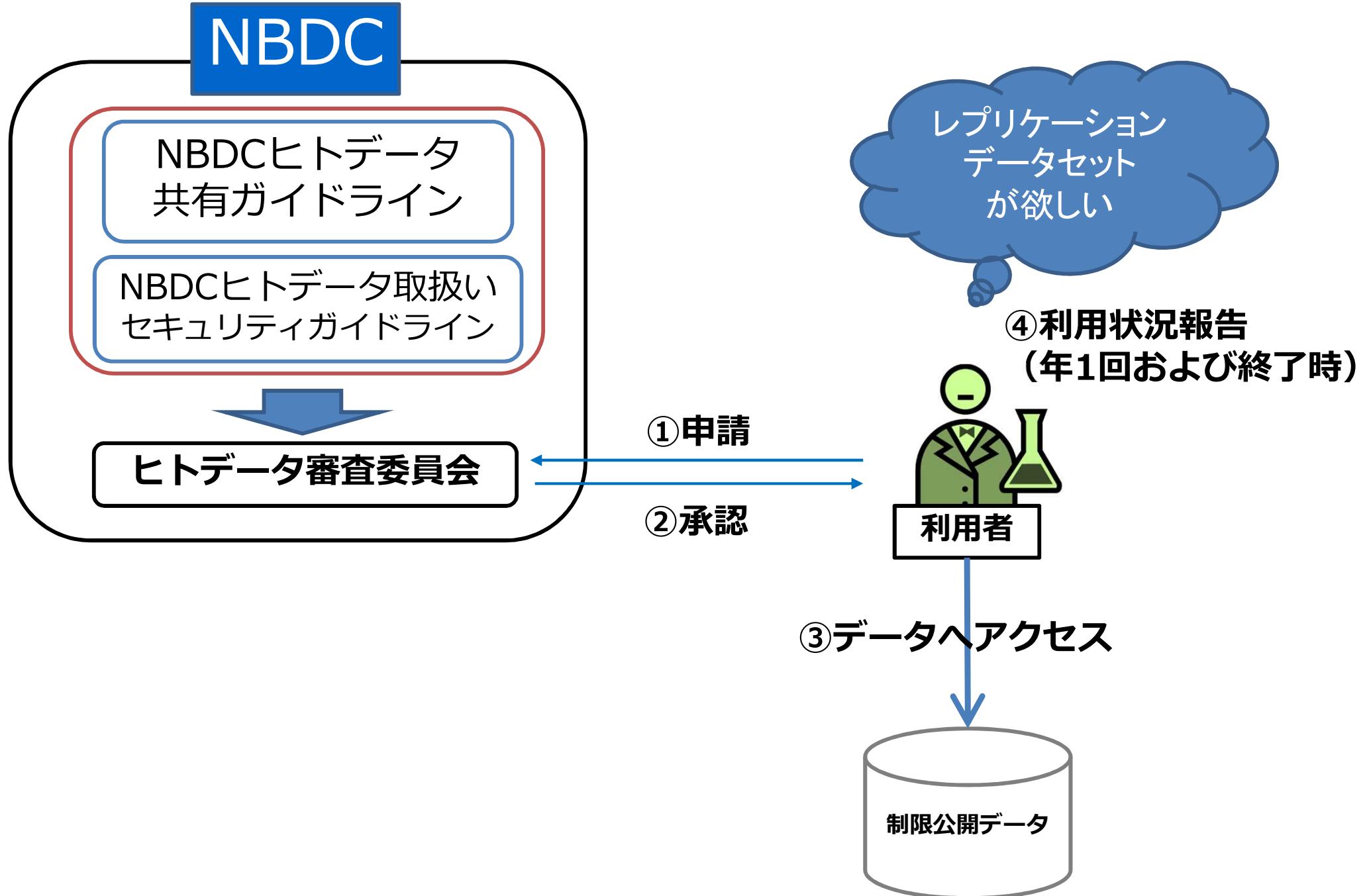
BBJ第1コホート：1,026名

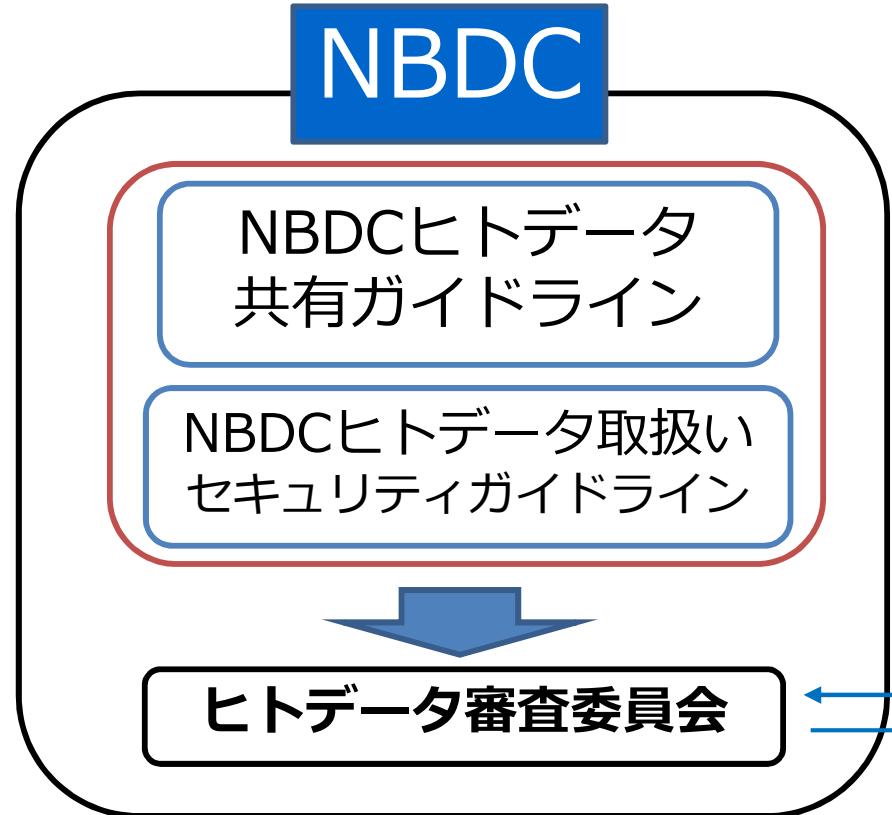
(日本人)

登録された対象疾患および形質：
心筋梗塞、2型糖尿病、アトピー性皮膚炎、
心房細動、BMI、開放隅角緑内障、
初潮・閉経年齢等々

プラットフォーム：
インベーダー法 (Hologic Japan社)
Human610-Quad BeadChip(Illumina社) 等の
Genotyping Arrayデータ

本年8月
Illumina社 HiSeq2500による
BBJ第一コホート 1,026名の
Whole Genome Sequencingを公開





NBDCと提供者間の協議、確認内容

- ・データ公開への同意
- ・機関の長によるデータ公開の許可
- ・データの分類（非制限 or 制限）
- ・データの公開予定日（エンバーゴ可）
- ・データ利用時の制限事項
- ・他の研究者にとっても有益な情報の付加 等

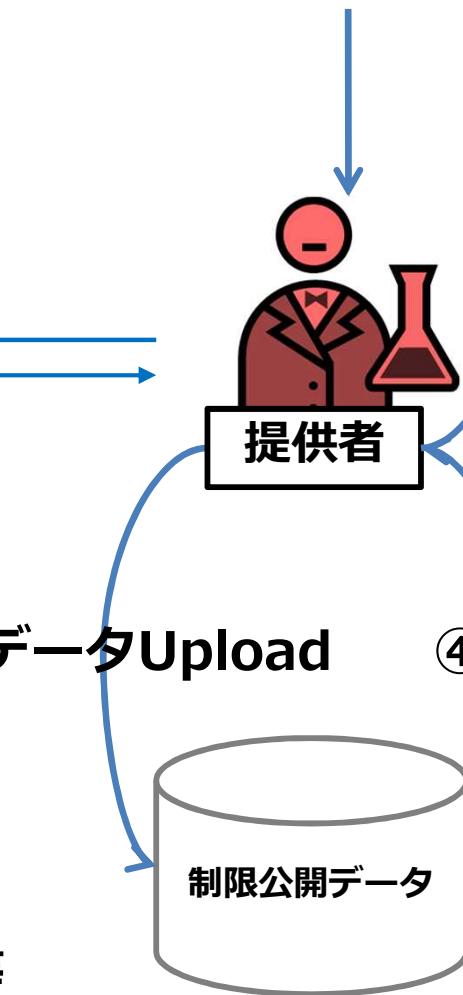
① ジャーナルの投稿規定や
Funding agencyからの要請

- ①申請
②承認

③データUpload
④accession ID付与

⑤投稿論文に
accession ID
を記載

等



日本人多様性データベース: TogoVar

- 運営元: NBDC, JST
- URL:
<https://togovar.biosciencedbc.jp/>
- 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、データの概要を把握可能に
- 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
- 現状germline variantのみ、somaticなし

『NBDCヒトDB』の概要を把握



NBDCヒトデータベース



Japanese Genotype-phenotype Archive

個人別の情報（個人情報）

制限公開（提供・利用審査あり）



個人別の情報でない（集計情報）

非制限公開

データ
提供者

研究プロジェクトA

同じ手法で
再解析

研究プロジェクトB



研究プロジェクトC

データ
利用者① 概要を
把握日本人大規模
バリエント頻度

② 利用申請

出典元 (© 2016 DBCLS TogoTV)

ワンストップ検索

ゲノム配列の個人による違い（バリアント）に関する
さまざまな条件を用いて、国内外のDBや文献情報などの
ワンストップ検索を可能に

7番染色体



ClinVar (NCBI)

位置 : chr7:127254587
関連する疾患 : 2型糖尿病
疾患感受性 : あり

ExAC (ブロード研究所)

位置 : chr7:127254587
アレル頻度 :
0.000304573



iJGVD 3.5KJPN

(東北メディカル・メガバンク機構)

位置 : chr7:127254587
アレル頻度 : 0.0233

HGVD(京都大学)

位置 :
chr7:127254587
アレル頻度 : 0.0272809

<https://biosciencedbc.jp/>

TogoVarID: tgv30913364

位置 : chr7:127254587

関連する疾患 : 2型糖尿病

疾患感受性 : あり

アレル頻度(iJGVD 3.5KJPN) : 0.0233

アレル頻度(HGVD) : 0.0272809

アレル頻度(ExAC) : 0.0003045

関連論文 : 73

A missense mutation of Pax4 gene ...

<https://togovar.biosciencedbc.jp/variant/tgv30913364>



ワンストップ検索



出典元 (© 2016 DBCLS TogoTV)



TogoVar／検索の対象データベース

データベース名 (運営組織)	説明	対象人数
NBDCヒトデータベース (JST-NBDCと国立遺伝学研究所DDBJセンターの共同運営) • JGA-NGSデータセット (全エクソーム) • JGA-SNPデータセット (SNP Chip)	主に日本の研究者からの個人ゲノムデータのリポジトリ	125人 (全エクソーム) 183,884人 (既知SNP)
Integrative Japanese Genome Variation Database (iJGVD) (東北メディカル・メガバンク機構)	ゲノムコホート (東北地方中心)	3,554人 (全ゲノム)
Human Genetic Variation Database (HGVD) (京都大学)	ゲノムコホート (滋賀県長浜市を中心)	1,208人 (全エクソーム)
Exome Aggregation Consortium(ExAC) (ブロード研究所)	配列決定プロジェクトの再解析データ (約20プロジェクト)	60,706人 (全エクソーム)
ClinVar (NCBI)	バリアントの疾患関連性	
PubTator (NCBI) Colil(DBCLS)	バリアント (rs番号) が出現する文献情報	

JGA-NGS/JGA-SNPデータの由来

JGA-NGS

https://togovar.biosciencedbc.jp/doc/datasets/jga_ngs

集約されたデータの由来

JGAID	ヒトDB	研究題目	対象集団	サンプルサイズ	データ提供者
JGAD000000000004	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 アストロサイトーマ	6	斎藤 延人 (P-DIRECT)
JGAD00000000106	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 オリゴデンドログリオーマ	16	斎藤 延人 (P-DIRECT)
JGAD00000000112	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 小脳グリオーマ	17	斎藤 延人 (P-DIRECT)
JGAD00000000014	hum0021	精神神経疾患の原因解明および診断法・治療法の開発に関する研究	健常一卵性双生児	6	加藤 忠史 (ゲノム支援)
JGAD00000000036	hum0035	固形腫瘍における遺伝子異常の網羅的解析	多数の固形腫瘍性疾患検体	23	滝田 順子 (P-DIRECT)
JGAD00000000038	hum0040	ヒト胎盤におけるインプリント制御領域およびインプリント遺伝子の同定	ヒト胎盤および母体血	48	有馬 隆博 (IHEC)
JGAD00000000060	hum0066	癌の再発・転移に関与するnon-coding RNAの同定とその機序の解明	大腸がん	9	三森 功士
合計					125

JGA-SNP

https://togovar.biosciencedbc.jp/doc/datasets/jga_snp

集約されたデータの由来

JGAID	ヒトDB	研究題目	対象集団	サンプルサイズ	データ提供者
JGAD00000000123	hum0014	オーダーメイド医療の実現プログラム	健常者及び罹患者	182,557	久保 充明 (BBJ)
JGAD00000000018	hum0028	オーダーメイド医療の実現プログラム	健常者	908	久保 充明 (BBJ)
JGAD00000000130/131	hum0082	日本人健常者におけるゲノム全域のSNP解析	健常者	419	徳永 勝士
Total					183,884

検索ボックス

- rs番号
- 位置検索、範囲検索 (hg19)
- 遺伝子名 (あいまい検索)
- 関連疾患名 (あいまい検索)

検索結果

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

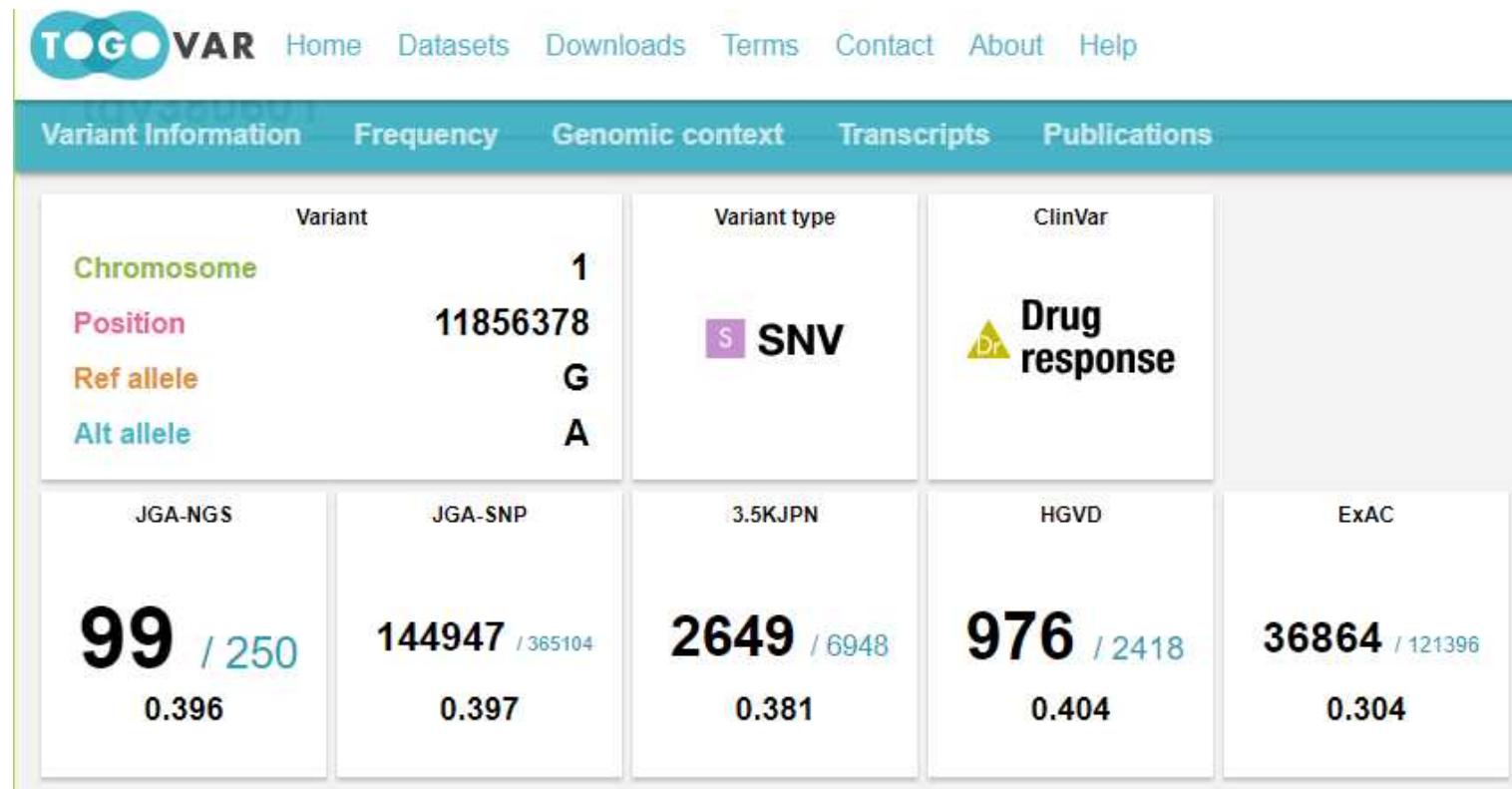
ID	Position	Ref	Alt	Symbol	Frequency ⓘ	Consequence	SIFT ⓘ	Polyphen ⓘ	ClinVar conditions
S tgv48208253	13:32889080	A	G	ZAR1L		intron_variant			Breast-ovarian cancer, familial
S tgv48208258	13:32889363	G	A	ZAR1L		intron_variant			Breast-ovarian cancer, familial
S tgv48208267	13:32889669	C	T	BRCA2		5_prime_UTR_variant			Fanconi anemia Hereditary breast and ovarian cancer Breast-ovarian cancer, familial
S tgv48208277	13:32889803	G	A	BRCA2		splice_region_variant 5_prime_UTR_variant			Breast-ovarian cancer, familial
S tgv48208285	13:32889968	G	A	BRCA2		intron_variant			Breast-ovarian cancer, familial
S tgv48208287	13:32889996	C	T	BRCA2		intron_variant			Breast-ovarian cancer, familial

フィルタ機能

- データセット
- データセット+頻度
- バリアントタイプ
- ClinVarの情報

Alt 7 - 565 - -
 Ref 250 - 7082 - -
 Freq 0.028 - 0.08 - -

<https://biostaticdb.jp/>



TogoVar ID tgv380601

refSNP ID rs1801133

Variation SNV

Chromosome 1

Position 11856378

バリエントの
基本的な情報や
ClinVarに登録されたpathogenicityの情報、
各データセットにおける
アリル頻度を表示



TogoVar

Home Datasets Downloads Terms Contact About Help

Variant Information

Frequency

Genomic context

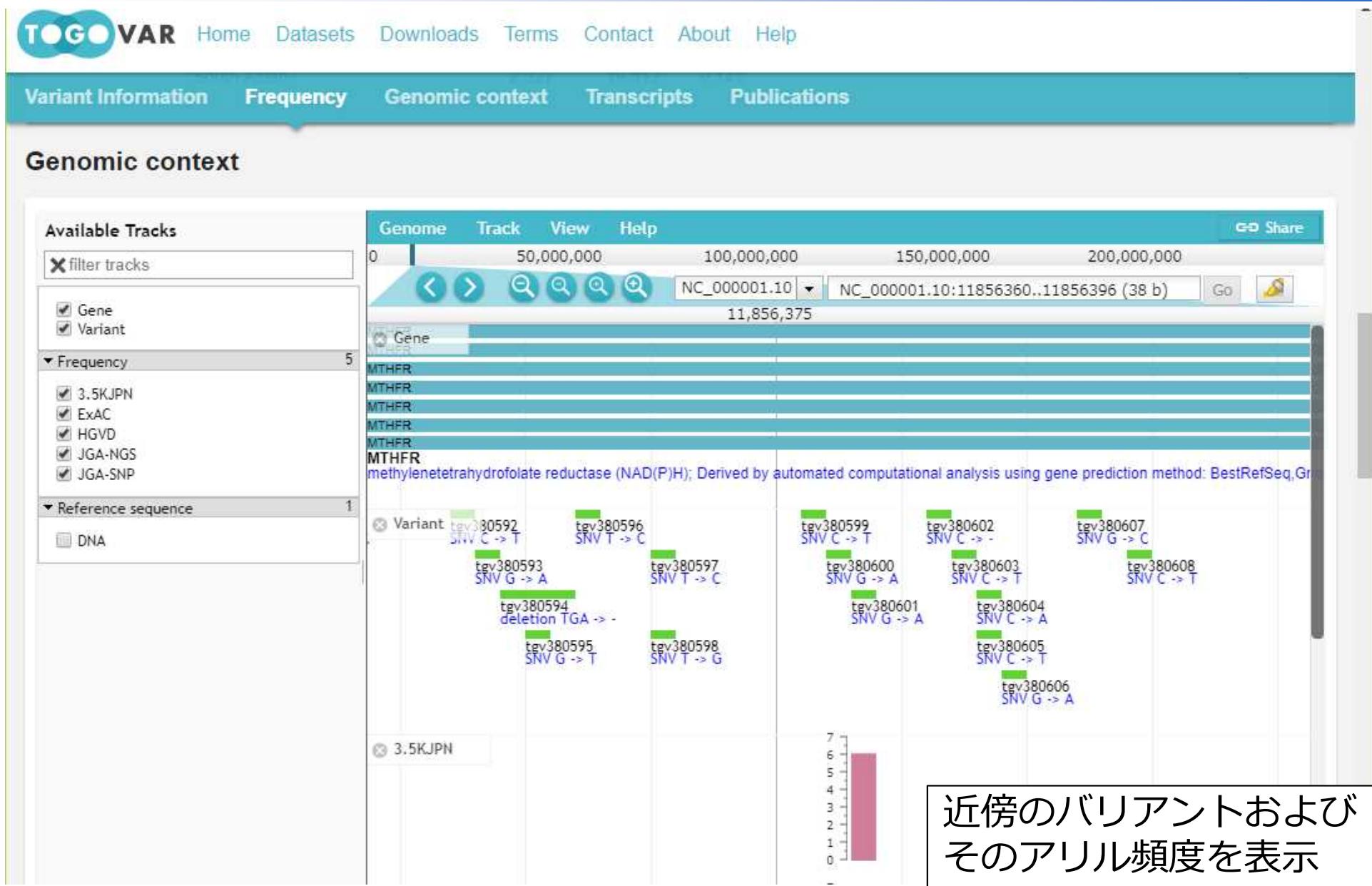
Transcripts

Publications

Frequency

Dataset	Population	Allele count		Alt allele freq	Genotype count			Filter status
		Alt	Total		Alt / Alt	Alt / Ref	Ref / Ref	
JGA-NGS	Japanese	99	250	0.396				PASS
JGA-SNP	Japanese	144,947	365,104	0.397	29,471	86,005	67,076	N/A
3.5 KJPN	Japanese	2,649	6,948	0.381				PASS
HGVD	Japanese	976	2,418	0.404				PASS
ExAC	Total	36,864	121,396	0.304				PASS
	African	1,169	10,404	0.112				-
	East Asian	2,640	8,650	0.305				
	European (Finnish)	1,473	6,614	0.223				
	European (Non-Finnish)	23,025	66,730	0.345				
	Latino	5,952	11,578	0.514				
	Other	278	908	0.306				
	South Asian	2,327	16,512	0.141				

各データセットからの
アリルカウント、アリル頻度、
ジェノタイプカウント、
NGSのバリアントコールのクオリティ
を表示



TOGO VAR Home Datasets Downloads Terms Contact About Help

Variant Information Frequency Genomic context Transcripts Publications

Transcripts

Transcript ID	Gene symbol	Consequence	SIFT	Polyphen
ENST00000376590	MTHFR	• missense_variant	0	0.9409999999999999
ENST00000376585	MTHFR	• missense_variant	0.02	0.987
ENST00000376583	MTHFR	• missense_variant	0.02	
ENST00000376592	MTHFR	• missense_variant	0	

トランскriプト毎の
SIFT, Polyphen2のスコアを表示

Publications (rs1801133)

PMID	Reference	Year	Cited by	Diseases
27452984 (PubTator)	PACsin2 polymorphism is associated with thiopurine-induced hematological toxicity in children with acute lymphoblastic leukaemia undergoing maintenance therapy. Smid A, Karas-Kuzelicki N, Jazbec J, Mlinaric-Rascan I <i>Sci Rep.</i> 2016 Jul 25;6:30244.	2016	2	<ul style="list-style-type: none"> MeSH:D064420 drug-related side effect and adverse reaction MeSH:D054198 precursor cell lymphoblastic leukemia-lymphoma MeSH:D013280 stomatitis MeSH:D006402 hematologic disease

PubMedに登録された論文のabstractに当該バリアントの記述がある論文と関連疾患名を表示。論文の引用数も表示。

The screenshot shows the TogoTV website interface. At the top, there's a navigation bar with links for DBCLS, Research, Services, Contact, and About. Below that is the TOGO TV logo and a sub-header "生命科学系DB・ツール使い倒し系チャンネル". A banner at the top says "2018-08-25 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". On the left, there's a sidebar with "目的別に検索" (Search by purpose) options: 講習会 実習資料 (AJACS) and ゲノム・核酸 配列解析. The main content area shows a video player with a play button and a progress bar (0:18 / 8:36). A large blue callout box points to the video player with the text: "TogoVar (https://togovar.biosciencedbc.jp/) は ヒトゲノムに存在するバリアントに関する 情報について、分子生物学的な情報や頻度、 関連論文を知ることが出来るサイトです。". Below the video player, there's a link to "YouTube版を視聴できない方はオリジナル版ファイル(mov形式)をダウンロードして、ご覧ください。". Further down, there's a detailed description of TogoVar, mentioning its collaboration with NBDC and other institutions, and its purpose of providing variant frequency information and related papers. At the bottom, there's a section for "見どころダイジェスト" (Key features) with three numbered items: 1. 検索とフィルタリングの方法と検索結果の見方 (0:42), 2. バリアント詳細ページの見方 (3:10), 3. アルコール代謝に関連する有名なバリアントの人種別割合を調べる (6:30). A note at the bottom states: "この記事は、クリエイティブ・コモンズ表示ライセンスの下で利用可能です。利用にあたっては、下記のクレジットを必ず表示してください。 © 2016 DBCLS TogoTV / CC-BY-4.0".

- 操作方法等の解説動画 (8分36秒)
 - Youtube、TogoTV

頻度データ(tsvファイル)の一括ダウンロードURL

- (<https://togovar.biosciencedbc.jp/downloads>)

The screenshot shows the 'Downloads' section of the Togovar website. The 'Downloads' menu item is highlighted with a red box. The page displays a table with two columns for each chromosome from 1 to 9, listing 'Variation frequency' and 'Molecular annotation' files, along with their sizes.

Chromosome	Variation frequency	Molecular annotation
1	chr 1 frequency [tsv] (110.0 MB)	chr 1 molecular annotation [tsv] (113.0 MB)
2	chr 2 frequency [tsv] (100.0 MB)	chr 2 molecular annotation [tsv] (111.0 MB)
3	chr 3 frequency [tsv] (81.0 MB)	chr 3 molecular annotation [tsv] (95.0 MB)
4	chr 4 frequency [tsv] (70.0 MB)	chr 4 molecular annotation [tsv] (74.0 MB)
5	chr 5 frequency [tsv] (69.0 MB)	chr 5 molecular annotation [tsv] (74.0 MB)
6	chr 6 frequency [tsv] (71.0 MB)	chr 6 molecular annotation [tsv] (70.0 MB)
7	chr 7 frequency [tsv] (68.0 MB)	chr 7 molecular annotation [tsv] (73.0 MB)
8	chr 8 frequency [tsv] (59.0 MB)	chr 8 molecular annotation [tsv] (66.0 MB)
9	chr 9 frequency [tsv] (54.0 MB)	chr 9 molecular annotation [tsv] (53.0 MB)

[ハンズオン4] TogoVar

[basic]

- rs671を検索し、データセットごとの頻度の差を確認して下さい
- 3.5KJPN(ToMMo)とHGVDとともに頻度が5%以上あるSNPの総数を確認して下さい。

[advanced]

欧米人では、何かの疾患においてpathogenicと報告され、頻度0.1%以下のrare variantが、日本人集団においては1%以上のcommon variantとなっているものを探してください。

1. 國際的なデータベース

1. gnomAD、dbSNP、GWAS-catalogの取り込み
2. ClinVarの定期更新
3. eQTLの表示 (GTEX)

2. 国内のデータベース

1. JENGER(BBJ)、MGeND（日本版ClinVar）との連携
2. 3.5KJPN v.2(chrMT、chrXの追加)の取り込み
3. 日本人 1万人のWGSのデータ提供、再解析
4. LD情報の取り込み

3. 検索内容の追加

1. HGVS表記（例. CCDS4702.1:c.123C>T）
2. GRCh38対応

ご清聴いただき、誠にありがとうございました。

Appendix

Japanese ENcyclopedia of GEnetic associations by Riken (JENGER)

URL: <http://jenger.riken.jp/>

- ・バイオバンク・ジャパン（以下BBJ）のGWASデータを集約した多型データベース

ClinVar

URL: <https://www.ncbi.nlm.nih.gov/clinvar/>

- ・疾患とバリアントの関連を専門家のマニュアルキュレーションを施したデータベース

Medical Genomics Japan Variant Database (MGeND)

URL: <https://mgend.med.kyoto-u.ac.jp/>

- ・日本版ClinVar
- ・約9,000バリアントに臨床的意義を付与
- ・対象疾患：がん/希少疾患/感染症/認知症/難聴

DBKERO

URL: <http://kero.hgc.jp/>

- ・薬剤応答データを検索可能なデータベース

Haploreg

URL: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

- ・米国Broad研究所
- ・検索対象のSNPだけでなく、高い連鎖不平衡状態にあるSNPも検索可能