

多型データベースの紹介 (TogoVarなど)

令和元年8月7日 @AJACS番町3

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)
三橋 信孝

多型/バリアントとは

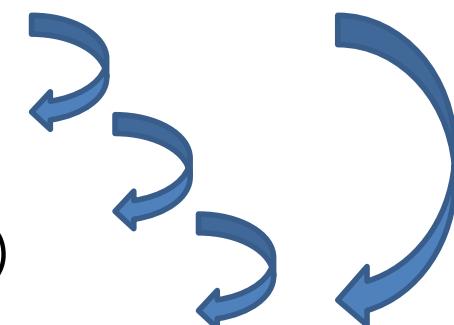
- 多型(polymorphism) / バリアント (variant)
 - ヒトゲノムにおいて観察されるreferenceと異なる塩基配列のこと
 - 多型:集団内での頻度1%以上
- ゲノムと形質との関連を明らかにするうえで重要な基礎データ

ゲノム、エピゲノム

トランск립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

Types of genetic variation

Genetic variation is commonly divided into three main forms:

Single base-pair substitution

There are also known as single nucleotide polymorphisms (SNPs) and can be any nucleic acid substitution:

- Transition
 - interchange of the purine (Adénine/Guanine)
 - or pyrimidine (Cytosine/Thymine) nucleic acids
- Transversion
 - interchange of a purine and pyrimidine nucleic acid (Figure 3)

Move the slider to the right to see a SNP :

DNA with SNP

ACTGACGCATGCATCAT**T**CATGC

Figure 3 SNPs result from the substitution of a single base-pair. In this example we have a transversion event substituting a Thymine nucleic acid in place of a Guanine.

Insertion or deletion, also known as 'indel'

Insertion or deletion of a single stretch of DNA sequence that can range from two to hundreds of base-pairs in length (Figure 4).

Reference	ACTGACGCATGCATCATGCATGC
Insertion	ACTGACGCATGG T A CATCATGCATGC
Deletion	ACTGACG--TGCATCATGCATGC

Figure 4 Indels affect a string of base-pairs. In this example the insertion string shows that GTA has been inserted, and the deletion string shows a deletion of CA.

一塩基多型

(SNP: Single Nucleotide Polymorphism)
(個人当たり～400万)

18番目のGがTに置換された

挿入(insertion)・欠失(deletion)
(個人当たり～70万)
(長さ～数百bp)

<https://www.ebi.ac.uk/training/online/course/human-genetic-variation-i-introduction-2019/what-genetic-variation/types-genetic-variation>

多型/バリエントの種類

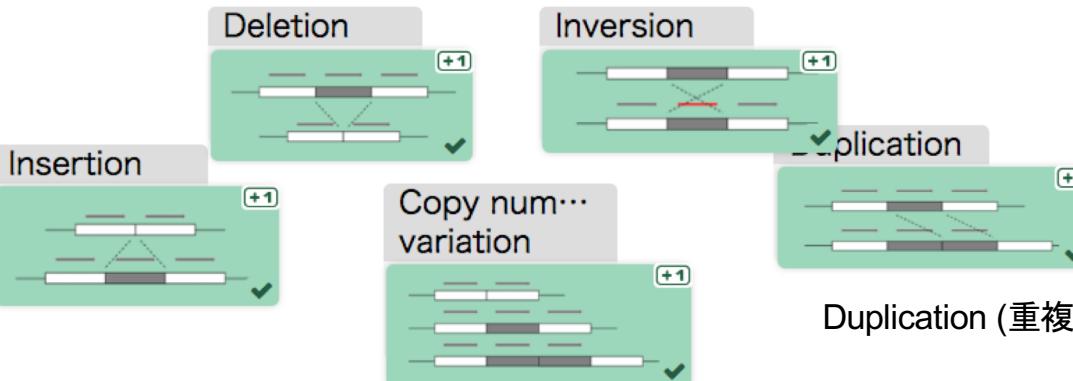
構造多型 (SV: Structural Variant) : 長いものを指す(50bp～)
(個人当たり 1～2万)

Structural variation

Typically used to describe genetic variation that occurs over a larger DNA sequence. This category of genetic variation includes both copy number variation and chromosomal rearrangement events. Discover the five most common types of structural variants in this drag-and-drop game:

Drag and drop the images into the correctly labelled area

Inversion (逆位)



欠失と重複をCopy Number Variation (CNV)ともいう

★ 5 / 5

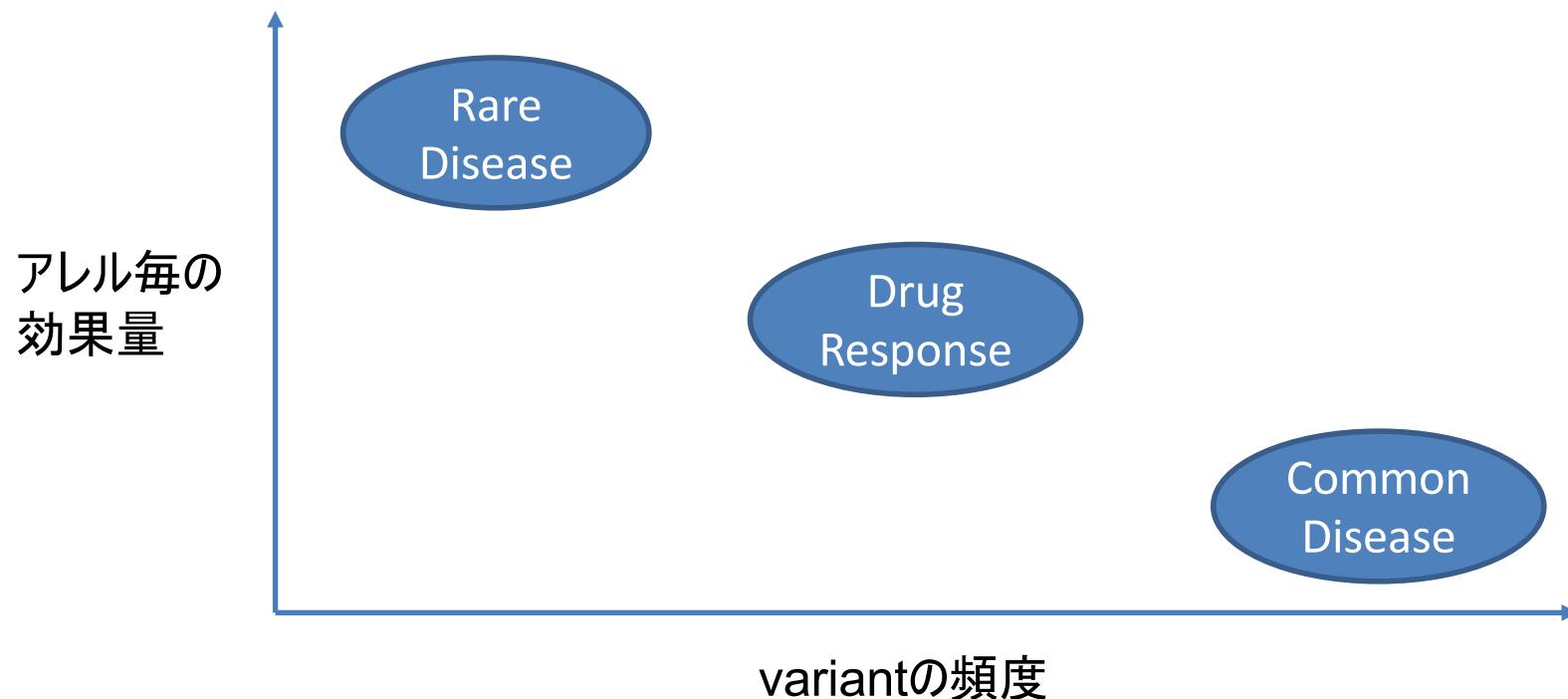
今回の対象外

<https://www.ebi.ac.uk/training/online/course/human-genetic-variation-i-introduction-2019/what-genetic-variation/types-genetic-variation>

4

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

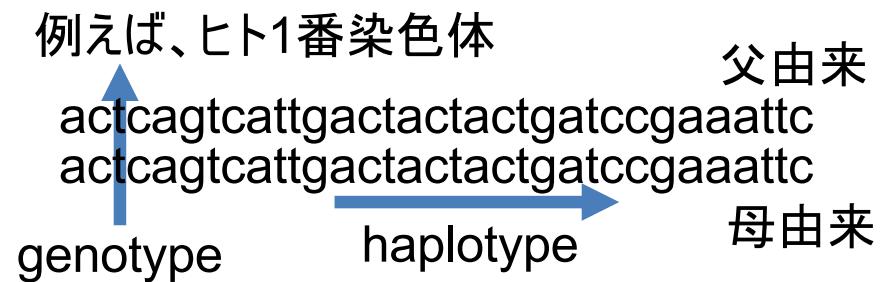


疾患と関連する遺伝子同定の研究概要

- ありふれた疾患 (common disease)
Rare VariantはPrimaryの探索から外されることが多い。
 - 稀少疾患 (rare disease)
Common 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

- ・サンガーシーケンサー
 - 配列情報
- ・SNP Chip
 - 60-90万箇所のgenotype情報
- ・次世代(第2世代)シーケンサー(例:Illumina)
 - リード長: 数十bから数百b
 - 全exome領域や全genome領域等のgenotype情報、haplotype情報
- ・第3世代シーケンサー(例:PacBio)
 - リード長: 数kbから数十kb
 - 特定の領域のgenotype情報、haplotype情報、構造多型の検出

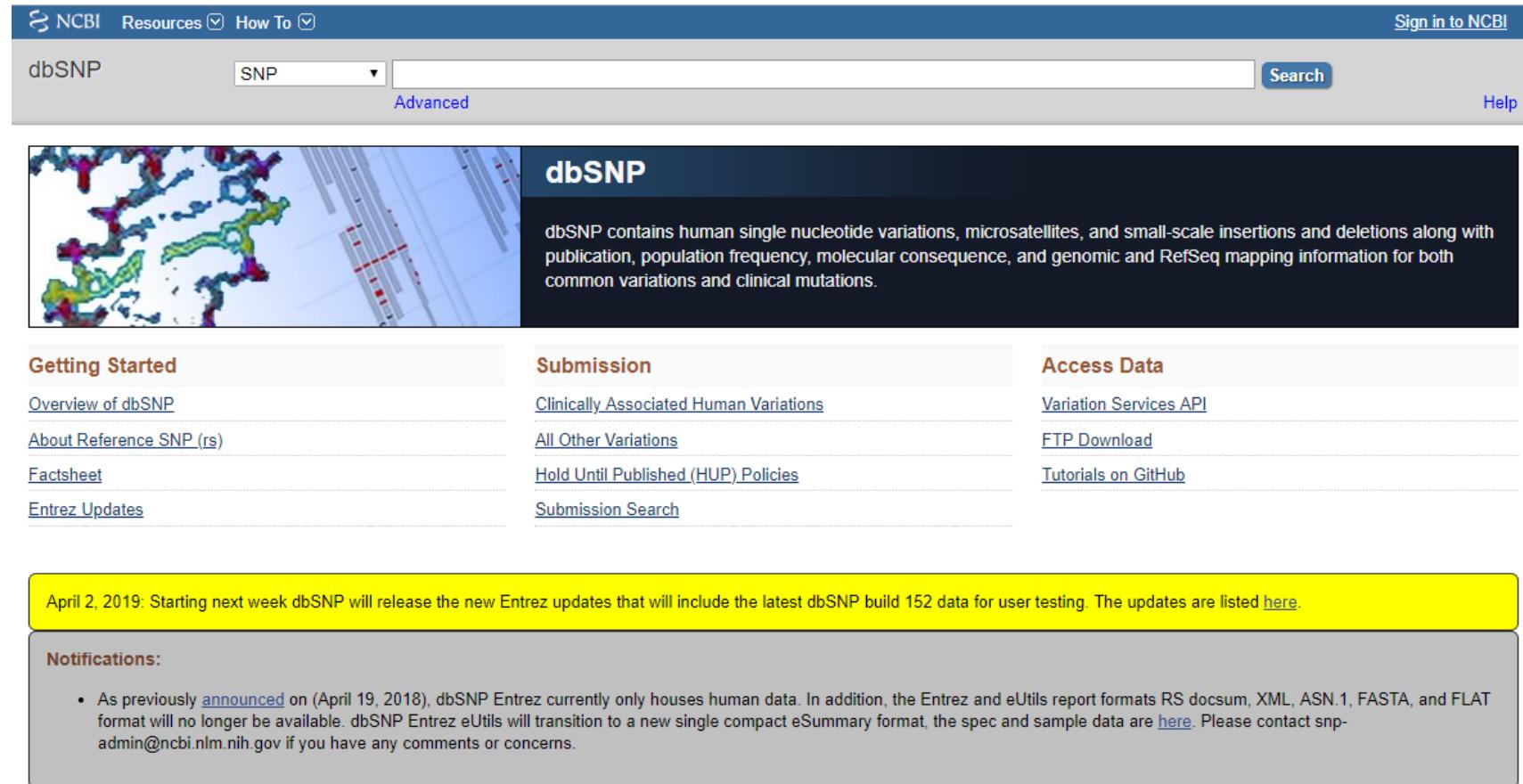


多型データベースの紹介 1:dbSNP

dbSNP

- 運営元：米国NCBI（National Center for Biotechnology Information）
- 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(refSNP)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
 - 6億バリアント

dbSNPの画面説明、操作



The screenshot shows the dbSNP homepage. At the top, there's a navigation bar with links for NCBI, Resources, How To, Sign in to NCBI, and Help. Below the navigation is a search bar with dropdown menus for "dbSNP" and "SNP", and a "Search" button. There's also an "Advanced" link. The main content area features a large image of a genome map on the left and a dark panel on the right with the dbSNP logo and a brief description of the database's content. Below this are three columns: "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), and "Access Data" (Variation Services API, FTP Download, Tutorials on GitHub). A yellow banner at the bottom left contains a message about upcoming Entrez updates. A "Notifications:" section below it contains a bullet point about changes to dbSNP Entrez.

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](#)

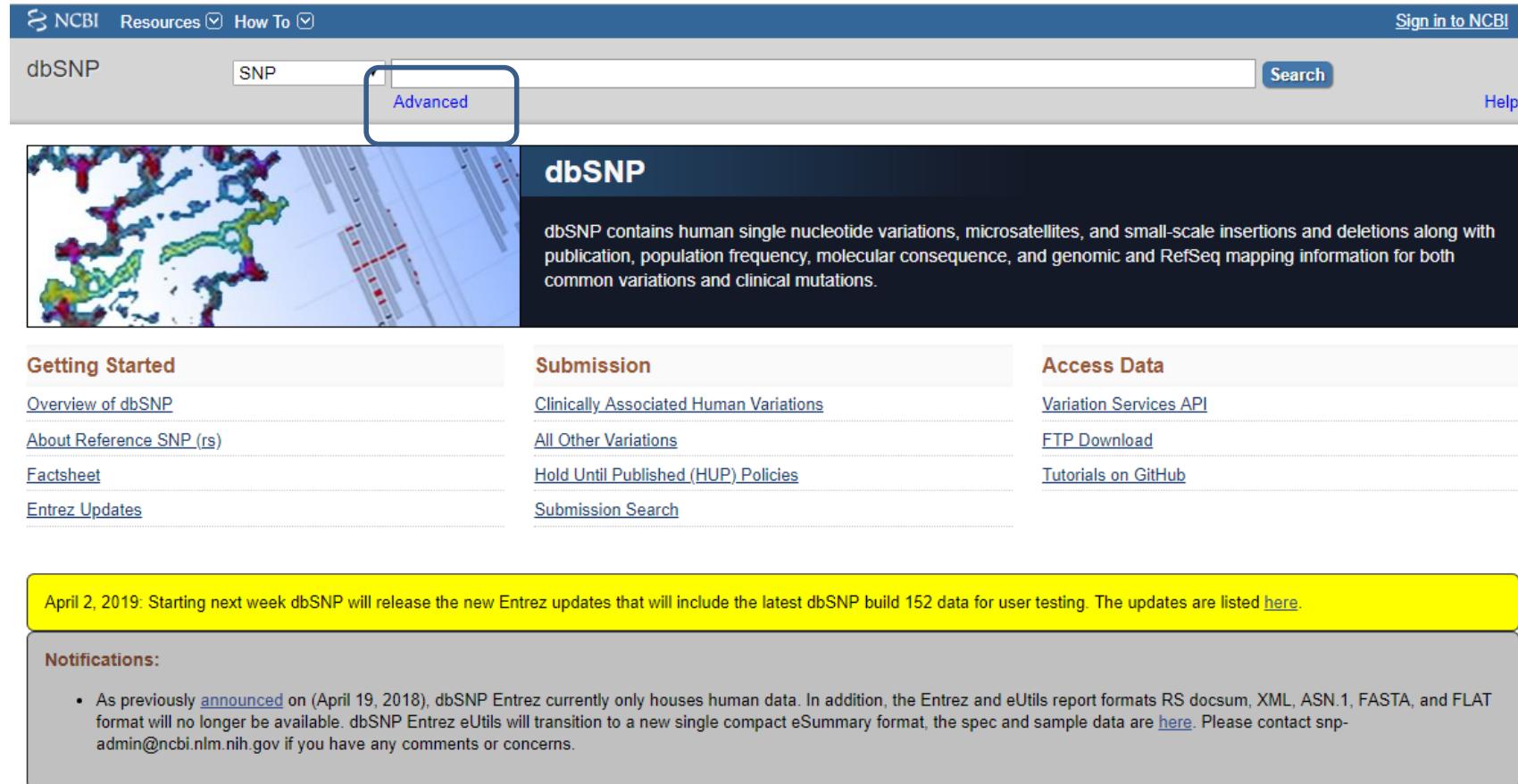
[dbGaP](#)

[ClinVar](#)

[GTR](#)

検索対象語を入力だけ→何を対象に調べているのか？

dbSNPにおける拡張検索機能



The screenshot shows the dbSNP homepage. At the top, there is a navigation bar with links for NCBI, Resources, How To, Sign in to NCBI, and Help. Below the navigation bar, the dbSNP logo is on the left, followed by tabs for SNP and Advanced (which is highlighted with a blue box). A search bar and a "Search" button are to the right of the tabs. The main content area features a map-like visualization on the left and a dark panel on the right containing the dbSNP logo and a brief description of the database's content. Below this, there are three columns: "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), and "Access Data" (Variation Services API, FTP Download, Tutorials on GitHub). A yellow banner at the bottom of the main content area contains a message about upcoming Entrez updates. A "Notifications:" section below the banner contains a bullet point about changes to dbSNP Entrez data formats.

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		Show index list
AND	Accession	Show index list
Search	All Fields	Show index list
	Allele	
	Base Position	
History	Chromosome	Download history Clear history
Search	Clinical Significance	
#4	Contig Position	
#3	Create Build ID	
#2	Filter	
#1	Function Class	
	Gene Description	
	Gene Name	

You are here: NCBI > Variation >

GETTING STARTED

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

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 (PDCD1 rs11568821)

NCBI Resources How To Sign in to NCBI

dbSNP SNP PDCD1 rs11568821 Search Create alert Advanced 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

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

Search See more...

過去の検索内容の表示

Recent activity Turn Off Clear

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

See more...

検索候補が表示

Search results Items: 2

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 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, 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]

2. Variant type: SNV Alleles: C>G,T Chromosome: NT_187527.1:63858 Gene: LOC105373977 (GeneView), PDCD1 (GeneView)

dbSNPの検索結果(詳細画面)

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Log in

dbSNP Short Genetic Variations

Search for rs Example: rs268

Search

旧表示形式
での表示も可能

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

Current Build 152
Released October 2, 2018

ClinVarへのリンク

Variation Viewer

Variant Details

- Clinical Significance
- Frequency
- Aliases
- Submissions
- History
- Publications

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

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

dbSNPの旧表示形式

NCBI dbSNP Short Genetic Variations

Search small variations in dbSNP or large structural variations in dbVar

Search Entrez | dbSNP ▾ for Go

Reference SNP (refSNP) Cluster Report: rs11568821 ** With other allele **

RefSNP	Allele	HGV Names	Links
Organism: human (<i>Homo sapiens</i>)	SNV: single nucleotide variation	CM000664.2:g.241851760>T NC_000002.11:g.242793912C>G	
Molecule Type: Genomic	RefSNP Alleles: A/C/G (REV)	NC_000002.11:g.242793912C>T NC_000002.12:g.241851760>G	
Created/Updated in build: 120/151	Allele Origin: C:germline G:germline	NC_000002.12:g.241851760>T NG_012110.1:g.12147G>A	
Map to Genome Build: 108/Weight 1	Ancestral Allele: G	NG_012110.1:g.12147G>C NM_005018.2:c.627+189G>A	
Validation Status:	Variation Viewer:	NM_005018.2:c.627+189G>C NT_187527.1:g.63858C>G	
Citation: PubMed LitVar NEW	MAF/MinorAlleleCount: T=0.0409/205 (1000 Genomes) T=0.0746/9373 (TOPMED)	...more	

SNP Details are organized in the following sections: GeneView Map Submission Fasta Resource Diversity Validation

Integrated Maps (Hint: click on 'Chr Pos' to see variant in the new NCBI variation viewer)

Assembly	Annotation Release	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh38.p7 (Alt loci)	108	2	NA	NT_187527.1	63858	NA	C	NA	view	mapup
GRCh38.p7	108	2	241851780	NT_005403.18	147355745	Rev	C	Fwd	view	mapup
GRCh37.p13	105	2		NT_005418.13	1984780	Rev	C	Fwd	view	blast

GeneView via analysis of contig annotation: [PDCD1](#) programmed cell death 1

View more variation on this gene (click to hide).

Clinical Source: in gene region cSNP has frequency double hit Go

Primary Assembly Mapping

Assembly	SNP to Chr	Chr	Chr position	Contig	Contig position	Allele
GRCh38.p7	Rev	2	241851780	NT_005403.18	147355745	C

RefSeqGene Mapping

RefSeqGene	Gene (ID)	SNP to RefSeqGene	Position	Allele
NG_012110.1	PDCD1 (5133)	Fwd	12147	G

Gene Model(s)

Function	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
nearGene-3	NA	XM_008712573.2	NA	NA ⇒ NA	NA	NA	NA

NC_000002.12 | Find: rs11568821 | 241,851,710 - 241,851,780

Genes, NCBI Homo sapiens Annotation Release 109, 2018-03-27

Suspect_variations, dbSNP Build 151 (Homo sapiens Annotation Release 108)

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

Function Class frame shift stop gained

Global MAF Custom range...

Validation Status by-cluster by-frequency

[Clear all](#)

[Show additional filters](#)

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,in

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,

dbSNPの遺伝子ビュー

NCBI

dbSNP
Short Genetic Variations

dbVar ClinVar GaP PubMed Nucleotide Protein

Search small variations in dbSNP or large structural variations in dbVar

Search Entrez **dbSNP** for Go

Have a question about dbSNP? Try searching the SNP FAQ Archive! Go

SNP linked to Gene (geneID:5133) Via Contig Annotation

The SNP GeneView page only reports human variation on GRCh38. A new Variation Viewer is available to view the gene PDCD1 variations in GRCh37/p13 or GRCh38, and will replace SNP GeneView later this year. Please visit the [Help Page](#) or [YouTube](#) for available features and send your comments and suggestions to NCBI [helpdesk](#).

Send rs# on all gene models to Batch Query Download all rs# to file.

Gene Model (mRNA alignment) information from genome sequence

Total gene model (contig mRNA transcript):				4		
mRNA	transcript	protein	mRNA orientation	Contig	Contig Label	List SNP
NM_005018.2	minus strand	NP_005009.2	reverse	NT_005403.18	GRCh38.p7	<- currently shown
NM_005018.2	minus strand	NP_005009.2	reverse	NT_187527.1	GRCh38.p7	View SNP on GeneModel
XM_017004293.1	minus strand	XP_016859782.1	reverse	NT_005403.18	GRCh38.p7	View SNP on GeneModel
XM_006712573.2	minus strand	XP_006712636.1	reverse	NT_005403.18	GRCh38.p7	View SNP on GeneModel

Clinical Source in gene region cSNP has frequency double hit refresh

gene model Contig Label Contig mRNA protein mRNA orientation transcript SNP count
(contig mRNA transcript): GRCh38.p7 NT_005403.18 NM_005018.2 NP_005009.2 reverse minus strand 298, coding

Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Heterozygosity	Validation	MAF	Allele origin	3D	Clinically Associated	Clinical Significance	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos	PubMed
241851063	930	rs1350210358	N.D.							missense	T	Phe [F]	1	288		
241851063	930	rs1452312939	N.D.							contig reference	C	Leu [L]	1	288		
241851068	925	rs547824816	0.000		0.0002					frame shift	-	Ser [S]	1	288		
241851068	925	rs547824816	0.000		0.0002					contig reference	C	Leu [L]	1	288		
241851068	925	rs547824816	0.000		0.0002					nonsense	A		2	286		
241851068	925	rs547824816	0.000		0.0002					contig reference	G	Trp [W]	2	286		

見やすいとは言い難い -> Variant Viewer

dbSNPの遺伝子ビュー(Variation Viewer)

NCBI Resources How To Sign in to NCBI

Variation Viewer

Homo sapiens: GRCh38.p12 (GCF_000001405.38) Chr 2 (NC_000002.12): 241,848,978 - 241,859,811

Pick Assembly Search

Region PCD1 NM_005018.2 Gene Transcript Exons: click an exon above to zoom in, mouse over to see details

Reset All Share this page FAQ Help Version 2.1.4

YouTube

Search examples: PCD1

Genes Other features

Name Location

PCD1 Chr2: 241,849,881 - 241,858,908
NT_187527.1: 61,979 - 71,004
C274 Chr9: 5,450,503 - 5,470,567
PCD1LG2 Chr9: 5,510,438 - 5,571,282
VSR Chr10: 71,747,556 - 71,773,580
TNF Chr6: 31,575,567 - 31,578,336
TNF NT_13891.3: 3,053K - 3,056K
TNF NT_167244.2: 2,909K - 2,911K

User Data and Track Hubs History

Assembly Region Details

Features of Interest Other sequence representations for assembly region(s) in view

Sequence ID	Type
Region REGION120	
NT_187523.1	alt
NT_187527.1	alt
NT_187647.1	alt

2 GRC genome issues in this view. Add Track

Variation Data

Filter by

Source database dbSNP (2893) dbVar (404)

In ClinVar Yes (107) No (3190)

Most severe clinical significance pathogenic (66) pathogenic-likely-pathogenic (0) likely-pathogenic (8) drug-response (0) confers sensitivity (0) More...

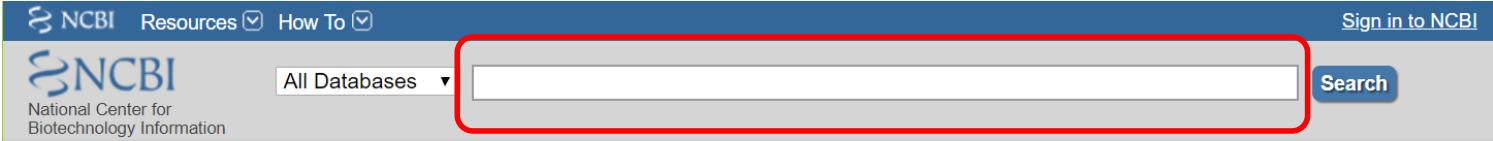
Variant type single nucleotide variant (2700)

Variant ID Location Variant type Gene Molecular consequences Most severe clinical significance 1000G MAF GO-ESP MAF ExAC MAF Publications

nsv429565 12,784 - 242,147,305 copy number variation LOC112268441 and 3863 more Not-Provided 1
nsv3885544 14,238 - 242,106,609 copy number variation LOC112268441 and 3861 more Pathogenic 1
nsv3874648 15,672 - 242,157,305 copy number variation LOC112268441 and 3863 more Pathogenic 1
nsv1004728 36,551 - 242,063,877 copy number variation LOC112268441 and 3860 more Pathogenic 1
nsv535502 36,551 - 242,063,877 copy number variation LOC112268441 and 3860 more Pathogenic 1
nsv2772356 50,659,294 - 242,160,331 copy number variation LOC112268441 and 3094 more Pathogenic 1
nsv3910630 50,659,294 - 242,160,331 copy number variation LOC112268441 and 3094 more Pathogenic 1
nsv2779041 110,076,072 - 242,160,331 copy number variation DPP10 and 2011 more Uncertain-Significance 1
nsv3923963 110,076,072 - 242,160,331 copy number variation DPP10 and 2011 more Uncertain-Significance 1
nsv2770864 130,387,022 - 242,160,331 copy number variation SPAG16 and 1698 more Uncertain-Significance 1
nsv3924199 130,387,022 - 242,160,331 copy number variation SPAG16 and 1698 more Uncertain-Significance 1
nsv1008042 183,229,089 - 242,063,877 copy number variation LOC105376755 and 982 more 1
nsv536073 183,229,089 - 242,063,877 copy number variation LOC105376755 and 982 more 1
nsv3890989 188,818,195 - 242,065,208 copy number variation PARD3B and 933 more Pathogenic 1
nsv531671 188,818,195 - 242,065,208 copy number variation PARD3B and 933 more Pathogenic 1

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



The screenshot shows the NCBI homepage. At the top, there is a navigation bar with links for NCBI Resources, How To, and Sign in to NCBI. Below the navigation bar is the NCBI logo and the text "National Center for Biotechnology Information". A search bar is prominently displayed, with a red box highlighting the search input field. To the right of the search bar is a "Search" button. On the left side of the page is a sidebar with a "Resource List (A-Z)" menu containing links to various databases: NCBI Home, All Resources, Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation. The main content area features a "Welcome to NCBI" section with a brief introduction and links to About the NCBI, Mission, Organization, and NCBI News & Blog. Below this are six sections: 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), and Research (Explore NCBI research and collaborative projects). To the right of the main content is a "Popular Resources" sidebar listing links to PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. At the bottom right is a "NCBI News & Blog" section with several news items.

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、、、

NCBIの統合検索機能を利用した情報の取得

NCBI Databases

Results found in 31 databases for: CYP2D6

Literature

Bookshelf	615
MeSH	5
NLM Catalog	11
PubMed	7,058
PubMed Central	8,971

Genes

Gene	84
GEO DataSets	12
GEO Profiles	50,222
HomoloGene	2
PopSet	12
UniGene	10

Genetics

ClinVar	70
dbGaP	0
dbSNP	3,122
dbVar	302
GTR	83
MedGen	48
OMIM	17

Proteins

Conserved Domains	0
Identical Protein Groups	14
Protein	579
Protein Clusters	0
Sparcle	2
Structure	24

Genomes

Assembly	0
BioCollections	0
BioProject	7
BioSample	54
Genome	7
Nucleotide	1,706
Probe	101
SRA	98
Taxonomy	0

Chemicals

BioSystems	210
PubChem BioAssay	3,822
PubChem Compound	1
PubChem Substance	357

[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クエンス(WES)データを併せて解析
 - 日本人は76名分のみ
- 次世代シークエンサーにおけるCoverageの情報トラックあり
- 7つの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

Search for a gene or variant or region

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

Search by gene, region, or variant

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

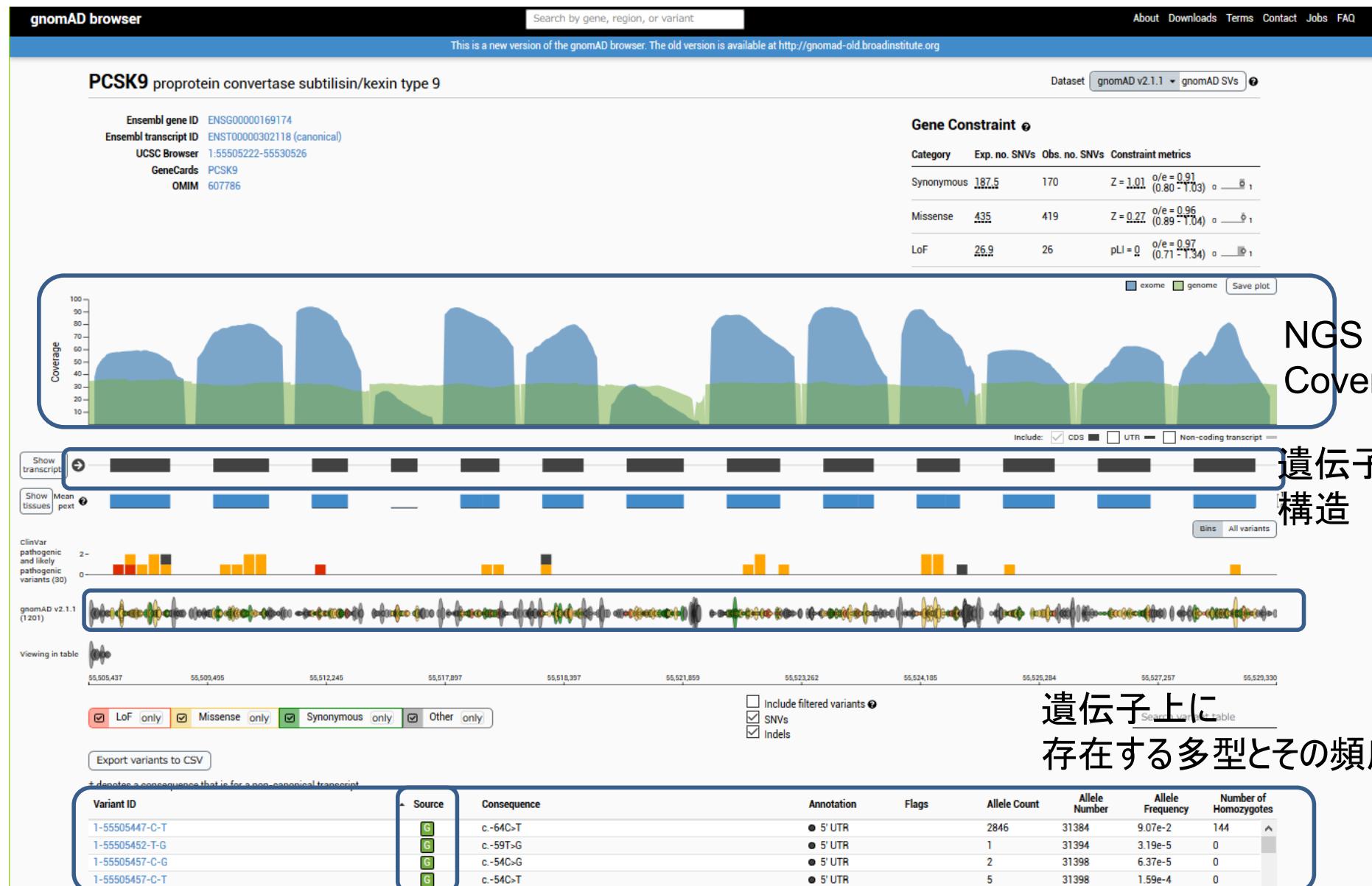
Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)

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)



gnomADの検索結果(バリアント)

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

Insertion: 1-55505552-A-ACTG

Filter	Exomes	Genomes	Total
Allele Count	Pass	Pass	24095
Allele Number	19607	4488	192344
Allele Frequency	0.1217	0.1436	0.1253
Popmax Filtering AF (95% confidence)	0.1723	0.1994	

This variant is multiallelic. Other alt alleles are:

- 1-55505552-A-AATGCTG
- 1-55505552-A-ACTGCTG
- 1-55505552-A-ACTGCTGCTG
- 1-55505552-A-ACTGCTGCTGCTG
- 1-55505552-ACTG-A
- 1-55505552-ACTGCTG-A

Annotations

This variant falls on 2 transcripts in 1 gene.

inframe insertion

- PCSK9
 - ENST00000302118 *
HGVS: p.Leu23dup
 - ENST00000452118
HGVS: p.Leu23dup

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	3359	17352	306	0.1936
Ashkenazi Jewish	1417	8730	82	0.1623
South Asian	3250	23124	183	0.1405
European (non-Finnish)	10076	78928	489	0.1277
Other	665	5572	21	0.1193
Overall	1556	13538	61	0.1149
Korean	297	2562	13	0.1159
East Asian	1073	9380	43	0.1144
Japanese	2	40	0	0.05000
Male	835	7092	33	0.1177
Female	721	6446	28	0.1119
European (Finnish)	1736	18904	57	0.09183
Latino	2036	26196	52	0.07772
Total	24095	192344	1251	0.1253

Include: Exomes Genomes

Dataset gnomAD v2.1.1

- gnomAD v2.1.1 (141,456 samples)
- gnomAD v2.1.1 (controls) (60,146 samples)
- gnomAD v2.1.1 (non-cancer) (134,107 samples)
- gnomAD v2.1.1 (non-neuro) (114,704 samples)
- gnomAD v2.1.1 (non-TOPMed) (136,743 samples)

Report

- Report this variant
- Request additional information

Age Distribution

Heterozygotes Homozygotes

Population毎、性別毎の頻度情報
・ExACよりも粒度が細かい

頻度の再計算が可能

- ・全サンプル
- ・コントロールサンプル
- ・非がんサンプル
- ・非神経疾患サンプル
- ・TOPMedに含まれないサンプル

Trans-Omics for Precision Medicine Program (TOPMed)

[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、東アジア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

Japanese Multi Omics Reference Panel (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
全ゲノムシーケンスデータから
- 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 Tohoku University website. The main header features a green background with a wavy pattern and two hands. The text "Welcome to Japanese Multi Omics Reference Panel." is displayed. Below this, there are sections for Phenome (represented by two people icons), Metabolome (represented by a chemical structure icon), Proteome (represented by a protein icon), Transcriptome (represented by a DNA helix icon), Methylome (represented by a DNA helix icon), Genome Variation (represented by a DNA helix icon, highlighted with a blue border), and Genome Sequence (represented by a DNA helix icon). A sidebar on the right provides information about the jMorp release 201905, dated May 8th, 2019, which includes a 3.5KJPv2 Genotype Frequency dataset calculated from 3.5K Japanese individuals. It also lists publications by Tadaka et al. and bioRxiv 529529.

jMorp

Sequence | Variation | Proteome | Metabolome | Repository | Downloads | Help | Licht Toyo-oka

Welcome to
Japanese Multi Omics
Reference Panel.

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](#).

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. 2018 Jan 4;46(D1):D551-D557. [\[CrossRef\]](#)

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> [\[CrossRef\]](#)

More

jMorp release 201905 / LastUpdate: May 8th, 2019 / Conditions of Use
Tohoku Medical Megabank Organization, Tohoku University [\[CrossRef\]](#)

jMorpの検索方法

The screenshot shows the jMorp search interface. At the top, there is a navigation bar with links for Sequence, Variation, Proteome, Metabolome, Repository, Downloads, Help, and a user account. Below the navigation bar, the jMorp logo and the text "jMorp ~Japanese Multi Omics Reference Panel~" are displayed. The main search area is titled "Genomic Variants". It features three search options: "Search by gene name", "Search by rs#", and "Search by region (GRCh37/hg19)". Below these options is a "Gene name" input field with a placeholder "Gene name" and examples listed below it: ALDH2, NFE2L2, GATA1. To the right of the input field is a green "Search" button. A large blue arrow points upwards from the bottom of the page towards the search button. Below the search area, there is a footer with the text "jMorp release 201806 / LastUpdate: May 8th, 2018 / Conditions of Use" and "Tohoku Medicoal Megabank Organization, Tohoku University".

遺伝子名、rs番号、領域を指定し、検索可能

jMorpの検索結果

jMorp -Japanese Multi Omics Reference Panel-

Search by gene: ALDH2 GRCh37/hg19

Showing unfiltered version.

4148 variants found

Filter by keyword

ToMMoにおける頻度

gnomADにおける頻度

Type	Position	Ref/Alt	rs#	Annotation	MeanDepth (162PE)	Gene	ToMMo 3.5KJPNv2 Allele Frequency Panel	gnomAD AFR.	gnomAD AMR.	gnomAD ASJ.	gnomAD EAS.	gnomAD NFE.
SNV	12:112199811	A/T		upstream_gene_variant	22.0/22.0	ALDH2	0.0001					
SNV	12:112199828	G/T	rs1022625657	upstream_gene_variant	21.9/21.9	ALDH2	0.0010					
SNV	12:112199876	C/T	rs944612066	upstream_gene_variant	22.5/22.5	ALDH2	0.0001					0.0001
SNV	12:112199884	T/C	rs539004647	upstream_gene_variant	22.4/22.4	ALDH2	0.0006					
SNV	12:112199912	G/A	rs914760172	upstream_gene_variant	22.3/22.3	ALDH2	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	Ref/Alt	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	rs671	missense_variant (p.Glu504Lys)	ALDH2	18.4/18.4	V1&V2		0.1959	0.0002	0.0012		0.2671	

BlastHit1 BlastHit2 BlastHit3 BlastHit4 BlastHit5 BlastHit6 BlastHit7 BlastHit8 BlastHit9 BlastHit10

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: 5I13
e-value: 0.0000
Sequence Identity: 100.00

5I13_entity_1 (1/1)



Chain:A (1/8)

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)とは

- 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



NGS Bioinformatics
Kyoto Course



Page views: 3,289,095

Total downloads: 6,259

Welcome to Human Genetic Variation Database

Search database

Gene name/ID GO

dbSNP rsID GO

Pathogenic Variation GO

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

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 (Genome Wide Association Study)

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(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 Search icon

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

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 a DNA helix composed of colored spheres. 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, examples of search terms are listed: "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

- T Traits 10
- G 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

多型データベースの紹介: GWAS-catalog

Associations 1409

ある形質で行われた関連解析の結果

Variant and risk allele	P-value	P-value annotation	RAF	OR	Beta	CI	Mapped gene	Reported trait	Trait(s)	Study accession	Location
rs2111836-G	7×10^{-6}		0.22	1.32	-	[1.17-1.44]	MTCL1	Breast cancer	breast carcinoma	GCST002735	18:8716323
rs13025833-A	1×10^{-7}		0.52	1.2	-	[1.13-1.33]	CTNNA2	Breast cancer	breast carcinoma	GCST002735	2:79716982
rs1078806-C	2×10^{-6}	(Pooled P value)	0.39	1.43	-	[NR]	FGFR2	Breast cancer	breast carcinoma	GCST001831	10:121579461
rs166870-T	3×10^{-7}	(HR+ HER2-)	0.13	2.3	-	[1.67-3.15]	AC026826.2, RNU6-667P	Disease-free survival in breast cancer	breast carcinoma, disease free survival	GCST002847	15:79777420
rs10825036-G	4×10^{-7}	(HR- HER2-)	0.32	2.26	-	[1.34-3.81]	RNA5SP318, AL365496.1	Disease-free survival in breast cancer	disease free survival, breast carcinoma	GCST002847	10:53506471

Showing 1 to 5 of 1409 rows 5 ▲ rows per page 1 2 3 4 5 ... 282 >

Studies 87

関連解析の研究概要

First author	Study accession	Publication date	Journal	Title	Reported trait	Trait(s)	Discovery sample number and ancestry	Replication sample number and ancestry	Association count	Summary statistics
Haryono SJ	GCST002735	2015-01-01	Asian Pac J Cancer Prev	A pilot genome-wide association study of breast cancer susceptibility loci in Indonesian.	Breast cancer	breast carcinoma	• 135 Asian unspecified	-	2	NA

[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
- 遺伝子「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

日本人ゲノム多様性統合データベース: 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



NGSデータ由來
頻度情報
(JGA-NGS)

SNP-Chipデータ由來
頻度情報
(JGA-SNP)

日本人大規模
バリエント頻度

① 概要を
把握

データ
利用者



出典元 (© 2016 DBCLS TogoTV)

② 利用申請

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

▲ 注目するバリアント

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

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)

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)
Japanese Multi Omics Reference Panel(jMorp) (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	齊藤 延人 (P-DIRECT)
JGAD000000000106	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	齊藤 延人 (P-DIRECT)
JGAD000000000112	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	齊藤 延人 (P-DIRECT)
JGAD000000000014	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)
JGAD000000000036	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)
JGAD000000000038	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)
JGAD000000000060	hum0066	癌の再発・転移に関する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
JGAD000000000123	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)
JGAD000000000018	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)
JGAD000000000130/131	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	

TogoVar 検索結果例（一覧検索画面）

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

TogoVar ID	RefSNP ID	Frequency	Consequence	SIFT	PolyPhen	Clinical significance
tgv67071948	rs7792	1: 10271	SNV			
tgv67071949	rs13563	1: 10271	SNV			
tgv67071950	rs14094	1: 10271	SNV			
tgv67071951	rs201732991	1: 10271	SNV			
tgv67071952	rs775928745	1: 10231	Deletion			
tgv67071953	rs1267031179	1: 10248	Deletion			
tgv67071954	rs1035171912	1: 10248	AAAC...gt>	Deletion		
tgv67071955	rs145427775	1: 10291	SNV			
tgv67071956	rs060413313	1: 10297	SNV			
tgv67071957	rs112750067	1: 10321	SNV			
tgv67071958	rs1035171912	1: 10327	T>C	Deletion		
tgv67071959	rs150969722	1: 10330	C>A	SNV		
tgv67071960	rs1357792609	1: 10334	T>A	SNV		
tgv67071961	rs1351390918	1: 10334	T>A	SNV		
tgv67071962	rs1193008993	1: 10336	A>C	SNV		
tgv67071963	rs1471210572	1: 10340	T>C	SNV		
tgv67071964	rs1363828207	1: 10343	CCCT...gtgt	SNV		
tgv67071965	rs1363828207	1: 10346	T>C	SNV		
tgv67071966	rs1363828207	1: 10348	ACCC...gtgt	SNV		
tgv67071967	rs1363828207	1: 10349	CCCT...gtgt	SNV		
tgv67071968	rs1015856060	1: 10352	ACCC...gtgt	SNV		
tgv67071969	rs1015856060	1: 10354	C>A	SNV		
tgv67071970	rs1015856060	1: 10384	CCCT...gtgt	SNV		
tgv67071971	rs1015856060	1: 10384	C>A	SNV		

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
SNV or INDEL...
- ・臨床的意義
Pathogenic, Benign...
- ・missense変異有害性予測スコア
PolyPhen-2, SIFT

Filters

Dataset	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> JGA NGS	4,679,025
<input checked="" type="checkbox"/> JGA SNP	1,249,724
<input checked="" type="checkbox"/> 3.5KJPN	64,675,495
<input checked="" type="checkbox"/> HGVD	554,461
<input checked="" type="checkbox"/> ExAC	10,195,872
<input checked="" type="checkbox"/> ClinVar	443,512

Alternative allele frequency

0 ~ 1
 Invert range

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

for all datasets for any dataset

Variant calling quality

Exclude filtered out variants in all datasets

Variant type

Variant type	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> SNV	63,066,161
<input checked="" type="checkbox"/> Insertion	5,383,629
<input checked="" type="checkbox"/> Deletion	6,240,052
<input checked="" type="checkbox"/> Indel	7,358
<input checked="" type="checkbox"/> Substitution	1,740

Clinical significance

Clinical significance	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> Not in ClinVar	74,255,428
<input checked="" type="checkbox"/> Pathogenic	112,579
<input checked="" type="checkbox"/> Likely pathogenic	59,343
<input checked="" type="checkbox"/> Uncertain significance	353,079
<input checked="" type="checkbox"/> Likely benign	175,644
<input checked="" type="checkbox"/> Benign	94,928
<input checked="" type="checkbox"/> Conflicting interpretations of pathogenicity	6,047

Variant report
tgv47264307 dbSNP
rs671

• 位置、ジェノタイプ、SNPタイプ

Variant type SNV Ref / Alt G>A
Position 12:112241766 (GRCh37) hgvs 12:g.112241766G>A

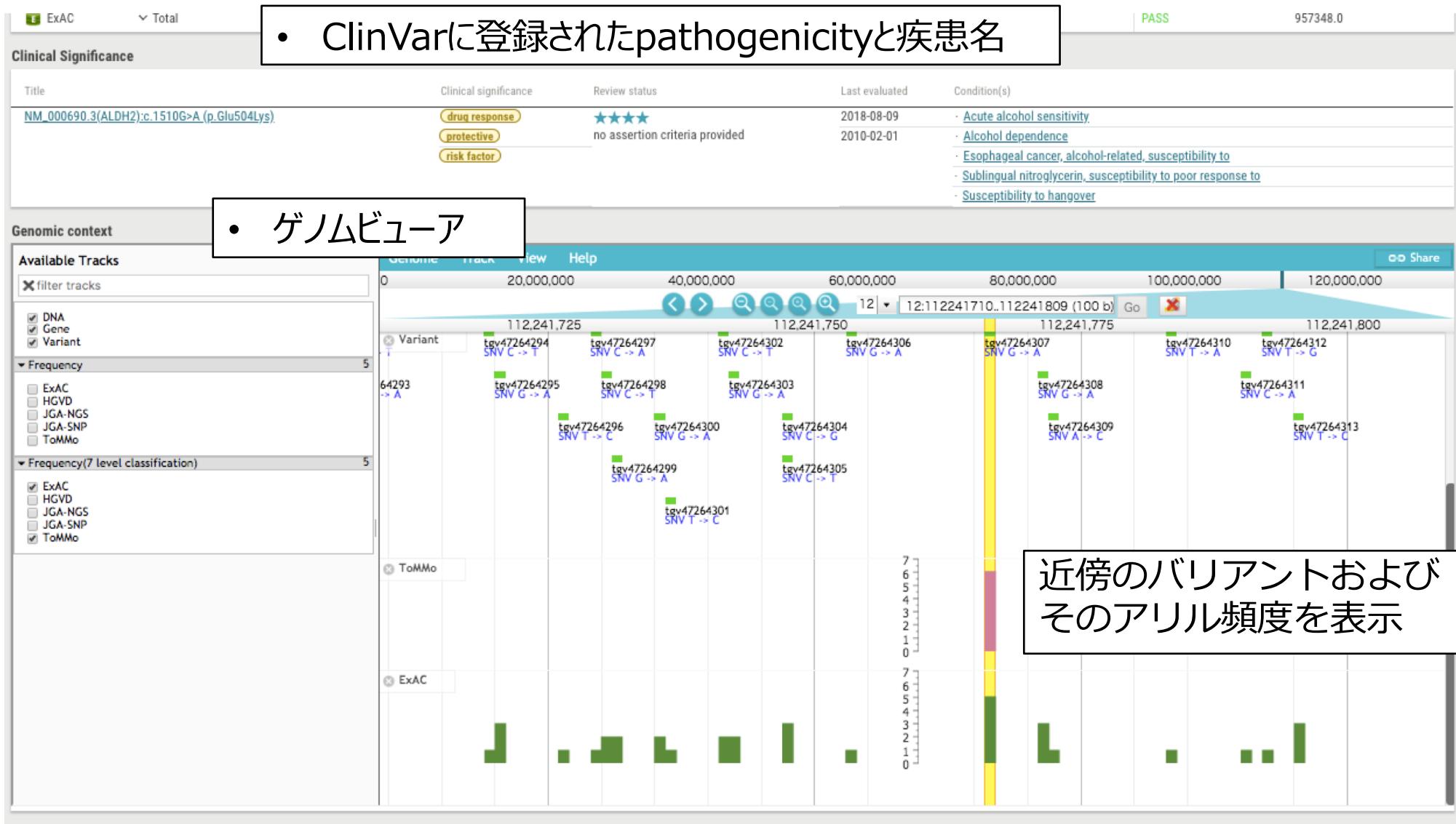
Other overlapping variant(s)
No other variants on the same location

• 同じ領域に存在する他のバリアント

Frequency

Dataset	Population	Allele count			Genotype count			Filter status	Quality score
		Alt	Total	Frequency	Alt / Alt	Alt / Ref	Ref / Ref		
JGA-NGS	Japanese	52 /	250	0.208				PASS	451.77
JGA-SNP	Japanese	90,026 /	365,930	0.246		11,778	66,470	104,717	-
3.5k JPN	Japanese	1,392 /	7,104	0.196				PASS	
HGVD	Japanese	451 /	1,890	0.239				PASS	
ExAC	^ Total	1,878 /	88,224	0.021				PASS	957348.0
	African/African American	1 /	7,614	1.313e-4				-	
	American	8 /	7,302	0.001				-	
	East Asian	1,860 /	6,992	0.266				-	
	Finnish	0 /	4,554	0.000				-	
	Non-Finnish European	3 /	48,388	6.200e-5				-	
	Other	3 /	668	0.004				-	
South Asian	2 /	10,706	0.001						

- データセット（民族毎）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



Gene

HGNC/Approved name	aldehyde dehydrogenase 2 family member
HGNC/Approved symbol	ALDH2
HGNC/Alias name	

Transcripts

Transcript ID	Gene symbol	Consequence type	HGVS(cDNA)	HGVS(Amino acid seq.)	SIFT	PolyPhen
ENST0000261733	ALDH2	missense_variant	ENST0000261733.2:c.1510G>A	ENSP0000261733.2:p.Glu504Lys	0.0	0.874 Possibly Damaging
ENST0000416293	ALDH2	missense_variant	ENST0000416293.3:c.1369G>A	ENSP0000403349.3:p.Glu457Lys	0.0	0.566 Possibly Damaging
ENST0000548536	ALDH2	3_prime_UTR_variant NMD_transcript_variant	ENST0000548536.1:c.*1386G>A			
ENST0000549106	ALDH2	3_prime_UTR_variant NMD_transcript_variant	ENST0000549106.1:c.*89G>A			

Publications

Showing 1 to 10 of 16

PMID	Reference	Year	Cited by	MeSH
28036260 (PubTator)	Effects of alcohol consumption, ALDH2 rs671 polymorphism, and Helicobacter pylori infection on the gastric cancer risk in a Korean population. Yang S, Lee J, Choi IJ, Kim YW, Ryu KW, Sung J, Kim J <i>Oncotarget.</i> 2017 Jan 24;8(4):6630-6641.	2017	1	<ul style="list-style-type: none"> D016481 Helicobacter infection D009369 tumor D007239 infection D013274 gastric cancer
28038378 (PubTator)	The causal effects of alcohol on lipoprotein subfraction and triglyceride levels using a Mendelian randomization analysis: The Nagahama study. Tabara Y, Arai H, Hirao Y, Takahashi Y, Setoh K, Kawaguchi T, Kosugi S, Ito Y, Nakayama T, Matsuda F <i>Atherosclerosis.</i> 2017 Feb;257:22-28.	2017	1	関連疾患名
28040078	A missense single nucleotide polymorphism in the ALDH2 gene, rs671, is associated with hip fracture.			<ul style="list-style-type: none"> D014947 wound and injury D050723 bone fracture

トランスクリプト毎の Molecular, Consequence, SIFT, Polyphen2のスコア

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

引用数

解説動画あります (TogoTV)

The screenshot shows a search result for 'TogoVar' on the NBDC website. The search bar at the top contains 'TogoVar'. Below it, there's a list of results:

- togotv - YouTube** (link: <https://www.youtube.com/user/togotv/feed>)
- TogoVar でヒトゲノムに存在するパリアントに関する情報を探べる** (link: <https://togovar.biosciencedbc.jp/>)

The main content area displays the details of the second result:

TogoVar (https://togovar.biosciencedbc.jp/) は
ヒトゲノムに存在するパリアントに関する
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。

YouTubeで視聴できない方はavi形式のファイルをダウンロードして、ご覧ください。

TogoVar (日本人ゲノム多様性統合データベース) は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC)と大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用 ...

3日前 - TogoVar (日本人ゲノム多様性統合データベース) は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC)と大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフ ...

この動画を引用する際は DOI をご利用ください。 DOI: 10.7303/biopub/2018_158

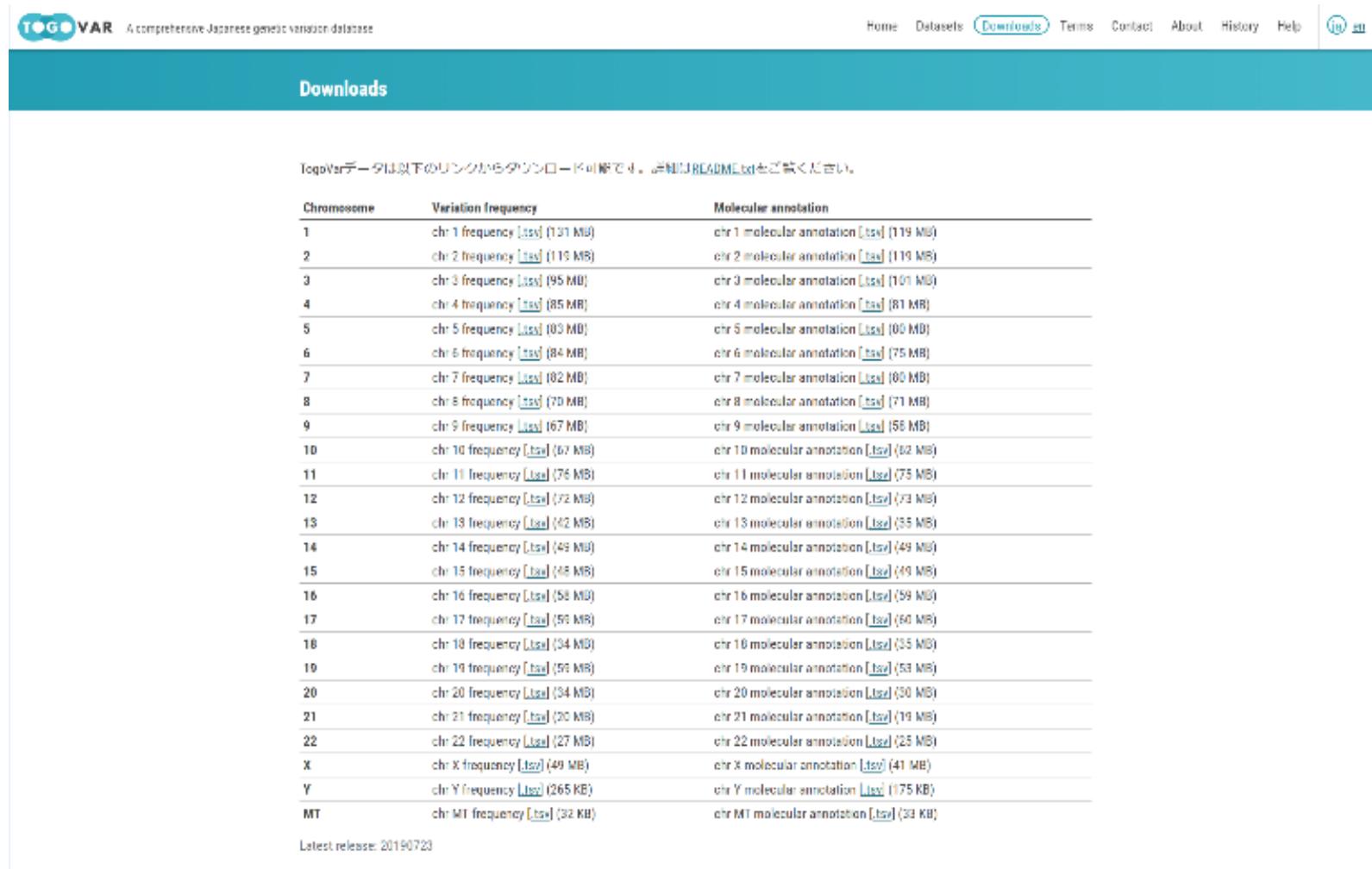
この記事は、クリエイティブ・コモンズ・ライセンスの下で利用可能です。利用上あたっては、下記のラジットを必ず表示してください。
© NBDC, JST Licensed Under CC 表示 4.0 国際

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

一括ダウンロードできます

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 page header includes the Togovar logo and navigation links for Home, Datasets, Downloads (which is highlighted in blue), Terms, Contact, About, History, and Help. Below the header is a teal bar with the word 'Downloads'. A message in Japanese at the top of the main content area says: 'Togovarデータは以下のリンクからダウンロード可能で、詳細なREADME.txtをご覧ください。' (Togovar data can be downloaded from the following links. Please refer to the detailed README.txt for more information.)

Chromosome	Variation frequency [tsv]	Molecular annotation [tsv]
1	chr 1 frequency [tsv] (131 MB)	chr 1 molecular annotation [tsv] (119 MB)
2	chr 2 frequency [tsv] (119 MB)	chr 2 molecular annotation [tsv] (119 MB)
3	chr 3 frequency [tsv] (95 MB)	chr 3 molecular annotation [tsv] (101 MB)
4	chr 4 frequency [tsv] (85 MB)	chr 4 molecular annotation [tsv] (81 MB)
5	chr 5 frequency [tsv] (103 MB)	chr 5 molecular annotation [tsv] (90 MB)
6	chr 6 frequency [tsv] (84 MB)	chr 6 molecular annotation [tsv] (75 MB)
7	chr 7 frequency [tsv] (82 MB)	chr 7 molecular annotation [tsv] (80 MB)
8	chr 8 frequency [tsv] (70 MB)	chr 8 molecular annotation [tsv] (71 MB)
9	chr 9 frequency [tsv] (67 MB)	chr 9 molecular annotation [tsv] (58 MB)
10	chr 10 frequency [tsv] (67 MB)	chr 10 molecular annotation [tsv] (62 MB)
11	chr 11 frequency [tsv] (76 MB)	chr 11 molecular annotation [tsv] (75 MB)
12	chr 12 frequency [tsv] (72 MB)	chr 12 molecular annotation [tsv] (73 MB)
13	chr 13 frequency [tsv] (42 MB)	chr 13 molecular annotation [tsv] (35 MB)
14	chr 14 frequency [tsv] (49 MB)	chr 14 molecular annotation [tsv] (49 MB)
15	chr 15 frequency [tsv] (48 MB)	chr 15 molecular annotation [tsv] (49 MB)
16	chr 16 frequency [tsv] (58 MB)	chr 16 molecular annotation [tsv] (59 MB)
17	chr 17 frequency [tsv] (55 MB)	chr 17 molecular annotation [tsv] (60 MB)
18	chr 18 frequency [tsv] (34 MB)	chr 18 molecular annotation [tsv] (35 MB)
19	chr 19 frequency [tsv] (55 MB)	chr 19 molecular annotation [tsv] (53 MB)
20	chr 20 frequency [tsv] (34 MB)	chr 20 molecular annotation [tsv] (30 MB)
21	chr 21 frequency [tsv] (20 MB)	chr 21 molecular annotation [tsv] (19 MB)
22	chr 22 frequency [tsv] (27 MB)	chr 22 molecular annotation [tsv] (25 MB)
X	chr X frequency [tsv] (49 MB)	chr X molecular annotation [tsv] (41 MB)
Y	chr Y frequency [tsv] (265 KB)	chr Y molecular annotation [tsv] (175 KB)
MT	chr MT frequency [tsv] (32 KB)	chr MT molecular annotation [tsv] (33 KB)

Latest release: 20190720

[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でともに頻度が5%以上あるSNPの総数を確認して下さい。

[advanced]

- Takenouchi-Kosaki syndromeで検索し、頻度が公共DBで報告されているか確認してください。報告されていない場合は、その理由を考えてください。

将来的な機能強化（予定）

1. 國際的なデータベース

1. gnomAD、GWAS-catalogの取り込み
2. ClinVarの定期更新
3. eQTL (Expression quantitative trait loci)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. 日本人8000人分のWGSのデータ提供、再解析
 1. LD(Linkage Disequilibrium)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

NBDCヒトデータベースとは？

NBDC National Bioscience Database Center NBDCヒトデータベース English サイト内検索 検索

ホーム データの利用 データの提供 ガイドライン 機関外サーバ NBDCヒトデータ審査委員会 成果発表 お問い合わせ FAQ

NBDCヒトデータベースについて

ヒトに関するデータは、次世代シーケンサーはじめとした解析技術の発達に伴って膨大な量が産生されつつあり、それらを整理・格納して、生命科学の進展のために有効に活用するためのルールや仕組みが必要です。

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)では、個人情報の保護に配慮しつつヒトに関するデータの共有や利用を推進するために、ヒトに関する様々なデータを共有するためのプラットフォーム『NBDCヒトデータベース』を設立するとともに、国立遺伝学研究所 DNA Data Bank of Japan (DDBJ)と協力して、ヒトに関するデータを公開しています。

本Webサイトを通して、ヒトに関するデータの利用及びヒトに関するデータの提供を行なうことができます。

なお、本データベースの目的・意義、扱うデータの種類、データ利用者の範囲、責任者については[こちら](#)をご覧ください。

新着情報

2019/05/28 理化学研究所 生命医科学研究センターからの制限公開データ (Type I) を公開しました (hum0160)

2019/05/09 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 (hum0141)

▶ [ニュース一覧へ](#)

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

NBDC Human Database Beacon is a member of GA4GH Beacon Network.

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

利用可能な研究データ一覧

データ利用方法は[こちら](#)をご覧下さい。

全 104 件 Copy CSV Excel 一覧内検索:

Research ID	研究題目	公開日	データの種類	研究方法	手法	参加者 (対象集団)	提供者	アクセス制限
hum0173.v1 JGAS00000000171	胎児期に始まる子どもの健康と発達に関する調査	v1:2019/04/23	NGS (RRBS)	メチル化 解析	Illumina (NextSeq)	臍帯組織 (正常出産) : 30検体 (日本人)	森千里	制限 (Type I)
hum0165.v1 JGAS00000000162	家族性骨髄異形成症候群の遺伝子解析研究	v1:2019/05/07	NGS (Exome)	配列決定	Illumina (HiSeq 2500)	家族性骨髄異形成症候群: 2症例 (1家系) (日本人)	古屋淳史	制限 (Type I)

<https://humandbs.biosciencedbc.jp/>

研究者の産出したデータを、他の研究者が再利用できるサービス

59

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年欧洲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)

NBDCヒトデータベース

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

昨年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ヒトデータベースのデータ例

J-ADNI:hum0043

NBDC Research ID: hum0043.v1

研究内容の概要

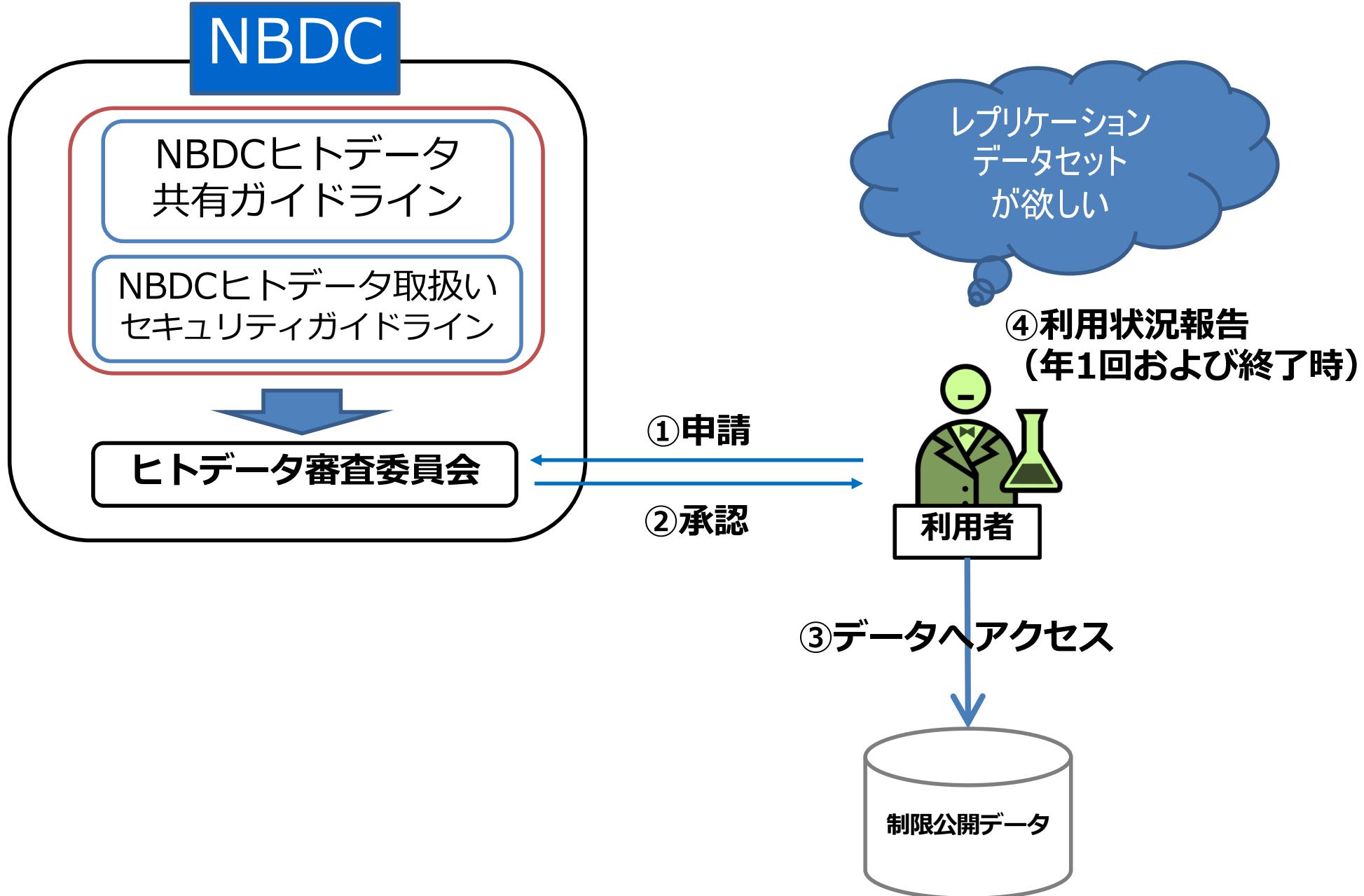
目的 : アルツハイマー病（AD）治療薬の薬効評価基準の最適化を行うために、ADの病態を忠実に反映するサロゲートマーカー (surrogate marker)を定めておく必要がある。J-ADNI研究は、MRIやPETなどの画像サロゲートマーカーの長期的変化に関する一定の基準値を作成するとともにその妥当性を証明するために臨床/神経心理検査データ、血液・脳脊髄液サンプルを並行して収集することを目的とする。

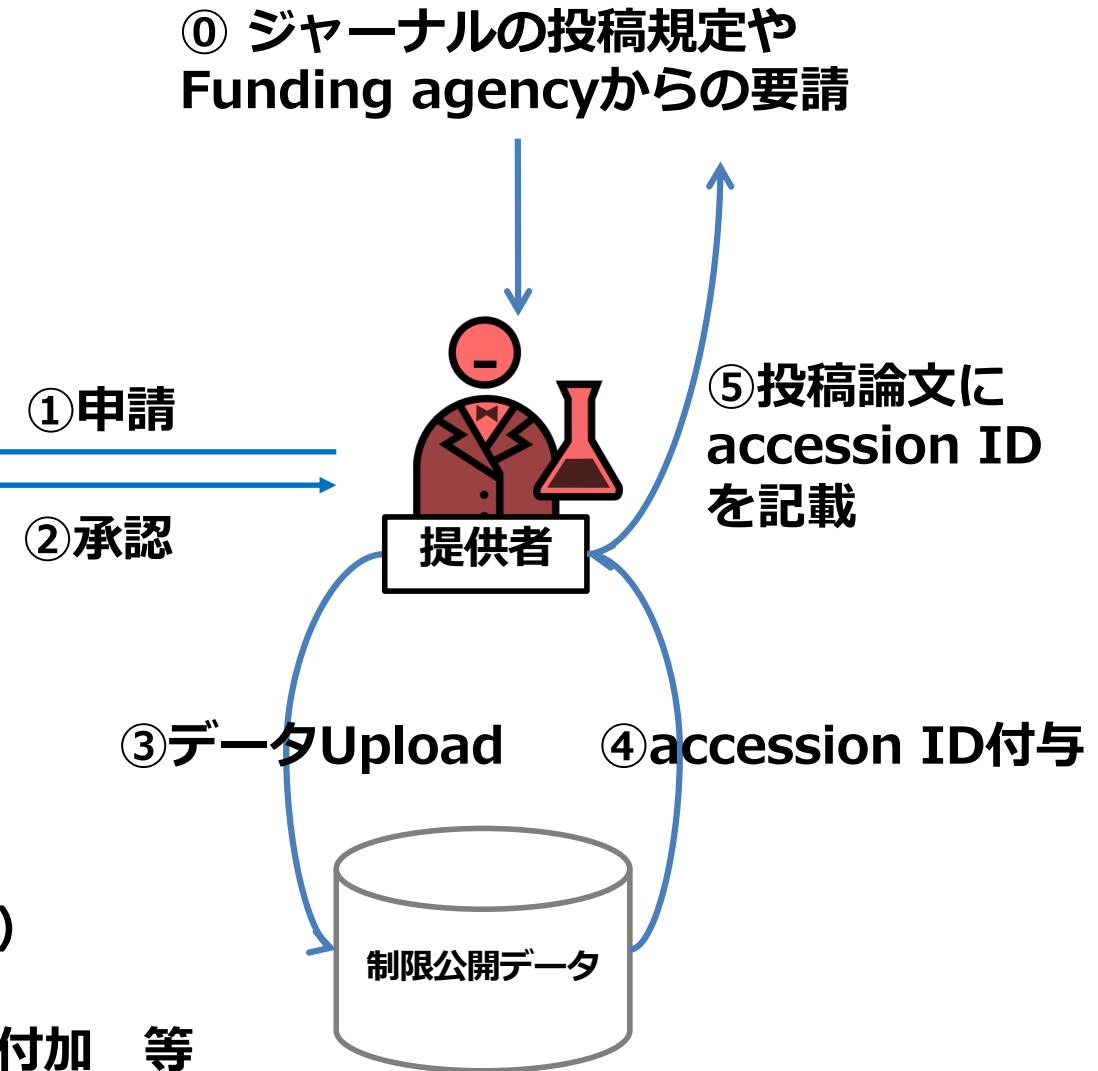
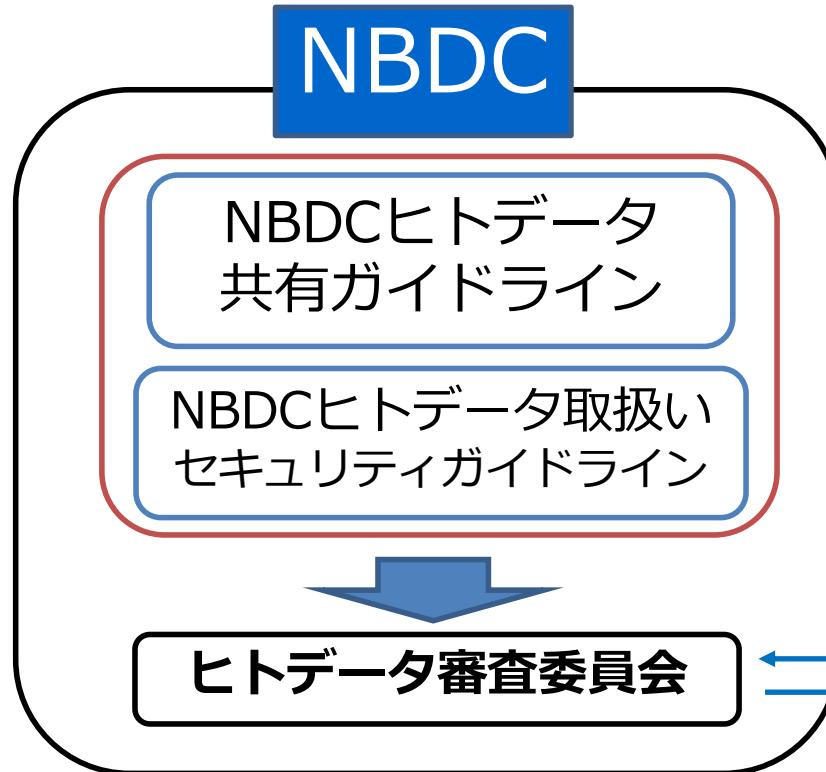
対象 :

537名

(高齢者健常群 149名、軽度認知機能障害 234名、
早期アルツハイマー病 154名)

- 臨床・心理学検査結果
- 検査結果
ApoE遺伝子型、末梢血液一般検査
- MRI画像、PET画像





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

ご清聴いただき、誠にありがとうございました。

Appendix

その他の多型データベース

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