

# ヒトDB・ヒト多型解析 (TogoVar、HGVDなど)

豊岡 理人

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

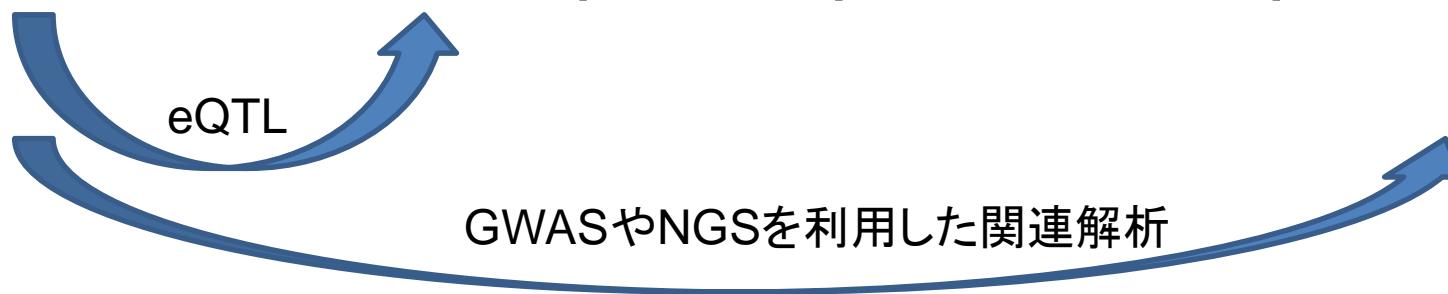
2021年 1月13日  
於 AJACSオンライン5



# 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

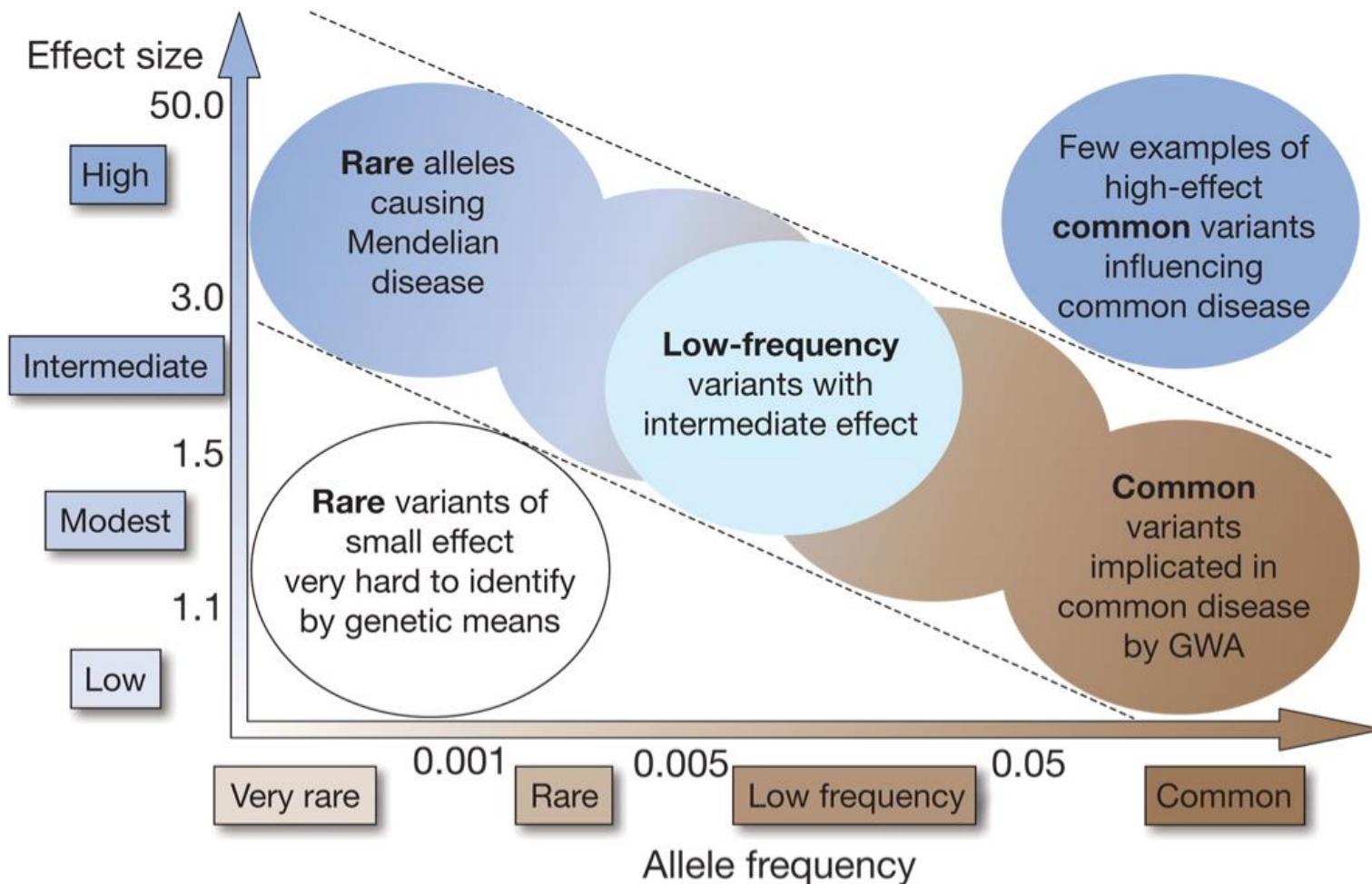
gene → transcript → protein → phenotype



# 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、トランスポゾン、レトロトранスポゾン
- 構造多型 (SV: Structural Variant)

# バリエントの頻度と形質への影響との関係



(Manolio, T. A, et al. (2009). Finding the missing heritability of complex diseases. Nature, 461(7265), 747–753. )

# 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。

- 希少疾患

CommonのVariantはPrimaryの探索から外されることが多い。

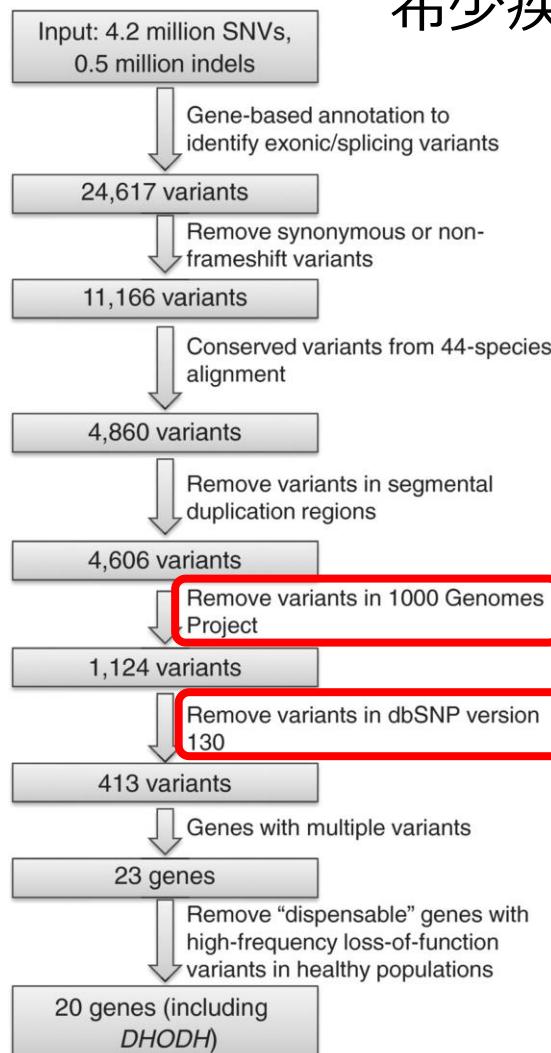
- Trio解析

- 発症者（子） + 未発症（父母）の3人から発症者における  
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

# バリアントの頻度情報を公開する意義

## 希少疾患の責任遺伝子探索における公共DBの利用例



トリオ解析等から出て来た解析対象のバリアントは膨大



公共データベースに存在するバリアントを絞り込みに利用  
(希少疾患関連のサンプルは含まれていない前提)

絞り込みには公共データベースを参照する必要あり  
(1000 Genomes Project, dbSNP , etc. )

ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data  
Nucleic Acids Res. 2010;38(16):e164. doi:10.1093/nar/gkq603

# 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

例えば、ヒト1番染色体

actcagtcattgactactactgatccgaaattc  
actcagtcattgactactactgatccgaaattc

genotype

haplotype

父由来

母由来

- SNP Chip

- 60-90万箇所のgenotype情報

- 次世代シーケンサー

- 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： 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/バリエ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
  - ALFA: Allele Frequency Aggregator

# 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**

- dbSNP 20th Anniversary
- Overview of dbSNP
- About Reference SNP (rs)
- Factsheet
- Entrez Updates (May 26, 2020)

**Submission**

- How to Submit
- Hold Until Published (HUP) Policies
- Submission Search

**Access Data**

- Web Search
- eUtils API
- Variation Services
- FTP Download
- Tutorials on GitHub

**ALFA Project Inaugural Release March 10, 2020**

The goal is to provide allele frequency from more than 1 million dbGaP subjects with regular updates. Visit the project [page](#) for more information or view the introduction video below. Please provide your feedback by completing this short 3 min [survey](#).



**How to Search dbSNP. Additional search terms are [here](#).**

|   |  |
|---|--|
| All of dbSNP (then use filters on results page)   | all[sb]  |
| dbSNP RefSNP ID   | Single: 328: Multiple 328 226,200                    |
| Gene  | Gene symbol PTEN[Gene Name] or gene ID 4023[Gene_ID] |
| Genomic location of a single position or range on GRCh38. See the <a href="#">announcement</a> and the <a href="#">guide</a> for using GRCh38 coordinates | 6[Chromosome] AND (1500000:3000000[Base_Position])   |

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

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**

- dbSNP 20th Anniversary
- Overview of dbSNP
- About Reference SNP (rs)
- Factsheet
- Entrez Updates (May 26, 2020)

**Submission**

- How to Submit
- Hold Until Published (HUP) Policies
- Submission Search

**Access Data**

- Web Search
- eUtils API
- Variation Services
- FTP Download
- Tutorials on GitHub

**i ALFA Project Inaugural Release March 10, 2020**

The goal is to provide allele frequency from more than 1 million dbGaP subjects with regular updates. Visit the project [page](#) for more information or view the introduction video below. Please provide your feedback by completing this short 3 min [survey](#).

ALFA CoLab Presentation, ASHG2020

NCBI ALFA

National Library of Medicine

How to Search dbSNP. Additional search terms are [here](#).

|   |  |
|---|--|
| All of dbSNP (then use filters on results page)   | all[sb]  |
| dbSNP RefSNP ID   | Single: 328: Multiple 328 226,200                    |
| Gene  | Gene symbol PTEN[Gene Name] or gene ID 4023[Gene ID] |
| Genomic location of a single position or range on GRCh38. See the <a href="#">announcement</a> and the <a href="#">guide</a> for using GRCh37 coordinates | 6[Chromosome] AND (1500000-3000000[Base Position])   |

# dbSNPの拡張検索の組み立て

use builder below to create your search

The screenshot shows the dbSNP search builder interface. On the left is a sidebar with various search fields: Accession, All Fields (which is selected and highlighted in blue), Allele, Base Position, Chromosome, Clinical Significance, Contig Position, Create Build ID, Filter, Function Class, Gene Description, Gene Name (with a note about dbSNP 141), Genotype, Global Minor Allele Frequency, Local SNP ID, LocusLink ID, and Method Class. The main area has a 'Query' input field containing 'rs11568821 Sort by: SNP\_ID'. Below it are two more entries: 'Sort by: SNP\_ID' and '03962 Sort by: SNP\_ID'. At the bottom, there's a section titled 'Associated Databases (dbSNP)' with two columns: 'POPULAR' (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP) and 'FEATURED' (Genetic Testing Registry, GenBank, Reference Sequences, Gene Expression Omnibus, Genome Data Viewer, Human Genome, Mouse Genome).

検索対象を使用しない場合は、検索語を All Fields に対して検索する

# dbSNPの検索結果 (*PDCD1* rs11568821)

NCBI Resources How To Sign in to NCBI

dbSNP SNP PDCD1 rs11568821 Create alert Advanced 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

**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

2. Variant type: SNV Alleles: C>G,T Chromosome: NT\_187527.1:63858 Gene: LOC105373977 (GeneView), PDCD1 (GeneView)

**rs606231173 has merged into rs11568821 [Homo sapiens]**

**検索式の表示**

Find related data Database: Select Find items

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

# dbSNPの検索結果（詳細画面）

National Library of Medicine  
National Center for Biotechnology Information

**dbSNP** Short Genetic Variations

Search for terms  Search Examples: rs268, BRCA1 and [more](#) Advanced search

**i Welcome to the Reference SNP (rs) Report**  
All alleles are reported in the [Forward orientation](#). Click on the [Variant Details tab](#) for details on Genomic Placement, Gene, and Amino Acid changes.  
HGVS names are in the [HGVS tab](#).

Reference SNP (rs) Report

[Switch to classic site](#) Download [Facebook](#) [Twitter](#) [Email](#) [Help](#)

**rs11568821** Current Build 154 Released April 21, 2020

|                |   |                       |   |
|----------------|---|-----------------------|---|
| Organism       | Homo sapiens  | Clinical Significance | Reported in <a href="#">ClinVar</a>                         |
| Position       | chr2:241851760 (GRCh38.p12)   | Gene : Consequence    | PDCD1: Intron Variant<br>LOC105373977: 2KB Upstream Variant |
| Alleles        | C>G / C>T   | Publications          | <a href="#">37 citations</a><br><a href="#">LitVar</a>      |
| Variation Type | SNV Single Nucleotide Variation   | Genomic View          | <a href="#">See rs on genome</a>                            |
| Frequency      | T=0.074645 (9373/125568, TOPMED)<br>T=0.07214 (2260/31330, GnomAD)<br>T=0.08953 (1591/17770, ALFA Project) ( <a href="#">+ 9 more</a> ) |                       |   |

**Variant Details**

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

Gene: [PDCD1](#), programmed cell death 1 (minus strand)

| Molecule type               | Change                                       | Amino acid[Codon] | SO Term                             |
|-----------------------------|--|-------------------|-------------------------------------|
| PDCD1 transcript            | NM_005018.3:c.627+189G>C                     | N/A               | Intron Variant                      |
| PDCD1 transcript variant X1 | <a href="#">XM_017004293.1:c.627+189G...</a> | N/A               | Intron Variant                      |
| PDCD1 transcript variant X2 | XM_006712573.2:c.                            | N/A               | Genic Downstream Transcript Variant |

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(ALFA)  
HGVS表記  
Submission情報  
rs番号の変遷  
文献情報  
近傍のゲノム配列  
をTabで切り替えられます

ClinVarへのリンク

# ALFA: Allele Frequency Aggregator の紹介

dbGaPにあるデータセットを統一的なプロトコールで解析した頻度情報データセット  
<https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/>

## Input and Output Counts

| Input              | Count              |
|--------------------|--------------------|
| Studies            | 42                 |
| Subjects           | 98,494             |
| Genotypes          | 551,345,630,054    |
| Genotypes Excluded | 689,601,274 (0.1%) |

| Output             | Count       |
|--------------------|-------------|
| Total RefSNPs      | 446,909,594 |
| Exist in dbSNP 153 | 442,539,572 |
| Novel              | 4,370,022   |

| Input Assay Source | Subjects* |
|--------------------|-----------|
| Exome              | 23,224    |
| Genomes            | 5,621     |
| SNP Arrays         | 87,232    |

\* Subject counts for different assay source can be overlapping.

| Output Population | BioSample ID | Subjects | Total Site Count | MAF = 0     | MAF >= 0.01 | 0.01 > MAF >= 0.001 | MAF < 0.001 | Singleton  |
|-------------------|--------------|----------|------------------|-------------|-------------|---------------------|-------------|------------|
| European          | SAMN10492695 | 82,475   | 445,828,538      | 413,136,224 | 8,616,684   | 4,946,828           | 432,265,026 | 15,684,278 |
| African American  | SAMN10492698 | 3,555    | 445,809,613      | 428,151,478 | 17,105,759  | 545,913             | 428,157,941 | 5,164,009  |
| African Others    | SAMN10492696 | 114      | 445,691,099      | 438,202,529 | 7,470,894   | 17,676              | 438,202,529 | 3,251,737  |
| African (Note 1)  | SAMN10492703 | 3,669    | 445,809,758      | 427,767,482 | 17,490,705  | 544,974             | 427,774,079 | 5,249,871  |
| East Asian        | SAMN10492697 | 153      | 445,720,538      | 442,454,208 | 3,246,297   | 20,033              | 442,454,208 | 1,646,717  |
| South Asian       | SAMN10492702 | 2,459    | 445,079,683      | 440,460,693 | 4,603,791   | 7,562               | 440,468,330 | 2,361,549  |
| Other Asian       | SAMN10492701 | 84       | 445,465,162      | 442,628,722 | 2,828,952   | 7,488               | 442,628,722 | 1,741,650  |
| Asian (Note 2)    | SAMN10492704 | 237      | 445,757,479      | 441,170,505 | 4,548,632   | 38,342              | 441,170,505 | 2,079,552  |
| Latin American 1  | SAMN10492699 | 354      | 3,176,558        | 1,403,168   | 1,752,470   | 20,920              | 1,403,168   | 254,597    |
| Latin American 2  | SAMN10492700 | 3,801    | 3,194,467        | 1,173,905   | 1,949,095   | 68,415              | 1,176,957   | 206,300    |
| Other             | SAMN11605645 | 5,499    | 445,829,307      | 434,443,847 | 10,314,708  | 450,584             | 435,064,015 | 3,604,601  |
| Total (Note 3)    | SAMN10492705 | 98,494   | 445,835,693      | 404,944,004 | 9,245,548   | 8,393,159           | 428,196,986 | 18,638,600 |

# dbSNPでの遺伝子名を検索語にした検索

NCBI Resources How To

dbSNP SNP PDCD1 Create alert Advanced

Clinical Significance: benighn, risk factor

Validation Status: by-ALFA, by-cluster, by-frequency

Publication: LitVar Annotated, PubMed Cited, PubMed Linked

Function Class: inframe deletion, inframe insertion, intron, missense, non coding transcript variant, synonymous

Variation Class: del, delins, ins, mnv

Annotation: ..

Display Settings: Summary, 20 per page, Sorted by SNP\_ID

Send to: << First < Prev Page 1 of 158 Next > Last >>

**Search results**

Items: 1 to 20 of 3150

rs11568821 [*Homo sapiens*]

1.

|                         |   |
|-------------------------|---|
| Variant type:           | SNV   |
| Alleles:                | C>G,T [Show Flanks]   |
| Chromosome:             | 2:241851760 (GRCh38)  |
|                         | 2:242793912 (GRCh37)  |
| Canonical SPDI:         | NC_000002.12:241851759:C:G,NC_000002.12:241851759:C:T   |
| Gene:                   | PDCD1 ( <a href="#">Varview</a> ), LOC105373977 ( <a href="#">Varview</a> )   |
| Functional Consequence: | intron_variant,upstream_transcript_variant,downstream_transcript_variant,genic_downstream_transcript_varia...   |
| Clinical significance:  | risk-factor   |
| Validated:              | by frequency,by alfa,by cluster   |
| MAF:                    | T=0.089533/1591 ( <a href="#">ALFA</a> )  |
|                         | T=0.040935/205 (1000Genomes)  |
|                         | T=0.071667/43 (NorthernSweden)  |
|                         | ...more   |
| HGVS:                   | NC_000002.12:g.241851760C>G, NC_000002.12:g.241851760C>T,<br>NC_000002.11:g.242793912C>G, NC_000002.11:g.242793912C>T,<br>NG_012110.1:g.12147G>C, NG_012110.1:g.12147G>A, NT_187527.1:g.63858C>G,<br>NT_187527.1:g.63858C>T |
|                         | ...more   |
| PubMed                  | LitVar  |

# dbSNPの遺伝子ビュー(Variation Viewer)

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

## Variation Viewer

Homo sapiens (human)

Assembly: GRCh38.p12 (GCF\_000001405.38) • Chr 2 (NC\_000002.12)

Search assembly: rs11568821 Examples ▾

Pick Assembly: GRCh38.p12 Annotation Release 109

User Data and Track Hubs: - no added tracks or track hubs Options ▾

Supported File/Data types: Clinical, dbSNP b154 v2 Live RefSNPs, dbSNP b154 v2

History: --- select gene --- Select gene from history

Assembly Region Details: Features of Interest Other sequence representations for assembly region(s) in view - none - 2 GRC genome issues in this view. Add Track

Variation Data: Filter by Source database: dbSNP (6246), dbVar (568) In ClinVar: Yes (183), No (6631) Most severe clinical significance: pathogenic (132)

Download edit columns Items 1 - 30 of 6,814 << First < Prev Page 1 of 228 Next > Last >

| Variant ID | Location             | Variant type          | Gene                 | Molecular consequences | Most severe clinical significance | 1000G MAF | GO-ESP MAF | ExAC MAF | Publications |
|------------|----------------------|-----------------------|----------------------|------------------------|-----------------------------------|-----------|------------|----------|--------------|
| nsv2776363 | 12,771 - 241,841,232 | copy number variation | H2ACP2 and 3864 more |                        | Pathogenic                        |           |            |          | 1            |
| nsv3882615 | 12,771 - 241,841,232 | copy number variation | H2ACP2 and 3864 more |                        | Pathogenic                        |           |            |          |              |
| nsv429565  | 12,784 - 242,147,305 | copy number variation | H2ACP2 and 3877 more |                        |                                   |           |            |          | 1            |
| nsv3885544 | 14,238 - 242,106,609 | copy number variation | H2ACP2 and 3875 more |                        | Pathogenic                        |           |            |          |              |
| nsv3874648 | 15,672 - 242,157,305 | copy number variation | H2ACP2 and 3877 more |                        | Pathogenic                        |           |            |          |              |

dbSNPおよびdbVar, ClinVarの情報を genome browser上で確認可能

# dbSNP以外からの利用方法

NCBI Resources How To Sign in to NCBI

All Databases  Search

**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、、、

# NCBIの統合検索機能を利用した情報の取得

The screenshot shows the NCBI search results for the query "CYP2D6". The search bar at the top contains "CYP2D6". Below the search bar, it says "Results found in 27 databases".

**GENE** section:

- CYP2D6 – cytochrome P450 family 2 subfamily D member 6**
- Homo sapiens (human)
- Also known as: CPD6, CYP2D, CYP2D7AP, CYP2D7BP, CYP2D7P2, CYP2D8P2, CYP2DL1, CYPIID6, P450-DB1, P450C2D, P450DB1
- GeneID: 1565
- RefSeq transcripts (2) RefSeq proteins (2) RefSeqGene (1) PubMed (989)
- Orthologs, Genome Browser, BLAST, Download buttons

**RefSeq transcripts** section:

| Transcript     | Isoform | Length (nt) |
|----------------|---------|-------------|
| NM_001025161.3 | 2       | 1,435       |
| NM_000106.6    | 1       | 1,588       |

**RefSeq Sequences** section:

- Literature**
  - Bookshelf (719)
  - MeSH (5)
  - NLM Catalog (11)
  - PubMed (7,621)
  - PubMed Central (11,469)
- Genes**
  - Gene (96)
  - GEO DataSets (13)
  - GEO Profiles (50,222)
  - HomoloGene (2)
  - PopSet (12)
- Proteins**
  - Conserved Domains (0)
  - Identical Protein Groups (16)
  - Protein (615)
  - Protein Clusters (0)
  - Protein Family Models (0)
  - Structure (24)
- Genomes**
  - Assembly (0)
  - BioCollections (0)
- Clinical**
  - ClinicalTrials.gov (0)
  - ClinVar (2,279)
- PubChem**
  - BioAssays (4,317)
  - Compounds (0)

At the bottom, there are navigation links for "tecpub" and "Research\_proposal...zip", and a "すべて表示" button.

# 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集団におけるアリル頻度の違いを表示

# 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
  - v.2 (GRCh 37)
    - 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
  - v.3 (GRCh 38)
    - 76,156名の全ゲノムシークエンスデータ  
(mtは、56,434名)

# gnomADの画面説明、操作



gnomAD browser gnomAD v2.1.1 Search About News Downloads Terms Publications Contact FAQ  
gnomAD v3.1 released!

# gnomAD

genome aggregation database

gnomAD v2.1.1 Search by gene, region, or variant

Please note that gnomAD v2.1.1 and v3.1 have substantially different but overlapping sample compositions and are on different genome builds. For more information, see the FAQ "Should I switch to the latest version of gnomAD?"

**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 v2 data set (GRCh37/hg19) 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 genetic studies. The v3.1 data set (GRCh38) spans 76,156 genomes, selected as in v2. The gnomAD Principal Investigators and groups that have contributed data to the current release are listed [here](#).

All data here are released for the benefit of the wider biomedical community, without restriction on use - see the [terms of use](#). Sign up for our [mailing list](#) for future release announcements.

# gnomADの検索結果(遺伝子)

gnomAD browser

gnomAD v2.1.1 ▾ Search

About News Downloads Terms Publications Contact FAQ

gnomAD v3.1 released!

## PCSK9 proprotein convertase subtilisin/kexin type 9

Dataset gnomAD v2.1.1 ▾ gnomAD SVs v2.1 ▾ ?

|                                       |                                 |
|---------------------------------------|---------------------------------|
| <b>Genome build</b>                   | GRCh37 / hg19                   |
| <b>Ensembl gene ID</b>                | ENSG00000169174.9               |
| <b>Ensembl canonical transcript</b> ⓘ | ENST00000302118.5               |
| <b>Region</b>                         | 1:55505221-55530525             |
| <b>References</b>                     | Ensembl, UCSC Browser, and more |

### Constraint ⓘ

| Category   | Exp. SNVs | Obs. SNVs | Constraint metrics                         |
|------------|-----------|-----------|--|
| Synonymous | 187.5     | 170       | Z = 1.01<br>o/e = 0.91 (0.8 - 1.03) 0 — 1  |
| Missense   | 435       | 419       | Z = 0.27<br>o/e = 0.96 (0.89 - 1.04) 0 — 1 |
| pLoF       | 26.9      | 26        | pLI = 0<br>o/e = 0.97 (0.71 - 1.34) 0 — 1  |

Per-base mean depth of coverage

Include:  CDS  UTR  Non-coding transcript

Show transcripts

Show tissues Mean pext ⓘ

ClinVar pathogenic and likely pathogenic variants (16)

gnomAD v2.1.1 (1197)

Viewing in table

55,505,436 55,509,494 55,512,244 55,517,896 55,518,396 55,521,858 55,523,261 55,524,184 55,525,283 55,527,256 55,529,329

pLoF only  Missense only  Synonymous only  Other only  Exomes  SNVs  Genomes  Indels  Filtered variants ?

遺伝子上に存在する多型とその頻度

# NGS Coverage

# 遺伝子の構造

# 遺伝子上に 存在する多型とその頻度

# 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  |
|--------------------------------------|-------------------------------------|-------------------------------------|--------|
| <input checked="" type="checkbox"/>  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |        |
| Allele Count                         | 19607                               | 4488                                | 24095  |
| Allele Number                        | 161082                              | 31262                               | 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-ACCTCTGCTG
- \* 1-55505552-A-ACCTCTCTGCTG
- \* 1-55505552-ACTG-A
- \* 1-55505552-ACCTGCTG-A

**Annotations**

This variant falls on 2 transcripts in 1 gene.

inframe insertion

- \* PCSK9
  - \* ENST00000302118 \*
  - HGVSp: p.Leu23dup
  - \* ENST00000462118
  - HGVSp: 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          | 5672          | 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

**References**

- \* dbSNP (rs371488778)
- \* UCSC
- \* ClinVar

**Report**

- \* Report this variant
- \* Request additional information

**Dataset** gnomAD v2.1.1

- gnomAD v2.1.1 (all samples)
- gnomAD v2.1.1 (controls)
- gnomAD v2.1.1 (non-cancer)
- gnomAD v2.1.1 (non-neuro)
- gnomAD v2.1.1 (non-neuro)
- gnomAD v2.1.1 (non-TOPMed)
- gnomAD v2.1.1 (TOPMed)

サブセットでの頻度の表示切り替え可能

Population毎、  
性別毎の  
頻度情報

# 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/202008/>
- 日本人の約8,300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

 jMorp

|  Sequence |  Variation |  Methylome |  Transcriptome |  Proteome |  Metabolome |  Repository |  GWAS | Downloads | Help | Login



Welcome to  
Japanese Multi Omics  
Reference Panel.

|   |   |
|---|---|
|    | <b>Phenome</b><br>To be provided              |
|    | <b>Metabolome</b>                             |
|    | <b>Proteome</b>                               |
|   | <b>Transcriptome</b><br>IMM (iMethyl) & ToMMo |
|  | <b>Methylome</b><br>IMM (iMethyl)             |
|  | <b>Genome Variation</b>                       |
|  | <b>Genome Sequence</b>                        |

**jMorp release 202008**

November 17th, 2020  
**New Articles were published.**  
New articles about [metabolome GWAS](#) and [jMorp database](#) have been published on Communications Biology and Nucleic Acids Research respectively.

November 12nd, 2020  
**Iso-Seq and GWAS summary statistics datasets have been added.**  
[Transcriptome (Iso-Seq)] We released transcriptome data for lymphoblastoid cell lines (LCLs) using long-read sequencing technology (PacBio Isoform-Sequencing; Iso-Seq).  
[GWAS summary statistics] We have added GWAS summary statistics of metabolome-GWAS study (TGA000003). Manhattan plots have been added in each compound page in the metabolome layer, which have been reported in this study.

August 31st, 2020  
**JG2 Japanese Reference Genome, 8.3KJPN allele frequency panel, and 25K Metabolome**  
[Genome Sequence] We updated the Japanese reference genome, JG2, which is constructed by integrating six de novo phased assemblies from three Japanese male individuals.  
[Genome Variation] 8.3KJPN, allele and genotype frequency panels from 8,380 Japanese

# jMorpの検索方法

The screenshot shows the jMorp search interface. At the top, there is a navigation bar with icons for Sequence, Variation, Methylome, Transcriptome, Proteome, Metabolome, Repository, Downloads, Help, and Login. Below the navigation bar, the jMorp logo is on the left, followed by the text "jMorp ~Japanese Multi Omics Reference Panel~". On the right of the header are "Statistics" and "Downloads" buttons. The main content area has a yellow background and features a large green search bar. Inside the search bar, there are three input fields: "Search by gene name", "Search by rs#", and "Search by region (GRCh37/hg19)". Below these fields is a "Gene symbol" input field, which is highlighted with a blue rounded rectangle and has a large blue arrow pointing upwards from below it. Underneath the "Gene symbol" field, the text "Examples: ALDH2, NFE2L2, GATA1" is displayed. To the right of the "Gene symbol" field is a green "Search" button. At the bottom of the page, there is a footer bar with the text "jMorp release 202008 / LastUpdate: November 17th, 2020 / Conditions of Use", "Tohoku Medical Megabank Organization, Tohoku University", and "Contact: imorp [at] omics.megabank.tohoku.ac.jp".

Genomic Variants (8.3KJPN)

Search by gene name    Search by rs#    Search by region (GRCh37/hg19)

Gene symbol

Examples: ALDH2, NFE2L2, GATA1

Search

遺伝子名、rs番号、領域を指定し、検索可能

jMorp release 202008 / LastUpdate: November 17th, 2020 / Conditions of Use  
Tohoku Medical Megabank Organization, Tohoku University  
Contact: imorp [at] omics.megabank.tohoku.ac.jp

# jMorpの検索結果

Sequence | Variation | Methylome | Transcriptome | Proteome | Metabolome | Repository | Downloads | Help | Login

jMorp ~Japanese Multi Omics Reference Panel~

Search by gene: ALDH2 (8.3KJPN) GRCh37/hg19

gnomADにおける頻度

ToMMoにおける頻度

4122 variants found

Filter by keyword

Browser Plot

| Type | Position ↑   | Ref/A <sub>t</sub> | rs#          | Annotation ↑          | Gene ↓ | MeanDepth (162PE) | JPA. | ClinVar Annotation | ToMMo 8.3KJPN | gnomAD AFR. | gnomAD AMR. | gnomAD ASJ. | gnomAD EAS. |
|------|--------------|--------------------|--------------|-----------------------|--------|-------------------|------|--------------------|---------------|-------------|-------------|-------------|-------------|
| SNV  | 12:112199811 | A/T                | rs1445408349 | upstream_gene_variant | ALDH2  | 22.0/22.0         |      |                    | 0.0002        |             |             |             |             |
| 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      |             |             |             |
| 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      |             |             |             |
| SNV  | 12:112199914 | G/A                | rs553976397  | upstream_gene_variant | ALDH2  | 22.2/22.2         |      |                    |               | 0.0002      |             |             |             |
| SNV  | 12:112199922 | A/G                | rs999175995  | upstream_gene_variant | ALDH2  | 22.2/22.2         |      |                    |               | 0.0002      |             |             |             |
| SNV  | 12:112199946 | G/A                | rs1186433838 | upstream_gene_variant | ALDH2  | 21.7/21.6         |      |                    |               | 0.0001      |             |             |             |

# 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<br>(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<br>(p.Glu504Lys) | ALDH2 | 18.4/18.4            | V1&V2 |  | 0.1959      | 0.0002      | 0.0012      |              | 0.2671      |  |

Blast-It! Blast-It!2 Blast-It!3 Blast-It!4 Blast-It!5 Blast-It!6 Blast-It!7 Blast-It!8 Blast-It!9 Blast-It!10

**Blast result:**

mRNA info: Homo sapiens aldehyde dehydrogenase 2 family (mitochondrial) (ALDH2), transcript variant 1, mRNA.  
mRNA change: NM\_000690.3(o.1510G>A)\_NP\_000681.2(p.504E>K)  
Query: ALDH2 (GeneID: 217) GF | NM\_000690.3 GF | NP\_000681.2 GF  
Subject: 51338  
e-value: 0.0000  
Sequence Identity: 100.00

5H13\_entity\_1 (1/1)

Chain:A (1/8)

Single:

- Secondary Structure: E
- Relative ASA: 0.374

BioUnit:1

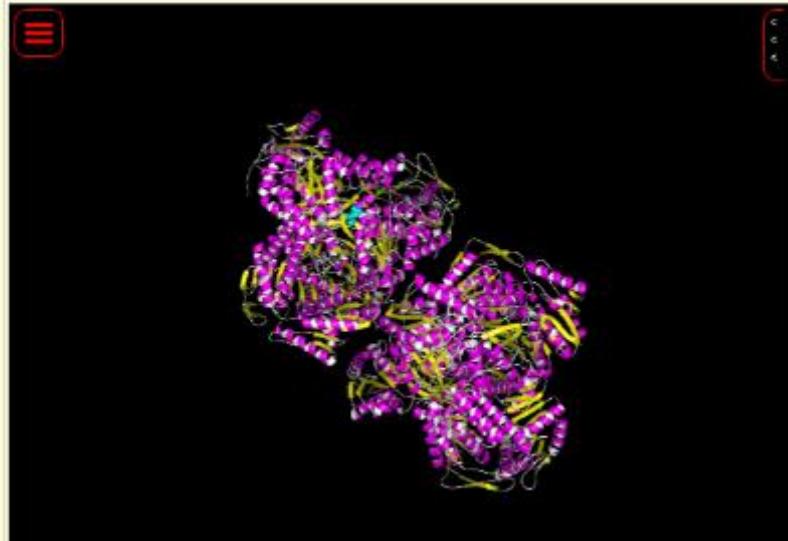
- delta Relative ASA: 0.295

BioUnit:2

- delta Relative ASA: NA

Mapped Position

- auth\_asym\_id: A, auth\_seq\_id: 437
- label\_asym\_id: A, label\_seq\_id: 504



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)

# 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



### 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?

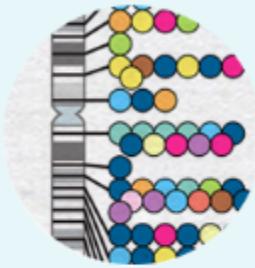
- 08/02/2017 HGVD version 2.3 is now downloadable.
- 04/17/2017 New HLA typing software 'HLA-HD' is released. [link](#)

# 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(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

Home Diagram Download Documentation About EMBL-EBI NIH National Human Genome Research Institute

## 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

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.

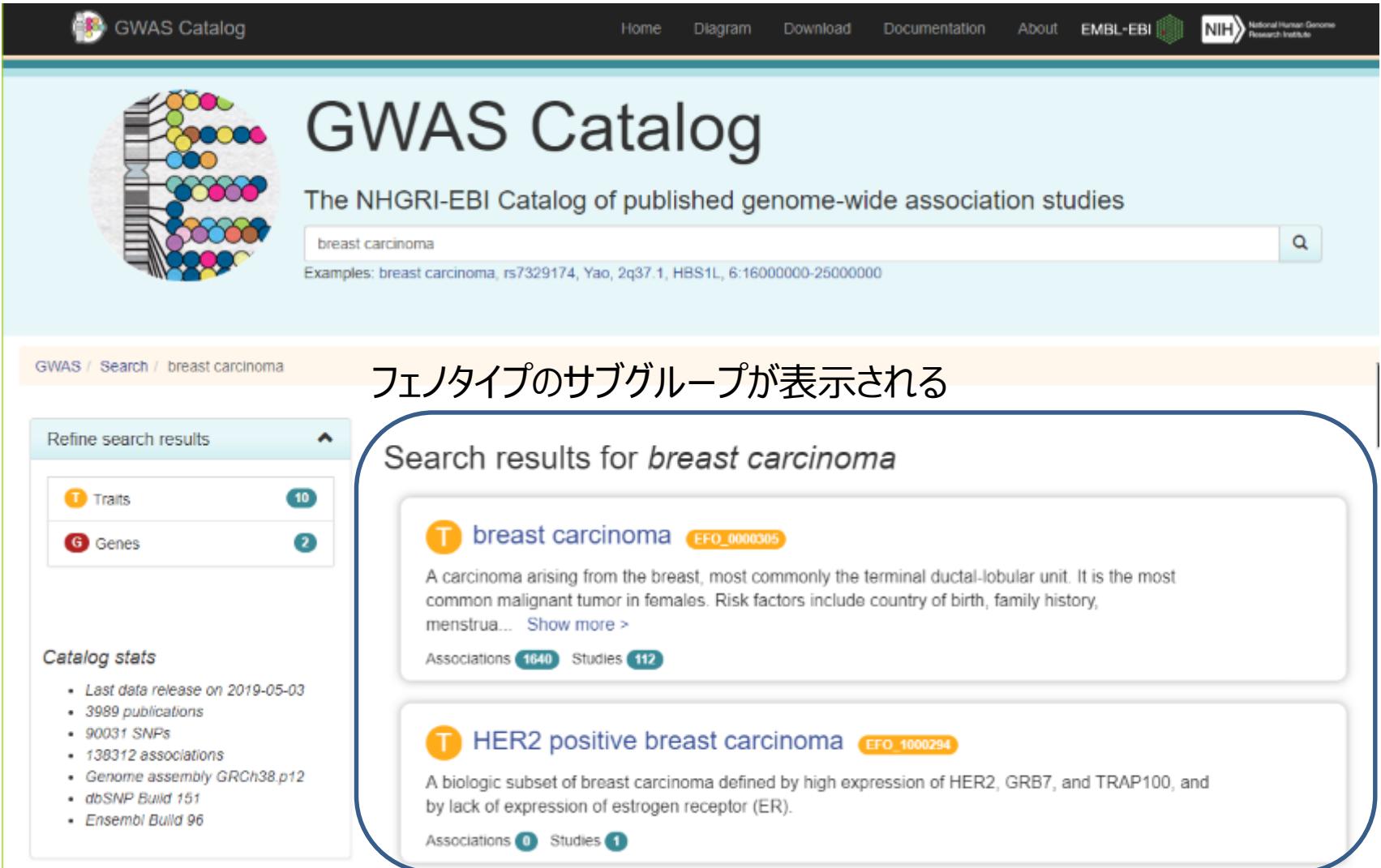
# 多型データベースの紹介 GWAS-catalog

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. At the top, there is a navigation bar with links to Home, Diagram, Download, Documentation, About, EMBL-EBI, and NIH. Below the navigation bar, there is a search bar with the placeholder "breast carcinoma" and a magnifying glass icon. Below the search bar, there is a text input field with the placeholder "Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000". On the left side, there is a sidebar titled "Refine search results" with two categories: "Traits" (10) and "Genes" (2). Below the sidebar, there is a section titled "Catalog stats" with a list of statistics: Last data release on 2019-05-03, 3989 publications, 90031 SNPs, 138312 associations, Genome assembly GRCh38.p12, dbSNP Build 151, and Ensembl Build 96. The main content area is titled "Search results for *breast carcinoma*". It displays two results: "breast carcinoma" (EFO\_0000305) and "HER2 positive breast carcinoma" (EFO\_1000294). Both results show a brief description, the number of associations (1640 and 0 respectively), and the number of studies (112 and 1 respectively).

フェノタイプのサブグループが表示される

GWAS / Search / breast carcinoma

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

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

# 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             | $7 \times 10^{-6}$ |                    | 0.22 | 1.32 | -    | [1.17-1.44] | MTOL1                    | Breast cancer                             | breast carcinoma                           | GCST002735      | 18:8716323   |
| rs13025833-A            | $1 \times 10^{-7}$ |                    | 0.52 | 1.2  | -    | [1.13-1.33] | CTNNA2                   | Breast cancer                             | breast carcinoma                           | GCST002735      | 2:79716982   |
| rs1078806-C             | $2 \times 10^{-6}$ | (Pooled P value)   | 0.39 | 1.43 | -    | [NR]        | FGFR2                    | Breast cancer                             | breast carcinoma                           | GCST001831      | 10:121579461 |
| rs166870-T              | $3 \times 10^{-7}$ | (HR+ HER2-)        | 0.13 | 2.3  | -    | [1.67-3.15] | AC026826.2,<br>RNU6-667P | Disease-free survival<br>in breast cancer | breast carcinoma,<br>disease free survival | GCST002847      | 15:79777420  |
| rs10825036-G            | $4 \times 10^{-7}$ | (HR- HER2-)        | 0.32 | 2.26 | -    | [1.34-3.81] | RNA5SP318,<br>AL365496.1 | Disease-free survival<br>in breast cancer | disease free survival,<br>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                 |

# 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ヒトデータベースとは？

**IBDC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を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再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 (DDJ)**と協力して、ヒトに関するデータを公開しています。

本Webサイトを通じて、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 球化研究所 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]

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

RCh37 e.g. 12:112241766 A Search Example: ALDH2 Variant (GRCh37, '12:112241766 A')

## 可能な研究データ一覧

再利用方法は[こちら](#)をご覧下さい。

14 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<br>(対象集団)                  | 提供者   | アクセス制限         |
|----------------------------|------------------------|---------------|---------------|------------|-----------------------|--------------------------------|-------|----------------|
| m0173.v1<br>AS000000000171 | 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<br>(RRBS) | メチル化<br>解析 | Illumina<br>(NextSeq) | 臍帯組織(正常出<br>産) : 30検体<br>(日本人) | 森 千里  | 制限<br>(Type I) |
| m0165.v1                   | 家族性骨髓異形成症候群            | v1:2019/05/07 | NGS           | 配列決定       | Illumina              | 家族性骨髓異形成症候<br>群                | 古屋 淳史 | 制限             |

<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

# 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

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)

# 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）

個人別の情報でない（集計情報）

制限公開（提供・利用審査あり）

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

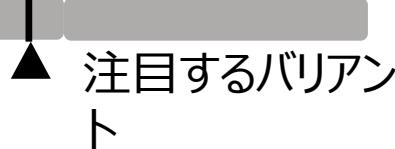
② 利用申請

出典元 (© 2016 DBCLS TogоЖ)

# 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

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) : 0.000304573

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)

# JGA-NGS/JGA-SNPデータの由来

## JGA-NGS

[https://togovar.biosciencedbc.jp/doc/datasets/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<br>(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<br>(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<br>(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<br>(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<br>(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<br>(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](https://togovar.biosciencedbc.jp/doc/datasets/jga_snp)

### 集約されたデータの由来

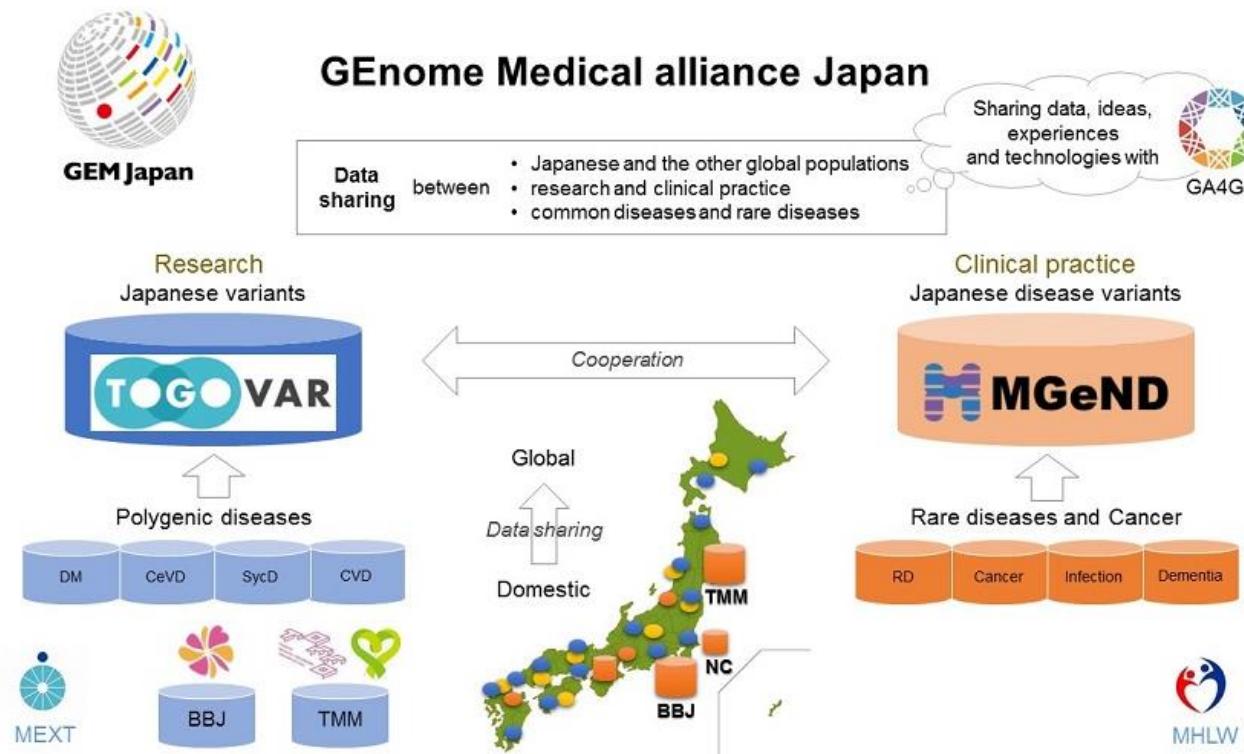
| JGAID               | ヒトDB    | 研究題目                  | 対象集団     | サンプルサイズ | データ提供者         |
|---------------------|---------|-----------------------|----------|---------|----------------|
| JGAD00000000123     | hum0014 | オーダーメイド医療の実現プログラム     | 健常者及び罹患者 | 182,557 | 久保 充明<br>(BBJ) |
| JGAD00000000018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

# NBDC TogoVar / その他の検索対象データ

| データベース名   | 対象人数                | 説明   |
|---|---------------------|--|
| JGA-NGS   | 125人 (WES)          | NBDCヒトデータベースに寄託されたデータからのアリル頻度情報<br>(JGA-SNPの大部分はBBJ) |
| JGA-SNP   | 183,884人 (SNP Chip) |  |
| ToMMo 4.7KJPN Allele Frequency Panel            | 4,773人 (WGS)        | 東北メディカル・メガバンク機構が収集したゲノムコホート（東北地方中心）からのアリル頻度情報        |
| Human Genetic Variation Database (HGVD)         | 1,208人 (WES)        | 長浜コホートのサンプルを中心としたゲノムコホートからのアリル頻度情報                   |
| Exome Aggregation Consortium (ExAC)             | 60,706人 (WES)       | 約20プロジェクトからのデータを再解析したアリル頻度情報                         |
| ClinVar (NCBI)                                  |                     | バリアントの疾患関連性  |
| PubTator (NCBI)<br>LitVar(NCBI)<br>Colil(DBCLS) |                     | バリアントに関連する文献情報                                       |

# Genome Medical alliance Japan (GEM Japan)

「GEM Japan」は、データシェアリングを進めながらゲノム医療の実現を目指すAMEDの各事業に関わる大学、研究所、病院等と日本全国規模で協力体制を築き、臨床情報と個人ゲノム情報のデータシェアリングと研究利用を促進し、ゲノム医療の実現を目指すものです。



([https://wwwAMED.go.jp/aboutus/collaboration/ga4gh\\_gem\\_japan.html](https://wwwAMED.go.jp/aboutus/collaboration/ga4gh_gem_japan.html) より)

日本人全ゲノム解析に基づく日本人アレル頻度情報の公開

<https://gr-sharingdbs.biosciencedbc.jp/agd>

NBDC National Bioscience Database Center

## AMEDゲノム制限共有データベース (AGD)

サイト内検索 検索

ホーム AGDホーム データの利用 データの提供 ガイドライン (AGD用) NBDCヒトデータ審査委員会

Amed Genome group sharing Database (AGD) は国立研究開発法人日本医療研究開発機構 (AMED) ゲノム制限共有データの共有を目的としたデータベースです。

データ利用方法は[こちらをご覧下さい。](#)

登録件数 : 7件 (研究データの登録があり次第、下のリストに追加されます)

| Research ID                    | 研究題目                            | 公開日                            | データの種類                  | 研究方法                             | 手法                                       | 参加者<br>(対象集団)               | 提<br>供<br>者 | セキユ<br>リティ<br>レベル |
|--------------------------------|---------------------------------|--------------------------------|-------------------------|----------------------------------|--|-----------------------------|-------------|-------------------|
| agd0001.v1<br>AGDS_00000000002 | 難聴の遺伝子解析と臨床応用に関する研究             | v1:2018/07/02                  | NGS<br>(Target Capture) | 63 / 69<br>遺伝子<br>領域<br>配列決<br>定 | Life technologies<br>(Ion PGM/IonProton) | 非症候群性難聴：<br>1051症例<br>(日本人) | 宇佐美真一       | 制限<br>(Type I)    |
| agd0002.v1<br>AGDS_00000000001 | B型肝炎に関する統合的臨床ゲノムデータベースの構築を目指す研究 | v1:2018/06/08                  | SNP-chip                | ゲノム<br>ワイド<br>関連解<br>析           | Affymetrix<br>(Axiom ASI)                | B型肝炎ワクチン接種者：1,193名<br>(日本人) | 満上雅史        | 制限<br>(Type I)    |
| agd0003.v2<br>AGDS_00000000003 | リンチ症候群の拾い上げ及び遺伝子診断に関する多施設共同研究   | v1:2020/03/23<br>v2:2020/04/13 | NGS<br>(Target)         | 配列決<br>定                         | Illumina<br>(MiSeq)                      | リンチ症候群疑い症例：100+82症例         | 赤木泰         | 制限<br>(Type I)    |

東北メディカルメガバンク(東北大学、岩手医科大学)、バイオバンクジャパン、理化学研究所の共同研究による参考パネル構築

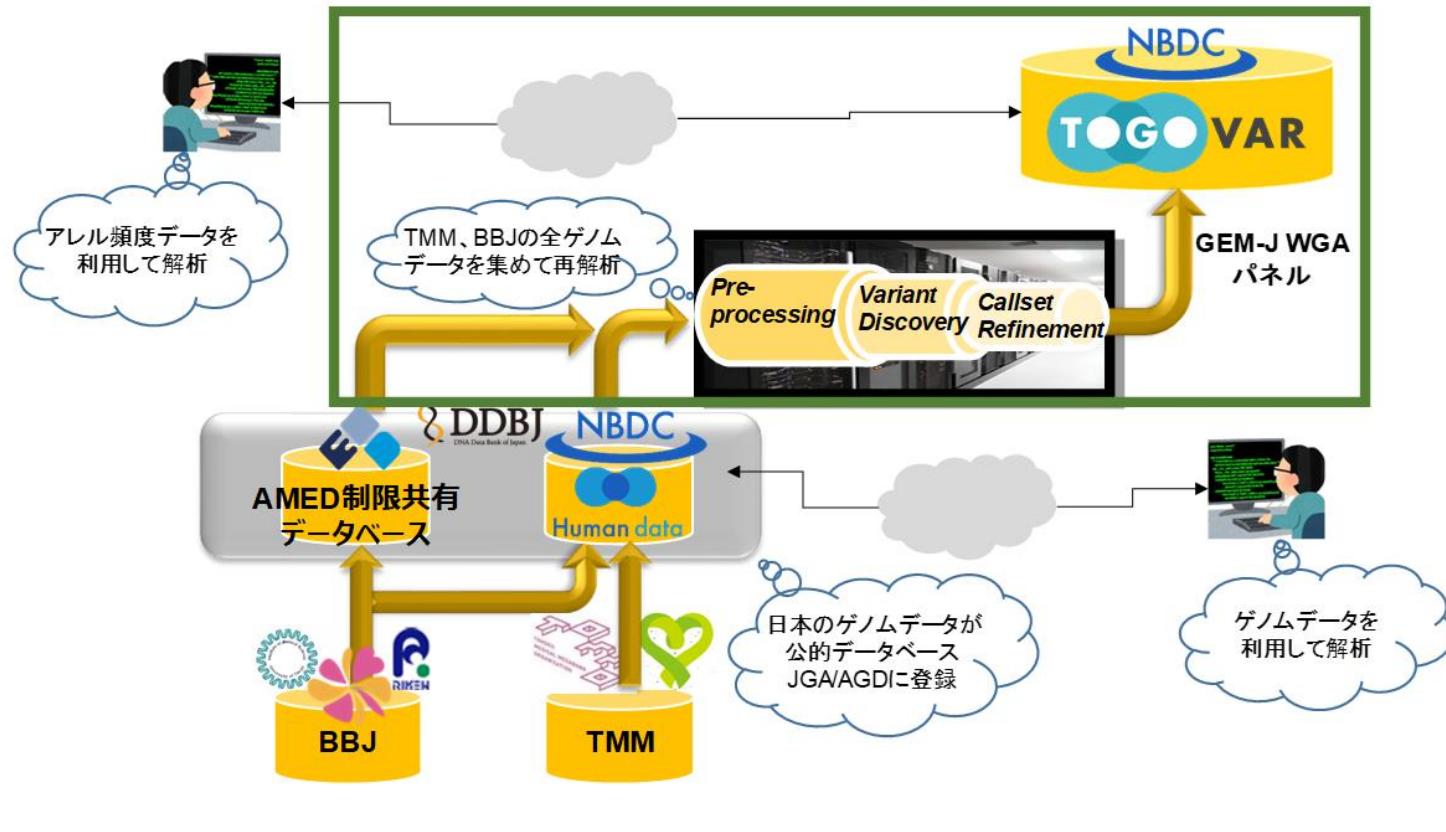


図. TMM、BBJ の全ゲノムデータを再解析して GEM-J WGA パネルを作成するまでの流れ (プレスリリースより)

NBDCは事務局として参画

| コホート名   | 人数<br>(JGA/AGD※9データID・人数)  |
|---|--|
| 東北メディカル・メガバンク計画による宮城県と岩手県でのコホート調査への協力者                        | 4,307  |
| 独立行政法人国立病院機構長崎医療センターにおける協力者                                   | 188  |
| オーダーメイド医療実現化プロジェクトおよびオーダーメイド医療の実現プログラム参加者<br>(バイオバンク・ジャパン協力者) | 2,857<br>(JGAD00000000220・768、<br>AGDS_00000000005・2,089)                  |
| 理化学研究所 生命医科学研究センターにおける協力者                                     | 257<br>(JGAD00000000117・17、<br>JGAD00000000228・220、<br>JGAD00000000233・20) |
| 合計  | 7,609  |

(プレスリリースより)

GATK Best Practiceを用いたjoint variant call、  
 1000 genomes projectと併せたPCAによる genetic background の  
 品質管理実施後のサンプル数

|       | SNV（一塩基多様性） |               | INDEL（挿入欠失配列） |               |
|-------|-------------|---------------|---------------|---------------|
|       | 総数          | 新規検知数<br>(内数) | 総数            | 新規検知数<br>(内数) |
| 常染色体  | 76,768,387  | 35,660,425    | 10,202,908    | 4,152,671     |
| X 染色体 | 2,898,518   | 1,420,888     | 410,435       | 164,077       |

※ 新規検知数: dbSNP152に含まれないバリアントの数

PCAでPlatform毎やサイト毎のクラスタリングが観察  
 → VCFのFILTER値にNotHighConfidenceRegionのフラグを付与

**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    | Position | Consequence  | SIFT               | PolyPhen           | Clinical significance |
|-------------|--------------|----------|--------------|--------------------|--------------------|-----------------------|
| tgv67071948 | rs7792       | 1: 10271 | SNV          | Intergenic variant |                    |                       |
| tgv67071949 |              | 1: 10271 | SNV          | Intergenic variant |                    |                       |
| tgv67071950 | rs1366       | 1: 10248 | Deletion     | Intergenic variant |                    |                       |
| tgv67071951 | rs1409       | 1: 10248 | AAAC... >G   | Deletion           | Intergenic variant |                       |
| tgv67071952 | rs20173260   | 1: 10271 | SNV          | Intergenic variant |                    |                       |
| tgv67071953 | rs775928745  | 1: 10231 | C >T         | Deletion           | Intergenic variant |                       |
| tgv67071954 | rs1267031179 | 1: 10248 | A >T         | Deletion           | Intergenic variant |                       |
| tgv67071955 |              | 1: 10248 | AAAC... >G   | Deletion           | Intergenic variant |                       |
| tgv13       | rs145427775  | 1: 10291 | C >T         | SNV                | Intergenic variant |                       |
| tgv14       | rs060413313  | 1: 10297 | C >T         | SNV                | Intergenic variant |                       |
| tgv19       | rs5756       | 1: 10321 | C >T         | SNV                | Intergenic variant |                       |
|             |              | 1: 10327 | T >C         | Deletion           | Intergenic variant |                       |
| tgv67071957 | rs112750067  | 1: 10327 | T >C         | SNV                | Intergenic variant |                       |
| tgv67071958 |              | 1: 10328 | A >T         | Deletion           | Intergenic variant |                       |
| tgv67071960 | rs1035171912 | 1: 10330 | C >A         | SNV                | Intergenic variant |                       |
| tgv67071959 | rs150969722  | 1: 10330 | C >T         | Deletion           | Intergenic variant |                       |
| tgv67071962 | rs1357792689 | 1: 10334 | T >A         | SNV                | Intergenic variant |                       |
| tgv67071961 | rs1351390918 | 1: 10334 | T >C         | Deletion           | Intergenic variant |                       |
| tgv67071963 | rs1193008993 | 1: 10336 | A >C         | SNV                |                    |                       |
| tgv67071964 | rs1471210572 | 1: 10340 | T >C         | SNV                |                    |                       |
| tgv67071965 |              | 1: 10343 | CCCT... >GAG | Deletion           |                    |                       |
| tgv67071966 |              | 1: 10346 | T >C         | SNV                |                    |                       |
| tgv67071967 | rs1363828207 | 1: 10348 | ACCC... >TGT | Deletion           |                    |                       |
| tgv67071968 |              | 1: 10349 | CCCT... >ATG | Deletion           |                    |                       |
| tgv67071969 |              | 1: 10352 | ACCC... >ATG | Insertion          |                    |                       |
| tgv67071970 | rs1015856060 | 1: 10354 | C >A         | SNV                |                    |                       |
| tgv67071971 |              | 1: 10384 | CCCT... >GAG | Deletion           | Intergenic variant |                       |
| tgv67071972 | rs73674772   | 1: 10376 | T >C         | Deletion           | Intergenic variant |                       |

**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

**Filters**

**Dataset**

- All (74,698,940)
- JGA NGS (4,679,025)
- JGA SNP (1,249,724)
- 3.5KJPN (64,675,495)
- HGVD (354,461)
- ExAC (10,195,872)
- ClinVar (443,512)

**Alternative allele frequency**

**Variant calling quality**

**Variant type**

**Clinical significance**

Variant report  
tgv47264307 dbSNP  
rs671

Variant type SNV

Position 12:112241766 (GRCh37)

Ref / Alt G&gt;A

hgvs 12:g.112241766G&gt;A

## Other overlapping variant(s)

No other variants on the same location

## Frequency

| Dataset  | Population | Allele count<br>Alt | Total   | Frequency |        | Genotype count<br>Alt / Alt | Alt / Ref | Ref / Ref | Filter status | Quality score |
|----------|------------|---------------------|---------|-----------|--------|-----------------------------|-----------|-----------|---------------|---------------|
|          |            |                     |         |           |        |                             |           |           |               |               |
| 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      |

## Clinical Significance

| Title                                      | Clinical significance       | Review status                          | Last evaluated           | Condition(s)  |
|--|-----------------------------|--|--------------------------|---|
| NM_000690.3(ALDH2):c.1510G>A (p.Glu504Lys) | drug response<br>protective | ★★★★<br>no assertion criteria provided | 2018-08-09<br>2010-02-01 | • Acute alcohol sensitivity<br>• Alcohol dependence |

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

また、PubMedに登録された論文のabstractに当該バリエ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。

# 解説動画あります (TogoTV)

The screenshot shows the TogoTV website interface. At the top, there's a navigation bar with links to DBCLS, Research, Services, Contact, and About. Below that is the TOGO-TV logo and a sub-header: 生命科学系DB・ツール使い始め系チャンネル. A search bar is present above a main content area. The main content area has a title: 2018-08-25 TogoVar でヒトゲノムに存在するパリアントに関する情報を調べる. Below the title, there's a video player showing a YouTube video. A large blue callout bubble points to the video player with the text: 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を、分子生物学的な情報や強度、関連論文を知ることが出来るサイトです。 The video player also shows the URL <https://togovar.biosciencedbc.jp/ja/20180825.html>.

- 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, Terms, Contact, About, History, Help, and user account information. The main content area displays a table with two columns: 'Chromosome' and 'Variation frequency'. The 'Variation frequency' column lists chromosomes 1 through MT, each associated with a 'frequency [tsv]' link and a file size in parentheses. The 'Molecular annotation' column lists chromosomes 1 through MT, each associated with a 'molecular annotation [tsv]' link and a file size in parentheses. A note at the top of the table reads: 'Togovarデータは以下のリンクからダウンロード可能です。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            | Molecular annotation                      |
|------------|--------------------------------|---|
| 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] (102 MB) | chr 7 molecular annotation [tsv] (90 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] (56 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] (59 MB) | chr 17 molecular annotation [tsv] (50 MB) |
| 18         | chr 18 frequency [tsv] (34 MB) | chr 18 molecular annotation [tsv] (35 MB) |
| 19         | chr 19 frequency [tsv] (59 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: 20190723

## Downloads

TogoVarデータは以下のリンクからダウンロード可能です。

- [GEM Japan Whole Genome Aggregation \(GEM-J WGA\) バルの頻度情報 \(VCFフォーマット\)](#)  
日本各地から収集した7,000人分の全ゲノム配列情報(Whole Genome Sequence: WGS)を集団の大規模バリエント頻度データセットです。全ゲノム配列情報も制限公開データとして実施されました。[\[詳細\]](#) [\[ダウンロード\]](#)
- [GEM-J WGA以外のバリエントの頻度情報 \(JGA-NGS, JGA-SNP, ExAC, 4.7KJPN, HGVD\)\(タブ\)](#)  
JGA-NGSとJGA-SNPは、NBDCヒトデータベース / Japanese Genotype-phenotype Archive (ジムデータ)を集約した頻度データです。[\[詳細\]](#) [\[ダウンロード\]](#)

## Index of /downloads/gem\_j\_wga/

|               | filename                             | last modified     | size |
|---------------|--------------------------------------|-------------------|------|
|               | README.txt                           | 20-Jul-2020 00:53 | 1289 |
| BBJ RIKEN TMM | 20200309 autosome.chr1.vqsr 99.5 ..  | 19-Jun-2020 10:06 | 498M |
| BBJ RIKEN TMM | 20200309 autosome.chr10.vqsr 99.5 .. | 19-Jun-2020 10:10 | 201K |
| BBJ RIKEN TMM | 20200309 autosome.chr11.vqsr 99.5 .. | 19-Jun-2020 10:06 | 291M |
| BBJ RIKEN TMM | 20200309 autosome.chr12.vqsr 99.5 .. | 19-Jun-2020 10:11 | 117K |
| BBJ RIKEN TMM | 20200309 autosome.chr13.vqsr 99.5 .. | 19-Jun-2020 10:06 | 293M |
| BBJ RIKEN TMM | 20200309 autosome.chr14.vqsr 99.5 .. | 19-Jun-2020 10:11 | 117K |
| BBJ RIKEN TMM | 20200309 autosome.chr15.vqsr 99.5 .. | 19-Jun-2020 10:11 | 285M |
| BBJ RIKEN TMM | 20200309 autosome.chr16.vqsr 99.5 .. | 19-Jun-2020 10:06 | 116K |
| BBJ RIKEN TMM | 20200309 autosome.chr17.vqsr 99.5 .. | 19-Jun-2020 10:06 | 201M |
| BBJ RIKEN TMM | 20200309 autosome.chr18.vqsr 99.5 .. | 19-Jun-2020 10:11 | 85K  |
| BBJ RIKEN TMM | 20200309 autosome.chr19.vqsr 99.5 .. | 19-Jun-2020 10:06 | 199M |
| BBJ RIKEN TMM | 20200309 autosome.chr20.vqsr 99.5 .. | 19-Jun-2020 10:11 | 79K  |
| BBJ RIKEN TMM | 20200309 autosome.chr21.vqsr 99.5 .. | 19-Jun-2020 10:06 | 194M |
| BBJ RIKEN TMM | 20200309 autosome.chr22.vqsr 99.5 .. | 19-Jun-2020 10:11 | 73K  |
| BBJ RIKEN TMM | 20200309 autosome.chr23.vqsr 99.5 .. | 19-Jun-2020 10:06 | 218M |
| BBJ RIKEN TMM | 20200309 autosome.chr24.vqsr 99.5 .. | 19-Jun-2020 10:11 | 72K  |
| BBJ RIKEN TMM | 20200309 autosome.chr25.vqsr 99.5 .. | 19-Jun-2020 10:06 | 192M |
| BBJ RIKEN TMM | 20200309 autosome.chr26.vqsr 99.5 .. | 19-Jun-2020 10:11 | 70K  |
| BBJ RIKEN TMM | 20200309 autosome.chr27.vqsr 99.5 .. | 19-Jun-2020 10:06 | 160M |
| BBJ RIKEN TMM | 20200309 autosome.chr28.vqsr 99.5 .. | 19-Jun-2020 10:11 | 62K  |
| BBJ RIKEN TMM | 20200309 autosome.chr29.vqsr 99.5 .. | 19-Jun-2020 10:06 | 166M |
| BBJ RIKEN TMM | 20200309 autosome.chr30.vqsr 99.5 .. | 19-Jun-2020 10:11 | 49K  |
| BBJ RIKEN TMM | 20200309 autosome.chr2.vqsr 99.5 ..  | 19-Jun-2020 10:06 | 513M |
| BBJ RIKEN TMM | 20200309 autosome.chr3.vqsr 99.5 ..  | 19-Jun-2020 10:12 | 212K |
| BBJ RIKEN TMM | 20200309 autosome.chr4.vqsr 99.5 ..  | 19-Jun-2020 10:06 | 134M |
| BBJ RIKEN TMM | 20200309 autosome.chr5.vqsr 99.5 ..  | 19-Jun-2020 10:12 | 50K  |
| BBJ RIKEN TMM | 20200309 autosome.chr6.vqsr 99.5 ..  | 19-Jun-2020 10:06 | 92M  |
| BBJ RIKEN TMM | 20200309 autosome.chr7.vqsr 99.5 ..  | 19-Jun-2020 10:12 | 28K  |
| BBJ RIKEN TMM | 20200309 autosome.chr8.vqsr 99.5 ..  | 19-Jun-2020 10:06 | 95M  |
| BBJ RIKEN TMM | 20200309 autosome.chr9.vqsr 99.5 ..  | 19-Jun-2020 10:12 | 30K  |
| BBJ RIKEN TMM | 20200309 autosome.chr10.vqsr 99.5 .. | 19-Jun-2020 10:06 | 41M  |
| BBJ RIKEN TMM | 20200309 autosome.chr11.vqsr 99.5 .. | 19-Jun-2020 10:12 | 173K |
| BBJ RIKEN TMM | 20200309 autosome.chr12.vqsr 99.5 .. | 19-Jun-2020 10:06 | 402M |

頻度情報(VCFフォーマット)は、ダウンロードして利用が可能。

データダウンロード用URL:

[https://togovar.biosciencedbc.jp/downloads/gem\\_j\\_wga/](https://togovar.biosciencedbc.jp/downloads/gem_j_wga/)

データ概要および利用条件URL :

[https://togovar.biosciencedbc.jp/doc/datasets/gem\\_j\\_wga](https://togovar.biosciencedbc.jp/doc/datasets/gem_j_wga)

ご清聴いただき、誠にありがとうございました。