

# TogoVarの活用事例で学ぶ バリエントデータベースの使い方

大学共同利用法人 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  
ライフサイエンス統合データベースセンター(DBCLS)


三橋 信孝

# 対象者と発表内容

---

- **対象者：日本人一般集団のアレル頻度データの利用例を知りたい**
  - 発表内容：TogoVarのアレル頻度を参照した論文を紹介します
    - 一般集団：特定の疾患や条件で選別せず、地域や国の住民から広く集めた集団
- **対象者：バリエーションを解釈するためのアノテーション情報を取得したい**
  - 発表内容：以下の情報の検索方法を TogoVar を使って説明します
    - 臨床的意義：ClinVarなどのDB由来の病原性分類（Pathogenic/Benign 等）
    - 関連解析結果：GWAS などの 疾患・形質との関連（効果量・p値等）
    - 有害性予測：SIFT/PolyPhen-2/AlphaMissense による影響度・有害性スコア
    - 文献情報：PubMed 等に基づく 関連論文（PMID・要約・リンク）
    - NBDC統合化推進プログラムとの連携
      - ハプロタイプ（JoGo）、プロテオーム（jPOST）、糖鎖（Glycosmos）、マウス（MoG+）

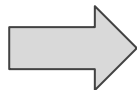
# アレル頻度・ジェノタイプ頻度とは

個人レベルのゲノムデータ  Human data  
(NBDCヒトデータベースに利用申請が必要)

頻度データ  
個人情報でないので利用申請不要



位置	1	2	3	4	5	
Ref	T	G	C	C	G	
個体1	T	G	C	C	G	母由来
	T	G	C	-	G	父由来
個体2	T	A	C	C	G	母由来
	T	G	C	C	G	父由来
個体3	T	A	C	-	G	母由来
	T	A	C	-	G	父由来
個体4	T	A	C	-	G	母由来
	T	G	C	C	G	父由来
個体5	T	G	C	C	G	母由来
	T	A	C	C	G	父由来



ID	位置	アレルの種類		アレル数		
		Ref	Alt	Ref	Alt	Freq
SNV1	2	G	A	5	5	5/10
DEL1	4	C	—	6	4	4/10

バリエーションのゲノム上の位置と参照アレルは選択する参照ゲノムで変わる

- [GRCh37\(hg19\)](#)→古
- [GRCh38\(hg38\)](#)→現在
- [CHM13\(T2T-CHM13\)](#)→新

ID	位置	ジェノタイプの種類			ジェノタイプ数		
		Ref/Ref	Ref/Alt	Alt/Alt	Ref/Ref	Alt/Ref	Alt/Alt
SNV1	2	G/G	G/A	A/A	1	3	1
DEL1	4	C/C	C/—	—/—	2	2	1

SNV: single nucleotide variant、DEL: deletion  
Ref (Reference、参照) : 参照ゲノムと同じアレル  
Alt (Alternate、代替) : 参照ゲノムと異なるアレル

個人レベルのゲノム:NBDCヒトデータベース/Japanese Genotype-phenotype Archive

[illegible]

## データ加工について

データ利用者の利便性向上のため、NBDCヒトデータベースに制限公開データとして登録されたデータ（元データ）に対して一定のワークフローにより加工した、アライメントデータ・バリエーションコールデータ・統計データを、ヒトデータベース審査委員会によって元データの利用を許可されたデータ利用者が希望する場合、加工データも併せて閲覧・利用することを可能にしています。加工データは、元のデータに紐づく形で配置され、Analysis と Dataset の title に「Processed by JGA」と記載されています。なお、加工データを含む解析結果を論文等で公表する際は、元データのアクセスシオ番号を記載して下さい。

### 加工方法

- ・生殖系列の全ゲノムシーケンスデータの加工

## GATK best practice - Germline short varia

ワークフローのソースコード ([jga-analysis](#))

加工済みデータ一覧

対象データセット				
JGADデータセット	研究タイトル	サンプル数	データ加工日	備考
JGAD0000282	がん腫解析に基づく諸がんの発生率・発症率の国際比較の解析 (hum0068)	21	Per-sample : 2021-12-27 QC : 2022-04-06 JoinH-chart : 2022-01-26	癌腫発症率の統計と対照腫瘍発症率の統計 (matched control) へのがんゲノムワイドエクスプレッション解析データ。非腫瘍腫瘍由来体細胞 (Germline) を対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00006070
JGAD0000235	シークエンス解析によるがんゲノム研究 : 大腸がん (hum0159)	10	Per-sample : 2021-12-27 QC : 2022-04-06 JoinH-chart : 2022-01-31	癌腫発症率の統計と対照腫瘍由来の体細胞 (matched control) へのがんゲノムワイドエクスプレッション解析データ。非腫瘍腫瘍由来体細胞 (Germline) を対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00006689
JGAD0000234	シークエンス解析によるがんゲノム研究 : 肝臓癌 (hum0161)	33	Per-sample : 2021-12-27 QC : 2022-01-06 JoinH-chart : 2022-01-31	癌腫発症率の統計と対照腫瘍由来の体細胞 (matched control) へのがんゲノムワイドエクスプレッション解析データ。非腫瘍腫瘍由来体細胞 (Germline) を対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00006888
JGAD0000335	1) 腫瘍細胞の分類・分枝とゲノム情報を利用した研究 癌細胞の分枝解析とゲノム情報を利用した研究 消化管癌性炎症性疾患における遺伝子発達の様態 (hum0201)	14	Per-sample : 2021-12-27 QC : 2022-01-06 JoinH-chart : 2022-01-08	癌腫発症率の統計と対照腫瘍由来の体細胞 (matched control) へのがんゲノムワイドエクスプレッション解析データ。非腫瘍腫瘍由来体細胞 (Germline) を対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00006887
JGAD0000220	オーダーメイド医療の実現プログラム (hum0014)	1,026	Per-sample : 2022-01-16 QC : 2022-02-24  JoinH-chart : 2021-01-30 (autosome and chrX PAR regions; BOTHSEX); 2023-02-26 (chrX non-PAR regions; FEMALE samples); 2023-03-02 (chrX non-PAR regions; MALE samples)	2003年から2009年度にバイオバンク・ジャパンに登録された1,026名のゲノムスケールエクスプレッション解析データを対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00006990
JGAD0000229	オーダーメイド医療の実現プログラム (hum0014)	1,026	Per-sample : 2022-01-16 QC : 2022-02-24  JoinH-chart : 2021-01-30 (autosome and chrX PAR regions; BOTHSEX); 2023-02-26 (chrX non-PAR regions; FEMALE samples); 2023-03-02 (chrX non-PAR regions; MALE samples)	2003年から2009年度にバイオバンク・ジャパンに登録された1,026名のゲノムスケールエクスプレッション解析データを対象にデータ加工を実施した。 加工データ@dataset1 : JGAD00007058

- メリット

- 生データ (fastq, BAMなど) にアクセスできる。
- 一部データセットでGATKでバリエントコールしたVCFデータが利用可能 (右側の画面)

- コスト

- 利用申請が必要、セキュリティガイドラインに準拠した解析環境が必要

■ 機関外サーバ@遺伝研スパコン等を利用すれば手間が軽減される



# 遺伝統計学の自習教材

東京大学大学院医学系研究科遺伝情報学  
理化学研究所IMS  
大阪大学大学院医学系研究科遺伝統計学 システム遺伝学チーム

HOME | RESEARCH | PEOPLE | SOFTWARE | PUBLICATION | JOBS | CONTACT

## 遺伝情報学

Department of Genome Informatics

## 遺伝統計学

Department of Statistical Genetics

## システム遺伝学

終わりに  
次代を担う若手研究者の人材育成



遺伝統計学・夏の学校 (2025)

遺伝統計学 HYBRID @東京大学

「基礎編」は、  
今夏で一旦終了!! 夏の学校

夏休みの思い出に、  
自分の世界に挑んでみませんか?

[tps://genome.m.u-tokyo.ac.jp/recruitment.html](https://genome.m.u-tokyo.ac.jp/recruitment.html)

・生命科学の幅広い研究分野で、情報解析技術が必要な時代に。  
・東京大学・大阪大学医学部の系統講義で、ゲノムデータ解析演習を実施。

東京大学 岡田随象先生

[https://genome.m.u-tokyo.ac.jp/school\\_2025.html](https://genome.m.u-tokyo.ac.jp/school_2025.html)

## ー 講義資料を公開しました ー

夏の学校の講義や演習で使用した資料一式を一般公開しています。下記からダウンロード可能です。

- ・事前準備資料 [スライド](#) [事前準備資料一式](#)
- ・8/29 (金) 午前：遺伝学入門 [スライド](#)
- ・8/29 (金) 午前：統計学入門 [スライド](#)
- ・8/29 (金) 午後：Linux入門 [スライド](#)
- ・8/29 (金) 午後：Program入門 [スライド](#) [演習データ一式](#)
- ・8/29 (金) 午後：Webツール入門 [スライド](#)
- ・8/30 (土) 午前：統計ソフトウェアR入門 [スライド](#) [演習データ一式](#)
- ・8/30 (土) 午後：GenomeData入門1 (ゲノムデータ解析入門) [スライド](#) [演習データ一式](#)
- ・8/30 (土) 午後：GenomeData入門2 (ゲノムワイド関連解析) [スライド](#) [演習データ一式](#)
- ・8/30 (土) 午前：GenomeData入門3 (Genotype Imputation解析) [スライド](#) [演習データ一式](#)
- ・8/31 (日) 午前：GenomeData入門4 (自然選択圧解析) [スライド](#) [演習データ一式](#)
- ・8/31 (日) 午後：GenomeData入門5 (Polygenic Risk Score解析) [スライド](#) [演習データ一式](#)
- ・8/31 (日) 午後：GenomeData入門6 (LDSC解析) [スライド](#) [演習データ一式](#)
- ・8/31 (日) 午後：GenomeData入門7 (シングルセル解析) [スライド](#) [演習データ一式](#)

## ー 開催概要 ー

### 遺伝統計学・夏の学校

夏の短期セミナーとして、「遺伝統計学・夏の学校」を開催します。

ゲノム研究に馴染みのない初心者の方を対象に、遺伝統計学の初歩的な講義から、統計解析やプログラミング入門、ゲノム・オミクスデータ解析の演習を実施します。

夏休みの思い出に、ゲノムの世界に触れてみませんか？

昨今のハイオインフォマティクス学習環境の充実を考慮し、「基礎編」の実施は今回で一旦終了となります。

- ・日時：  
令和7年8月29日 (金) ~ 8月31日 (日)  
実施時間帯：9:00~16:30
- ・開催形式：  
ハイブリッド ([東京大学FUKUTAKE Learning Theater](#))  
※ 東京大学本郷キャンパス構内からのハイブリッド形式です。オンライン形式でのWeb参加も可能です。
- ・対象：  
大学学部生以上  
※ 大学院生・勤務医・社会人の方も参加可能ですが、**初心者向けの内容**となります。  
※ 保護者の了解が得られれば高校生も参加可能です。

遺伝統計学の基本概念やデータ解析手法を入門から最近の話題までわかりやすく説明

# 遺伝統計学 夏の学校の教材例

## ① 遺伝統計学における関連解析

TGCTGACTAA  
TGCTGACTAA



疾患発症  
リスク  
×1.5

アレル頻度情報

	G	T	Total
ケース	0.60	0.40	1.00
コントロール	0.50	0.50	1.00

- 疾患発症リスクを定量化するためには、どうすればいいでしょうか？
- 疾患を発症した人(ケース)と、発症していない人(コントロール)とで、各遺伝子多型のアレル頻度を比較することで達成できます。

6

概念の直感的な説明と解析コマンドの実例

## ② PLINKを使ったゲノムワイド関連解析

○:ゲノムワイド関連解析の実施(ロジスティック回帰分析)  
./plink --bfile 1KG\_EUR\_QC --out 1KG\_EUR\_QC\_Pheno1 --pheno  
phenotype1.txt --logistic --ci 0.95

※Cygwinの場合plinkをplink.exeに変えてください



出力ファイル: 1KG\_EUR\_QC\_Pheno1.assoc.logistic

※Macユーザーの方は、"plink\_mac\_20210606.zip"を解凍して、  
Mac OS用のPLINK実行ファイルに置き換えて実行してください。  
※Macユーザーの方は、演習ファイルを置いたディレクトリを適宜指  
定してください。

- "--pheno"で、関連解析に用いる形質ファイルを指定します。  
(ped/famファイルに元々書き込まれていた形質情報より優先されます。)
- "--logistic"で、各SNPにおけるロジスティック回帰分析を実施します。
- "--ci"で、効果サイズの信頼区間を追加で出力します。

21

## ② PLINKを使ったゲノムワイド関連解析


出力ファイル: 1KG\_EUR\_QC\_Pheno1.assoc.logistic

- "CHR":染色体番号
- "SNP":SNP ID
- "BP":染色体上の位置(base pair)
- "A1":アレル1
- "TEST":ジェノタイプ効果
- "NMISS":サンプル数
- "OR":(アレル1の)オッズ比
- "SE":効果サイズ(OR対数値)のSE
- "L95":オッズ比95%CI下限
- "U95":オッズ比95%CI上限
- "STAT":統計量
- "P":P値

22

# Genome Aggregation Database (gnomAD: ノマド)

**gnomAD browser** gnomAD v4.1.0 About Team Federated Stats Policies Publications Blog Changelog Data Forum Contact Help/FAQ

  
Genome Aggregation Database

gnomAD v4.1.0 rs671  
rs671

- Download gnomAD data
- Read gnomAD publications
- Find co-occurrence of two variants
- Browse tandem repeats in gnomAD
- Locate features not yet in gnomAD v4

Please note that the gnomAD v3 genomes are now part of gnomAD v4. For more information, see "Should I switch to the latest version of gnomAD?"

**Examples**

- Gene: PCSK9
- Transcript: ENST00000302118
- Variant: 1-55051215-G-GA
- Structural variant region: 19-11078371-11144910
- Copy number variant region: 1-55039447-55064852
- Mitochondrial variant: M-8602-T-C
- Short tandem repeat locus: ATXN1
- Regional missense constraint (gnomAD v2, GRCh37): GRIN2A
- Variant co-occurrence (gnomAD v2, GRCh37): 1-55505647-G-T and 1-55523855-G-A

**gnomAD browser** gnomAD v4.1.0 Search About Team Federated Stats Policies Publications Blog Changelog Data Forum Contact Help/FAQ

**SNV: 12-111803962-G-A(GRCh38)** Copy variant ID Gene page Dataset gnomAD v4.1.0

Filters	Exomes	Genomes	Total
Allele Count	10123	1193	11316
Allele Number	1456236	152174	1608410
Allele Frequency	0.006951	0.007840	0.007036
Orpmax Filtering AF (95% confidence)	0.2420	0.2141	0.2398
Number of homozygotes	1239	141	1380

**External Resources**

- dbSNP (rs671)
- UCSC
- ClinVar (18390)
- ClinGen Allele Registry (CA128085)
- All of Us

**Feedback**  
[Report an issue with this variant](#)

**Genetic Ancestry Group Frequencies**  
gnomAD HGDP 1KG Local Ancestry

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency	
Overall	10844	44508	1373	0.2436	
East Asian	XY	5456	2172	693	0.2513
	XX	5388	22796	680	0.2364
Remaining		362	62186	7	0.005821
Admixed American		28	59338	0	0.0004719
South Asian		23	90310	0	0.0002547
African/African American		13	74902	0	0.0001736
Middle Eastern		1	6052	0	0.0001652
European (Finnish)		3	63634	0	0.00004714
European (non-Finnish)		42	1177106	0	0.00003568
Ashkenazi Jewish		0	29462	0	0.000
Amish		0	912	0	0.000
XX		5658	809810	685	0.006987
XY		5658	798600	695	0.007085
Total		11316	1608410	1380	0.007036

Include: ☒ Exomes ☒ Genomes

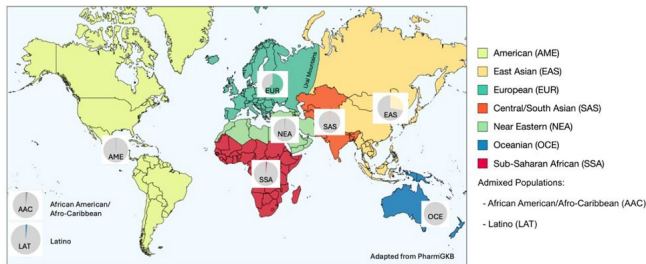
世界中の様々な祖先集団のアレル頻度

# なぜTogoVarが必要か？

ゲノム・バリエントデータは欧州祖先集団に偏っている

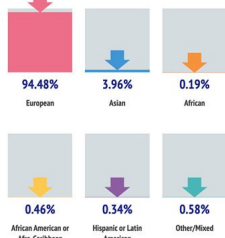
数千人から数万人規模の様々な日本人対象ゲノム研究から収集

Biased representation and missing global diversity in pharmacogenomics

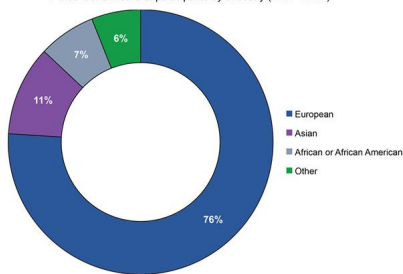


Total GWAS participants diversity

Version 1.0.0 Last check for data: 2024-09-09 00:22:35.



Percent of clinical trial participants by ancestry (2015 - 2019)



BioBank Japan

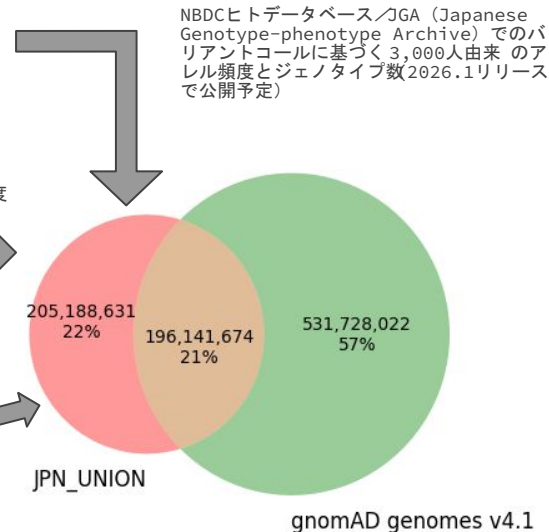


6万人由来のアレル頻度



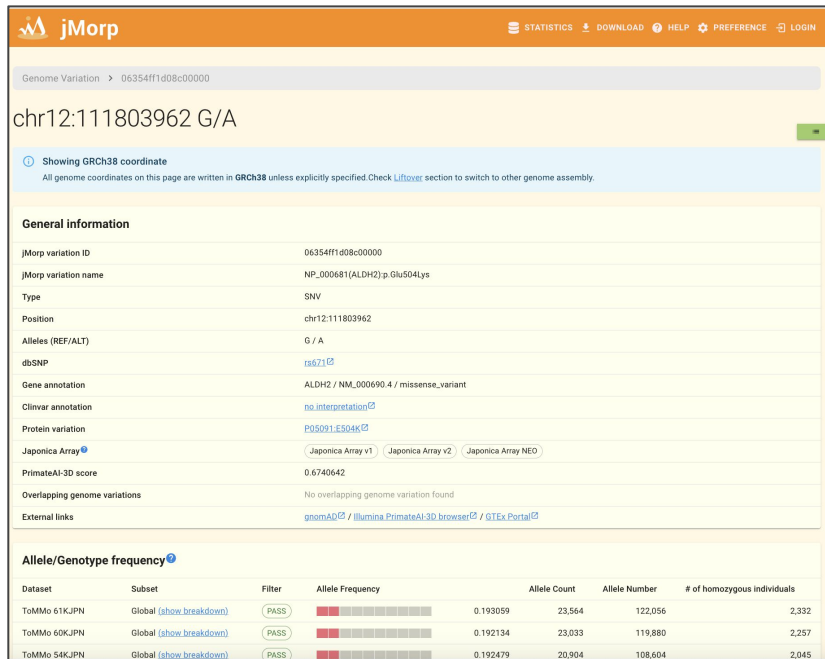
National Center  
Biobank Network

9,000人由来のアレル頻度と遺伝子型数

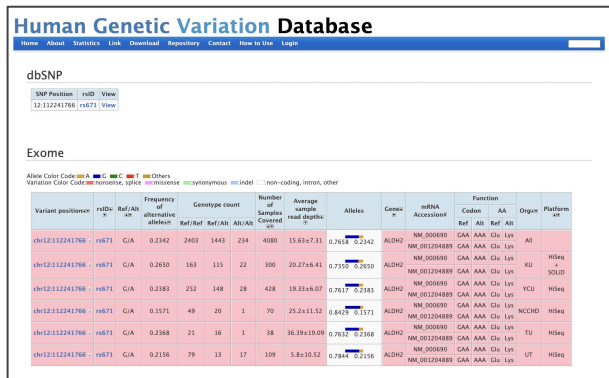


TogoVarには、gnomAD (Genome Aggregation Database) に含まれていない 日本人集団に特異的な常染色体上のバリエントが 2億件

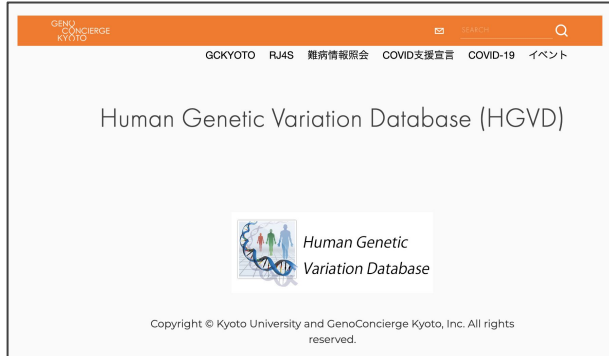
# 国内のバリエーション頻度データベース



jMorp: 東北メディカル・メガバンク (ToMMo)



無償版



有償版

HGVD: 京都大学 & ジェノコンシェルジュ京都



# TogoVarに収載されている頻度データセット

データセット名	祖先集団	サンプル数	解析手法	参照ゲノム	データセットの由来・概要
JGA-SNP	日本人	183,884	SNPアレイ	GRCh37 GRCh38*	JGAに寄託された複数研究の SNPアレイデータ
JGA-WES	日本人	125	全エクソーム	GRCh37 GRCh38*	JGAに寄託された複数研究の全エクソームデータを再解析
JGA-WGS	日本人	78	全ゲノム	GRCh38	JGAに寄託された複数研究の全ゲノムデータを再解析
GEM-J WGA	日本人	7,609	全ゲノム	GRCh37 GRCh38*	AMEDによる多機関共同プロジェクト( GEM Japan)での統合解析( ToMMo、理研、バイオバンクジャパン)
HGVD	日本人	1,208	全エクソーム	GRCh37	京都大学ながはま 0次コホート等
NCBN	日本人	11,794	全ゲノム	GRCh38	ナショナルセンター・バイオバンク ネットワーク (NCBN)と1000 Genomes の統合解析
ToMMo 8.3KJPN	日本人	8,380	全ゲノム	GRCh37	東北メディカル・メガバンク機構 (ToMMo) (最新の61KJPNに近々更新予定)
ToMMo 54KJPN	日本人	54,302		GRCh38	
gnomAD v2.1.1	世界中	15,708 125,748	全ゲノム 全エクソーム	GRCh37	Broad Instituteなどが主導する、複数の研究コホートを統合解析した国際的コンソーシアム
gnomAD v4.1	世界中	76,215 730,947	全ゲノム 全エクソーム	GRCh38	Broad Instituteなどが主導する、複数の研究コホートを統合解析した国際的コンソーシアム

太字：TogoVarが1次公開レポジトリ。JGA-SNPのTogoImputationの結果を搭載予定(2026年度)

# 日本人一般集団のアレル頻度データの利用例

# 利用例の探し方 (Google Scholar)

The screenshot shows a Google Scholar search for 'TogoVar'. The search bar at the top contains 'TogoVar' and a magnifying glass icon. Below the search bar, the results are listed. The first result is 'Identification of 5'untranslated region variants in genes involved in neurodevelopmental disorders' by T Hayashi, S Miyamoto, Y Endo, K Shimizu, et al., published in the Journal of Human Genetics in 2026. The second result is 'Genotype-Phenotype Correlations of Li-Fraumeni Syndrome in Japan Children's Cancer Group LFS20 Study Cohort' by F Yamazaki, Y Nakano, M Sanada, et al., published in Cancer in 2025. The third result is 'Scientific production on data repositories and open science published in the Web of Science database: Methodi Ordinatio and content analysis' by SA Rodrigues-Junior, MV Teixeira, published in Transinformação in 2025. The fourth result is 'JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes' by M Nagasaki, T Katayama, Y Moriya, et al., published in Nucleic Acids in 2026. The left sidebar contains filters for '記事' (Articles), '期間指定なし' (No date specified), '2026 年以降' (After 2026), '2025 年以降' (After 2025), '2022 年以降' (After 2022), '期間を指定...' (Specify period...), '関連性で並べ替え' (Sort by relevance), '日付順に並べ替え' (Sort by date), 'すべての言語' (All languages), '英語と日本語のページを検索' (Search for English and Japanese pages), 'すべての種類' (All types), '総説論文' (Review articles), '特許を含める' (Include patents), '引用部分を含める' (Include citation parts), and 'アラートを作成' (Create alert).



ORIGINAL ARTICLE | [Open Access](#) | [CC](#) [BY](#) [NC](#) [ND](#)

## Genotype-Phenotype Correlations of Li-Fraumeni Syndrome in Japan Children's Cancer Group LFS20 Study Cohort

[Fumito Yamazaki](#), [Yoshiko Nakano](#), [Masashi Sanada](#), [Hiroki Kurahashi](#), [Shunsuke Miyai](#), [Arisa Ueki](#), [Yuko Watanabe](#), [Daisuke Hasegawa](#), [Shuhei Karakawa](#), [Toshifumi Ozaki](#) ... [See all authors](#) ↓

First published: 29 December 2025 | <https://doi.org/10.1111/cas.70302> | [VIEW METRICS](#)

All variants were interpreted using *TP53*-specific guidelines from the American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP) [10] for all non-CNV variants, while CNVs were classified according to the ACMG CNV criteria [11]. Variant interpretation incorporated functional data from Kato [12], Giacomelli [13], and Kotler [14], population data from the Genome Aggregation Database (gnomAD) [15], hotspot data from Cancerhotspots [16], and in silico functional prediction algorithms of Align-GVGD [17] and BayesDel [18]. TogoVar [19] demonstrated more in-depth information on Japanese population data. PVS1 criteria were determined based on guidance from Tayoun et al. [20]. The *TP53* database [21] was used to collect clinical, functional, and in silico data. Classic LFS criteria [2], Chompret criteria [5] were applied to evaluate previous reports. We integrated functional data, population frequencies, mutational hotspots, in silico predictions, and published literature, and applied weighted *TP53*-specific ACMG/AMP guidelines to ultimately classify the pathogenicity of each variant [10]. ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) variant classifications, as updated on April 12, 2025, were used as a reference in our interpretation. Additionally, all missense variants were assessed with a machine-learning approach named AlphaMissense [22], though these results were not incorporated into the *TP53* ACMG/AMP guidelines.

<https://onlinelibrary.wiley.com/doi/full/10.1111/cas.70302>

[https://scholar.google.co.jp/scholar?scisbd=2&q=TogoVar&hl=ja&as\\_sdt=0.5](https://scholar.google.co.jp/scholar?scisbd=2&q=TogoVar&hl=ja&as_sdt=0.5)



# 神経発達障害における5'非翻訳領域(5'-UTR)病的バリエント同定

- 背景
  - 5'-UTRのバリエントと遺伝性疾患である神経発達障害の関係が不明
- 実施内容
  - 病的バリエントが同定されていなかった神経発達障害患者のゲノム配列を決定
  - 5'-UTRに存在する上流オープンリーディングフレーム( uORF)へ影響するバリエントを、5'-UTRバリエント注釈ツール UTRannotator を用いて探索
- 結果
  - ATRX関連疾患患者の ATRX の上流に母親由来の 1塩基重複(c.-138dup)がコーディング領域( CDS)と重なり合うフレーム外(out-of-frame)のuORF形成を引き起こすと予測。ATRX関連疾患の症状。RNAシーケンスでATRX発現低下
  - POU3F3関連疾患(Snijders Blok-Fisher症候群)のPOU3F3遺伝子の5'-UTRにde novo SNV: c.-303C>Aを同定。
- TogoVarのデータが使われたところ
  - レアバリエントの定義: 以下のデータセットで minor allele frequency < 0.005
    - in-house control exome database,
    - ToMMo 60KJPN
    - **GEM-J WGA**
    - gnomAD v4.1.0 database

# 日本人SCA疑い例に対する全エクソーム解析でELOVL4・ELOVL5・GRM1の新規変異を同定

- 背景・目的
  - 脊髄小脳変性症 (SCD)のうち、優性遺伝型は一般に脊髄小脳失調症 (SCA)と呼ばれる。
  - 既知のリピート伸長 (SCA1/2/3/6/7/8/31/36など)が陰性でも、既知原因遺伝子の未知変異を持つ未診断例があり得るため、全エクソーム解析 (WES)で既知SCA関連遺伝子を系統的に探索。
- 実施内容
  - 既知のリピート伸長がない174例のSCA疑い患者にWESを実施。既知SCA関連遺伝子群から候補変異を抽出
  - 候補条件として「データベース頻度 < 0.001」「ミスセンス/ナンセンス/indel」「リード深度>4」などを用い、AlphaMissense/CADD/SIFT/PolyPhen-2/MutationTaster の5種の予測アルゴリズムで病原性を評価。
- 主な結果
  - ELOVL4、ELOVL5、GRM1 の3遺伝子で新規SNVを5例に同定でき、174例のうち約3%でWESにより原因遺伝子候補を特定できた。
- TogoVarのデータが使われたところ
  - レアバリエーションの定義: 以下のデータセットで minor allele frequency < 0.001
    - gnomAD
    - **NCBN (National Center Biobank Network) → TogoVar を介して取得**

# α1-アンチトリプシン欠乏症 (AATD) の日本人集団での有病率の推定

— — —

- 背景
  - α1-アンチトリプシン欠乏症 (AATD) は、若年で肺気腫を引き起こし、労作時の呼吸困難や咳・痰など、COPD (慢性閉塞性肺疾患) の症状を呈する指定難病である。
  - AATD はヨーロッパや北米では高い有病率 (1,500~3,000 人に1人) を示す一方、日本では1,000万人あたり約2人と非常に稀とされている。しかし、AATD の認知度が低いため、十分に診断されていない可能性がある。
- 実施内容
  - 本研究では、日本人集団を対象とした公開バリエーションデータベースを用いて、AAT を産生する SERPINA1 遺伝子の5つの原因バリエーションの日本人におけるアレル頻度を調査し、有病率を推定した。
- 結果
  - 平均アレル頻度 ( $8.56 \times 10^{-4}$ ) から、日本人における AATD の有病率を137万人に1人と推定した。
  - この結果が過去の疫学調査と一致することから、日本人では AATD が稀な疾患であることが再確認された。
- TogoVar のデータが使われたところ
  - TogoVar に収載されている3つのデータセット (**GEM-J WGA**、**ToMMo 54KJPN**、**HGVD v2.3**) のアレル頻度を利用

# ミスセンスバリアントの病原性と遺伝形式予測ツールの開発

— — —

- 背景と実施内容
  - 筆者らはミスセンスバリアントの病原性と遺伝形式(病的常染色体劣性、病的常染色体優性、良性)を機械学習で予測する MOI-Pred (Mode of Inheritance Predictor) と ConMOI (Consensus Mode of Inheritance) を開発した。
- 結果
  - 29,981人の電子健康記録(EHR)データを用いた実世界の検証によって、既存の方法を上回る性能を示した。
- TogoVarのデータが利用されたところ
  - ツールの評価に必要な良性ベンチマークデータの一部として **GEM Japan Whole Genome Aggregation panel v.1 (GEM-J WGA)に含まれる 1,010件のミスセンスバリアント**
  - ツールの学習に用いた gnomADデータベースに含まれない日本人一般集団に高頻度のバリアントであり、良性ベンチマークに適したデータであった。

出典: Petrazzini, B. O., Balick, D. J., Forrest, I. S., Cho, J., Rocheleau, G., Jordan, D. M., & Do, R. (2024). Ensemble and consensus approaches to prediction of recessive inheritance for missense variants in human disease. Cell reports methods, 4(12), 100914. <https://doi.org/10.1016/j.crmeth.2024.100914>

# TogoVarに収載されている頻度データセット

データセット名	祖先集団	サンプル数	解析手法	参照ゲノム	データセットの由来・概要
JGA-SNP	日本人	183,884	SNPアレイ	GRCh37 GRCh38*	JGAに寄託された複数研究の SNPアレイデータ
JGA-WES	日本人	125	全エクソーム	GRCh37 GRCh38*	JGAに寄託された複数研究の全エクソームデータを再解析
JGA-WGS	日本人	78	全ゲノム	GRCh38	JGAに寄託された複数研究の全ゲノムデータを再解析
GEM-J WGA	日本人	7,609	全ゲノム	GRCh37 GRCh38*	AMEDによる多機関共同プロジェクト( GEM Japan)での統合解析 (ToMMo, 理研、バイオバンクジャパン)
HGVD	日本人	1,208	全エクソーム	GRCh37	京都大学ながはま0次コホート等
NCBN	日本人	11,794	全ゲノム	GRCh38	ナショナルセンター・バイオバンク ネットワーク (NCBN)と1000 Genomes の統合解析
ToMMo 8.3KJPN	日本人	8,380	全ゲノム	GRCh37	東北メディカル・メガバンク機構 (ToMMo) (最新の61KJPNに近々更新予定)
ToMMo 54KJPN	日本人	54,302		GRCh38	
gnomAD v2.1.1	世界中	15,708 125,748	全ゲノム 全エクソーム	GRCh37	Broad Instituteなどが主導する、複数の研究コホートを統合解析した国際的コンソーシアム
gnomAD v4.1	世界中	76,215 730,947	全ゲノム 全エクソーム	GRCh38	Broad Instituteなどが主導する、複数の研究コホートを統合解析した国際的コンソーシアム

赤字：利用例に含まれるデータセット、太字：TogoVarが1次公開レポジトリ

# バリエントを解釈するための アノテーション情報を取得

# バリエーションアノテーションツール Variant Effect Predictor (VEP)

The screenshot shows the Ensembl VEP website. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is present with the text "Search all species...". Below the navigation bar, there's a section titled "Using this website" with tabs for Annotation and prediction, Data access, API & software, and About us. The main content area is titled "Ensembl Variant Effect Predictor (Ensembl VEP)" and contains a description of the tool's purpose: "Ensembl VEP predicts the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on gene transcripts and protein sequence, as well as regulatory regions. It reports reference data including gene and variant phenotype associations and population allele frequencies to facilitate variant prioritisation and interpretation." Below this, there's a list of features and a section for "What's new in release 1157". At the bottom, there are three boxes representing different interfaces: "Web interface", "Command line tool", and "REST API", each with a brief description and a "Documentation" link.

3つのUI (webフォーム、コンテナ、web API)

他のアノテーションツール

- ANNOVAR (<https://annovar.openbioinformatics.org/>)
- SNPEff (<https://pcingola.github.io/SnpEff/>)

The screenshot shows the "Ensembl Variant Effect Predictor (VEP) Plugins" page. It includes a description of the plugin system and a list of available plugins. The list is organized into categories and includes details such as the plugin name, description, category, external libraries, and developer.

Plugin	Description	Category	External libraries	Developer
<a href="#">AlphaMissense</a>	This plugin for the Ensembl Variant Effect Predictor (VEP) annotates missense variants with the pre-computed AlphaMissense pathogenicity scores. AlphaMissense is a deep learning model developed by Google DeepMind that predicts the pathogenicity of single nucleotide missense variants. <a href="#">more</a>	Pathogenicity predictions	-	Ensembl
<a href="#">AncestralAllele</a>	An Ensembl VEP plugin that retrieves ancestral allele sequences from a FASTA file. <a href="#">more</a>	Conservation	-	Ensembl
<a href="#">AVADA</a>	Automatic Variant evidence Database is a novel machine learning tool that uses natural language processing to automatically identify pathogenic genetic variant evidence in full-text primary literature about monogenic disease and convert it to genomic coordinates. <a href="#">more</a>	Phenotype data and citations	<a href="#">List: More links</a> <a href="#">qw(unix)</a>	Ensembl
<a href="#">BayesDel</a>	This is a plugin for the Ensembl Variant Effect Predictor (VEP) that adds the BayesDel scores to Ensembl VEP output. <a href="#">more</a>	Pathogenicity predictions	-	Ensembl
<a href="#">Biosum62</a>	This is a plugin for the Ensembl Variant Effect Predictor (VEP) that looks up the BIOSUM 62 language matrix score for the reference and alternative amino acids predicted for a missense mutation. It adds one new entry to the output, BIOSUM62, which is the associated score. <a href="#">more</a>	Conservation	-	Ensembl
<a href="#">CADD</a>	An Ensembl VEP plugin that retrieves CADD scores for variants from one or more tabix-indexed CADD data files. <a href="#">more</a>	Pathogenicity predictions	-	Ensembl

豊富なプラグイン(例: AlphaMissense)

# バリエントの表記方法

---

例: GRCh38の12番染色体111803962番目の塩基GがAに置換した一塩基置換 (SNV)

- HGVS表記 (文献によく出現する)
  - ALDH2:p.Glu504Lys
    - ALDH2遺伝子のアミノ酸配列の504番目の残基がGluがLysに置換
  - NM\_000690:c.1510G>A
    - cDNA配列 NM\_000690の1510番目のGがAに置換
- dbSNP RefSNP number (バリエントIDのデファクト)
  - rs671
    - <https://www.ncbi.nlm.nih.gov/snp/rs671>
    - <https://grch38.togovar.org/variant/rs671>
- 染色体番号—染色体の先頭塩基からの位置—参照アレル—代替アレル
  - 12-111803962-G-A
    - [https://gnomad.broadinstitute.org/variant/12-111803962-G-A?dataset=gnomad\\_r4](https://gnomad.broadinstitute.org/variant/12-111803962-G-A?dataset=gnomad_r4)
    - <https://grch38.togovar.org/variant/12-111803962-G-A>



# バリエーションの表記方法の変換ツール Variant recoder

## Variant Recoder

New job

Species:



Homo\_sapiens

Assembly: GRCh38.p14

[Change species](#)

If you are looking for VR for Human GRCh37, please go to [GRCh37 website](#).

Name for this job (optional):

Input data:

Either paste data:

```
ENSP00000355627.4:p.Met259Thr  
ENSP00000443818.1:p.Gly229Asp  
ENSP00000439902.1:p.Glu13Terf
```

Examples: [Variant ID](#), [HGVS genomic](#), [HGVS transcript](#), [HGVS protein](#), [SPDI](#)

Or upload file:

ファイルを選択

選択されていません

Results:

- ☒ SPDI
- ☒ HGVS Genomic
- ☒ HGVS Transcript
- ☒ HGVS Protein
- ☒ VCF format
- ☒ Variant Identifier
- ☐ Variant synonyms
- ☐ MANE Select

```
./variant_recoder --id "AGT:p.Met259Thr" --pretty  
  
[  
  {  
    "warnings" : [  
      "Possible invalid use of gene or protein identifier"  
    ],  
    "c" : {  
      "input" : "AGT:p.Met259Thr",  
      "id" : [  
        "rs699",  
        "CM920010",  
        "COSV64184214"  
      ],  
      "hgvsq" : [  
        "NC_000001.11:g.230710048A>G"  
      ],  
      "hgvs" : [  
        "ENST00000366667.6:c.776T>C",  
        "ENST00000679684.1:c.776T>C",  
        "ENST00000679738.1:c.776T>C",  
        "ENST00000679802.1:c.776T>C",  
        "ENST00000679854.1:n.1287T>C",  
        "ENST00000679957.1:c.776T>C",  
        "ENST00000680041.1:c.776T>C",  
        "ENST00000680783.1:c.776T>C",  
        "ENST00000681269.1:c.776T>C",  
        "ENST00000681347.1:n.1287T>C",  
        "ENST00000681514.1:c.776T>C",  
        "ENST00000681772.1:c.776T>C",  
        "NM_001382817.3:c.776T>C",  
        "NM_001384479.1:c.776T>C"  
      ],  
      "hgvsp" : [  
        "ENSP00000355627.5:p.Met259Thr",  
        "ENSP00000505981.1:p.Met259Thr",  
        "ENSP00000505063.1:p.Met259Thr",  
        "ENSP00000505184.1:p.Met259Thr",  
        "ENSP00000506646.1:p.Met259Thr",  
        "ENSP00000504866.1:p.Met259Thr",  
        "ENSP00000506329.1:p.Met259Thr",  
        "ENSP00000505985.1:p.Met259Thr",  
        "ENSP00000505963.1:p.Met259Thr",  
        "ENSP00000505829.1:p.Met259Thr",  
        "NP_001369746.2:p.Met259Thr",  
        "NP_001371408.1:p.Met259Thr"  
      ],  
      "spdi" : [  
        "NC_000001.11:230710047:A:G"  
      ],  
    },  
  ],  
]
```

# HGVS表記 ALDH2:p.Glu504Lys を条件にTogoVarを検索

https://grch38.togovar.org

**TOGOVAR** A comprehensive Japanese genetic variation database GRCh38

Simple search Advanced search

ALDH2:p.Glu504Lys


Disease: Breast-ovarian cancer, familial 2 Gene: ALDH2 refSNP: rs114202595 TogoVar: tgv56616325 Position(GRCh38): 16:48224287 Region(GRCh38): 10:71510986-71617219

HGVSc: NM\_000690:c.1510G>A HGVSp: ALDH2:p.Glu504Lys

HGVS表記を手入力(またはExampleのHGVSpをクリック)

# HGVS表記 ALDH2:p.Glu504Lys を条件にTogoVarを検索

— — —

← → ↺ <https://grch38.togovar.org/?mode=simple&term=ALDH2%3Ap.Glu504Lys>    | N 仕事用 ⋮

**TOGOVAR** A comprehensive Japanese genetic variation database GRCh38 ▾ Home Datasets Downloads API Terms Contact About History Help Configuration Logi

**Simple search** **Advanced search**


ALDH2:p.Glu504Lys ✕ C

Disease: Breast-ovarian cancer, familial 2 Gene: ALDH2 refSNP: rs114202595 TogoVar: tgv56616325 Position(GRCh38): 16:48224287 Region(GRCh38): 10:71510986-71617219 HGVS: NM\_000690:c.1510G>A HGVS: ALDH2:p.Glu504Lys

**Results** The number of available variations is 1 out of 1. Download JSON CSV TSV

Translate HGVS representation 'ALDH2:p.Glu504Lys' to '12:111803962:G>A' **アレル頻度分布** **臨床的意義** **病原性、有害性予測**

Possible invalid use of gene or protein identifier 'ALDH2' as HGVS reference; ALDH2:p.Glu504Lys

TogoVar ID	RefSNP ID	Position	Ref / Alt	Type	Gene	Alt frequency	Consequence	Clinical significance	AlphaMissense	SIFT	PolyPhen
<a href="#">tgv47264307</a>	<a href="#">rs671</a>	12:111803962	<span>G</span> <span>A</span>	SNV	<a href="#">ALDH2</a>		Missense variant +2	<span>P</span> <a href="#">AMED syndrome, digenic</a> +5	0.8864	0	0.709

↓ バリエントレポートページ

↓ 遺伝子レポートページ

↓ 疾患レポートページ

**Statistics / Filters** Reset

**Dataset**

<input checked="" type="checkbox"/>	All	1
<input checked="" type="checkbox"/>	WGS GEM-J WGA	1
<input checked="" type="checkbox"/>	WGS JGA-WGS	1
<input checked="" type="checkbox"/>	WES JGA-WES	1
<input checked="" type="checkbox"/>	SNP JGA-SNP	1

# バリエントレポートページ(2/6)

<https://grch38.togovar.org/variant/rs671#frequency>

<https://grch38.togovar.org/variant/12-111803962-G-A#frequency>

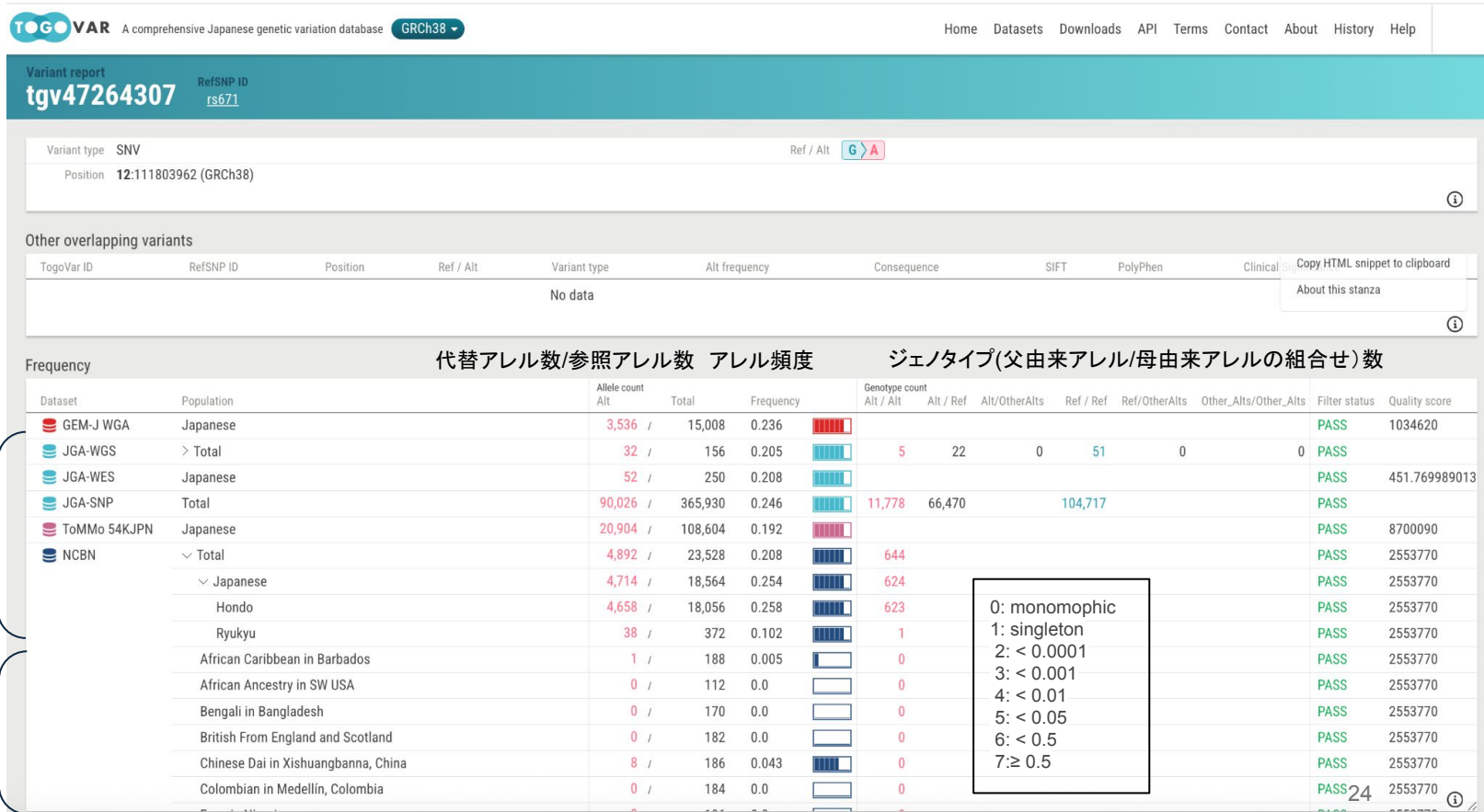
GRCh38上の位置

領域が重複する  
バリエント

祖先集団別の頻度

日本人集団  
(代替アレル頻度高い)

非日本人集団  
(代替アレル頻度低い)



# バリエントレポートページ(2/6)

<https://grch38.togovar.org/variant/rs121244447#frequency>

<https://grch38.togovar.org/variant/1-23825385-G-A#frequency>

Variant report **tgvr746871** RefSNP ID: **rs121244447** 1:23825385 SNV Synonymous variant

## Frequency

Dataset	Population	Allele count Alt / Total	Frequency	Genotype count Alt / Alt	Alt / Ref	Alt/OtherAlts	Ref / Ref	Ref/OtherAlts	Other_Alts/Other_Alts	Filter status	Quality score
GEM-J WGA	Japanese	1 / 14,646	6.828e-5							PASS	210.39999389648438
JGA-WGS	> Total	1 / 28	0.036	0	1	0	13	0	0	PASS	
ToMMo 54KJPN	Japanese	1 / 108,604	9.000e-6							PASS	320.6600036621094
NCBN	> Total	2 / 23,542	8.495e-5	0						PASS	1376.510009765625
gnomAD Exomes	> Total	4 / 1,409,212	2.838e-6							PASS	

## Clinical significance

Title	Clinical significance	Condition
<a href="#">SNV[hg38]chr1:23825385-23825385[gG&gt;A]</a>		others

## Clinical significance

Title	VCV review status	RCV review status	Clinical significance	Last evaluated	Condition(s)
<a href="#">NM_000191.3(HMGCL):c.31C&gt;T(p.Arg11Ter)</a>	★★★ criteria provided, multiple submitters, no conflicts	★★★ criteria provided, multiple submitters, no conflicts	Pathogenic	2024-01-29	<a href="#">Deficiency of hydroxymethylglutaryl-CoA lyase</a>
		★★★ no assertion criteria provided	Pathogenic	2020-11-23	<a href="#">Long chain 3-hydroxyacyl-CoA dehydrogenase deficiency</a>
		★★★ criteria provided, single submitter	Pathogenic	2017-05-18	<a href="#">Inborn genetic diseases</a>

MedGen

Full Report

C0268601 has been replaced by C1533587, showing C1533587

**Deficiency of hydroxymethylglutaryl-CoA lyase (HMGCLD)**

MedGen UID: 735832 • Concept ID: C1533587 • Disease or Syndrome

**Synonyms:** 3-hydroxy-3-methylglutaric aciduria; 3-Hydroxy-3-Methylglutaryl-Coenzyme A Lyase Deficiency; Defect in leucine metabolism; HMG-CoA LYASE DEFICIENCY; HMGCLD; HYDROXYMETHYLGLUTARIC ACIDURIA

**SNOMED CT:** Hydroxymethylglutaric aciduria (410059004); 3-hydroxy-3-methylglutaryl-CoA lyase deficiency (410059004); 3-hydroxy-3-methylglutaric aciduria (410059004); HMG-CoA lyase deficiency (410059004)

**Modes of inheritance:** Autosomal recessive inheritance (Orphanet)

**Gene (location):** HMGCL (1p36.11)

**Monarch Initiative:** MONDO:0009520

**OMIM®:** 246450

**Orphanet:** ORPHA20

**Definition**

3-Hydroxy-3-methylglutaryl-CoA lyase deficiency (HMGCLD) is a rare autosomal recessive disorder with the cardinal manifestations of metabolic acidosis without ketonuria, hypoglycemia, and a characteristic pattern of elevated urinary organic acid metabolites, including 3-hydroxy-3-methylglutaric, 3-methylglutaric, and 3-hydroxyisovaleric acids. Urinary levels of 3-methylcrotonylglycine may be increased. Dicarboxylic aciduria, hepatomegaly, and hyperammonemia may also be observed. Presenting clinical signs include irritability, lethargy, coma, and vomiting (summary by Gibson et al., 1988). [from OMIM]

一般集団では ALT/ALT がない → 希少な常染色体劣性遺伝病の原因バリエントの可能性が排除できない

# バリエントレポートページ(3/6)

<https://grch38.togovar.org/variant/rs671#clinical-significance-mgend>

<https://grch38.togovar.org/variant/12-111803962-G-A#clinical-significance-mgend>

Variant report  
tg<sup>v</sup>47264307

臨床的意義 (MGEND:Medical Genomics Japan Variant Database)

Clinical significance MGEND

Title	Clinical significance	Condition
No data		

Clinical significance ClinVar

Title	VCV review status	RCV review status	Clinical significance	Last evaluated	Condition(s)
NM_000690.4(ALDH2):c.1510G>A (p.Glu504Lys)	★★★★★ reviewed by expert panel	★★★★★ no assertion criteria provided	Pathogenic	2010-02-01	AMED syndrome, digenic
		★★★★★ no assertion criteria provided	drug response	2021-02-12	Alcohol sensitivity, acute
		★★★★★ no assertion criteria provided	protective	2010-02-01	Alcohol dependence
		★★★★★ no assertion criteria provided	risk factor	2010-02-01	Esophageal cancer, alcohol-related, susceptibility to
		★★★★★ no assertion criteria provided	risk factor	2010-02-01	Sublingual nitroglycerin, susceptibility to poor response to
		★★★★★ no assertion criteria provided	risk factor	2010-02-01	Susceptibility to hangover

臨床的意義 (ClinVar)

Clinical significance:

- Pathogenic(病的) → Likely pathogenic → Likely benign → Benign (良性)
- Uncertain significance (不明、VUS)

Review status:

- 星4つ practice guideline
- 星3つ reviewed by expert panel
- 星2つ criteria provided, multiple submitters, no conflicts
- 星1つ criteria provided, conflicting classifications / criteria provided, single submitter
- 星なし no assertion criteria provided / no classification provided / no classification for the individual variant

MGEND All rs671 Search

MGEND data only All data Dataset Download Help News About

SNV/INDEL (0) CNV (0) FUSION/SV (0)

Search Word

Target data : MGEND data only  
Category :  
Search word : rs671  
Filtering :  
Filtering  
Showing 0 to 0 of 0 entries 25 entries per page

Variant name AA change Entry Origin Type Annotation Annotation

No data available in table

<https://mgend.jih.s.go.jp/search/search?query=rs671>

国立健康危機管理研究機構

NIH National Library of Medicine  
National Center for Biotechnology Information

ClinVar Genetic variation as it relates to human health

About Access Submit Stats FTP Help

NM\_000690.4(ALDH2):c.1510G>A (p.Glu504Lys) Cite Follow Print Download

We've updated the full submission template; previous versions are no longer supported. Download the latest version from the ClinVar FTP site and read more about the available templates.

Reviewed by expert panel  
★★★★★  
Mar 2021 by Pharmacogenetics

Drug response

for ethanol response: Toxicity  
Classification is based on the expert panel submission  
Mar 2021 by Pharmacogenetics

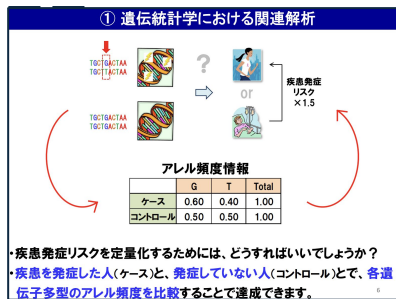
No data submitted for somatic clinical impact

No data submitted for oncogenicity

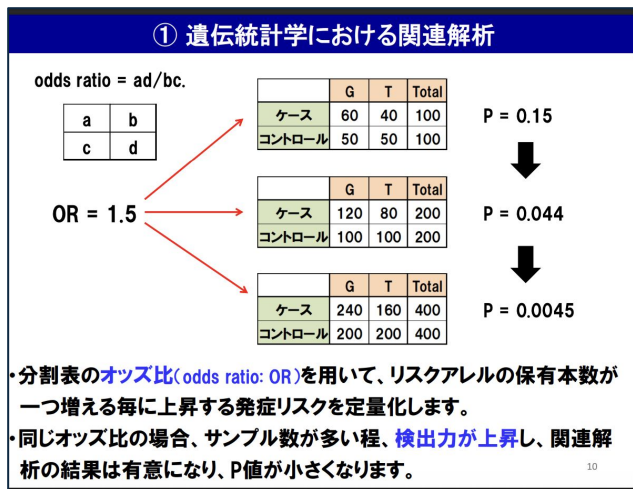
Variant Details

Identifiers: NM\_000690.4(ALDH2):c.1510G>A (p.Glu504Lys)  
Variation ID: 18390 Accession: VCV000018390.10  
Type and length: single nucleotide variant, 1 bp  
Location: Cytogenetic: 12q24.12 12-111803962 (GRCh38) [NCBI UCSC] 12-112241766 (GRCh37) [NCBI UCSC]  
Timeline in ClinVar: First in ClinVar Last submission Last evaluated  
Germline Apr 4, 2013 Jan 11, 2025 Mar 24, 2021

# ゲノムワイド関連解析



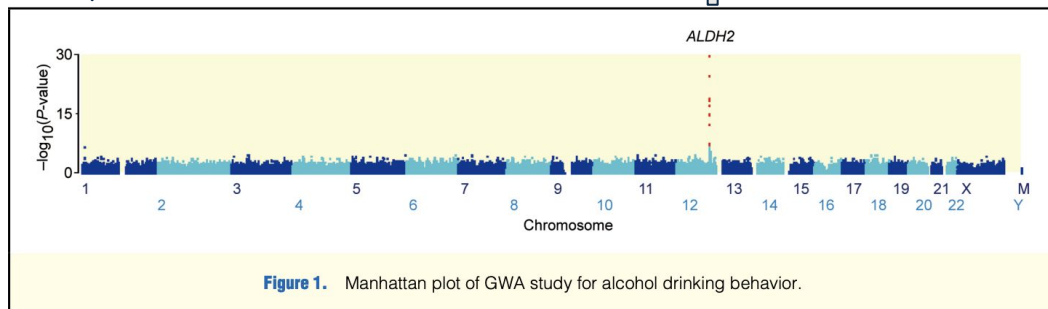
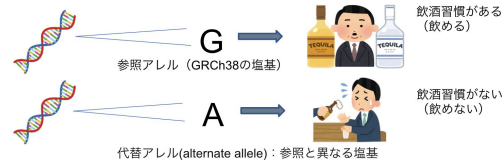
全ゲノムの各バリエーションのアレル頻度差を統計的に検定



ゲノムバリエーションは表現型と関連がある

例) バリエーションとアルコール代謝(表現型)の関連

ヒト参照塩基配列(GRCh38)の12番染色体111803962番目の塩基



A cluster of 12 SNPs on 12q24 were found to **significantly ( $P < 5 \times 10^{-8}$ )** associate with drinking behavior in stage 1, among which **rs671 (a Glu-to-Lys substitution at position 504) in the ALDH2 gene showed the strongest association (odds ratio (OR)=0.16,  $P = 3.6 \times 10^{-211}$  in the joint analysis).**

出典: Takeuchi, Fumihiko et al. "Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population." Circulation journal : official journal of the Japanese Circulation Society vol. 75,4 (2011): 911-8. doi:10.1253/circj.cj-10-0774



# バリエントレポートページ(4/6)

[#genome-assosiation-study](https://grch38.togovar.org/variant/rs671)

<https://grch38.togovar.org/variant/12-111803962-G-A#genome-association-study>

ゲノムワイド関連解析: GWAS Catalog ( $p < 1e - 5$  を収載)を表示

Variant report

tg<sup>v</sup>47264307

RefSNP ID

rs671

no assertion criteria provided

Genome-wide association study

Search for keywords...

Show 5 entries

rs# and risk allele	RAF	P-value	OR	CI	Beta	Beta unit	Trait(s)	PubMed ID	Study accession	Discovery sample description	Replication sample description
rs671-2	0.000e+0	0e+0	0.000e+0	[1.79-1.84]	1.815e+0	unit decrease	alcohol drinking	31959922	GCST009799	83,713 Japanese ancestry ever drinkers, 81,371 Japanese ancestry never drinkers	NA
rs671-2	0.000e+0	0e+0	0.000e+0	[0.41-0.45]	4.300e-1	unit decrease	alcohol consumption measurement	31959922	GCST009800	58,610 Japanese ancestry individuals	NA
rs671-A	2.502e-1	0e+0	0.000e+0	[0.17-0.18]	1.746e-1	unit decrease	serum gamma-glutamyl transferase...	34594039	GCST90018734	133,471 East Asian ancestry individuals	NA
rs671-A	0.000e+0	0e+0	0.000e+0	[0.17-0.18]	1.745e-1	unit decrease	serum gamma-glutamyl transferase...	34594039	GCST90018954	344,104 European ancestry individuals, 133,471 East Asian ancestry individuals	NA
rs671-2	7.500e-1	4e-211	6.250e+0	[5.56-7.14]	0.000e+0		drinking behavior	21372407	GCST000994	733 Japanese ancestry cases, 729 Japanese ancestry controls	2,794 Japanese ancestry drinkers, 1,521 Japanese ancestry chance drinkers, 1,351 Japanese ancestry non-drinkers

1 2 3 4 5 > >>

Page 1 of 16

A cluster of 12 SNPs on 12q24 were found to **significantly ( $P < 5 \times 10^{-8}$ )** associate with drinking behavior in stage 1, among which **rs671 (a Glu-to-Lys substitution at position 504) in the ALDH2 gene showed the strongest association (odds ratio (OR)=0.16,  $P = 3.6 \times 10^{-211}$ )** in the joint analysis).



# バリエントレポートページ(5/6)

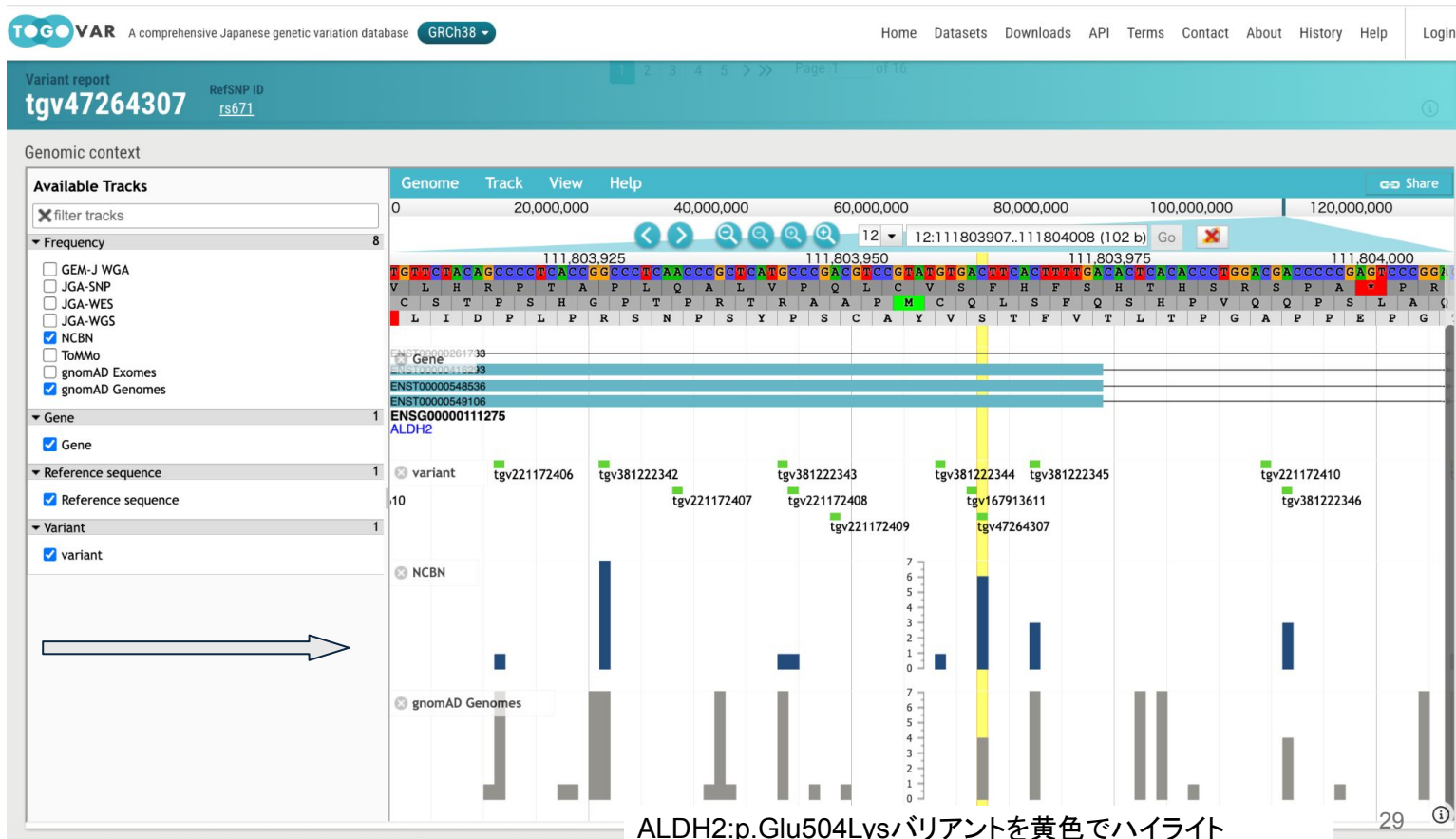
<https://grch38.togovar.org/variant/rs671#genomic-context>

<https://grch38.togovar.org/variant/12-111803962-G-A#genomic-context>

## ゲノムビューア

日本人集団  
それ以外の集団との  
アレル頻度の比較

0: monomorphic  
1: singleton  
2: < 0.0001  
3: < 0.001  
4: < 0.01  
5: < 0.05  
6: < 0.5  
7: ≥ 0.5



\_\_\_\_\_

<https://grch38.togovar.org/variant/12-111803962-G-A#gene>

このバリエントが言及されている文献情報

30

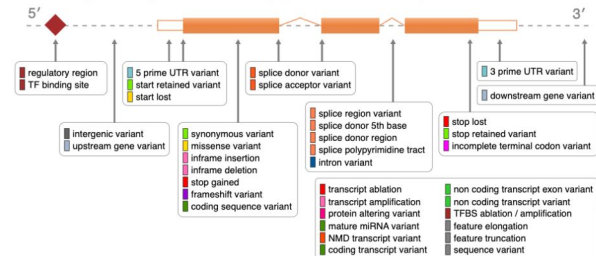
# Molecular consequences

## Ensembl Variation - Calculated variant consequences

For each variant that is mapped to the reference genome, we identify all overlapping Ensembl transcripts. We then use a rule-based approach to predict the effects that each allele of the variant may have on each transcript. The set of consequence terms, defined by the [Sequence Ontology](#) (SO), that can be currently assigned to each combination of an allele and a transcript is shown in the table below. Note that each allele of each variant may have a different effect in different transcripts.

This approach is applied to all germline variants and somatic mutations stored in the Ensembl databases. The resulting consequence type calls, along with information determined as part of the process, such as the cDNA and CDS coordinates, and the affected codons and amino acids in coding transcripts, are stored in the Ensembl Variation database and displayed on our website. For human and mouse variants any overlap with regulatory features is also displayed. For structural variants consequence terms are calculated on the fly for display on our website or API access. You can use this pipeline to annotate your own data via [Ensembl VEP](#). By default, Ensembl VEP will include upstream and downstream annotations for variants within 5kb of a nearby feature, see [--distance](#) in [Ensembl VEP options](#).

See below a diagram showing the location of each display term relative to the transcript structure:



## VEPがtranscriptごとに計算する

The terms in the table below are shown in order of severity (more severe to less severe) as estimated by Ensembl, and this ordering is used on the website summary views. This ordering is necessarily subjective and API and Ensembl VEP users can always get the full set of consequences for each allele and make their own severity judgement. The [IMPACT rating](#) is a separate rating given for compatibility with other variant annotation tools (e.g. [snpEff](#)).

* SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO:0001907	Feature elongation	HIGH
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO:0001908	Feature truncation	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE

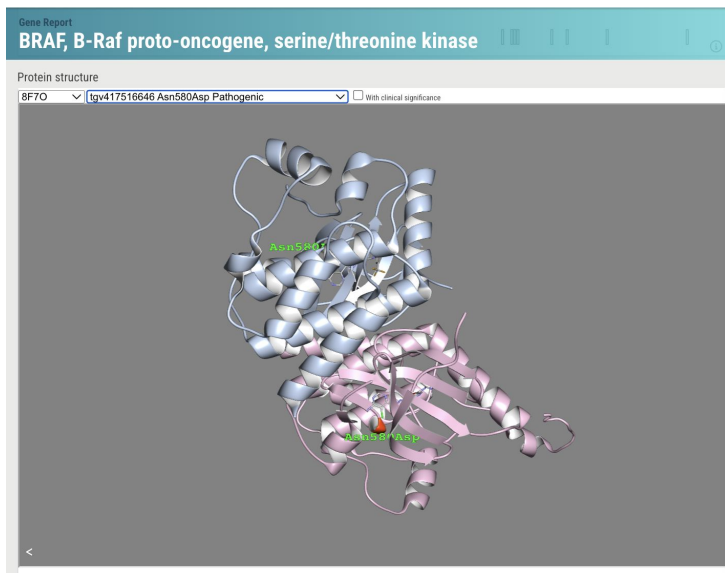
## Consequence

<input checked="" type="checkbox"/> All	845,473,347
<input checked="" type="checkbox"/> Transcript variant	
<input checked="" type="checkbox"/> Coding variant	
<input checked="" type="checkbox"/> Coding sequence variant	415,722
<input checked="" type="checkbox"/> Feature elongation	0
<input checked="" type="checkbox"/> Feature truncation	0
<input checked="" type="checkbox"/> Frameshift variant	12,099,044
<input checked="" type="checkbox"/> Incomplete terminal codon variant	12,404
<input checked="" type="checkbox"/> Inframe deletion	2,731,055
<input checked="" type="checkbox"/> Inframe insertion	1,126,690
<input checked="" type="checkbox"/> Missense variant	170,985,848
<input checked="" type="checkbox"/> NMD transcript variant	357,834,908
<input checked="" type="checkbox"/> Protein altering variant	76,521
<input checked="" type="checkbox"/> Synonymous variant	75,712,606
<input checked="" type="checkbox"/> Start lost	494,648
<input checked="" type="checkbox"/> Stop gained	6,470,885
<input checked="" type="checkbox"/> Stop lost	238,408
<input checked="" type="checkbox"/> Start retained variant	7,672
<input checked="" type="checkbox"/> Stop retained variant	140,627
<input checked="" type="checkbox"/> Non-coding variant	
<input checked="" type="checkbox"/> Splice variant	
<input checked="" type="checkbox"/> Transcript ablation	772
<input checked="" type="checkbox"/> Transcript amplification	0
<input checked="" type="checkbox"/> Regulatory variant	
<input checked="" type="checkbox"/> Mature miRNA variant	28,499
<input checked="" type="checkbox"/> Regulatory region ablation	0
<input checked="" type="checkbox"/> Regulatory region amplification	0
<input checked="" type="checkbox"/> Regulatory region variant	0
<input checked="" type="checkbox"/> TF binding site variant	40,049,364
<input checked="" type="checkbox"/> TFBS ablation	271,533
<input checked="" type="checkbox"/> TFBS amplification	0
<input checked="" type="checkbox"/> Intergenic variant	
<input checked="" type="checkbox"/> Downstream gene variant	0
<input checked="" type="checkbox"/> Intergenic variant	272,608,855
<input checked="" type="checkbox"/> Upstream gene variant	31 0

# 遺伝子レポートページ

PDB/AlphaFold 上での ClinVar バリアント表示

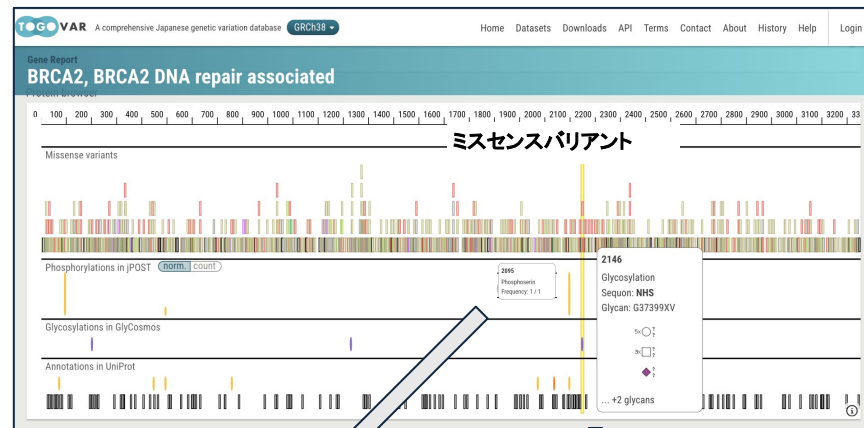
<https://grch38.togovar.org/gene/BRAF#protein-structure>



注: バリアントの影響を考慮した立体構造ではない。

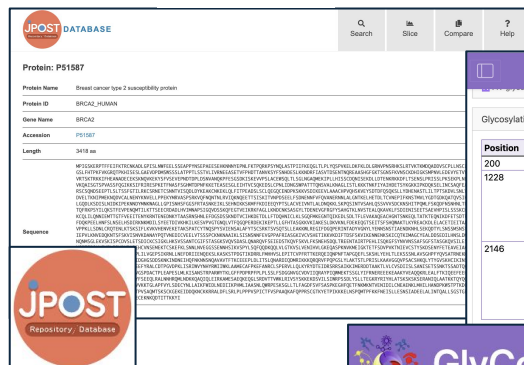
アミノ酸配列上での並列表示

<https://grch38.togovar.org/gene/BRCA2#protein-browser>



リン酸化部位

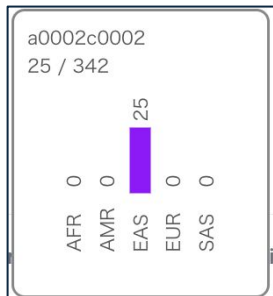
糖鎖修飾部位



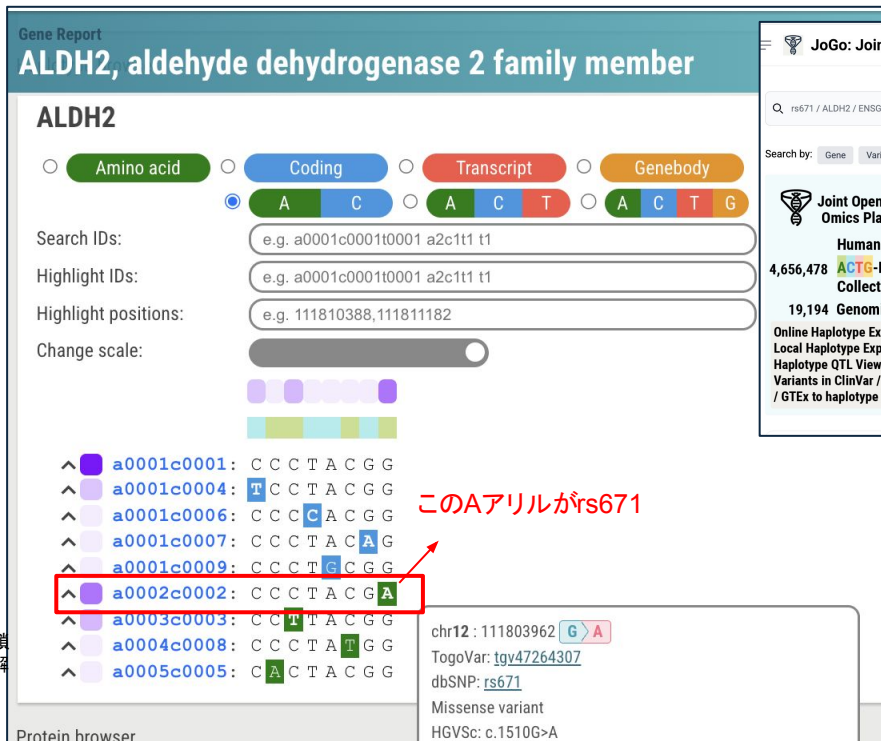
Position	Description	PubMed ID	GlyTouCan ID	Source
200				GlyGen
1228				GlyGen
2146			G03238UC	GlyGen
			G03238BL	GlyGen

# 遺伝子レポートページ(ハプロタイプ)

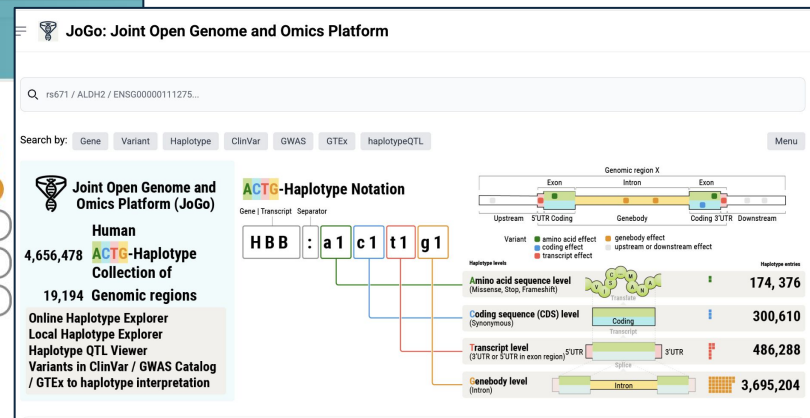
## ハプロタイプ頻度



5大陸の258人から得られた全ゲノム長鎖型シーケンサ由来の塩基配列情報を解析しました(このうち108人はJoGoが測定)



<https://grch38.togovar.org/gene/ALDH2#haplotype-explorer>



JoGo (Joint Open Genome and Omics: ジョーゴ; <https://jogo.csml.org>) でヒトの19,194遺伝子について、4,656,478個のハプロタイプを決定

Nagasaki M, Katayama T, Moriya Y, et al. JoGo 1.0: the ACTG hierarchical nomenclature and database covering 4.7 million haplotypes across 19,194 human genes. Nucleic Acids Res. Published online November 29, 2025. doi:10.1093/nar/gkaf1232



# 遺伝子レポートページ: ヒトバリエントに対応するマウスバリエントの検索例

## ヒトTLR4遺伝子

自然免疫において、病原体のタンパク質をセンシングするトル様受容体

### • TLR4:p.Asp299Gly(rs4986790)

- ミスセンス変異: Aspartic acid → Glycine at position 299 で、TLR4 の外側ドメインに影響を与える。
- LPS 応答性の低下や炎症応答変化 (Figueroa L. et al. J Immunol. 2012 May 1;188(9):4506-15. doi: 10.4049/jimmunol.1200202.)

日本人集団では検出されていないバリエント

TOGVAR A comprehensive Japanese genetic variation database GRCh38

Simple search Advanced search

TLR4:p.Asp299Gly

Handwritten: 手順1. TLR4:p.Asp299Glyを入力

Results The number of available variations is 1 out of 1.

Handwritten: アリル頻度概要

Download JSON CSV TSV

Alternative allele frequencies

Dataset	Alt	Total	Frequency
WGS GEM-J WGA	/	/	/
WGS JGA-WGS	/	/	/
WGS JGA-WES	/	/	/
SNP JGA-SNP	/	/	/
WGS ToMMo	/	/	/
54KJPN	/	/	/
WGS NCBI	296 /	23,534	0.0126
WGS gnomAD	10,842 /	152,192	0.064

TOGVAR A comprehensive Japanese genetic variation database

Handwritten: 日本人集団では検出されていないバリエント

Frequency

Variant	Population	Alt count	Total	Frequency	Alt/Total
rs4986790	Japanese	0	18,522	0.0	0.0
rs4986790	African Caribbean in Barbados	0	188	0.043	0.043
rs4986790	African Ancestry in CHL USA	0	112	0.054	0.054
rs4986790	Bengali in Bangladesh	0	178	0.108	0.108
rs4986790	British from England and Scotland	0	162	0.044	0.044
rs4986790	Chinese in Hainan, Guangdong, China	0	188	0.0	0.0
rs4986790	Colombian in Medellin, Colombia	0	184	0.040	0.040

Clinical significance: None

Search for keywords:

Handwritten: TogoVar IDから詳細なバリエント情報へ

ClinVarの情報に基づいた慢性閉塞性肺疾患の早期発症型との関連性

手順2. TLR4遺伝子ページへ

### Gene Report

#### TLR4, toll like receptor 4

Mouse variants in MoG+ database

C3H/HeJ

Handwritten: 手順3. C3H/HeJ Strainを絞り込み

Link to MoG+ ALT match TogoVar ID rs Chr-pos-ref-alt

4:66759343-C-A C3H/HeJ Yes rs1829300071 9:117714269-C-A

missense\_variant 3\_prime\_UTR\_variant

Handwritten: Jackson研究所へ

C3H/HeJマウスは、LPS 非応答性の自然変異モデル

DOI: 10.1126/science.282.5396.2085

Beutler (Science 2008)

Popular

C3H/HeJ

Strain #000659

RRID:IMSR\_JAX:000659

Common Name: C3H [Also Known As: C3H Heston, C3]

C3H/HeJ mice, commonly called C3H, are used as a general purpose strain in a wide variety of research areas including cancer, infectious disease, sensorimotor, and cardiovascular biology research. A spontaneous mutation in *Tlr4* occurred in C3H/HeJ at the lipopolysaccharide response locus (mutation in toll like receptor 4 gene, *Tlr4*<sup>tm1.1</sup> making C3H/HeJ mice more resistant to endotoxin. C3H/HeJ mice are highly susceptible to infection by Gram-negative bacteria such as *Salmonella enterica*. The C3H substrains at The Jackson Laboratory are homozygous for the retinal degeneration 1...

Read More

# ファイルダウンロード / コマンドで取得

<https://grch38.togovar.org/doc/downloads>

---

## 1. 頻度データ (VCF: Variant Call Format)

GRCh38

Home Datasets Downloads

### Downloads

All variants in TogoVar

- Allele frequency
  - GRCh38: [VCF format \(README\)](#)
  - GRCh37: [VCF format \(README\)](#)
- Annotation
  - GRCh38: [tab-separated text \(README\)](#)
  - GRCh37: [tab-separated text \(README\)](#)

Last updated : 2024/12/3

The allele frequencies were aggregated by [dataset](#). Previous releases

Index of /downloads/release/c		
<a href="#">jga_wgs.vcf.gz</a>	03-Dec-2024 00:00	246M
<a href="#">jga_wgs.vcf.gz.tbi</a>	03-Dec-2024 00:00	2M

**\$ bcftools view -H -r 'chr12:111803962'**

**[https://grch38.togovar.org/downloads/release/current/grch38/frequency/vcf/jga\\_wgs/jga\\_wgs.vcf.gz](https://grch38.togovar.org/downloads/release/current/grch38/frequency/vcf/jga_wgs/jga_wgs.vcf.gz)**

```
chr12 111803962 . G A . PASS
AC=32;AN=156;AF=0.205128;AAC=5;ARC=22;AOC=0;RRC=51;ROC=0;OOC=0;HAC=0;HOC=0
```

## 2. アノテーションデータ (TSV)

Index of /downloads/release/current/gr		
<a href="#">README.txt</a>	15-Jul-2025 04:44	145B
<a href="#">chr_10_molecular_annotation.tsv.gz</a>	23-Apr-2025 08:17	1G
<a href="#">chr_10_molecular_annotation.tsv.gz.tbi</a>	23-Apr-2025 08:17	120K
<a href="#">chr_11_molecular_annotation.tsv.gz</a>	23-Apr-2025 03:23	1G
<a href="#">chr_11_molecular_annotation.tsv.gz.tbi</a>	23-Apr-2025 03:23	125K
<a href="#">chr_12_molecular_annotation.tsv.gz</a>	23-Apr-2025 06:26	1G

**\$ tabix**

**[https://grch38.togovar.org/downloads/release/current/grch38/annotation/chr\\_12\\_molecular\\_annotation.tsv.gz](https://grch38.togovar.org/downloads/release/current/grch38/annotation/chr_12_molecular_annotation.tsv.gz)**  
**12:111803962-111803962**

```
tgV47264307 rs671 12 111803962 G A ALDH2
NM_001204889.2,ENST00000548536,ENST00000416293,ENST00000261733,NM_000690.4,ENST00000549106
3_prime_UTR_variant,missense_variant,NMD_transcript_variant Deleterious0.0 Possibly Damaging 0.709 Likely pathogenic 0.8864
```

bcftoolsやtabixコマンドでpositionなどの検索条件を指定すれば目的のバリエーションの情報だけを取得できる

# 検索API

<https://grch38.togovar.org/api>



A comprehensive Japanese genetic variation database

GRCh38

Home

Datasets

Downloads

API

Terms

Contact

About

History

Help

Configuration

Login



## TogoVar API

/api/v1.yml

TogoVar API version 0.9.1 specification

Servers  
<https://grch38.togovar.org/api>

### search

POST /search/variant

#### Parameters

Name	Description
pretty	If set, response body is pretty
stat	If set to 0, do not return statistics
data	If set to 0, do not return data

#### Request body

Schema for searching variant

#### Examples:

List variants exist on ALDH2 (Aldehyde Dehydrogenase 2 family member).

Example Value Schema

```
{
  "query": {
    "gene": {
      "relation": "eq",
      "terms": [
        "ALDH2"
      ]
    }
  }
}
```

- List variants whose class is any of insertion, deletion and indel.
- List variants by identifiers (tg, rs)
- List variants exist on an exact position
- List variants exist in a certain range
- List variants range from 0.05 to 0.95 in frequency in JGA-WES.
- List variants by genotype count.
- ✓ List variants exist on ALDH2 (Aldehyde Dehydrogenase 2 family member).
- List variants associated with C2675520 (Breast-ovarian cancer, familial 2).
- List variants associated with MONDO\_0003582 (hereditary breast ovarian cancer syndrome).
- Paginate result over 10,000 records.
- List variants by prediction score
- List variants by prediction score unassigned or unknown

### Responses

#### Curl

```
curl -X 'POST' \
  'https://grch38.togovar.org/api/search/variant' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
    "query": {
      "gene": {
        "relation": "eq",
        "terms": [
          404
        ]
      }
    }
  }'
```

#### Request URL

<https://grch38.togovar.org/api/search/variant>

#### Server response

Code Details

200

#### Response body

```
{
  "data": [
    {
      "type": "SO_0001483",
      "chromosome": "12",
      "position": 111766887,
      "reference": "A",
      "alternate": "T",
      "symbols": [
        {
          "name": "ALDH2",
          "id": 404,
          "synonyms": []
        }
      ],
      "external_link": {
        "gnomad": [
          {
            "title": "12-111766887-A-T",
            "xref": "https://gnomad.broadinstitute.org/variant/12-111766887-A-T?dataset=gnomad_r4"
          }
        ]
      }
    }
  ]
}
```



# まとめ

- 対象者：日本人一般集団のアレル頻度データの利用例を知りたい
  - 発表内容：TogoVarのアレル頻度を参照した論文を紹介
  - 伝えたいこと
    - 日本人祖先集団とそれ以外の集団ではアレル頻度（含：バリエーションの有無）が違う
    - 日本人について言及するならば日本人集団のデータ（個人レベル、頻度データ）が必要
    - アレル頻度を参照いただいた際にはTogoVarへの言及をお願いいたします
- 対象者：バリエーションを解釈するためのアノテーション情報を取得したい
  - 発表内容：以下の情報の検索方法をTogoVarを使って説明
    - 臨床的意義：ClinVarなどのDB由来の病原性分類（Pathogenic/Benign 等）
    - 関連解析結果：GWAS などの 疾患・形質との関連（効果量・p値等）
    - 有害性予測：SIFT/PolyPhen-2/AlphaMissense による影響度・有害性スコア
    - 文献情報：PubMed 等に基づく 関連論文（PMID・書誌情報・リンク）
    - NBDC統合化推進プログラムとの連携
      - ハプロタイプ（JoGo）、プロテオーム（jPOST）、糖鎖（Glycosmos）、マウス（MoG+）
  - 伝えたいこと
    - 臨床情報から文献情報のワンストップ提供は使い勝手に勝負
    - NBDC統合化推進プログラムとの連携は現状は相互リンク。ユニークな試み。乞うご期待

# 最後に

---

- TogoVar現開発メンバー

- DBCLS

- 三橋 信孝
    - 守屋 勇樹
    - 川嶋 実苗
    - 川島 秀一
    - 片山 俊明

- 株式会社 カモノハシ

- 佐藤 大輔

- 株式会社 PENQE

- 永野 朗夫
    - 野田 夕紀子

- 謝辞

- データ作成・提供してくださった皆様
  - 共同研究、連携先データベースの皆様
  - NBDCの皆様
    - 高木 利久 データベース特別主監