

・○豊岡理人<sup>1</sup>、三橋信孝<sup>2</sup>、川嶋実苗<sup>1</sup>、片山俊明<sup>2</sup>、川島秀一<sup>2</sup>

1) 科学技術振興機構バイオサイエンスデータベースセンター

2) 情報・システム研究機構ライフサイエンス統合データベースセンター

# 概要

- ヒトゲノムのバリエーション頻度情報は疾患関連遺伝子探索において重要な基礎的情報
- コホート等のプロジェクトで収集・解析されプロジェクト毎にウェブで公開された日本人のバリエーション頻度情報を統合
- バリエーションに関連する頻度情報、疾患との関連、アノテーション情報および論文情報も統合して検索
- gnomADの頻度情報の取り込み、および、拡張検索機能を開発

# TogoVarで検索可能なバリエーション頻度情報

| データベース名       | 対象人数    | プラットフォーム | 説明                                                             |
|---------------|---------|----------|----------------------------------------------------------------|
| JGA-NGS       | 125     | WES      | NBDCヒトデータベースに寄託されたデータからのアليل頻度情報                               |
| JGA-SNP       | 183,884 | SNP chip |                                                                |
| GEM-J WGA     | 7,609   | WGS      | バイオバンク・ジャパン、理化学研究所、東北メディカル・メガバンク機構(ToMMo)の共同研究により解析されたアليل頻度情報 |
| ToMMo 8.3KJPN | 8,380   | WGS      | ToMMoが収集したゲノムコホート（東北地方中心）からのアليل頻度情報                           |
| HGVD          | 1,208   | 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ليل頻度情報                            |
| gnomAD        | 125,748 | WES      | 世界の様々な地域から収集した8集団におけるアليل頻度情報 ※日本人は76名のWESデータのみしか含まれない         |
|               | 15,708  | WGS      |                                                                |

# TogoVarで検索可能なアノテーション情報

| データベース名                  | 運営                | 説明                                                          |
|--------------------------|-------------------|-------------------------------------------------------------|
| Variant Effect Predictor | EMBL-EBI          | バリエーションのアノテーション情報を表示                                        |
| GWAS catalog             | NHGRI<br>EMBL-EBI | ゲノムワイド関連研究(Genome-Wide Association Study:GWAS)によって得られた結果を表示 |
| ClinVar                  | NCBI              | 疾患とバリエーションの関連情報を表示                                          |
| PubTator                 | NCBI              | 論文とバリエーションの関連情報を表示                                          |
| LitVar                   | NCBI              |                                                             |
| Colil                    | DBCLS             | 論文の引用情報を表示                                                  |

# 一覧検索画面

**TogoVar** A comprehensive Japanese genetic variation database

Home Datasets Downloads Terms Contact About History Help Configuration

Search for disease or gene symbol or rs... **① 検索入力欄**

Region(GRCh37/hg19): 10-73270743-73376976

**Results** The number of available variations is 10,000 out of 93,337,906. **② 検索結果表示欄**

| TogoVar ID                  | RefSNP ID                    | Position | Ref / Alt    | Type     | Gene | Alt frequency | Consequence        | SIFT | PolyPhen | Clinical sig |
|-----------------------------|------------------------------|----------|--------------|----------|------|---------------|--------------------|------|----------|--------------|
| <a href="#">tgv83272254</a> |                              | 1: 10117 | CCCT... 31bp | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv83272256</a> | <a href="#">rs144773400</a>  | 1: 10145 | A            | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv67071948</a> | <a href="#">rs779258992</a>  | 1: 10147 | C            | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv83272257</a> | <a href="#">rs1286868604</a> | 1: 10150 | CT           | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv83272259</a> |                              | 1: 10151 | T            | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv83272260</a> |                              | 1: 10152 | A            | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv83272264</a> | <a href="#">rs1164014856</a> | 1: 10168 | CTAA... 10bp | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv67071950</a> | <a href="#">rs1366371903</a> | 1: 10173 | CCTA... 5bp  | Deletion |      |               | Intergenic variant |      |          |              |
| <a href="#">tgv67071951</a> | <a href="#">rs1409475383</a> | 1: 10174 | CTAA         | Deletion |      |               | Intergenic variant |      |          |              |

**Filters** **③ フィルタ欄**

Dataset

|                                     |                   |            |
|-------------------------------------|-------------------|------------|
| <input checked="" type="checkbox"/> | All               | 93,337,906 |
| <input checked="" type="checkbox"/> | WGS GEM-J WGA     | 78,767,666 |
| <input checked="" type="checkbox"/> | WES JGA NGS       | 4,561,277  |
| <input checked="" type="checkbox"/> | SNP JGA SNP       | 1,249,723  |
| <input checked="" type="checkbox"/> | WGS ToMMo 4.7KJPN | 65,598,014 |
| <input checked="" type="checkbox"/> | WES HGVD          | 554,461    |
| <input checked="" type="checkbox"/> | WES ExAC          | 9,107,451  |
| <input checked="" type="checkbox"/> | Disease ClinVar   | 663,441    |

Alternative allele frequency

0 1 Invert range

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

☐ for all datasets ☒ for any dataset

- ① 検索入力欄 : rs番号、遺伝子名、疾患名、位置、範囲で検索可能
- ② 検索結果表示欄 : 検索結果を表示
- ③ フィルタ欄 : ②の結果を絞り込む

# 一変異画面 1

**TogoVar** A comprehensive Japanese genetic variation database

Home Datasets Downloads Terms Contact About History Help

Variant report

dbSNP  
tgvr47264307 rs671

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

**Other overlapping variants**

| TogoVar ID | Variant type | Ref / Alt | Alt frequency |
|------------|--------------|-----------|---------------|
|------------|--------------|-----------|---------------|

**Frequency**

| Dataset       | Population | Allele count<br>Alt | Total   | Frequency |
|---------------|------------|---------------------|---------|-----------|
| GEMJ 10K      | Japanese   | 3,536               | 15,008  | 0.236     |
| JGA-NGS       | Japanese   | 52                  | 250     | 0.208     |
| JGA-SNP       | Japanese   | 90,026              | 365,930 | 0.246     |
| ToMMo 4.7KJPN | Japanese   | 1,886               | 9,546   | 0.198     |
| HGVD          | Japanese   | 451                 | 1,890   | 0.239     |

**Clinical significance**

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

**Genome-wide association study**

Search for keywords...

| rs# and risk allele | RAF  | P-Value | OR    | CI               | Beta | Trait(s)           | PubMed ID | Study Details     | Discovery sample description                                | Replication sample description                                                                                  |
|---------------------|------|---------|-------|------------------|------|--------------------|-----------|-------------------|-------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| rs671-?             | 0.75 | 4e-211  | NA    | [5.56-7.14]      | 6.25 | drinking behavior  | 21372407  | Drinking behavior | 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 |
| rs671-2             | NR   | 4e-153  | 0.354 | [0.33-0.38] unit | NA   | coffee consumption | 31959922  | Coffee            | 152,634 Japanese ancestry individuals                       | NA                                                                                                              |

- 頻度情報、疾患との関連(ClinVar)、GWASの結果(GWAS-catalog)を表示

# 一変異画面2

The screenshot displays the Ensembl genome browser interface for the **ALDH2** gene. The top section shows the genomic context with a track view of the reference sequence and various annotations. The left sidebar lists available tracks, including Frequency, Gene, Reference sequence, and Variation. The right sidebar provides detailed information about the gene, transcripts, and publications.

**Gene Information:**

- Gene symbol: **ALDH2**
- Gene description: aldehyde dehydrogenase 2 family member
- Gene ID: ENSG00000111275

**Transcripts:**

| Transcript ID   | Gene symbol | Consequence type       | HGVSc(cDNA)                   | HGVSp(Amino acid seq.) | SIFT              | PolyPhen                |
|-----------------|-------------|------------------------|-------------------------------|------------------------|-------------------|-------------------------|
| ENST00000261733 | ALDH2       | missense_variant       | ENSP00000261733.2:p.Glu504Lys |                        | 0.000 Deleterious | 0.874 Possibly Damaging |
| ENST00000416293 | ALDH2       | missense_variant       | ENSP00000403349.3:p.Glu457Lys |                        | 0.000 Deleterious | 0.566 Possibly Damaging |
| ENST00000548536 | ALDH2       | 3_prime_UTR_variant    |                               |                        |                   |                         |
| ENST00000549106 | ALDH2       | NMD_transcript_variant |                               |                        |                   |                         |

**Publications:**

| PMID     | Reference                                                                                                                | Year       | Cited by |
|----------|--------------------------------------------------------------------------------------------------------------------------|------------|----------|
| 34243011 | Global colorectal cancer burden in 2020 and projections to 2040. Xi Y, Xu P. <i>Transl Oncol.</i> 2021-10;14(10):101174. | 2021-10-01 | 0        |
| 34228649 | Aldehyde dehydrogenase 2 protects against acute kidney injury by regulating autophagy via the Beclin-1 pathway.          | 2021       |          |

- ゲノムブラウザ、バリエーションが存在する遺伝子の情報
- バリエーションが存在するトランスクリプトの情報およびトランスクリプトに与える機能的影響のスコア値 (SIFT, Polyphen)
- バリエーションに関連する論文情報を表示

# 拡張検索画面

**TOGOVAR** A comprehensive Japanese genetic variation database

Home Datasets Terms Contact About Help ja en

Simple search **Advanced search** History Clear condition

Search query: `[disease:"cancer" | alt: { type:"freq", db:"all", gt:0.05 } & genotype: { genotype:"rr", type:"count", db:"bbj", gfe:10, lfe:100 } & variant: ["snv","insertion","deletion"] & significance:["p"]]`

**Datasets** **Builder**

Conditions:

- Disease: control ☒ Not
- Gene symbol: PC-1 ☐ Not
- Variant: ☒ SNV ☒ Insertion ☒ Deletion ☒ Indel ☐ Substitution
- Consequence: Intergenic variant ☐ Not
- Consequence: Splice region variant ☐ Not
- Clinical significance: Pathogenic ☐ Not

**Consequence**

|                    |                          |                         |                                                                                                                                                        |
|--------------------|--------------------------|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| Transcript variant | Coding variant           | Splice acceptor variant | Splice region variant                                                                                                                                  |
| Regulatory variant | Non-coding variant       | Splice donor variant    | A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron |
| Intergenic variant | Splice variant           | Splice region variant   |                                                                                                                                                        |
|                    | Transcript ablation      |                         |                                                                                                                                                        |
|                    | Transcript amplification |                         |                                                                                                                                                        |

**123,456,789** Variants: showing 1 to 25 of 343 variants in the gene "PDCD1", (filtered from 19,537,677 variants)

| TogoVar ID | RefSNP ID   | Position | Ref / Alt    | Type  | Symbol  | Alt frequency | Consequence        | SIFT  | PolyPhen | Clinical significance                                |
|------------|-------------|----------|--------------|-------|---------|---------------|--------------------|-------|----------|------------------------------------------------------|
| tgvr21     | rs55998931  | 1:10492  | C / T        | SNV   |         |               | Intergenic variant | 0.000 | 0.998    | Intellectual disability                              |
| tgvr27     | rs189107123 | 1:10611  | AGAG... 52bp | Ins   |         |               | Intergenic variant | 0.070 | 0.446    | Immunodeficiency 38 with basal ganglia calcification |
| tgvr29     | rs62635286  | 1:10623  | TCTC... 32bp | Del   |         |               | Intergenic variant | 0.027 | 0.000    | Thoracic aortic aneurysm and aortic dissection       |
| tgvr41     | rs62028691  | 1:13116  | TCT / G      | Indel | DDX11L1 |               | Intron variant     | 0.001 | 0.992    | Shprintzen-Goldberg syndrome                         |

(※ 画面は開発中のもの)

- 様々な検索語をAND、ORで結合した複雑な条件を組み合わせたクエリが作成可能



# 画面以外のデータ利用方法

- データのダウンロードが可能
  - <https://togovar.biosciencedbc.jp/doc/downloads>
- APIを公開予定
  - <https://togovar.biosciencedbc.jp/api>
- 一変異画面に埋め込まれた各可視化要素(スタンザ)は再利用が可能
  - DBCLSが開発したMetaStanzaを利用
  - 各スタンザの右下にあるiマークをクリックすると、埋め込みに必要な情報が表示される⇒ユーザーが独自のページを作成可能

# 今後の予定

- GRCh 38対応
- 疾患別頻度情報の表示
- 構造多型(SV)の表示
- その他、要望などありましたらご連絡下さい。

togovar [at] biosciencedbc [dot] jp