日本人ゲノムバリエーション統合データベース TogoVarの機能拡張

○三橋 信孝¹、川嶋 実苗²、片山 俊明¹、川島 秀一¹、守屋 勇樹¹、豊岡 理人³

1. 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(DBCLS)
2. 科学技術振興機構 NBDC事業推進部(NBDC)
3. 富山国際大学

TogoVarとは

- ヒトゲノムのバリエーション頻度情報の統合データベースで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

TogoVarで閲覧可能なバリエーション頻度

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リル頻度情報
JGA-SNP	183,884	SNP chip	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
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リル頻度情報
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)からのアリル頻度情報
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リル頻度情報
gnomAD	125,748	WES	世界の様々な地域から収集した8集団におけるアリル頻度情報 ※日本人は76名のWESデータのみしか含まれない

今回の機能拡張

トップページ: <https://togovar.biosciencedbc.jp/>

1. Advanced search (新規)

4. 検索API (新規)

The screenshot shows the TogoVar website interface. At the top, there is a navigation bar with links for Home, Datasets, Downloads, API, Terms, Contact, About, History, Help, and Configuration. Below the navigation bar, there are two tabs: Simple search and Advanced search. The search bar contains the text 'rs671'. Below the search bar, there is a summary of the search results: Disease: Breast-ovarian cancer, familial 2; Gene: ALDH2; refSNP: rs114202595; TogoVar: tgv421843; Position(GRCh37/hg19): 16:48258198; Region(GRCh37/hg19): 10:73270743-73376976; HGVS: c.1510G>A; HGVS: p.Glu504Lys. Below this, there is a 'Results' section stating 'The number of available variations is 1 out of 1.' A table displays the search results with columns for TogoVar ID, RefSNP ID, Position, Ref / Alt, Type, Gene, Alt frequency, Consequence, SIFT, PolyPhen, and Clinical significance. The first row shows: TogoVar ID: tgv47264307; RefSNP ID: rs671; Position: 12: 112241766; Ref / Alt: G > A; Type: SNV; Gene: ALDH2; Alt frequency: 100% (represented by a bar chart); Consequence: Missense variant; SIFT: 0 (+3); PolyPhen: 0.874 (+3); Clinical significance: P AMED syndrome, digenic (+5). To the right of the table is a 'Statistics / Filters' panel with a 'Reset' button. It shows a 'Dataset' filter with three options: All (1), WGS GEM-J WGA (1), WES JGA NGS (1), and SNP JGA SNP (1).

2. 遺伝子レポートページ (新規)

3. 疾患レポートページ (新規)

1. Advanced search

ANDとORを入れ子にした複雑な検索式をGUIで直感的に作成可能

例: 集団間で代替アレル頻度に差のあるバリエントを検索

((frequency(GEM-J WGA) >= 0.3) OR (frequency(ToMMo 8.3KJPN) >= 0.3))

AND

(frequency(gnomAD genomes) <= 0.05) OR (frequency(gnomAD exomes) <= 0.05))

日本人集団で高頻度

世界中の集団の平均値で低頻度

TOGOVAR A comprehensive Japanese genetic variation database

Home Datasets Downloads API Terms Contact About History Help Configuration

Simple search Advanced search

Alternative allele frequency/count GEM-J WGA 0.3 ~ 1 Exclude filtered out variants

Alternative allele frequency/count ToMMo 8.3KJPN 0.3 ~ 1 Exclude filtered out variants

Alternative allele frequency/count gnomAD genomes 0 ~ 0.05 Exclude filtered out variants

Alternative allele frequency/count gnomAD exomes 0 ~ 0.05 Exclude filtered out variants

Add condition Group Ungroup Delete

検索式ビルダー (Add conditionから追加)

検索結果: 条件変更時にリアルタイムに反映

Results The number of available variations is 10,000 out of 30,350.

TogoVar ID	RefSNP ID	Position	Ref / Alt	Type	Gene	Alt frequency	Consequence	SIFT
tgv2158	rs201365517	1: 612688	CTC	Deletion			Intron variant	
tgv67086890	rs1553174279	1: 974792	GC	Insertion	AGRN		Intron variant	
tgv22068	rs57527288	1: 1092563	G T	SNV			Intergenic variant	
tgv23923	rs35247267	1: 1123284	AA	Deletion	ITLL10		Intron variant	
tgv30173	rs112572774	1: 1223626	T C	SNV	SCNN1D		Intron variant	
tgv30244	rs1977348	1: 1225579	G C	SNV	SCNN1D		Intron variant	
tgv30526	rs2273276	1: 1226292	G A	SNV	SCNN1D		Synonymous variant	
tgv31621	rs2297234	1: 1233580	G T	SNV	ACAP3		Non coding transcript exon variant	

Statistics

Dataset

All	30,350
WGS GEM-J WGA	27,846
WES JGA NGS	18,641
SNP JGA SNP	2,987
WGS ToMMo 8.3KJPN	30,341
WES HGVD	320
WGS gnomAD genomes	30,338
WES gnomAD exomes	646
Disease ClinVar	194

Variant type

▼ Karyotype

操作手順は以下のヘルプページを参照
「日本人集団とその他の集団で代替アレル頻度が異なるバリエントを検索できますか」

1. Advanced search(続き)

- 従来のSimple searchが便利な場合
 - キーワードがひとつの場合
 - 遺伝子シンボル1つ(例:[ALDH2](#))
 - 疾患名1つ(例:[Acute Myeloid Leukemia](#))
 - TogoVar ID(例:[tgv47264307](#))
 - dbSNP rsID(例:[rs671](#))
 - SNVのポジション(例:[12:112241766](#))
- Simple searchのみが対応している検索条件
 - HGVS表記(例:[ALDH2:p.Glu504Lys](#))
 - 疾患名による検索(Advanced searchにも実装中)
 - SIFT、PolyPhen2の値によるファセット検索

2. 遺伝子レポートページ

1遺伝子に関するバリエーション情報を1ページに集約

- 例: ALDH2(hgnc_id: 404)

<http://togovar.biosciencedbc.jp/gene/404>

遺伝子DBへのリンク

当該遺伝子上のバリエーションの臨床的意義

当該遺伝子上のバリエーションのGWAS情報

当該遺伝子領域のゲノムビューア

当該遺伝子領域の全バリエーション

当該遺伝子領域のバリエーションが参照されている文献

Accession	Frequency	RefSeq	Type	Class	Phenotype	Comment	DBP	Alleles	Clinical significance
rs1042522	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042523	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042524	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042525	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042526	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042527	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042528	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042529	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042530	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042531	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042532	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042533	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042534	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042535	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042536	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042537	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042538	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042539	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	
rs1042540	0.000000	CG	SNP	SNP	2 gene OTR variant		SNP	C	

- U6-19 not provided
- 06-10 not provided
- 08-07 Download CSV
- 12-31 Download TSV
- 05-21 Copy HTML snippet to clipboard
- About this stanza

各表(スタンザ)の右端の iマークメニューに以下の機能

- ダウンロード機能
- スタンザを任意のwebサイトに埋め込むためのHTMLタグの表示 (Copy HTML snippet to clipboard)

3. 疾患レポートページ

1疾患に関するバリエーション情報を1ページに集約

例: 急性骨髄性白血病 (MedGen CUI: C0023467)

<https://togovar.biosciencedbc.jp/disease/C0023467>

当該疾患の MedGen での説明

疾患 DB へのリンク

当該疾患が関係するバリエーションとその臨床的意義

当該疾患が関係する GWAS の情報

TOGOVAR A comprehensive Japanese genetic variation database

Home Datasets Downloads API Terms Contact About History Help ja en

Disease Report

Acute myeloid leukemia

Title Acute myeloid leukemia

Definition CEBPA-associated familial acute myeloid leukemia (AML) is defined as the presence of a heterozygous germline CEBPA pathogenic variant in an individual with AML and/or family in which more than one individual has AML. In contrast, sporadic CEBPA-associated AML is defined as AML in which a CEBPA pathogenic variant(s) is identified in leukemic cells but not in the non-leukemic cells. In the majority of individuals, the age of onset of familial AML appears to be earlier than sporadic AML; disease onset has been reported in persons as young as age 1.8 years and up to age 50 years. The prognosis of CEBPA-associated familial AML appears to be favorable compared with sporadic CEBPA-associated AML. Individuals with CEBPA-associated familial AML who have been cured of their initial disease may be at greater risk of developing additional independent leukemic episodes, in addition to the risk of relapse from preexisting clones.

Links EFO: [EFO_000222](#) MedGen: [C0023467](#) MeSH: [D015470](#) MONDO: [MONDO:0018874](#)

Clinical significance

TogoVar ID	rs#	Position	Title	Clinical significance	Review status	Last evaluated
tg948213495	rs431825334	13:32914239	NM_000059.3(BRCA2):c.5747A>G (p.His1916Arg)	Uncertain significance	★★★★ criteria provided, multiple submitters, no conflicts	2016-12-07
tg948217024	rs587782860	13:32945216	NM_000059.3(BRCA2):c.8611G>C (p.Glu2871Gln)	Uncertain significance	★★★★ criteria provided, multiple submitters, no conflicts	2016-06-10
tg948214235	rs56019712	13:32915052	NM_000059.3(BRCA2):c.6560C>T (p.Pro2187Leu)	Uncertain significance	★★★★ criteria provided, conflicting interpretations	2016-11-07
tg917039039	rs569408054	4:55593396	NM_000222.2(KIT):c.1553C>T (p.Pro518Leu)	Uncertain significance	★★★★ criteria provided, multiple submitters, no conflicts	2018-10-31
tg917038078	rs144933028	4:55561810	NM_000222.2(KIT):c.200C>G (p.Thr67Ser)	Uncertain significance	★★★★ criteria provided, conflicting interpretations	2018-10-31

Genome-wide association study

rs# and risk allele	RAF	P-value	OR	CI	Beta	Beta unit	Trait(s)	PubMed ID	Study accession	Discovery sample description	Replication sample description
rs2600685-T (tg9565739) (rs2600683-C)	5.270e-1	1e-6	0.000e+0	"NA"	NaN		acute myeloid leukemia	27903959	GCST008413	175 European ancestry cases, 218 European ancestry controls	NA
rs7571171-G (tg9573105) (rs2541066-C)	0.000e+0	3e-10	0.000e+0	"NA"	NaN		acute myeloid leukemia	27903959	GCST008413	175 European ancestry cases, 218 European ancestry controls	NA
rs7571171-G (tg9573105) (rs2541066-C)	1.100e-1	1e-7	0.000e+0	"NA"	NaN		acute myeloid leukemia	27903959	GCST008413	175 European ancestry cases, 218 European ancestry controls	NA
rs7604033-C (tg9674438) (rs10497522-C)	0.000e+0	2e-12	0.000e+0	"NA"	NaN		acute myeloid leukemia	27903959	GCST008413	175 European ancestry cases, 218 European ancestry controls	NA
rs7604033-C (tg9674438) (rs10497522-C)	1.410e-1	4e-7	0.000e+0	"NA"	NaN		acute myeloid leukemia	27903959	GCST008413	175 European ancestry cases, 218 European ancestry controls	NA

- Download JSON
- Download CSV
- Download TSV
- Copy HTML snippet to clipboard
- About this stanza

遺伝子レポートページと同様

4. 検索API

APIページ: <https://togovar.biosciencedbc.jp/api/>

Swaggerを利用したAPIドキュメント上で検索例を実行可能

TogoVar API 0.9.1 OAS3

TogoVar API version 0.9.1 specification

Servers
<https://togovar.biosciencedbc.jp/api>

手順1. バリアントの検索 (/search/variant)を選択

- 他に遺伝子 (/search/gene)および疾患 (/search/disease)を検索するAPIがある

search

POST /search/variant

Parameters

Name	Description
pretty	If set, response body will be nicely formatted.

pretty
boolean
If set, response body will be nicely formatted.
(query)

Request body

Schema for searching variant

Examples:

List variants by identifiers (tgv, rs)

```
{
  "query": {
    "ids": [
      "tgv421843",
      "rs114202595"
    ]
  }
}
```

手順2. 「Try it out」をクリック

手順3. tgv, rsで検索するクエリ例を選択

JSONで記述されたクエリ例

手順4. Executeボタンを押下

Execute Clear

Responses

Curl

```
curl -X 'POST' \
  'https://togovar.biosciencedbc.jp/api/search/variant' \
  -H 'accept: application/json' \
  -H 'content-type: application/json' \
  -d '{
  "query": {
    "id": [
      "tgv421843",
      "rs114202595"
    ]
  }
}'
```

curlコマンドの例

Request URL

<https://togovar.biosciencedbc.jp/api/search/variant>

Server response

Code Details

200

Response body

```
{
  "data": [
    {
      "id": "tgv421843",
      "type": "SO_0001483",
      "chromosome": "1",
      "position": 12979845,
      "start": 12979845,
      "stop": 12979845,
      "reference": "T",
      "alternate": "C",
      "existing_variations": [
        "rs1890765"
      ],
      "symbols": [
        {
          "name": "PRAMEF7",
          "id": 28415,
          "synonyms": []
        }
      ]
    }
  ]
}
```

検索結果はJSONで返される

Response headers

```
access-control-allow-methods: GET,POST,HEAD
access-control-allow-origin: *
access-control-expose-headers:
access-control-max-age: 7200
```

今後の開発計画(開発中を含む)

- Simple/Advanced search 検索結果表のダウンロード機能
 - 現在の上限10,000レコード以上のデータをダウンロード可能に
- 構造多型(Structural variant, SV, 50bp以上)の追加と可視化
- 簡易認証機能(登録者公開データ公開のため)の実装
 - NBDCヒトデータベース/Japanese Genotype-phenotype Archive(JGA)のデータセット(例:[JGAD000220](#))毎のアレル頻度の表示
- データベース間連携
 - [Medical Genomics Japan Variant Database\(MGeND\)](#)
 - [バイオバンク横断検索システム](#)
 - [Mouse Genome database with high added value \(MoG+\)](#)

バリエーション頻度データ更新計画

データセット名	対象人数	プラットフォーム	GRCh38対応	今後の予定
JGA-NGS	125	WES	追加予定WGSデータはGRCh38でバリエーションコール。既存データ(WES)はLiftOver	ヒトDBのWGSデータをGATK4でバリエーションコール。データセット毎に集計したアレル頻度を公開予定(登録者公開データ)
JGA-SNP	183,884	SNP chip	LiftOverで座標変換	バイオバンクジャパン47疾患別アレル頻度公開予定(非制限公開)
GEM-J WGA	7,609	WGS	LiftOverで座標変換	特になし
ToMMo 8.3KJPN	8,380	WGS	ToMMo14JPN、38KJPNに更新	ToMMo14JPN、38KJPNに更新
HGVD	1,208	WES	LiftOverで座標変換	特になし
gnomAD	125,748	WES	gnomAD v3.1.2に更新	特になし