20

日本人ゲノムバリアント統合データベース 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

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

TogoVarとは

- ヒトゲノムのバリアント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アント頻度情報を統合
- バリアント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

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

TogoVarで閲覧可能なバリアント頻度

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

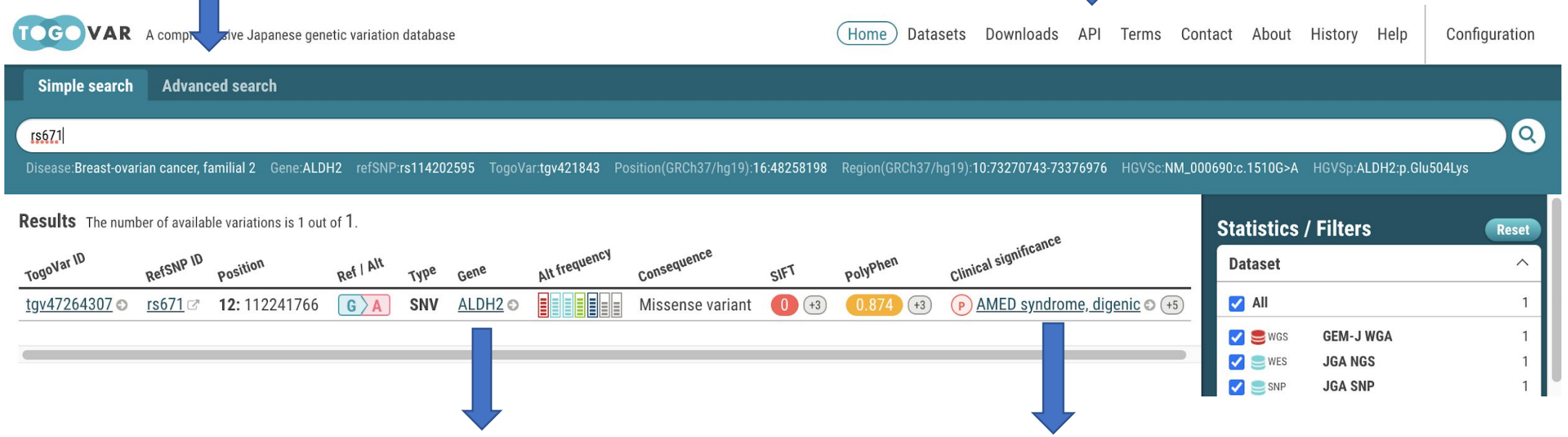
Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

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）



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）

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　ライフサイエンス統合データベースセンター

1. Advanced search

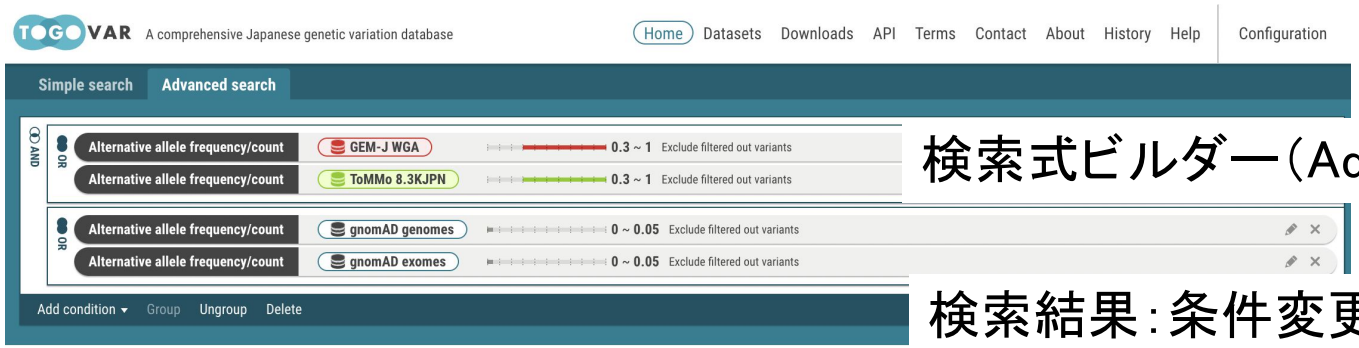
ANDとORを入れ子にした複雑な検索式をGUIで直感的に作成可能

例：集団間で代替アレル頻度に差のあるバリアントを検索

((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))

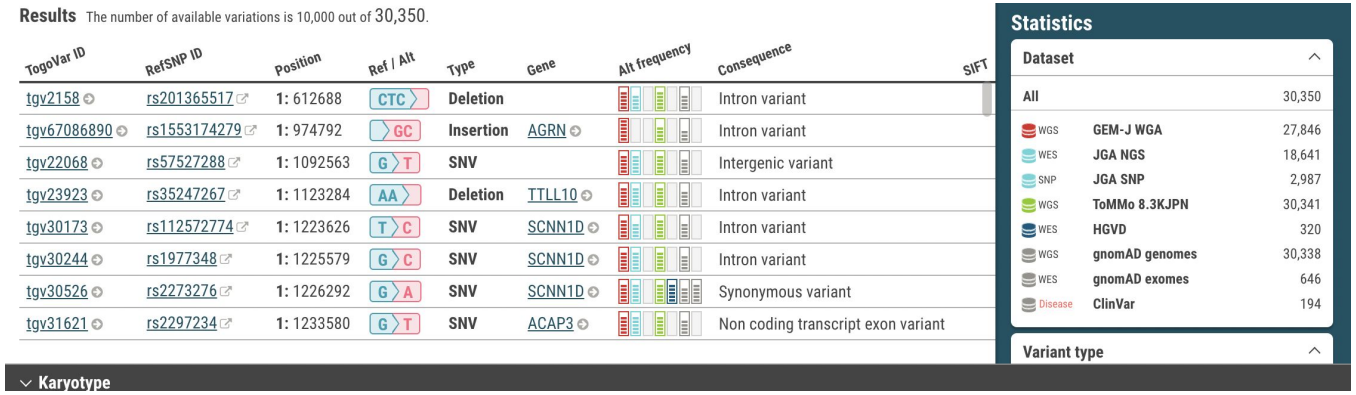
日本人集団で高頻度

世界中の集団の平均値で低頻度



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



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アントを検索できますか」

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

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

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

2. 遺伝子レポートページ

1遺伝子に関係するバリアント情報を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

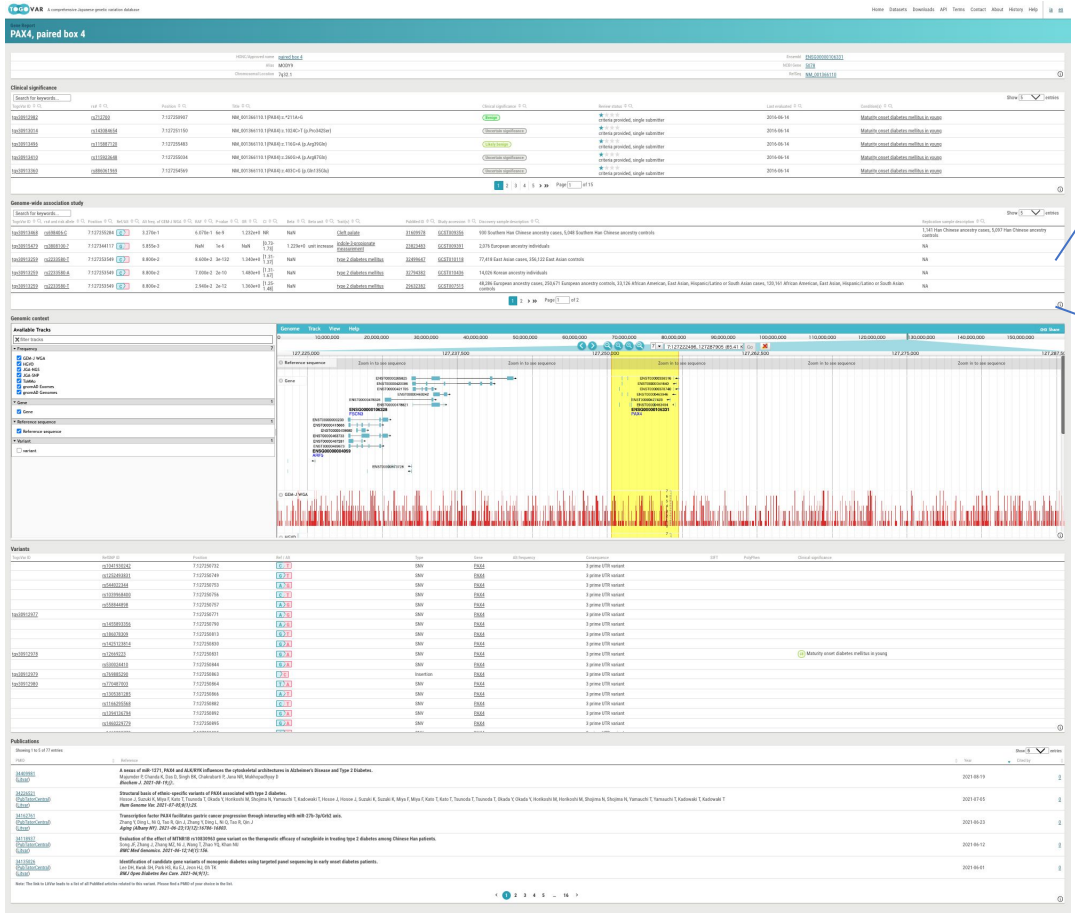
当該遺伝子上のバリアントの臨床的意義

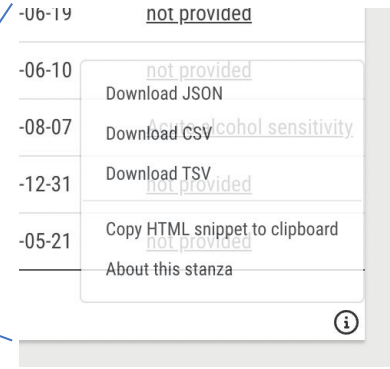
当該遺伝子上のバリアントのGWAS情報

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アント

当該遺伝子領域のバリアントが参照されている文献





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)

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

3. 疾患レポートページ

1疾患に関係するバリアント情報を1ページに集約

例：急性骨髄性白血病 (MedGen CUI: C0023467)

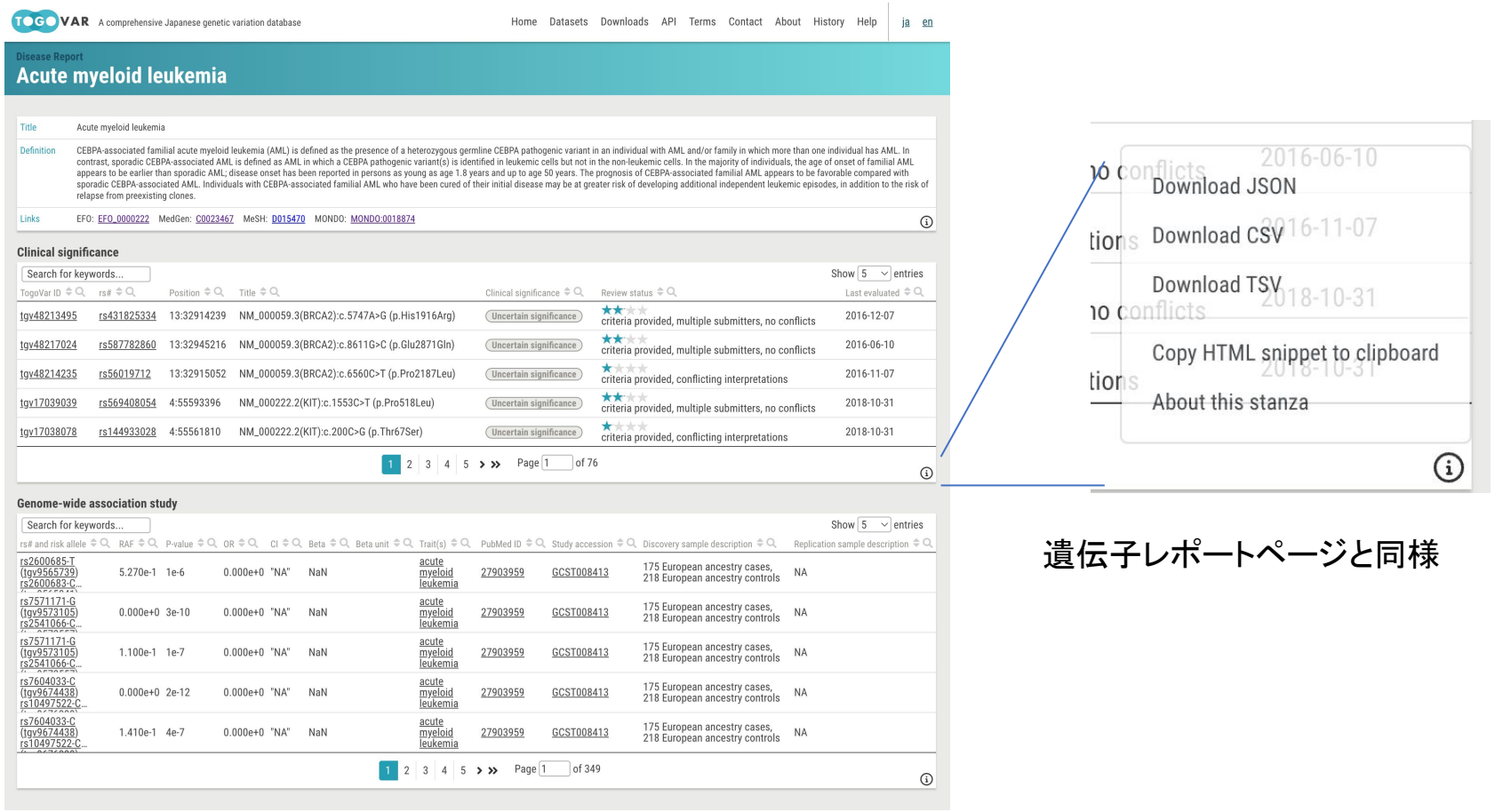
https://togovar.biosciencedbc.jp/disease/C0023467

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アントとその臨床的意義

当該疾患が関係するGWASの情報



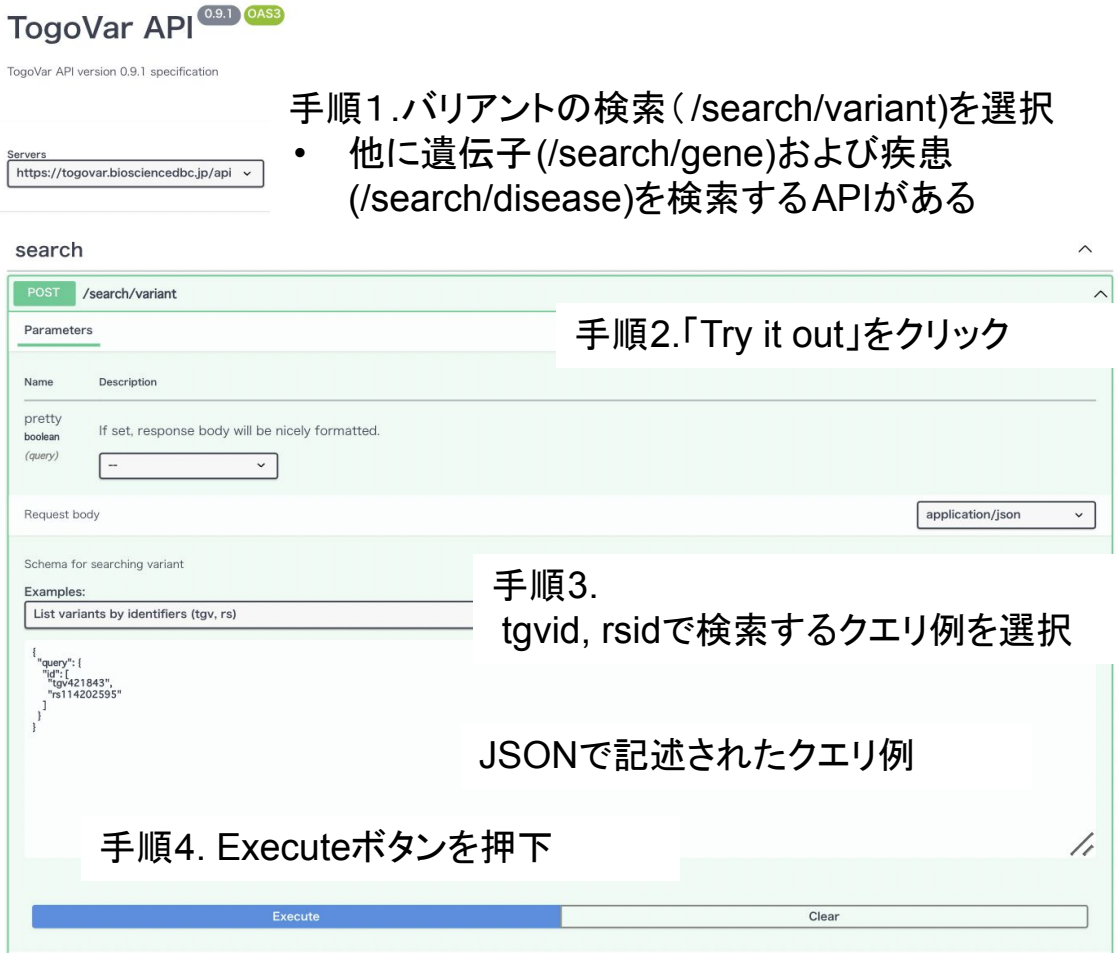
遺伝子レポートページと同様

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

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



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

手順2.「Try it out」をクリック

手順3.
tgvid, rsid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

curlコマンドの例

検索結果はJSONで返される

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022 ライフサイエンス統合データベースセンター

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+)

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　ライフサイエンス統合データベースセンター

バリアント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でバリアントコール。既存データ(WES)はLiftover	ヒトDBのWGSデータをGATK4でバリアントコール。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

Licensed under a Creative Commons 表示4.0国際ライセンス
©2022　ライフサイエンス統合データベースセンター