

日本人ゲノム多様性統合データベースTogoVarを使ってみる

TogoVarとは

- NBDCヒトデータベースから個人特定されない加工データ（頻度情報）を作成して、データの概要を把握可能に
- 国内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して、ワンストップで検索可能に

WSプログラム（以下の内容を2回実施）

1. TogoTVを使ったハンズオンデモ 10分
2. 今後の開発予定と話題提供 10分
3. 自由討論 5分

資料URL <https://bit.ly/2DNNGELE>

TogoTVを使ったハンズオンデモ 約10分

TogoTVの手順に沿ってTogoVarを使ってみます。
PCやタブレットをお持ちの方は一緒に操作できます。

質問等は近くにいる開発チームメンバーまでお願いします。

TogoVarでヒトゲノムに存在するバリエーションに関する情報を調べる (8分36秒)



生命科学系DB・ツール使い倒し系チャンネル

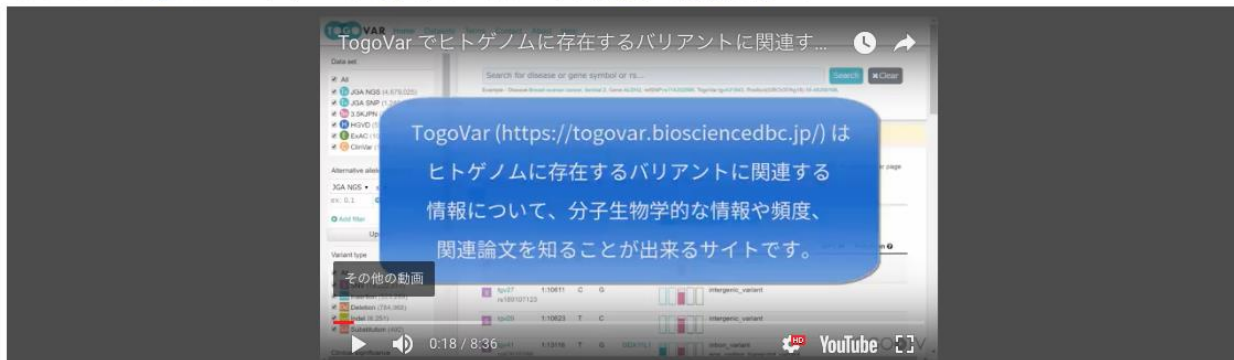
はじめの方へ 再生数ランキング お問い合わせ・番組をリクエスト

『統合TV』は、生命科学分野の有用なデータベースやツールの使い方を動画で紹介するウェブサイトです。

目的別に検索

- 講習会 実習資料 (AJACS)
- ゲノム・核酸 配列解析
- タンパク質 配列・構造解析
- 発現制御解析・可視化
- 文献・辞書・プログラミング
- 著名データベース
- その他講演・講習会
- 自由に使える画像を探す

2018-08-25 TogoVar でヒトゲノムに存在するバリエーションに関する情報を調べる



YouTube版を視聴できない方はオリジナル版ファイル(mov形式)をダウンロードして、ご覧ください。

TogoVar (日本人ゲノム多様性統合データベース) は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC) と大学共同利用機関法人 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) が共同開発する、日本人ゲノム配列の個人による違い (バリエーション) とそれに関する疾患情報や関連論文などを収集・整理したデータベースです。

TogoVarは、研究プロジェクト横断的に集約した日本人におけるバリエーションの頻度情報を提供します。日本人集団として、NBDCヒトデータベース/Japanese Genotype-phenotype Archive (JGA) に登録されている個人ごとのゲノムデータを集約して得られた頻度情報である JGA-NGSデータセット (125 人の WES (Whole Exome Sequencing) データ)、JGA-SNPデータセット (18 万人以上の SNP-chip データ)、東北メディカルメガバンクの Integrative Japanese Genome Variation Database (IJGVD)、京都大学の Human Genetic Variation Database (HGVD)、日本人以外の Population を含むものとして Broad Institute の Exome Aggregation Consortium (ExAC) などが検索対象のデータセットとして用意されています [TogoVar収録データセット一覧]。このように多種多様なデータベースに散在して収録されてきた遺伝子型や表現型に関連する情報を整理統合し、バリエーションを解釈するための情報をワンストップでわかりやすく提供しています。

今回は、TogoVarの基本的な機能やデータの見方について紹介します。

▶▶ 見どころダイジェスト

1. 検索とフィルタリングの方法と検索結果の見方 (0:42)
2. バリエーションの詳細ページの見方 (3:10)
3. アルコール代謝に関連する有名なバリエーションの人類別割合を調べる (6:30)

この動画を引用する際はDOIをご利用ください。 DOI: 10.7875/togotv.2018.153

この記事は、クリエイティブ・コモンズ表示ライセンスの下で利用可能です。利用にあたっては、下記のクレジットを必ず表示してください。
© 2016 DBCLS TogoTV / CC-BY 4.0



TogoVar TogoTV

すべて 地図 動画 画像 ショッピング もっと見る 設定 ツール

約 93 件 (0.40 秒)

togotv - YouTube

<https://m.youtube.com/user/togotv/feed>

<http://togotv.dbcls.jp/20180825.html> TogoVar (日本人ゲノム多様性統合データベース) は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC) と大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用 ...

TogoVar でヒトゲノムに存在するバリエーションに関する情報を ... - Togo...

togotv.dbcls.jp/ja/20180825.html

3 日前 - TogoVar (日本人ゲノム多様性統合データベース) は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC) と大学共同利用機関法人 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 ...

今後の開発予定と話題提供 約10分

1. 2018年度中に公開予定
1. 具体的な実装方法や公開時期は検討中
1. 取り組むべきか否か
1. TogoVarデータの再利用

1. 2018年度中に公開予定

1. 機能追加

- 頻度データ（tsvファイル）の一括ダウンロード
 - <https://togovar.biosciencedbc.jp/downloads>
- GRCh38やJRGのゲノム位置で検索可能にする
- HGVS表記（例：11:g.112241766G>A）での検索

1. 日本人バリエーションデータの追加・更新

- 3.5KJPNv2 へのバージョンアップ
- MGeND の臨床的意義(Clinical significance) 追加

1. その他リファレンスデータの追加

- gnomAD（12万WES、1万WGSのアレル頻度）追加
- dbSNP（約3億バリエーションのID）追加
- GTEx (eQTL: バリエーションと発現の関係) 追加

2. 具体的な実装方法や公開時期は検討中

1. 遺伝統計学的な切り口

- 連鎖不平衡係数 (r^2 や D') の追加
- アレイデータ (JGA-SNP) のインピュテーション
- ゲノムワイド関連解析(GWAS)データの追加
 - GWAS-Catalog、JENGERなど

1. 遺伝子、ゲノム、分子的な切り口

- タンパク質立体構造上でバリエーション位置を可視化
- ミスセンス変異の有害性予測ツールの予測値の追加
 - SIFT、PolyPhen2よりも高精度のツールが登場

1. 疾患、表現型的な切り口

- PubCaseFinder (症状と疾患名の対応) との連携
 - 症状⇔疾患⇔バリエーション

2. 具体的な実装方法や公開時期は検討中 (続き)

4. 文献中のバリエーション情報の抽出の高度化

- ー バリエーションが有害か否かまで自動的に抽出したい
 - 現在はバリエーションが出現する文献名を表示するのみ (NCBI PubTator利用)

4. エンドユーザー手持ちのバリエーションリストの一括処理

- ー DBに登録済みの既知のバリエーションのフィルタリング
 - 希少疾患では対象の患者のみに存在するバリエーションを優先的に解析したい
 - 本当に欲しいのはフィルタリング後のVUS(Variant Uncertain Significance)を解釈するための情報。1から4はそれに資するものにした

3. 取り組むべきか？

1. 遺伝子発現

- 組織別の発現データ
 - 特定の臓器で発現する遺伝子のバリエーションを検索

1. 比較ゲノム

- 疾患モデル生物（バイオリソースへのリンク）
- オースログや近縁種の遺伝子との系統樹

1. パスウェイ、エピゲノムなど

TogoVar単独ではしっかりしたものを作るのは難しい
TogoVarのデータを再利用しやすくして、各分野の専門家
にお任せする

4. TogoVarデータの再利用

JGA-NGS、JGA-SNPの日本人情報およびその他の頻度情報、論文情報が一括取得できるように

1. ダウンロードファイル

- tsvファイル (<https://togovar.biosciencedbc.jp/downloads>)
- vcfファイル
- ttlファイル (RDFデータ)

1. SPARQLList

- REST URLをSPARQLクエリに変換してRDFストア検索

1. SPARQL Endpoint

- TogoVar RDFストアをSPARQLで直接検索可能

4. TogoVarデータの再利用

どういうデータがAPIやSPARQLで取得できると使われるか？

入力データ種別	値例	出力結果
Disease	breast cancer	関連バリアントのリスト
Disease	breast cancer	関連 遺伝子 のリスト
Gene	PCK9	当該 遺伝子 上のバリアントのリスト
Gene	PCK 9	関連 疾患 のリスト
SNPID	rs7988901 tgv123450	当該バリアント 関連情報 （1バリアントビュー）
位置 + アレル	22-46661580-T-C	該当する 遺伝子関連情報 （1遺伝子ビュー）
ゲノム上の範囲	22-46661580-46661680	当該領域に存在する バリアント一覧
ゲノム上の範囲	22-46661580-46661680	当該領域に存在する 遺伝子一覧

自由討論 約5分

以下に限らず、ご意見をいただけますと幸いです。

ユーザーの皆様：

TogoVarに格納しているデータや機能について

開発者の皆様：

TogoVarとのシステム連携について

いただいたご意見と回答

いただいたご意見やご質問	回答
<p>1バリエーションのアレルやジェノタイプだけでなく、ハプロタイプ、ディプロタイプの情報も個人特定できない粒度でよいので表示してほしい。</p>	<p>どのくらいの近傍を表示すればよいのか、どのくらいの頻度なら個人特定できないと判断するのかなどの課題はありますが、検討したいと思います。</p>
<p>頻度情報再解析前の1次データを取得可能か。</p>	<p>JGA-NGS、JGA-SNPデータセットの1次データはNBDCヒトデータベースから取得可能です。</p>
<p>DNAメチル化の情報も表示してほしい。</p>	<p>エピゲノムデータの必要性を指摘していただきありがとうございます。検討したいと思います。</p>
<p>SIFT/PolyPhen2が表示されていなかったり、詳細画面で1バリエーションに対して複数の値が表示されたのはどういう意味か。</p>	<p>SIFT/PolyPhen2はアミノ酸置換を伴うmissenseバリエーションに対してのみ計算される値です。また、スプライシングバリエーションなどで、当該バリエーションを含む領域の翻訳アミノ酸配列が複数存在する場合は、各々に対して値を計算しています。検索結果の表には重大度が最大の値のみを表示しています。</p>