日本人ヒトゲノムを中心としたバリアントデータベース「TogoVar」

〇豊岡理人[1)]、三橋信孝[1)]、川嶋実苗[1)]、建石由佳[1)]、藤原豊史[2)]、片山俊明[2)]、川島秀一[2)]
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

•日本人ゲノム多様性統合データベース

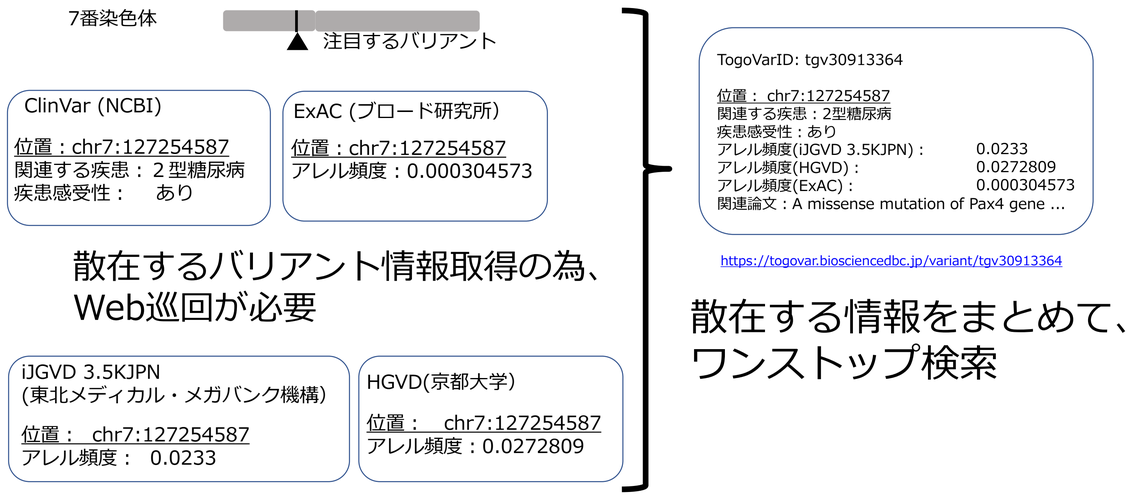
日本や海外で公開されている頻度情報、ゲノム多様性と疾患との関連情報を統合、ワンストップで検索可能に

•NBDCヒトデータベースへ登録、公開されたデータについて、個人特定されない加工データ（頻度情報）を提供、データの概要を把握可能に

•当面はgermline variantのみ扱う方針

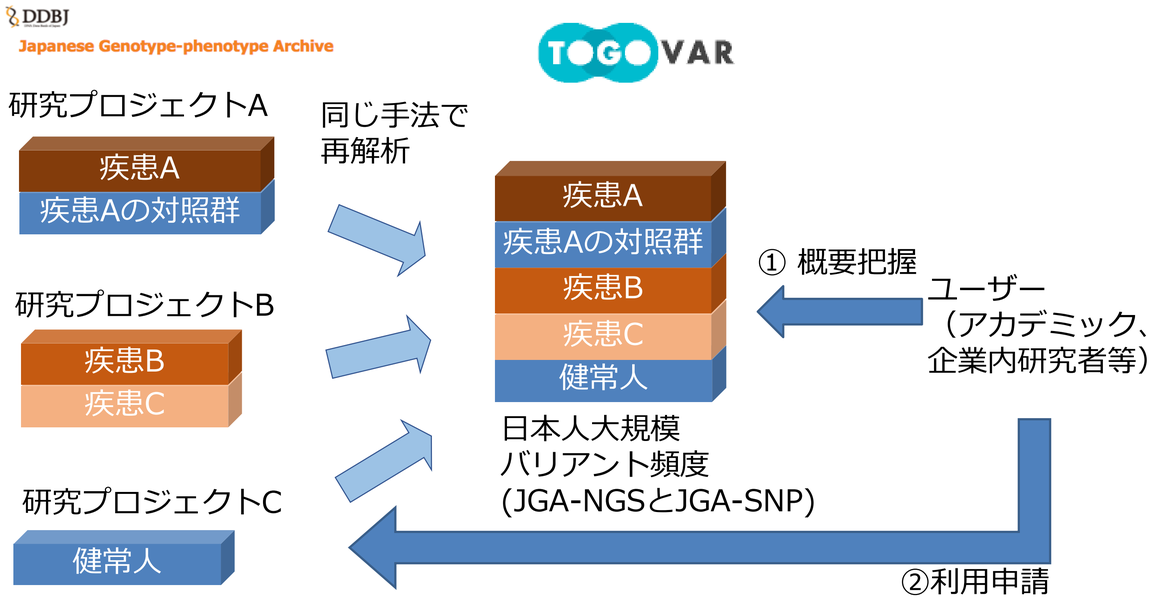
目的1：ワンストップ検索

多種多様なデータベースに散在して収録されてきたGenotypeやPhenotypeに関連する情報を整理統合し、バリアントを解釈するための情報をワンストップで提供



目的2：NBDCヒトDBから公開されているデータの把握

NBDCヒトデータベースに登録・公開された日本人のゲノムデータから集計した大規模なバリアントの頻度情報をTogoVarから公開



ワンストップ検索の対象データベース

データベース名 および　運営組織	説明	対象人数(解析対象)
JGA-NGSデータセット JGA-SNPデータセット （NBDC）	主に日本の研究者からの全エクソームとSNP Chipデータ	125人 （全エクソーム） 183,884人 （SNP Chip）
Japanese Multi Omics Reference Panel (jMorp) （東北メディカル・メガバンク機構）	ゲノムコホート （東北地方中心）	3,554人 （全ゲノム） **1%以下の頻度も公開**
Human Genetic Variation Database （京都大学）	ゲノムコホート （滋賀県長浜市を中心）	1,208人 （全エクソーム）
Exome Aggregation Consortium(ExAC) （ブロード研究所）	約２０プロジェクトからのゲノムデータを再解析	60,706人 （全エクソーム）
ClinVar（NCBI）	バリアントと疾患との関連	443,213 variants **2019年6月までのデータを取り込み**
PubTator (NCBI) Colil(DBCLS)	バリアント（rs番号）が出現する文献情報	

東北メディカル・メガバンクおよびClinVarのデータを更新

JGA-NGSデータセット作成時　解析パイプライン

プログラム	バージョン	説明
qcleaner†	4.1.0	FastQCと組み合わせて低品質リード配列の除去等
SAM tools	1.6	SAM/BAMファイル等を操作するコマンド群
FastQC	0.11.5	リード配列品質チェック
calcCoverage†	1.7	リード配列マッピングカバー率の計算
BWA	0.7.16a	リード配列をリファレンスゲノムにマッピング
Picard	2.13.2	BAMやVCFファイルを操作するコマンド群
GATK	3.8.0-ge9d806836	BAMからバリアントを検出
bedtools	2.13.3	BED関連ファイルの操作

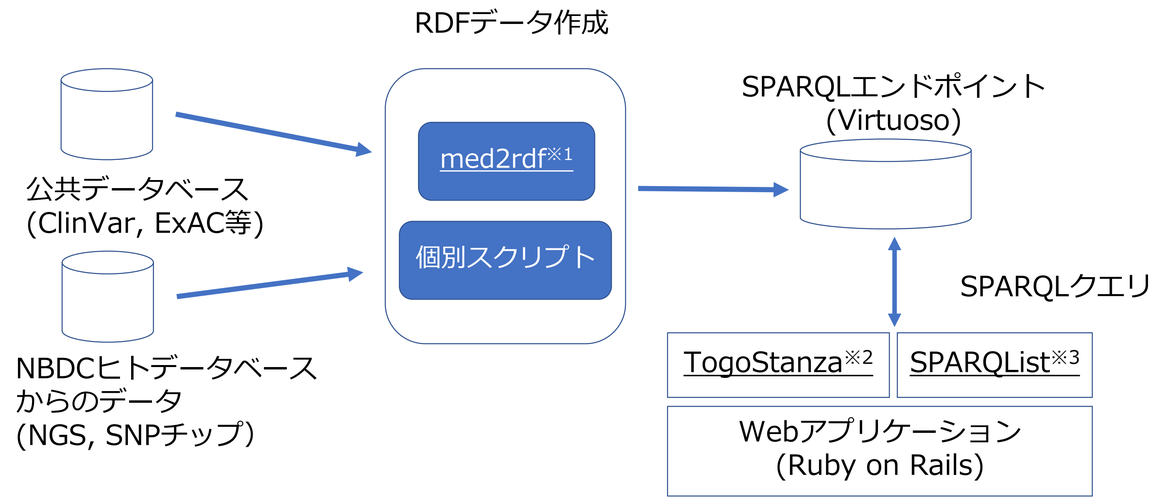
†印：アメリエフ社独自開発ソフト

解析パイプラインソフト
（アメリエフ社製 BioReT　ver.3.3.1.1）を使用

今後は、GATK Best Practice準拠のパイプラインを利用予定

その他、全てのバリアントについては
Variant Effect Predictor (VEP) を用いてアノテーション情報を付加

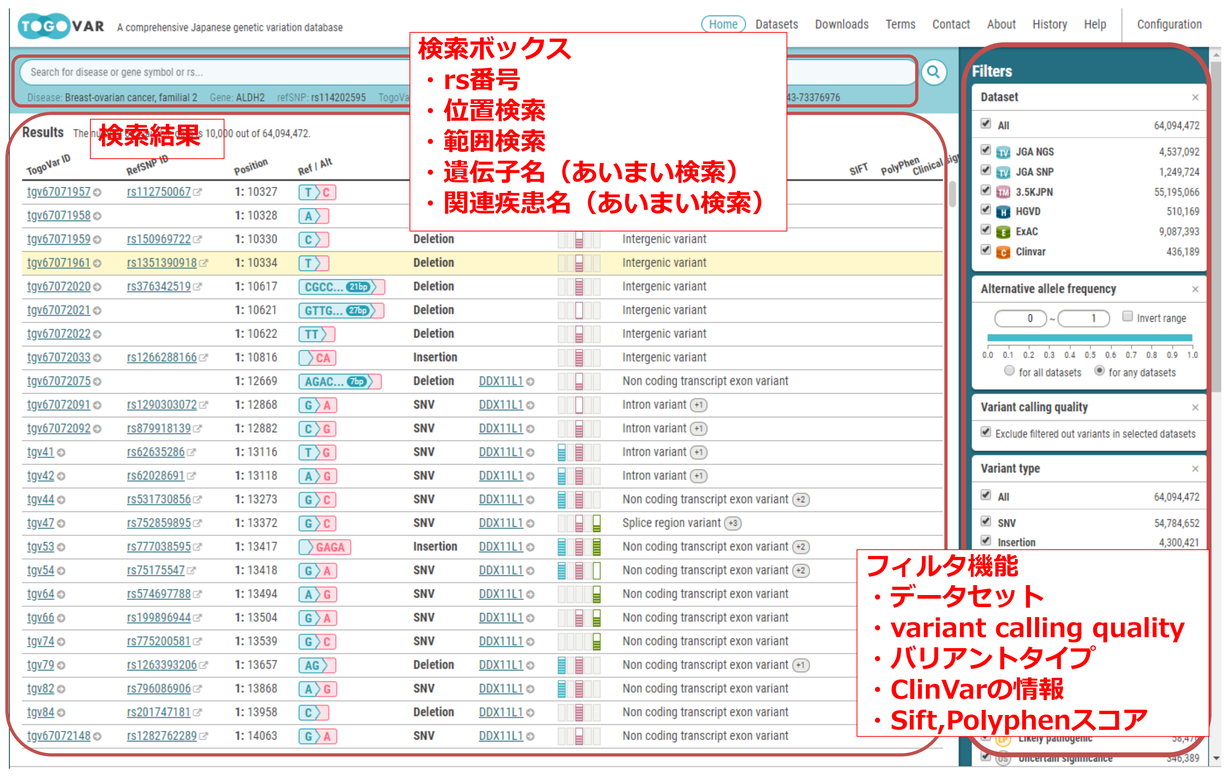
DBCLSのRDF基盤技術の利活用



利用した主なDBCLSのRDF基盤技術
※1 med2rdf: 公共データベースのRDFデータ化スクリプトのリポジトリ
※2 SPARQList：SPARQLクエリをREST APIとして提供するツール
※3 TogoStanza：　SPARQLクエリの結果を可視化するツール

全文検索エンジンとしてElasticsearchを利用

検索画面(http://togovar.biosciencedbc.jp)



2019年7月に機能更新
表示機能の追加
・画面項目、フィルタ項目のユーザーカスタマイズが可能に

フィルタ項目の追加
・次世代シークエンサのvariant callingのquality
・Sift, Polyphenのスコア
・Variant Effect PredictorのConsequence

トーゴ―の日2019
会期：2019年10月5日　　於：科学未来館