NBDC
National Bioscience Database Center

2018年10月5日

トーゴーの日シンポジウム2018
バイオデータベース：つないで使う

誰がためにTogoVarはある

三橋 信孝

バイオサイエンスデータベースセンター(NBDC)、JST
研究員

科学技術振興機構
Japan Science and Technology Agency

TOGOVAR

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

2017年トーゴーの日のおさらい

NBDCの今後の取組み：「作る」から「使う」へ

大型プロジェクト研究と連携し、未公開データのプロジェクト内共有を支援

応用につながる領域に焦点をあて、基礎研究データの統合を行う ✔

様々な利用者との協業により、統合データの利活用に取り組む

統合データベース：「作る」から「使う」へ　(高木 利久 2017年10月4日)
https://events.biosciencedbc.jp/images/togo2017/H29-s01_takagi.pdf

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

目次

1. TogoVar開発の経緯

2. TogoVar開発の現場

3. まとめ

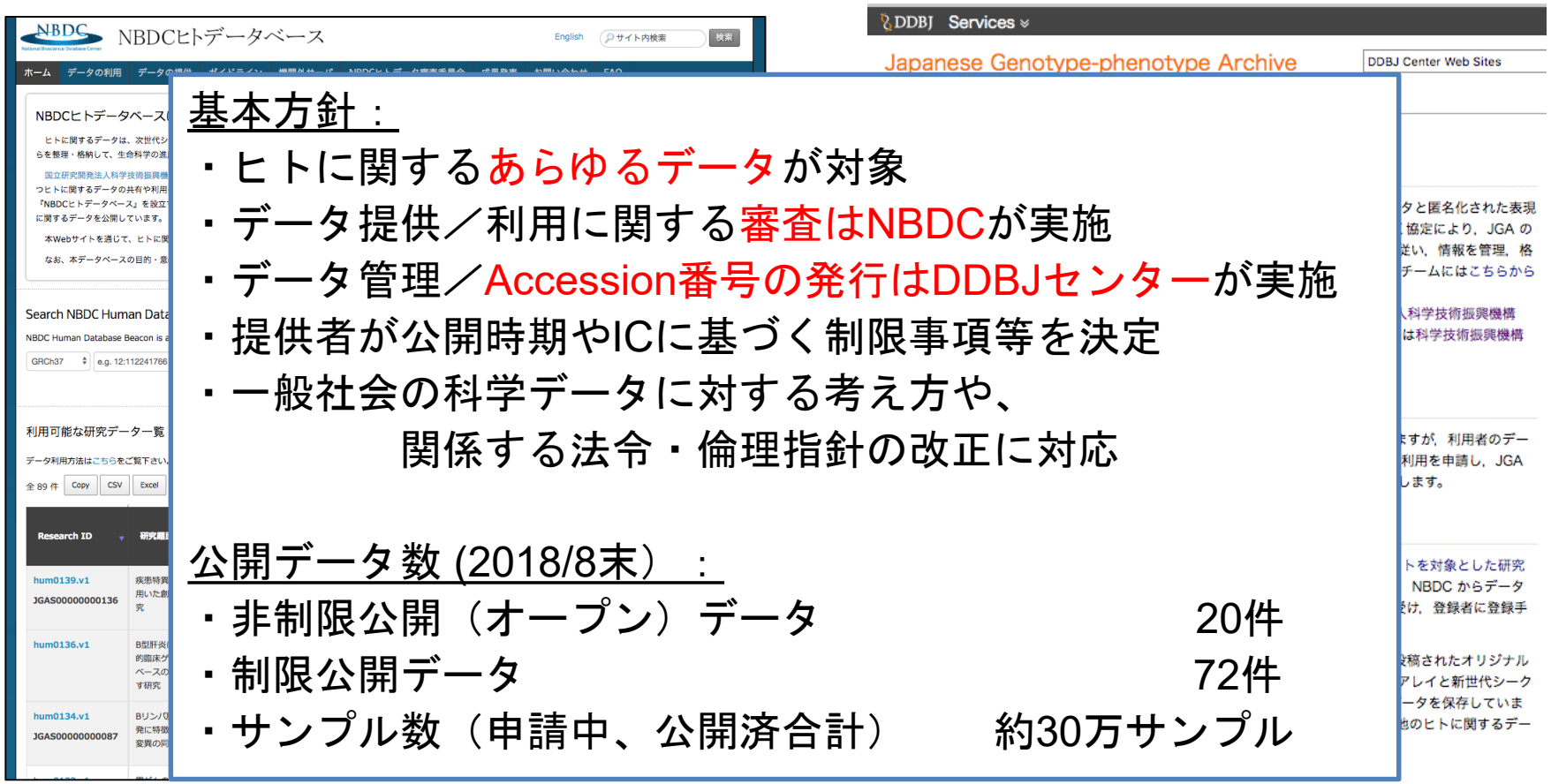
© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

1.TogoVar開発の経緯

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

NBDCヒトデータベース（2013年10月～）

・倫理的な配慮が必要なヒトに関するデータを収集・公開する枠組み
・2段階のアクセスレベルを設定（制限公開データと非制限公開）
・国立遺伝学研究所DDBJセンターが運営する
　　Japanese Genotype-phenotype Archive （JGA）と共同運営



基本方針：

・ヒトに関するあらゆるデータが対象
・データ提供／利用に関する審査はNBDCが実施
・データ管理／Accession番号の発行はDDBJセンターが実施
・提供者が公開時期やICに基づく制限事項等を決定
・一般社会の科学データに対する考え方や、
　　関係する法令・倫理指針の改正に対応

公開データ数 (2018/8末）：

・非制限公開（オープン）データ	20件
・制限公開データ	72件
・サンプル数（申請中、公開済合計）	約30万サンプル

© 2018 統合データベース講習会 Licensed Under CC 表示 4.0 国際

NBDCヒトデータベースを使いやすく

NBDCヒトデータベース

DDBJ Japanese Genotype-phenotype Archive

個人別の情報（個人情報）

制限公開（利用審査あり）

TOGOVAR

個人別の情報でない（集計情報）

非制限公開（利用審査なし）

データ提供者

出典元 (© 2016 DBCLS TogoTV)

研究プロジェクトA

NGSデータ

SNP-Chipデータ

研究プロジェクトB

NGSデータ

研究プロジェクトC

SNP-Chipデータ

同じ手法で再解析

NGSデータ由来頻度情報 (JGA-NGS)

SNP-Chipデータ由来頻度情報 (JGA-SNP)

日本人大規模バリアント頻度

① 概要を把握

データ利用者

出典元 (© 2016 DBCLS TogoTV)

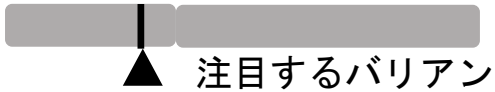
②利用申請

© 2018 統合データベース講習会 Licensed Under CC 表示 4.0 国際

バリアント情報のワンストップ検索

バリアントの解釈に必要な情報を収集・整理する作業から研究者を解放

7番染色体



ClinVar (NCBI)

位置：chr7:127254587
関連する疾患：２型糖尿病
疾患感受性：　あり

ExAC (ブロード研究所)

位置：chr7:127254587
アレル頻度：0.000304573

TogoVarID: tgv30913364

位置： chr7:127254587
関連する疾患：2型糖尿病
疾患感受性：あり
アレル頻度(iJGVD 3.5KJPN)：0.0233
アレル頻度(HGVD)：0.0272809
アレル頻度(ExAC)：0.000304573
関連論文：A missense mutation of Pax4 gene ...

https://togovar.biosciencedbc.jp/variant/tgv30913364

研究者・遺伝カウンセラー

iJGVD 3.5KJPN
(東北メディカル・メガバンク機構)

位置： chr7:127254587
アレル頻度： 0.0233

HGVD(京都大学)

位置： chr7:127254587
アレル頻度：0.0272809

ワンストップ検索

出典元 (© 2016 DBCLS TogoTV)

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

TogoVar／検索の対象データベース

データベース名 (運営組織)	説明	対象人数
NBDCヒトデータベース	主に日本の公的資金による研究にて生産された個人別データのリポジトリ	125人（全エクソーム） 183,884人（既知SNP）
Integrative Japanese Genome Variation Database (iJGVD)（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）
Human Genetic Variation Database (HGVD)（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）	配列決定プロジェクトの再解析データ（約２０プロジェクト）	60,706人（全エクソーム）
ClinVar（NCBI）	バリアントの疾患関連性	
PubTator (NCBI) Colil(DBCLS)	バリアント（rs番号）が出現する文献情報	

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

TogoVar検索例

検索ボックス
- rs番号
- 位置検索
- 領域検索
- 遺伝子名（あいまい検索）
- 関連疾患名（あいまい検索）

TOGOVAR Home Datasets Downloads Terms Contact About Help

Data set
- All
- JGA NGS (119)
- JGA SNP (55)
- 3.5KJPN (173)
- HGVD (50)
- ExAC (926)
- ClinVar (4,545)

Alternative allele frequency

JGA NGS

Add filter

Variant type
- All
- SNV (2,787)
- Insertion (482)
- Deletion (1,203)
- Indel (65)
- Substitution (8)

Clinical significance
- All
- Not in ClinVar (0)
- Pathogenic (2,242)

フィルタ機能
- データセット
- データセット＋頻度
- バリアントタイプ
- ClinVarの情報

Breast-ovarian cancer, familial 2 Search Clear

Example - Disease:Breast-ovarian cancer, familial 2, Gene:ALDH2, refSNP:rs114202595, TogoVar:tgv421843, Position(GRCh37/hg19):16:48258198, Region(GRCh37/hg19):10:73270743-73376976

検索結果

Display 25 variants per page

Showing 1 to 25 of 4,545 variants **related to disease "Breast-ovarian cancer, familial 2".** (filtered from 19,537,677 variants)

1 2 3 4 5

	Position	Ref	Alt	Symbol	Frequency	Consequence	SIFT	Polyphen	ClinVar conditions
48208253 988901	13:32889080	A	G	ZAR1L		intron_variant			Breast-ovarian cancer
tgv48208258 rs3092989	13:32889363	G	A	ZAR1L		intron_variant			Breast-ovarian cancer
tgv48208267 rs55880202	13								Fanconi anemia Hereditary breast and Breast-ovarian cancer
tgv48208269 rs567110692	13								Breast-ovarian cancer
tgv48208271 rs545321666	13								Fanconi anemia Hereditary breast and Breast-ovarian cancer

	JGA-NGS	JGA-SNP	3.5KJPN	HGVD	ExAC
Alt	7	-	565	-	-
Total	250	-	7082	-	-
Freq	0.028	-	0.08	-	-

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

公開前後に頂いた評価

- ExACやToMMoの頻度データがワンストップで閲覧でき、便利
- 同じバリアントの表記や位置が異なる時に、togovar IDのようなユニークなIDがあると便利
- ExAC/gnomADのように日本人集団でJoint callしたバリアント頻度情報は有用
- 日本人集団頻度情報は、希少疾患研究において日本人特有のバリアントの除去に使っている

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

公開前後に頂いた要望（一部）

- 参照データベースの追加
  - MGeND、gnomAD、GWAS catalog、HGMD、1000Genomes 、dbSNPなど
- 文献情報の充実
- システムの改良
  - HGVS表記（例：11:g.112241766G>A）での検索
  - リファレンスの他バージョンへの対応
  - 表示の工夫、フィルター追加
  - 日本人とその他の集団で頻度に差があるバリアントを明示
  - 日本人特有の配列への対応
  - データダウンロードやAPIの整備

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

２．TogoVar開発の現場

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

DBCLS基盤技術の利活用

RDF化済みデータベース

NBDC RDF Portal

med2rdf Project

ExACやClinVar

TOGOGENOME

Colil

引用情報

今回RDF化したデータベース

Japanese Genome Variation Database

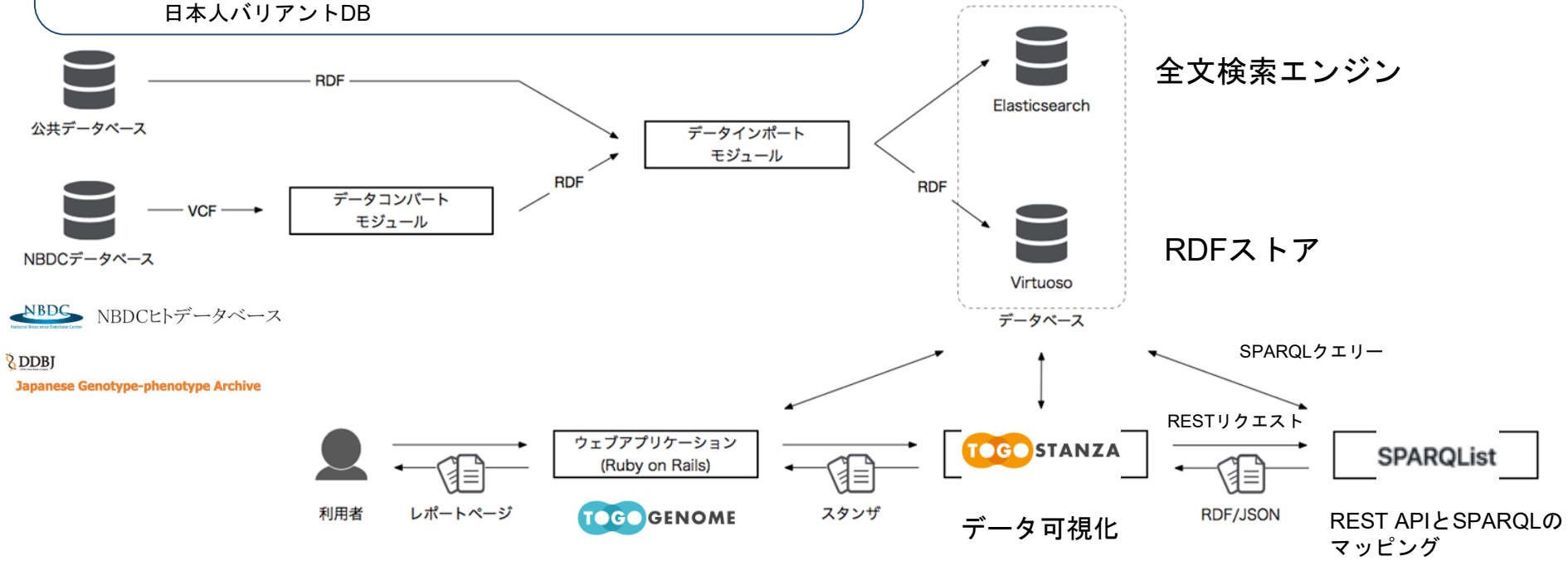
Human Genetic Variation Database

日本人バリアントDB

PubTator

バリアントと文献

バリアントデータの加工など
TogoVar特有の作業に集中できた



© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

TogoVar開発メンバー

NBDC

川嶋　実苗
人類遺伝学、NBDCヒトDB

建石　由佳
文献関連技術

豊岡　理人
人類遺伝学、変異データ統合

三橋　信孝
取りまとめ

宮崎　和典
NGSデータ再解析

DBCLS

片山　俊明
RDF技術、システム構築

川島　秀一
RDF技術、システム構築

藤原　豊史
文献関連技術

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

TogoVar普及に向けて各種支援

1. NBDC広報
   - 公開時プレスリリース
   - 学会シンポジウム開催（日本人類遺伝学会、日本生化学学会大会）
   - トーゴーの日（講演、ワークショップ）
   - AJACS講習会や学会出展等でのユーザーの獲得や意見の収集

2．AMEDバイオバンク課
   - 臨床ゲノム情報統合データベース整備事業の先生方へのデモ

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

3. まとめ

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

2017年トーゴーの日のおさらい

NBDCの今後の取組み：「作る」から「使う」へ

大型プロジェクト研究と連携し、未公開データのプロジェクト内共有を支援

応用につながる領域に焦点をあて、基礎研究データの統合を行う

→医科学領域に焦点を当て、 TogoVarの開発と公開

様々な利用者との協業により、統合データの利活用に取り組む

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際

誰がためにTogoVarはある...

1. エンドユーザーが利用し易いサービスの提供
   - NBDCヒトDBに登録されているデータの概要をわかり易く示すことで、利用したいデータにたどり着き易くする
   - 散在する様々な集団におけるバリアント情報のワンストップ検索を可能にする

2. 基盤技術開発の促進
   - TogoVarを通じ、RDFを使った基盤技術の有用性を示すと共に、より良い技術にするための改善点をフィードバックする

3. データの受け皿として
   - 日本人ゲノムの理解を促進するため、NBDCヒトDBと共に、国内ゲノムプロジェクトからのデータを効率よく共有するための受け皿となる

© 2018 TogoVar開発チーム Licensed Under CC 表示 4.0 国際