TogoVarの更新：疾患および国内地域別に集計されたアレル頻度およびジェノタイプ数の公開

○三橋信孝 (DBCLS)、川嶋実苗 (DBCLS)、守屋勇樹 (DBCLS)、豊岡理人 (富山国際大)、高木利久 (富山国際大)、川島秀一 (DBCLS)、片山俊明 (DBCLS)

過去1年間の新規公開データ

1. **日本人および他の集団間で比較可能なアレル頻度およびジェノタイプ数**
   - バイオバンクジャパン（BBJ）のSNPアレイの18万人の内訳（45疾患別、男女別）
   - ナショナルセンターバイオバンクネットワーク（NCBN）コントロール群の全ゲノム9千人（本土および琉球クラスター）とジョイントコールされた1000 Genomes (1KGP) の2.5千人
2. **バリアントの解釈のための注釈情報**
   - AlphaMissense（ミスセンスバリアントの病原性予測）

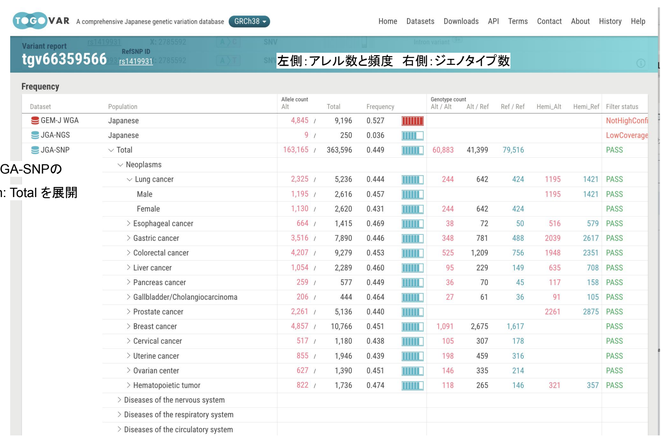
日本人および他の集団間で比較可能なアレル頻度およびジェノタイプ数

データセット	サンプル数	解析手法	参照ゲノム	対象集団など
JGA-NGS	125	WES	GRCh37 GRCh38*	NBDC ヒトデータベース / JGAに登録されたNGSを再解析
JGA-SNP	183,884	SNP chip	GRCh37 GRCh38*	NBDC ヒトデータベース / JGAに登録されたSNPアレイデータ **（BBJ 18万人の 45疾患別、男女別のアレル・ジェノタイプ数）**
GEM-J WGA	7,609	WGS	GRCh37 GRCh38*	東北大学東北メディカル・メガバンク機構（ToMMo）、岩手医科大学いわて東北メディカル・メガバンク機構（IMM）、理化学研究所および東京大学医科学研究所による共同研究
HGVD	1,208	WES	GRCh37	住民ゲノムコホート（ながはま0次コホート）
NCBN	9,290 +2,504	WGS	GRCh38	**がんや希少疾患の日本人コントロール群 (9290人)と1KGP(2504人のジョイントコール （ナショナルセンターバイオバンクネットワーク）**
ToMMo 8.3KJPN	8,380	WGS	GRCh37	住民ゲノムコホート（東北メディカルメガバンク）
ToMMo 54KJPN	54,302		GRCh38	
gnomAD v2.1.1	125,748	WGS WES	GRCh37	日本人以外の8つの遺伝的祖先集団
gnomAD v4.0	76,156	WGS	GRCh38	日本人以外の10の遺伝的祖先集団

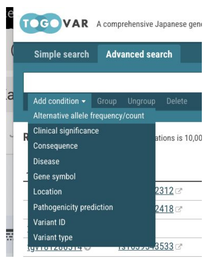
GRCh38* : GRCh37座標のliftover (transanno利用)によりGRCh38座標を計算

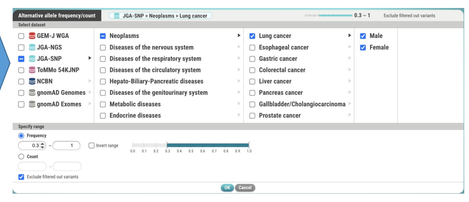
BBJ 45疾患別のアレル・ジェノタイプ数

バリアントレポートページ（例：https://grch38.togovar.org/variant/tgv66359566#frequency）に表示



Dataset: JGA-SNPの
Population: Total を展開

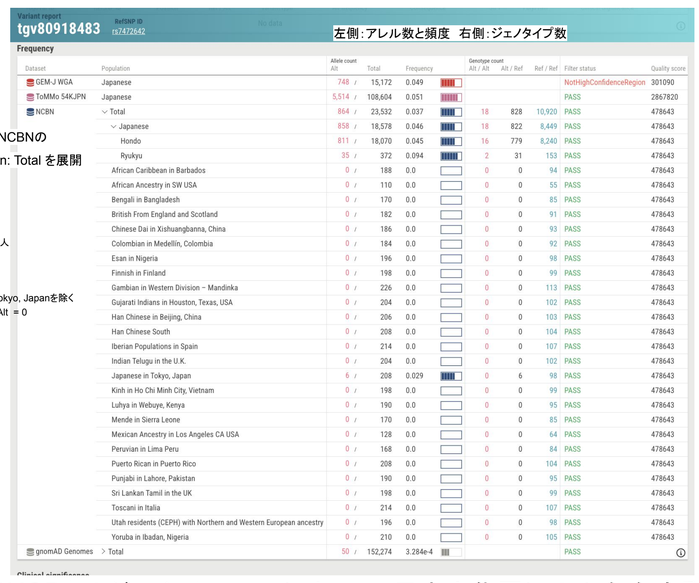




Advanced search の Alternative allele frequency/count から疾患別、男女別のアレル頻度やアレル数を条件に検索可能

NCBN日本人9千人と1KGP 2.5千人のジョイントコール

バリアントレポートページ（例：https://grch38.togovar.org/variant/tgv80918483#frequency）に表示



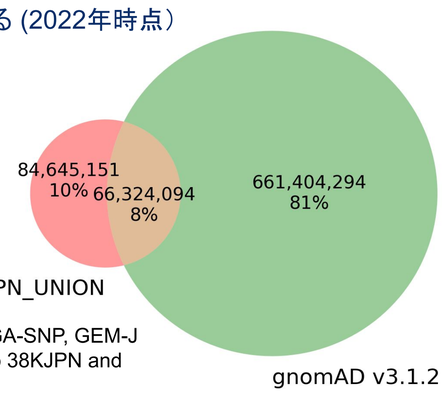
Dataset: NCBNの
Population: Total を展開

上：NCBN日本人（本土、琉球）
下：1KGP Japanese in Tokyo, Japanを除く全ての集団で Alt = 0

1KGPとのジョイントコールによって、日本人集団にのみ存在するバリアントであることがわかる

データセット間の包含関係

gnomAD v3にない8400万以上のバリアントが日本人集団で発見されている (2022年時点)



JPN_UNION:
JGA-NGS, JGA-SNP, GEM-J WGA, ToMMo 38KJPN and HGVD

ノーコードでRDFデータ作成（RDF-config利用）

手順1. 以下のファイルを作成
- model.yaml：RDFの定義
- convert.yml：変換したいデータととRDFのマッピング

手順2. rdf-config convertコマンドを実行してRDF(Turtle, JSON-LD、JSON Lines)を出力

参考サイト(RDF-config GitHub)
https://github.com/dbcls/rdf-config/blob/convert/doc/spec_ja.md#convertyaml

利用例：2バリアント間の連鎖不平衡係数データを TSVからJSON-LDに変換

```
dbcls3284:convert_togovar mitsuhashi$ cat
PASS.autosome_PAR_ploidy_2.biallelic.chr22.tsv
SNPA      SNPB      R2        Dprime
chr22:10510166:A:G        chr22:10525190:C:G        0.1       0.5
chr22:10510166:A:G        chr22:10577648:A:G        0.2       0.6
chr22:10577648:A:G        chr22:10611759:C:T        0.3       0.7
chr22:10577648:A:G        chr22:10640034:T:C        0.4       0.8
dbcls3284:convert_togovar mitsuhashi$
```

入力データ （TSV形式）

```
- LDRelation hco:22/GRCh38#10510166-A-G:
    - a: obo:SO_0001483 # SNV
    - tgvo:hasLDRelation+:
        - []:
            - tgvo:involves{1}:
                - snv: hco:22/GRCh38#10525190-C-G
            - tgvo:hasR2Value{1}:
                - r2_value: 0.1
            - tgvo:hasDPrimeValue{1}:
                - dprime_value: 0.5
```

config/model.yaml

```
LDRelation:
  - source("PASS.autosome_PAR_ploidy_2.biallelic.chr22.tsv")
  - csv("SNPA")
  - variables:
    - snv:
      - csv("SNPB")
    - r2_value:
      - csv("R2")
    - dprime_value:
      - csv("Dprime")
```

config/convert.yaml

```
$ bundle exec rdf-config --config config --format jsonld --convert
./PASS.autosome_PAR_ploidy_2.biallelic.chr22.tsv | jq
{
  "@context": {
    "obo": "http://purl.obo.org/obo/obo/cl-extend-ext#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "dct": "http://purl.org/dc/terms#",
    "dcat": "http://www.w3.org/ns/dcat#",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "prov": "http://www.w3.org/ns/prov#",
    "sio": "http://semanticscience.org/resource/",
    "dbpedia": "http://dbpedia.org/resource/Value",
    "dbpr": "http://dbpedia.org/ns/",
    "hgnc": "http://identifiers.org/hgnc/",
    "ncbigene": "http://identifiers.org/ncbigene/",
    "ncbitaxon": "http://purl.bioontology.org/ncbitaxon/",
    "dbsnp": "http://identifiers.org/dbsnp/",
    "ncbi": "http://identifiers.org/ncbi/",
    "ncbi": "http://ncbi.nlm.nih.gov/resources/",
    "ncbi": "http://ncbi.nlm.nih.gov/resources/variation/",
    "ncbi": "http://ncbi.nlm.nih.gov/resources/",
    "ncbi": "http://ncbi.nlm.nih.gov/resources/variation/",
    "hco": "http://purl.jp/bio/4/hco/",
    "tgvo": "http://purl.jp/bio/4/tgvo/",
    "LDRelation": "@id",
    "data": "@graph",
    "snv": {
      "@id": "tgvo:involves",
      "@type": "@id"
    },
    "r2_value": {
      "@id": "tgvo:hasR2Value"
    },
    "dprime_value": {
      "@id": "tgvo:hasDPrimeValue"
    }
  },
  "data": [
    {
      "LDRelation": "chr22:10510166:A:G",
      "@type": "obo:SO_0001483",
      "snv": [
        "chr22:10525190:C:G",
        "chr22:10577648:A:G"
      ],
      "r2_value": [
        "0.1",
        "0.2"
      ],
      "dprime_value": [
        "0.5",
        "0.6"
      ]
    },
    {
      "LDRelation": "chr22:10577648:A:G",
      "@type": "obo:SO_0001483",
      "snv": [
        "chr22:10611759:C:T",
        "chr22:10640034:T:C"
      ],
      "r2_value": [
        "0.3",
        "0.4"
      ],
      "dprime_value": [
        "0.7",
        "0.8"
      ]
    }
  ]
}
```

出力RDF （JSON-LD）

バリアントの解釈のための注釈情報

- 遺伝子名（例：ALDH2)
- dbSNP rsID（例：rs671)
- HGVS表記（例：ALDH2:p.Glu504Lys)
- 臨床的意義
  - キュレーション（ClinVar)
    - 疾患名（例：Acute Myeloid Leukemia)
  - 予測プログラム（AlphaMissense、PolyPhen、SIFT）
- アレル頻度と表現型（疾患）との相関解析 (GWAS Catalog)

まもなく公開予定

1. MGeND の臨床的意義
   - 疾患名をMedGenの語彙にマッピング
2. 文献情報へのスニペット追加
   - 文献中のバリアント名の出現位置をPubTator3から取得しPubAnnotationに格納
3. (MoG+) マウスバリアント
   - オーソログ遺伝子の配列アライメントによるヒトとマウスバリアントの対応づけに基づく統合
4. 利用者が指定したバリアントと連鎖不平衡にあるバリアントの検索
   - BBJ WGS 1000人のジェノタイプからr^2とD'を計算
   - ゲノムビューアで連鎖不平衡ブロックを可視化
5. データ更新
   - ToMMo 54KJPNから60KJPN
   - gnomAD v4.0からv4.1

今後の取り組み（取り込み中を含む）

1. 短いバリアント(<50bp)データの追加
   - BBJ WGS 2000人 (JGAに寄託されたNGSデータをGATKで再解析)
   - NCBN 9000人の内訳（都道府県別アレル、ジェノタイプ数）
   - TogoImputation でインピュテーションしたJGA-SNP
2. 構造多型（SV、>50bp）データの統合と公開
   - 日本人集団
     - TogoVar-repositoryに登録されたSV
     - JGA寄託されたNGSデータ再解析 (GATK-SVで解析)
   - 日本人集団以外
     - gnomAD SV 4.1
3. 注釈情報の追加
   - タンパク質立体構造上でのバリアントの可視化
   - スプライシングとバリアント（SpliceAI)
   - 遺伝子発現とバリアントの関係（eQTL from GTEx）

Licensed under CC-BY 4.0 ©2023 Nobutaka Mitsuhashi (DBCLS)