

多型データベースの紹介 4: GWAS-catalog (Genome Wide Association Study)

GWASの結果を集めた多型データベース GWAS-catalog

- 運営元：米国国立ヒトゲノム研究所(National Human Genome Research Institute) および 欧州バイオインフォマティクス研究所(EBI)
- URL: <https://www.ebi.ac.uk/gwas/>
- 今までGWASによって同定された形質(疾患、形質、薬剤副反応等)と関連が報告されたSNPをマニュアルキュレーション

多型データベースの紹介: GWAS-catalog



GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Search the catalog Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000

feedback

疾患名、rs番号、著者名、遺伝子名、領域で検索可能

Take a tour of the NEW search interface...

Search

Search the Catalog in a number of ways, including by trait, SNP identifier, publication, gene and genomic location.

Diagram

Explore an interactive visualisation of all SNP-trait associations with genome-wide significance ($p \leq 5 \times 10^{-8}$).

Download

Download a full copy of the GWAS Catalog in spreadsheet format as well as current and older versions of the GWAS diagram in SVG format.

多型データベースの紹介: GWAS-catalog

Show SNPs for	
Digestive system disease	333
Cardiovascular disease	1042
Metabolic disease	403
Immune system disease	1879
Nervous system disease	1269
Liver enzyme measurement	159
Lipid or lipoprotein measurement	993
Inflammatory marker measurement	448
Hematological measurement	4955
Body weights and measures	1726
Cardiovascular measurement	955
Other measurement	11855
Response to drug	325
Biological process	2155
Cancer	1415
Other disease	1633
Other trait	1281



多型データベースの紹介: GWAS-catalog

GWAS Catalog

Home Diagram Download Documentation About EMBL-EBI NIH National Human Genome Research Institute

 GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

breast carcinoma

Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000

Search results for *breast carcinoma*

T *breast carcinoma* [EFO_0000305](#)

A carcinoma arising from the breast, most commonly the terminal ductal-lobular unit. It is the most common malignant tumor in females. Risk factors include country of birth, family history, menstrua... [Show more >](#)

Associations 1640 Studies 112

T *HER2 positive breast carcinoma* [EFO_1000294](#)

A biologic subset of breast carcinoma defined by high expression of HER2, GRB7, and TRAP100, and by lack of expression of estrogen receptor (ER).

Associations 0 Studies 1

多型データベースの紹介: GWAS-catalog

Associations 1409

ある形質で行われた関連解析の結果

Variant and risk allele	P-value	P-value annotation	RAF	OR	Beta	CI	Mapped gene	Reported trait	Trait(s)	Study accession	Location
rs2111836-G	7×10^{-6}		0.22	1.32	-	[1.17-1.44]	MTCL1	Breast cancer	breast carcinoma	GCST002735	18:8716323
rs13025833-A	1×10^{-7}		0.52	1.2	-	[1.13-1.33]	CTNNA2	Breast cancer	breast carcinoma	GCST002735	2:79716982
rs1078806-C	2×10^{-6}	(Pooled P value)	0.39	1.43	-	[NR]	FGFR2	Breast cancer	breast carcinoma	GCST001831	10:121579461
rs166870-T	3×10^{-7}	(HR+ HER2-)	0.13	2.3	-	[1.67-3.15]	AC026826.2, RNU6-667P	Disease-free survival in breast cancer	breast carcinoma, disease free survival	GCST002847	15:79777420
rs10825036-G	4×10^{-7}	(HR- HER2-)	0.32	2.26	-	[1.34-3.81]	RNA5SP318, AL365496.1	Disease-free survival in breast cancer	disease free survival, breast carcinoma	GCST002847	10:53506471

Showing 1 to 5 of 1409 rows 5 ▲ rows per page 1 2 3 4 5 ... 282 >

Studies 87

関連解析の研究概要

First author	Study accession	Publication date	Journal	Title	Reported trait	Trait(s)	Discovery sample number and ancestry	Replication sample number and ancestry	Association count	Summary statistics
Haryono SJ	GCST002735	2015-01-01	Asian Pac J Cancer Prev	A pilot genome-wide association study of breast cancer susceptibility loci in Indonesian.	Breast cancer	breast carcinoma	• 135 Asian unspecified	-	2	NA

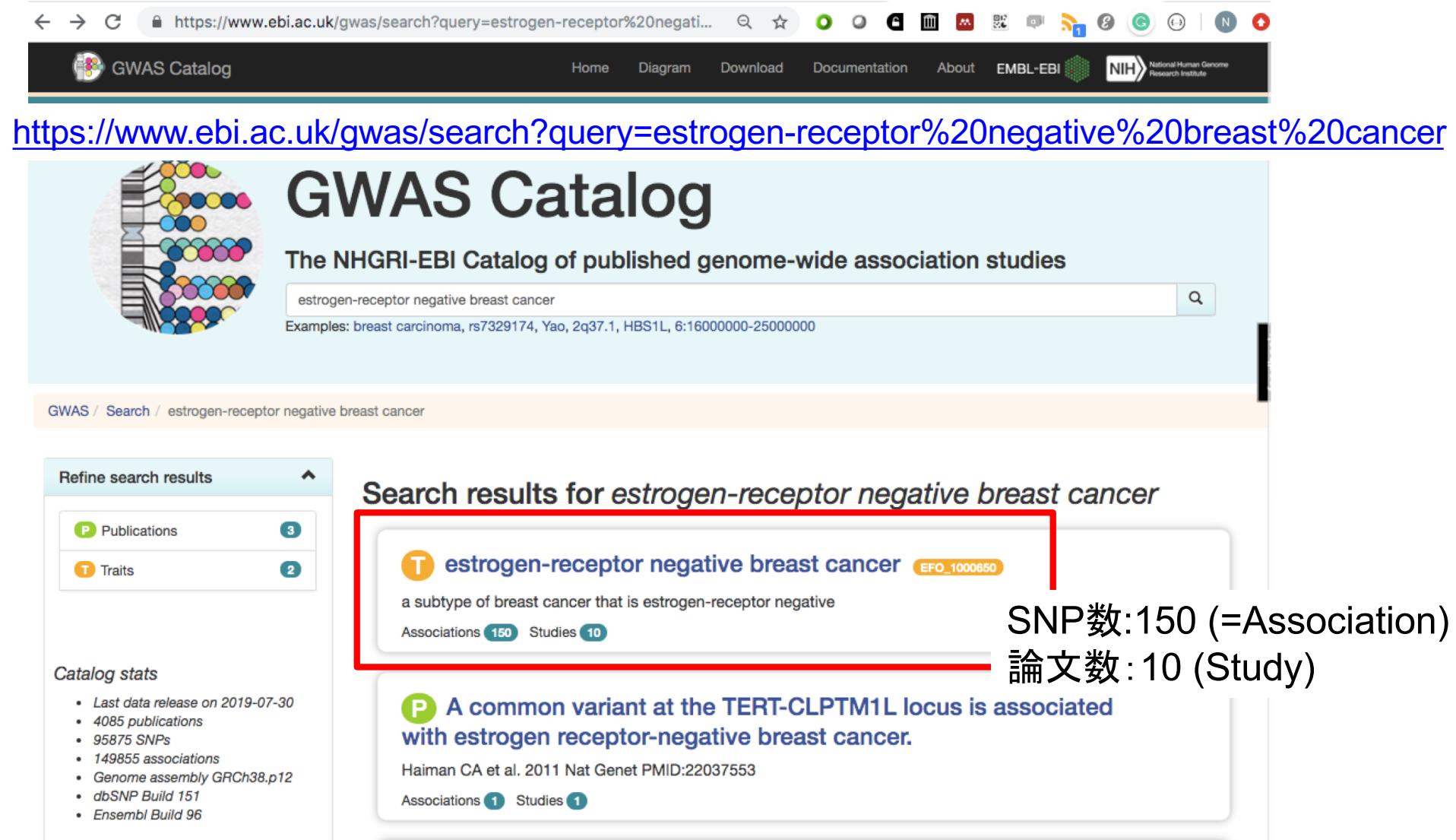
[ハンズオン3] GWAS-catalog

[basic]

- 疾患「estrogen-receptor negative breast cancer」と関連が報告されたSNPと論文数を調べましょう
- 遺伝子「CYP2C19」と関連が報告された形質およびその集団を確認しましょう

疾患「estrogen-receptor negative breast cancer」と関連が報告されたSNPと論文数を調べましょう

<https://www.ebi.ac.uk/gwas/search?query=estrogen-receptor%20negative%20breast%20cancer>



The screenshot shows the GWAS Catalog search interface. The search term "estrogen-receptor negative breast cancer" is entered in the search bar. The results page displays a summary of findings for this specific cancer subtype. A red box highlights the first result, which is a trait entry for "estrogen-receptor negative breast cancer". The entry includes the identifier EFO_1000650, a definition ("a subtype of breast cancer that is estrogen-receptor negative"), and counts for 150 associations and 10 studies. Below this, another result is shown: a publication abstract by Haiman CA et al. (2011) describing an association at the TERT-CLPTM1L locus.

GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

estrogen-receptor negative breast cancer

Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000

GWAS / Search / estrogen-receptor negative breast cancer

Refine search results

- Publications (3)
- Traits (2)

Catalog stats

- Last data release on 2019-07-30
- 4085 publications
- 95875 SNPs
- 149855 associations
- Genome assembly GRCh38.p12
- dbSNP Build 151
- Ensembl Build 96

Search results for estrogen-receptor negative breast cancer

T estrogen-receptor negative breast cancer EFO_1000650

a subtype of breast cancer that is estrogen-receptor negative

Associations 150 Studies 10

P A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer.

Haiman CA et al. 2011 Nat Genet PMID:22037553

Associations 1 Studies 1

SNP数:150 (=Association)
論文数: 10 (Study)

遺伝子「CYP2C19」と関連が報告された形質およびその集団を確認しましょう

Available data: [Associations 7](#) [Studies 7](#) [Traits 7](#) [Download Catalog data](#)

Variant and risk allele	P-value	P-value annotation	RAF	OR	Beta	CI	Mapped gene	Reported trait	Trait(s)	Study accession
rs4494250-A	3×10^{-7}		0.319	-	0.15 unit increase	[0.091-0.209]	CYP2C19, AL583836.1	Diastolic blood pressure	diastolic blood pressure	GCST006227
rs4494250-A	1×10^{-6}		0.318915546764454	-	0.24689345	[0.15-0.35]	CYP2C19, AL583836.1	Mean arterial pressure	mean arterial pressure	GCST006231
rs79						[NR]	CYP2C19, AL583836.1	Clopidogrel active metabolite levels	clopidogrel metabolite measurement	GCST004264
rs12						-	CYP2C19, AL583836.1	Acenocoumarol maintenance dosage	response to anticoagulant	GCST000436
rs199562446-T	2×10^{-9}		0.1252	-	0.3412 mmHg decrease	[0.23-0.45]	CYP2C19, AL583836.1	Systolic blood pressure	systolic blood pressure	GCST007267

Showing 1 to 5 of 7 rows [5](#) rows per page <https://www.ebi.ac.uk/gwas/genes/CYP2C19>

Study: GCST006227

GWAS / Studies / GCST006227

Study information

Reported trait	Diastolic blood pressure	Trait(s)	diastolic blood pressure	View in PubMed
Genotyping technology	Exome genotyping array	Platform [SNPs passing QC]	Illumina [247039]	View in Europe PMC
Discovery sample description	120,473 European ancestry individuals, 21,503 African American individuals, 4,586 Hispanic individuals	Replication sample description	154,543 European ancestry individuals, 26,183 South Asian ancestry individuals	
Discovery ancestry (country of recruitment)	120473 European (U.S., Germany, Netherlands, Iceland), 21503 African American or Afro-Caribbean (U.S.), 4586 Hispanic or Latin American (U.S.)	Replication ancestry (country of recruitment)	26183 South Asian (Bangladesh, Pakistan), 154543 European (France, Germany, Netherlands, Denmark, Finland, Republic of Ireland, Norway, Sweden, U.K., Greece, Italy, Spain, NR)	
PubMed ID	27618448	Publication date	2016-10-01	
First author	Liu C			
Full Summary Statistics	FTP Download or API access			
Title	Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci.			
Authors	Liu C, Kraja AT, Smith JA  , Brody JA, Franceschini N, Bis JC, Rice K, Morrison AC, Lu Y, Weiss S... Show more >			

<https://www.ebi.ac.uk/gwas/studies/GCST006227>

多型データベースの紹介 5: TogoVar

日本人ゲノム多様性統合データベース: TogoVar

- 運営元: NBDC, JST
- URL:
<https://togovar.biosciencedbc.jp/>
- 「NBDCヒトデータベース」を基に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
- 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
- 2018年6月7日公開
- 現状germline variantのみ、somaticなし

『NBDCヒトDB』の概要を把握



NBDCヒトデータベース



Japanese Genotype-phenotype Archive

個人別の情報（個人情報）

制限公開（提供・利用審査あり）



個人別の情報でない（集計情報）

自由に検索、閲覧できる

データ
提供者



研究プロジェクトA

NGSデータ
SNP-Chipデータ

同じ手法で
再解析

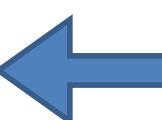
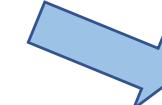


研究プロジェクトB

NGSデータ

研究プロジェクトC

SNP-Chipデータ



データ
利用者



① 概要を
把握



② 利用申請

出典元 (© 2016 DBCLS TogoTV)

ワンストップ検索

ゲノム配列の個人による違い（バリアント）に関する
さまざまな条件を用いて、国内外のDBや文献情報などの
ワンストップ検索を可能に

7番染色体



注目するバリアント

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



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

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) : 0.0003045

関連論文 : 73

A missense mutation of Pax4 gene ...

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



ワンストップ検索



出典元 (© 2016 DBCLS TogoTV)

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 (JST-NBDCと国立遺伝学研究所DDBJセンターの共同運営) ・ JGA-NGSデータセット(全エクソーム) ・ JGA-SNPデータセット(SNP Chip)	主に日本の研究者からの個人ゲノムデータのリポジトリ	125人 (全エクソーム) 183,884人 (既知SNP)
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)
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	

JGA-NGS/JGA-SNPデータの由来

JGA-NGS https://togovar.biosciencedbc.jp/doc/datasets/jga_ngs

集約されたデータの由来

JGAID	ヒトDB	研究題目	対象集団	サンプルサイズ	データ提供者
JGAD000000000004	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 アストロサイトーマ	6	齊藤 延人 (P-DIRECT)
JGAD000000000106	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 オリゴデンドログリオーマ	16	齊藤 延人 (P-DIRECT)
JGAD000000000112	hum0006	脳腫瘍のゲノム・遺伝子解析とその臨床病理学的意義の解明	脳腫瘍 小脳グリオーマ	17	齊藤 延人 (P-DIRECT)
JGAD000000000014	hum0021	精神神経疾患の原因解明および診断法・治療法の開発に関する研究	健常一卵性双生児	6	加藤 忠史 (ゲノム支援)
JGAD000000000036	hum0035	固形腫瘍における遺伝子異常の網羅的解析	多数の固形腫瘍性疾患検体	23	滝田 順子 (P-DIRECT)
JGAD000000000038	hum0040	ヒト胎盤におけるインプリント制御領域およびインプリント遺伝子の同定	ヒト胎盤および母体血	48	有馬 隆博 (IHEC)
JGAD000000000060	hum0066	癌の再発・転移に関するnon-coding RNAの同定とその機序の解明	大腸がん	9	三森 功士
合計					125

JGA-SNP https://togovar.biosciencedbc.jp/doc/datasets/jga_snp

集約されたデータの由来

JGAID	ヒトDB	研究題目	対象集団	サンプルサイズ	データ提供者
JGAD000000000123	hum0014	オーダーメイド医療の実現プログラム	健常者及び罹患者	182,557	久保 充明 (BBJ)
JGAD000000000018	hum0028	オーダーメイド医療の実現プログラム	健常者	908	久保 充明 (BBJ)
JGAD000000000130/131	hum0082	日本人健常者におけるゲノム全域のSNP解析	健常者	419	徳永 勝士
Total				183,884	

TogoVar 検索結果例（一覧検索画面）

Search for disease or gene symbol or rs...

TogoVar ID	RefSNP ID	Frequency	Consequence	SIFT	PolyPhen	Clinical significance
tgv67071948	rs7792		Intergenic variant			
tgv67071949	rs13563		Intergenic variant			
tgv67071950	rs14094		Intergenic variant			
tgv67071951	rs201732991	1: 10.77	SNV			
tgv67071952	rs775928745	1: 10231	Deletion			Intergenic variant
tgv67071953	rs1267031179	1: 10248	Deletion			Intergenic variant
tgv67071954	rs145427775	1: 10291	SNV			Intergenic variant
tgv67071955	rs060413313	1: 10297	SNV			Intergenic variant
tgv67071956	rs112750067	1: 10321	SNV			Intergenic variant
tgv67071957	rs1035171912	1: 10327	Deletion			Intergenic variant
tgv67071958	rs150964722	1: 10330	SNV			Intergenic variant
tgv67071959	rs1357792609	1: 10334	Deletion			Intergenic variant
tgv67071960	rs1351390918	1: 10334	SNV			Intergenic variant
tgv67071961	rs1193008993	1: 10336	SNV			Intergenic variant
tgv67071962	rs1471210572	1: 10340	Deletion			Intergenic variant
tgv67071963	rs1363828207	1: 10343	SNV			Intergenic variant
tgv67071964	rs1363828207	1: 10346	Deletion			Intergenic variant
tgv67071965	rs1015856060	1: 10348	SNV			Intergenic variant
tgv67071966	rs1015856060	1: 10349	Deletion			Intergenic variant
tgv67071967	rs1015856060	1: 10352	SNV			Intergenic variant
tgv67071968	rs1015856060	1: 10354	Deletion			Intergenic variant
tgv67071969	rs1015856060	1: 10384	SNV			Intergenic variant

検索ボックス

- rs番号
- 位置検索、範囲検索 (hg19)
- 遺伝子名 (あいまい検索)
- 関連疾患名 (あいまい検索)

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
SNV or INDEL...
- 臨床的意義
Pathogenic, Benign...
- missense変異有害性予測スコア
PolyPhen-2, SIFT

Filters

Dataset

Dataset	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> JGA NGS	4,679,025
<input checked="" type="checkbox"/> JGA SNP	1,249,724
<input checked="" type="checkbox"/> 3.5KJPN	64,675,495
<input checked="" type="checkbox"/> HGVD	554,461
<input checked="" type="checkbox"/> ExAC	10,195,872
<input checked="" type="checkbox"/> ClinVar	443,512

Alternative allele frequency

0 ~ 1 Invert range

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

For all datasets For any dataset

Variant calling quality

Exclude filtered out variants in all datasets

Variant type

Variant type	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> SNV	63,066,161
<input checked="" type="checkbox"/> Insertion	5,383,629
<input checked="" type="checkbox"/> Deletion	6,240,052
<input checked="" type="checkbox"/> Indel	7,358
<input checked="" type="checkbox"/> Substitution	1,740

Clinical significance

Clinical significance	Count
<input checked="" type="checkbox"/> All	74,698,940
<input checked="" type="checkbox"/> Not in ClinVar	74,255,428
<input checked="" type="checkbox"/> Pathogenic	112,579
<input checked="" type="checkbox"/> Likely pathogenic	59,343
<input checked="" type="checkbox"/> Uncertain significance	353,079
<input checked="" type="checkbox"/> Likely benign	175,644
<input checked="" type="checkbox"/> Benign	94,928
<input checked="" type="checkbox"/> Conflicting interpretations of pathogenicity	6,047

Variant report
tgv47264307 dbSNP
rs671

• 位置、ジェノタイプ、SNPタイプ

Variant type SNV Ref / Alt G>A
Position 12:112241766 (GRCh37) hgvs 12:g.112241766G>A

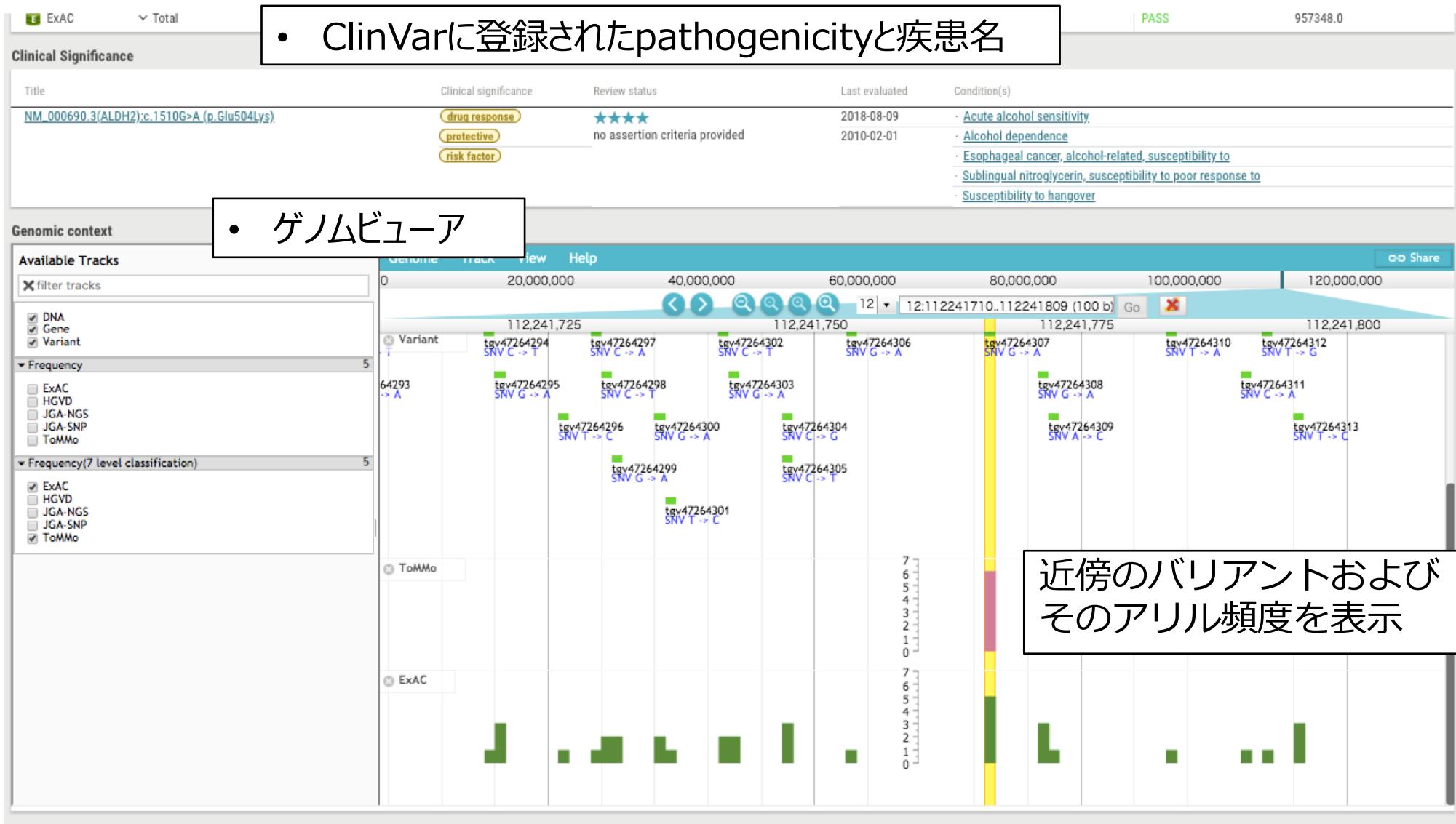
Other overlapping variant(s)
No other variants on the same location

• 同じ領域に存在する他のバリアント

Frequency

Dataset	Population	Allele count			Genotype count			Filter status	Quality score
		Alt	Total	Frequency	Alt / Alt	Alt / Ref	Ref / Ref		
JGA-NGS	Japanese	52 /	250	0.208				PASS	451.77
JGA-SNP	Japanese	90,026 /	365,930	0.246		11,778	66,470	104,717	-
3.5k JPN	Japanese	1,392 /	7,104	0.196				PASS	
HGVD	Japanese	451 /	1,890	0.239				PASS	
ExAC	^ Total	1,878 /	88,224	0.021				PASS	957348.0
	African/African American	1 /	7,614	1.313e-4				-	
	American	8 /	7,302	0.001				-	
	East Asian	1,860 /	6,992	0.266				-	
	Finnish	0 /	4,554	0.000				-	
	Non-Finnish European	3 /	48,388	6.200e-5				-	
	Other	3 /	668	0.004				-	
South Asian	2 /	10,706	0.001						

- データセット（民族毎）のアリルカウント、アリル頻度、ジェノタイプカウント、NGSのバリアントコールのクオリティ



Gene

HGNC/Approved name	aldehyde dehydrogenase 2 family member
HGNC/Approved symbol	ALDH2
HGNC/Alias name	

Transcripts

Transcript ID	Gene symbol	Consequence type	HGVS(cDNA)	HGVS(Amino acid seq.)	SIFT	PolyPhen
ENST0000261733	ALDH2	missense_variant	ENST0000261733.2:c.1510G>A	ENSP0000261733.2:p.Glu504Lys	0.0	0.874 Possibly Damaging
ENST0000416293	ALDH2	missense_variant	ENST0000416293.3:c.1369G>A	ENSP0000403349.3:p.Glu457Lys	0.0	0.566 Possibly Damaging
ENST0000548536	ALDH2	3_prime_UTR_variant NMD_transcript_variant	ENST0000548536.1:c.*1386G>A			
ENST0000549106	ALDH2	3_prime_UTR_variant NMD_transcript_variant	ENST0000549106.1:c.*89G>A			

Publications

Showing 1 to 10 of 16

PMID	Reference	Year	Cited by	MeSH
28036260 (PubTator)	Effects of alcohol consumption, ALDH2 rs671 polymorphism, and Helicobacter pylori infection on the gastric cancer risk in a Korean population. Yang S, Lee J, Choi IJ, Kim YW, Ryu KW, Sung J, Kim J <i>Oncotarget.</i> 2017 Jan 24;8(4):6630-6641.	2017	1	<ul style="list-style-type: none"> D016481 Helicobacter infection D009369 tumor D007239 infection D013274 gastric cancer
28038378 (PubTator)	The causal effects of alcohol on lipoprotein subfraction and triglyceride levels using a Mendelian randomization analysis: The Nagahama study. Tabara Y, Arai H, Hirao Y, Takahashi Y, Setoh K, Kawaguchi T, Kosugi S, Ito Y, Nakayama T, Matsuda F <i>Atherosclerosis.</i> 2017 Feb;257:22-28.	2017	1	関連疾患名
28040078	A missense single nucleotide polymorphism in the ALDH2 gene, rs671, is associated with hip fracture.			<ul style="list-style-type: none"> D014947 wound and injury D050723 bone fracture

トランスクリプト毎の Molecular, Consequence, SIFT, Polyphen2のスコア

PubMedに登録された論文のabstractに当該バリアントの記述がある論文

引用数

解説動画あります (TogoTV)

The screenshot shows two main pages from the NBDC website:

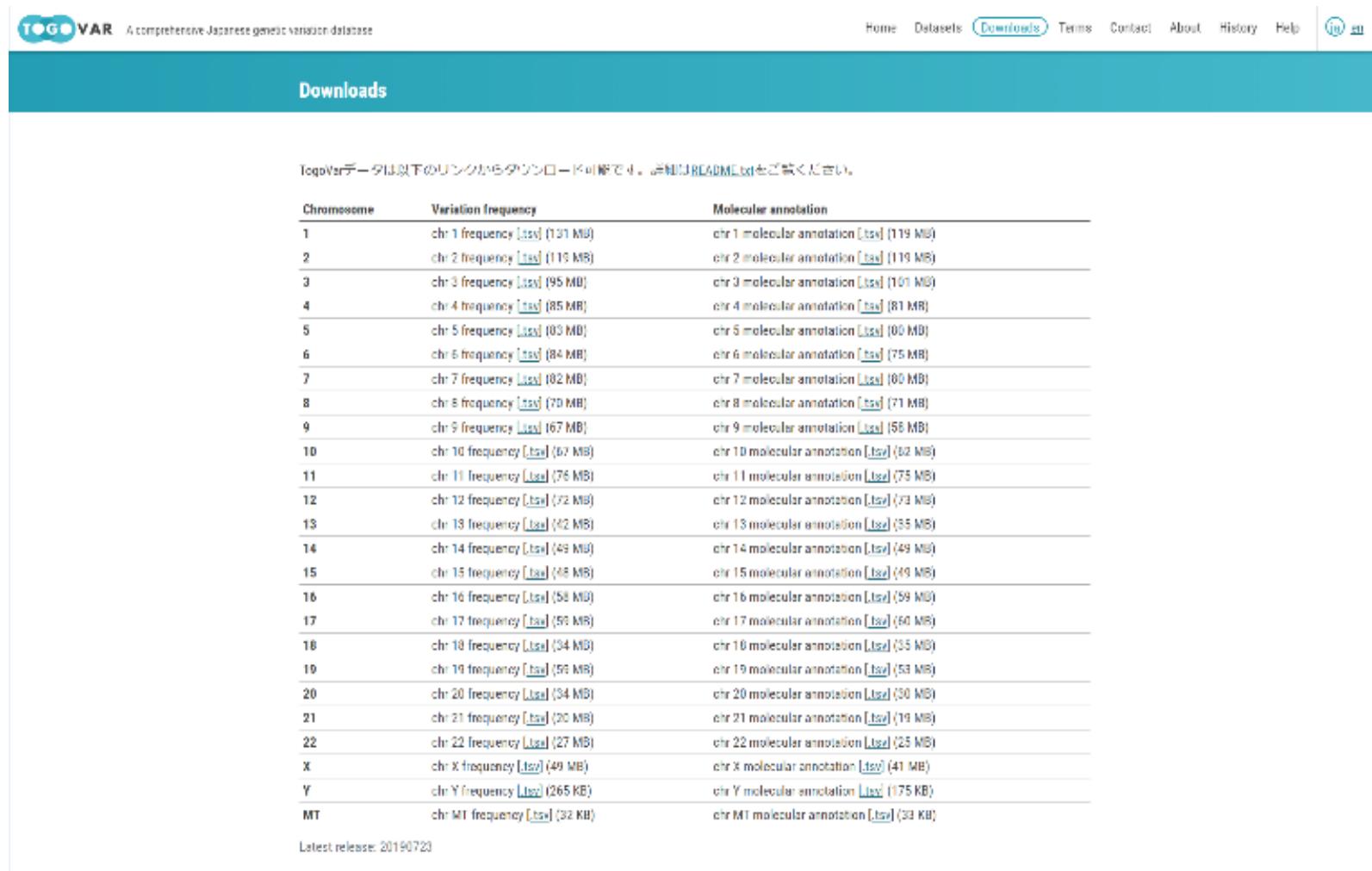
- TogoTV Page:** A search results page for "TogoVar TogoTV". It lists a single result: "togotv - YouTube" with the URL <https://im.youtube.com/useed/togotv/feed>. Below the link is a brief description: "TogoVar (日本人ゲノム多様性統合データベース)は、国立研究開発法人科学技術振興機構 バイオサイエンスデータベースセンター (NBDC)と大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用 ...".
- TogoVar Page:** A detailed page about TogoVar. The title is "2018-08-25 TogoVar でヒトゲノムに存在するパリアントに関する情報を調べる". The page content includes a screenshot of a browser showing the TogoVar interface, a large blue callout box with Japanese text explaining what TogoVar is, and several sections of explanatory text in Japanese.

- 操作方法等の解説動画 (8分36秒)
 - Youtube、TogoTV

一括ダウンロードできます

頻度データ（tsvファイル）の一括ダウンロードURL

- （<https://togovar.biosciencedbc.jp/downloads>）



The screenshot shows the 'Downloads' section of the Togovar database. The page header includes the Togovar logo and navigation links for Home, Datasets, Downloads (which is highlighted in blue), Terms, Contact, About, History, and Help. Below the header is a teal bar with the word 'Downloads'. A message in Japanese at the top of the main content area says: 'Togovarデータは以下のリンクからダウンロード可能で、詳細なREADME.txtをご覧ください。' (Togovar data can be downloaded from the following links. Please refer to the detailed README.txt for more information.)

Chromosome	Variation frequency [tsv]	Molecular annotation [tsv]
1	chr 1 frequency [tsv] (131 MB)	chr 1 molecular annotation [tsv] (119 MB)
2	chr 2 frequency [tsv] (119 MB)	chr 2 molecular annotation [tsv] (119 MB)
3	chr 3 frequency [tsv] (95 MB)	chr 3 molecular annotation [tsv] (101 MB)
4	chr 4 frequency [tsv] (85 MB)	chr 4 molecular annotation [tsv] (81 MB)
5	chr 5 frequency [tsv] (103 MB)	chr 5 molecular annotation [tsv] (90 MB)
6	chr 6 frequency [tsv] (84 MB)	chr 6 molecular annotation [tsv] (75 MB)
7	chr 7 frequency [tsv] (82 MB)	chr 7 molecular annotation [tsv] (80 MB)
8	chr 8 frequency [tsv] (70 MB)	chr 8 molecular annotation [tsv] (71 MB)
9	chr 9 frequency [tsv] (67 MB)	chr 9 molecular annotation [tsv] (58 MB)
10	chr 10 frequency [tsv] (67 MB)	chr 10 molecular annotation [tsv] (62 MB)
11	chr 11 frequency [tsv] (76 MB)	chr 11 molecular annotation [tsv] (75 MB)
12	chr 12 frequency [tsv] (72 MB)	chr 12 molecular annotation [tsv] (73 MB)
13	chr 13 frequency [tsv] (42 MB)	chr 13 molecular annotation [tsv] (35 MB)
14	chr 14 frequency [tsv] (49 MB)	chr 14 molecular annotation [tsv] (49 MB)
15	chr 15 frequency [tsv] (48 MB)	chr 15 molecular annotation [tsv] (49 MB)
16	chr 16 frequency [tsv] (58 MB)	chr 16 molecular annotation [tsv] (59 MB)
17	chr 17 frequency [tsv] (55 MB)	chr 17 molecular annotation [tsv] (60 MB)
18	chr 18 frequency [tsv] (34 MB)	chr 18 molecular annotation [tsv] (35 MB)
19	chr 19 frequency [tsv] (55 MB)	chr 19 molecular annotation [tsv] (53 MB)
20	chr 20 frequency [tsv] (34 MB)	chr 20 molecular annotation [tsv] (30 MB)
21	chr 21 frequency [tsv] (20 MB)	chr 21 molecular annotation [tsv] (19 MB)
22	chr 22 frequency [tsv] (27 MB)	chr 22 molecular annotation [tsv] (25 MB)
X	chr X frequency [tsv] (49 MB)	chr X molecular annotation [tsv] (41 MB)
Y	chr Y frequency [tsv] (265 KB)	chr Y molecular annotation [tsv] (175 KB)
MT	chr MT frequency [tsv] (32 KB)	chr MT molecular annotation [tsv] (33 KB)

Latest release: 20190720

[ハンズオン4] TogoVar

[basic]

- rs671を検索し、データセットごとの頻度の差を確認して下さい
- 3.5KJPN (ToMMo)とHGVDでともに頻度が5%以上あるSNPの総数を確認して下さい。

[advanced]

- Takenouchi-Kosaki syndromeで検索し、頻度が公共DBで報告されているか確認してください。報告されていない場合は、その理由を考えてください。

rs671を検索し、データセットごとの頻度の差を確認して下さい

TOGO VAR A comprehensive Japanese genetic variation database

Home Datasets Downloads Terms Contact About History Help Configuration

rs671

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

Results The number of available data is 1 out of 1 which is filtered with 74,698,940 all data.

TogoVar ID	RefSNP ID	Position	Ref / Alt	Type	Gene	Alt frequency	Consequence	SIFT	PolyPhen	Clinical significance
tgv47264307	rs671	12: 112241766	G > A	SNV	ALDH2	0.874	Missense variant	+1	RF	Esophageal cancer, alcohol-related, sus

Variant report tgv47264307 dbSNP rs671

Variant type SNV Position 12:112241766 (GRCh37) hgvs 12:g.112241766G>A

Other overlapping variant(s)
No other variants on the same location

Frequency

Dataset	Population	Allele count	Total	Frequency	Genotype count	Alt / Alt	Alt / Ref	Ref / Ref	Filter status	Quality score
JGA-NGS	Japanese	52	250	0.208	11,778	66,470	104,717	-	PASS	451.77
JGA-SNP	Japanese	90,026	365,930	0.246						
3.5k JPN	Japanese	1,392	7,104	0.196						
HGVD	Japanese	451	1,890	0.239						
ExAC	Total	1,878	88,224	0.021						

Clinical Significance

Title	Clinical significance	Review status	Last evaluated	Condition(s)
NM_000690.3(ALDH2):c.1510G>A (p.Glu504Lys)	drug response protective risk factor	★★★★ no assertion criteria provided	2018-08-09 2010-02-01	- Acute alcohol sensitivity - Alcohol dependence - Esophageal cancer, alcohol-related, susceptibility to - Sublingual nitroglycerin, susceptibility to poor response to

Preview

Detailed variant report page

Genes
Symbol: ALDH2

External links
refSNP: rs671
ClinVar: VCV000018390

Alternative allele frequencies

Dataset	Alt	Total	Frequency
JGA NGS	52	250	0.2080
JGA SNP	90,026	365,930	0.2460e-1
3.5JPN	1,392	7,104	0.1959
HGVD	451	1,890	2.386e-1
ExAC	1,878	88,224	2.129e-2

Consequence

Missense variant
A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved

3 prime UTR variant
A UTR variant of the 3' UTR

NMD transcript variant

→日本人や東アジア集団では、0.2前後ですが、ExACの平均では0.021と1/10以下の頻度。日本人はお酒に弱い(ClinVarではAcute alcohol sensitivityなど)

3.5KJPN (ToMMo)とHGVDとともに頻度が5%以上あるSNPの総数を確認して下さい。

58176

TOGO VAR A comprehensive Japanese genetic variation database

Search for disease or gene symbol or rs...

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

Results The number of available data is 10,000 out of 58,176 which is filtered with 74,698,940 all data.

TogoVar ID	RefSNP ID	Position	Ref / Alt	Type	Gene	Alt frequency	Consequence	SIFT	PolyPhen	Clinical sign
tgv457	rs201219564	1: 69270	A > G	SNV	OR4F5	██████	Synonymous variant			
tgv478	rs2691305	1: 69511	A > G	SNV	OR4F5	██████	Missense variant	0.82		
tgv528	rs776815449	1: 69849	G > A	SNV	OR4F5	██████	Stop gained			
tgv535	rs200676709	1: 69897	T > C	SNV	OR4F5	██████	Synonymous variant			
tgv6332	rs9988179	1: 865694	C > T	SNV	SAMD11	██████	Missense variant	0	+3	0.637 +4
tgv6660	rs28419423	1: 871215	C > G	SNV	SAMD11	██████	Synonymous variant			
tgv7053	rs4372192	1: 876499	A > G	SNV	SAMD11	██████	Non coding transcript exon variant	+2		
tgv7255	rs79037098	1: 877782	C > G	SNV	SAMD11	██████	Splice region variant	+3		
tgv7270	rs6672356	1: 877831	T > C	SNV	SAMD11	██████	Missense variant	+1		1 +2
tgv7709	rs7523549	1: 879317	C > T	SNV	SAMD11	██████	Synonymous variant			
tgv8234	rs2272757	1: 881627	G > A	SNV	NOC2L	██████	Synonymous variant	+1		
tgv8426	rs4970378	1: 883625	A > G	SNV	NOC2L	██████	Intron variant	+1		
tgv8767	rs3828047	1: 887801	A > G	SNV	NOC2L	██████	Synonymous variant	+1		
tgv8894	rs3748596	1: 888639	T > C	SNV	NOC2L	██████	Synonymous variant	+1		
tgv8898	rs3748597	1: 888659	T > C	SNV	NOC2L	██████	Missense variant	+1		1
tgv8942	rs13303056	1: 889158	G > C	SNV	NOC2L	██████	Splice region variant	+2		
tgv8943	rs13302945	1: 889159	A > C	SNV	NOC2L	██████	Splice region variant	+2		
tgv9382		1: 892460	G > C	SNV	NOC2L	██████	Intron variant	+1		
tgv9676	rs13303010	1: 894573	G > A	SNV	NOC2L	██████	Intron variant	+1		
tgv10069	rs4970441	1: 897325	G > C	SNV	KLHL17	██████	Synonymous variant	+1		
tgv10252	rs7549631	1: 897730	C > T	SNV	KLHL17	██████	Splice region variant	+2		
tgv10511	rs41285808	1: 898467	C > T	SNV	KLHL17	██████	Synonymous variant	+2		
tgv10644	rs117269332	1: 898852	C > T	SNV	KLHL17	██████	Synonymous variant	+2		
tgv10900	rs6677386	1: 899928	G > C	SNV	KLHL17	██████	Intron variant			
tgv10976	rs80351873	1: 900319	G > A	SNV	KLHL17	██████	Intron variant			
tgv11034	rs28705211	1: 900505	G > C	SNV	KLHL17	██████	Synonymous variant			
tgv11201	rs116147894	1: 902069	T > C	SNV	PLEKH1N1	██████	Intron variant			
tgv11223	rs28499371	1: 902128	C > T	SNV	PLEKH1N1	██████	Missense variant	0.08 +2	0.067 +2	

Filters

Dataset

- All 58,176
- 3.5KJPN 58,176
- HGVD 58,176
- JGA NGS 53,449
- ICA SNP 25,776
- ExAC 57,635
- Clinvar 8,870

Alternative allele frequency

0.05 ~ 1 Invert range
for all datasets for any dataset

Variant calling quality

Exclude filtered out variants in all datasets

Variant type

- All 58,176
- SNV 54,936
- Insertion 1,609
- Deletion 1,631
- Indel 0
- Substitution 0

Clinical significance

- All 58,176
- Not in ClinVar 49,306
- Pathogenic 37
- Likely pathogenic 5
- Uncertain significance 144
- likely benign 3,547
- Benign 15,837
- Conflicting interpretations of pathogenicity 25

Takenouchi-Kosaki syndromeで検索し、頻度が公共DBで報告されているか確認してください。報告されていない場合は、その理由を考えてください。

TOGO VAR A comprehensive Japanese genetic variation database

Home Datasets Downloads Terms Contact About History Help Configuration

Takenouchi-Kosaki syndrome

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

Results The number of available data is 6 out of 6 which is filtered with 74,698,940 all data.

TogoVar ID	RefSNP ID	Position	Ref / Alt	Type	Gene	Alt frequency	Consequence	SIFT	PolyPhen	Clinical significance
tgv697162	rs1064795845	1: 22405033	T > C	SNV	CDC42		Missense variant (+2)	0 (+1)	0.851 (+3)	LP not provided (+1)
tgv697429	rs864309721	1: 22412944	A > G	SNV	CDC42		Missense variant (+1)	1 (+4)	P not provided (+1)	
tgv697430	rs797044870	1: 22412949	A > G	SNV	CDC42		Missense variant (+1)	0.96 (+4)	P Takenouchi-Kosaki syndrome (+1)	
tgv697435		1: 22412995	G > T	SNV	CDC42		Missense variant (+1)	0.981 (+4)	P Takenouchi-Kosaki syndrome (+1)	
tgv697436		1: 22413000	T > C	SNV	CDC42		Missense variant (+1)	0.998 (+4)	P Takenouchi-Kosaki syndrome (+1)	
tgv697597		1: 22417945	G > A	SNV	CDC42		Missense variant	0.02 (+2)	0.976 (+2)	P Noonan-Like Syndrome Disorder (+1)

Filters

Dataset

- All 6
- JGA NGS 0
- JGA SNP 0
- 3.5KJPN 0
- HGVD 0
- ExAC 0
- Clinvar 6

Alternative allele frequency

0 ~ 1 Invert range
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
 for all datasets for any dataset

Variant calling quality

Exclude filtered out variants in all datasets

Variant type

- All 6
- SNV 6
- Insertion 0
- Deletion 0
- Indel 0
- Substitution 0

Clinical significance

- All 6
- Not in ClinVar 0
- P Pathogenic 9
- LP Likely pathogenic 3
- US Uncertain significance 0
- LB likely benign 0
- B Benign 0
- CI Conflicting interpretations of pathogenicity 0

Reportedなし

Reason: 稀少疾患 (rare disease)だから

Takenouchi-Kosaki syndromeで検索し、頻度が公共DBで報告されているか確認してください。報告されていない場合は、その理由を考えてください。

日本医療研究開発機構 難治性疾患実用化研究事業

CDC42阻害剤による武内・小崎症候群の治療法の開発研究班

ホーム 疾患概要 診断基準 研究概要 研究者 研究成果

■ CDC42遺伝子異常症、Takenouchi-Kosaki症候群とは

巨大血小板性血小板減少症、知的障害、特徴的顔貌、感音性難聴、脳構造異常、屈指、リンパ浮腫、反復性の感染症、甲状腺機能低下症などを特徴とする先天異常症候群の一つです。多くの症状が認められますが、症状の組み合わせから臨床的に診断することが可能な疾患です。CDC42という遺伝子の異常によって起こることが原因であることがわかっています。

一方で、CDC42遺伝子の異常を持つ方が、すべて同じ症状を持つようになるわけではありません。CDC42遺伝子の特定の部分に変異を持つ場合にのみ、上記のような特徴的な症状が現れると考えられています。

■ 有病率

正確な頻度は不明です。少なくとも国内に複数の患者さんがいることは確認されており、海外にも患者さんがいることは確認されています。診断がつかないままになっている患者さんも多くいると考えられています。

■ 病因

CDC42遺伝子の特定のアミノ酸置換変異で発症することがわかっています。CDC42遺伝子は1番染色体上にあります。2つあるCDC42遺伝子のうち、片方に変化があると発症する（常染色体優性）

<https://plaza.umin.ac.jp/tks/disease/>

将来的な機能強化（予定）

1. 國際的なデータベース

1. gnomAD、GWAS-catalogの取り込み
2. ClinVarの定期更新
3. eQTL (Expression quantitative trait loci)の表示 (GTEx)

2. 国内のデータベース

1. JENGER(BBJ)、MGeND (日本版ClinVar)との連携
2. 日本人8000人分のWGSのデータ提供、再解析
 1. LD(Linkage Disequilibrium)情報の取り込み

3. 検索内容の追加

1. HGVS表記 (例. CCDS4702.1:c.123C>T)
2. GRCh38対応

NBDCヒトデータベースとは？

NBDC National Bioscience Database Center NBDCヒトデータベース English サイト内検索 検索

ホーム データの利用 データの提供 ガイドライン 機関外サーバ NBDCヒトデータ審査委員会 成果発表 お問い合わせ FAQ

NBDCヒトデータベースについて

ヒトに関するデータは、次世代シーケンサーはじめとした解析技術の発達に伴って膨大な量が産生されつつあり、それらを整理・格納して、生命科学の進展のために有効に活用するためのルールや仕組みが必要です。

国立研究開発法人科学技術振興機構(JST)バイオサイエンスデータベースセンター(NBDC)では、個人情報の保護に配慮しつつヒトに関するデータの共有や利用を推進するために、ヒトに関する様々なデータを共有するためのプラットフォーム『NBDCヒトデータベース』を設立するとともに、国立遺伝学研究所 DNA Data Bank of Japan (DDBJ)と協力して、ヒトに関するデータを公開しています。

本Webサイトを通して、ヒトに関するデータの利用及びヒトに関するデータの提供を行なうことができます。

なお、本データベースの目的・意義、扱うデータの種類、データ利用者の範囲、責任者については[こちら](#)をご覧ください。

新着情報

2019/05/28 理化学研究所 生命医科学研究センターからの制限公開データ (Type I) を公開しました (hum0160)

2019/05/09 大阪大学免疫学フロンティア研究センター 実験免疫学からの制限公開データ (Type I) を公開しました (hum0141)

▶ ニュース一覧へ

Search NBDC Human Database Beacon for Alternative Alleles [API help]

NBDC Human Database Beacon is a member of GA4GH Beacon Network.

GRCh37 e.g. 12:112241766 A Search Example: ALDH2 Variant (GRCh37, '12:112241766 A')

利用可能な研究データ一覧

データ利用方法は[こちら](#)をご覧下さい。

全 104 件 Copy CSV Excel 一覧内検索:

Research ID	研究題目	公開日	データの種類	研究方法	手法	参加者 (対象集団)	提供者	アクセス制限
hum0173.v1 JGAS00000000171	胎児期に始まる子どもの健康と発達に関する調査	v1:2019/04/23	NGS (RRBS)	メチル化 解析	Illumina (NextSeq)	臍帯組織 (正常出産) : 30検体 (日本人)	森千里	制限 (Type I)
hum0165.v1 JGAS00000000162	家族性骨髄異形成症候群の遺伝子解析研究	v1:2019/05/07	NGS (Exome)	配列決定	Illumina (HiSeq 2500)	家族性骨髄異形成症候群: 2症例 (1家系) (日本人)	古屋淳史	制限 (Type I)

<https://humandbs.biosciencedbc.jp/>

研究者の産出したデータを、他の研究者が再利用できるサービス

73

NBDCヒトデータベースの必要性 研究データ公開・再利用の世界的潮流から

- オープンサイエンス
学術論文のオープン化 + 研究データの公開

2014年米国NIHのGenomics Data Sharing Policy
ヒトゲノム研究データのデータリポジトリへの登録を要求

2018年3月 AMEDデータマネージメントプラン提出の義務化

- FAIR原則
Findable(見つけられる)、Accessible(アクセスできる)、
Interoperable(相互運用できる)、Reusable (再利用できる)
- Global Alliance for Genomics and Health(GA4GH)
国際的な研究情報共有や統一プロトコルでのゲノム解析の仕組みや、
研究者認証等についての枠組みを討議するアライアンス

NBDCヒトデータベースの必要性 制限公開データベースの必要性（プライバシーの担保）

- 研究データのオープン化の懸念
 - プライバシー保全とのバランス
 - データ公開・利用についての適切な審査
 - →制限公開データベース
- 制限公開データベースの発足
 - 2007年米国NIH
 - Database of Genotypes and Phenotypes (dbGaP)
 - 2008年欧洲EMBL-EBI
 - European Genome-phenome Archive (EGA)
 - 2013年10月 NBDCおよび国立遺伝学研究所DDBJセンター
 - NBDCヒトデータベース/Japanese Genotype-phenotype Archive(JGA)

NBDCヒトデータベース

非制限公開データ
Open data

- ・頻度や統計量
 - ・特定の個人由来では無い試料の解析結果
- Webからダウンロード可能

制限公開データ
Controlled-Access data

- ・個人ごとのゲノムデータ
- 利用者を限定
(国内外のアカデミアや企業内研究者)

二重匿名化

各プロジェクト・実施機関

NBDCヒトデータベースのデータ例 バイオバンクジャパン (BBJ) : hum0014

心筋梗塞：1666症例、対照：3198名
健常者：934名 (JSNP)
35疾患：各約190症例 (JSNP)
食道癌：182症例 (JSNP)
ALS：92症例 (JSNP)
T2DM：9817症例、対照：6763名
T2DM：5646症例、対照：19,420名
AD：1472症例、対照：7966名
AF：8180症例、対照：28,612名
BMI：158,284名
Genotypeデータ：182,505名
POAG：3980症例、対照：18,815名
58臨床検査値：162,255名
初潮年齢データを有する女性：67,029名
閉経年齢データを有する女性：43,861名
BBJ第1コホート：1,026名
(日本人)

登録された対象疾患および形質：
心筋梗塞、2型糖尿病、アトピー性皮膚炎、
心房細動、BMI、開放隅角緑内障、
初潮・閉経年齢等々

プラットフォーム：
インベーダー法 (Hologic Japan社)
Human610-Quad BeadChip(Illumina社) 等の
Genotyping Arrayデータ

昨年8月
Illumina社 HiSeq2500による
BBJ第一コホート 1,026名の
Whole Genome Sequencingを公開

NBDCヒトデータベースのデータ例

J-ADNI:hum0043

NBDC Research ID: hum0043.v1

研究内容の概要

目的 : アルツハイマー病（AD）治療薬の薬効評価基準の最適化を行うために、ADの病態を忠実に反映するサロゲートマーカー（surrogate marker）を定めておく必要がある。J-ADNI研究は、MRIやPETなどの画像サロゲートマーカーの長期的変化に関する一定の基準値を作成するとともにその妥当性を証明するために臨床/神経心理検査データ、血液・脳脊髄液サンプルを並行して収集することを目的とする。

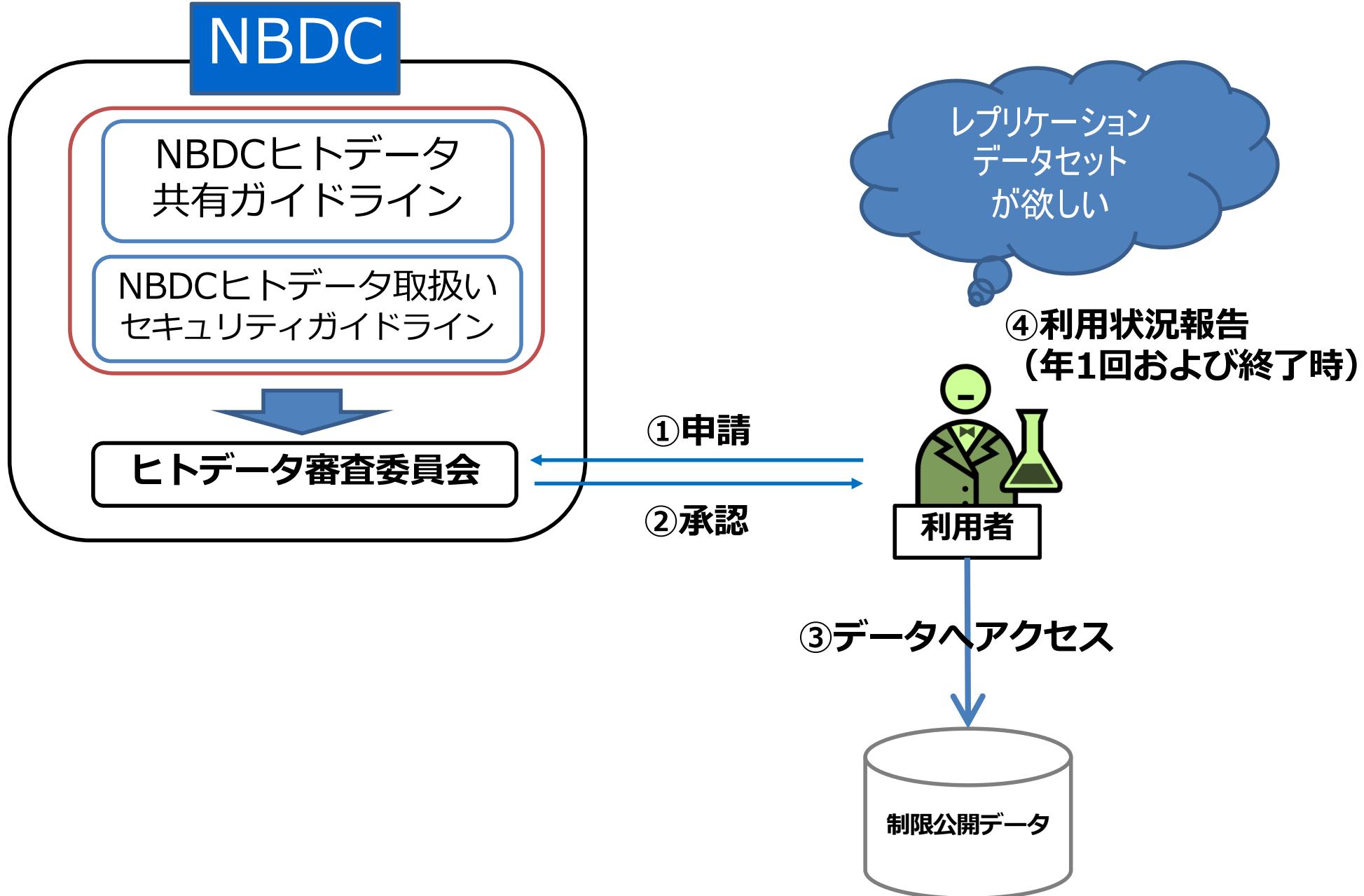
対象 :

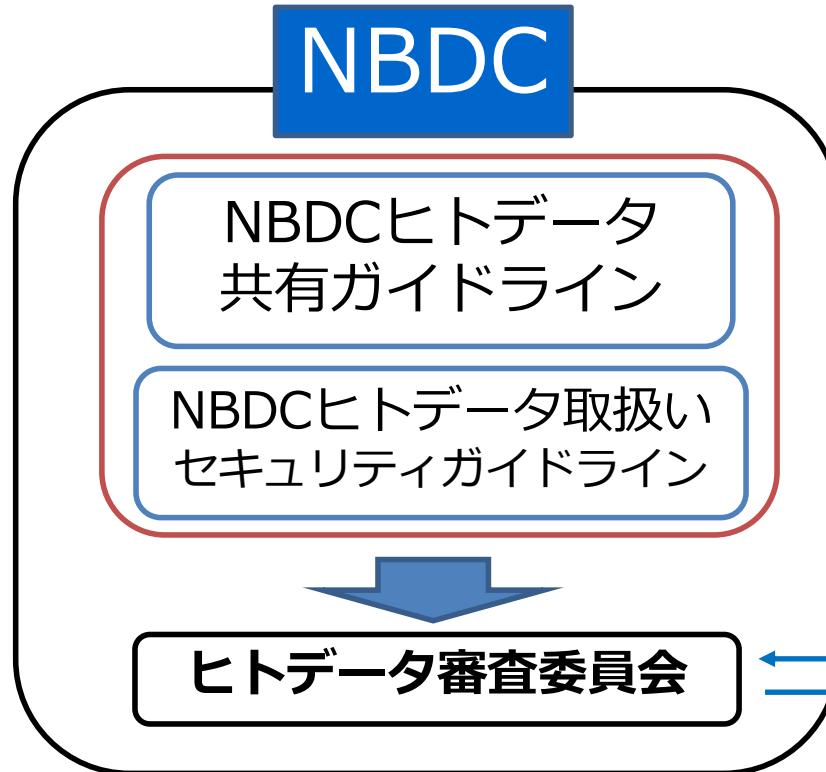
537名

(高齢者健常群 149名、軽度認知機能障害 234名、
早期アルツハイマー病 154名)

- 臨床・心理学検査結果
- 検査結果
ApoE遺伝子型、末梢血液一般検査
- MRI画像、PET画像

78





① ジャーナルの投稿規定や
Funding agencyからの要請

NBDCと提供者間の協議、確認内容

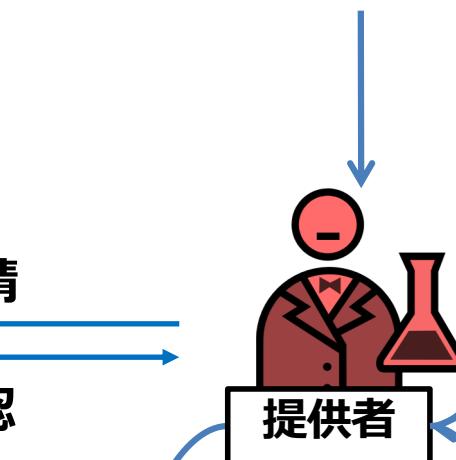
- ・データ公開への同意
- ・機関の長によるデータ公開の許可
- ・データの分類（非制限 or 制限）
- ・データの公開予定日（エンバーゴ可）
- ・データ利用時の制限事項
- ・他の研究者にとっても有益な情報の付加 等

- ①申請
②承認

③データUpload
④accession ID付与



⑤投稿論文に
accession ID
を記載



ご清聴いただき、誠にありがとうございました。

Appendix

その他の多型データベース

Japanese ENcyclopedia of GEnetic associations by Riken (JENGER)

URL: <http://jenger.riken.jp/>

- ・バイオバンク・ジャパン（以下BBJ）のGWASデータを集約した多型データベース

ClinVar

URL: <https://www.ncbi.nlm.nih.gov/clinvar/>

- ・疾患とバリアントの関連を専門家のマニュアルキュレーションを施したデータベース

Medical Genomics Japan Variant Database (MGeND)

URL: <https://mgend.med.kyoto-u.ac.jp/>

- ・日本版ClinVar
- ・約9,000バリアントに臨床的意義を付与
- ・対象疾患：がん/希少疾患/感染症/認知症/難聴

DBKERO

URL: <http://kero.hgc.jp/>

- ・薬剤応答データを検索可能なデータベース

Haploreg

URL: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

- ・米国Broad研究所
- ・検索対象のSNPだけでなく、高い連鎖不平衡状態にあるSNPも検索可能