### セマンティックウェブ技術を用いたバイオ リソース横断検索システム 1/7

#### トーゴーの日 シンポジウム2021

○櫛田達矢、臼田大輝、高田豊行、桝屋啓志

国立研究開発法人理化学研究所 バイオリソース研究センター(BRC)

#### 理研BRC:バイオリソースを総合的取り扱う世界最大級のリソースセンター

- ・実験マウス(例,近交系,トランスジェニック,KO,KI,疾患モデル,ゲノム編集マウス)
- ・植物材料(例,遺伝子,シロイヌナズナ種子(野生株,変異体),培養細胞)
- ・細胞株(例,幹細胞,臍帯血細胞,がん細胞株,疾患特異的iPS細胞)
- ・遺伝子材料(例,DNAクローン,ゲノムDNA,タンパク質産生・遺伝子発現用ベクター,組換えウイルス)
- ・微生物株(細菌(放線菌を含む),アーキア(古細菌),真菌(酵母と糸状菌),酵母様藻類)

#### BRCカタログシステム(https://brc.riken.jp)





- ・RDFベースのバイオリソース横断検索システム.
- ・Ensembl, UniProt, Orthologous Matrix (OMA)のオルソログ, DisGeNETの疾患-遺伝子間相互作用, KEGG Medicusのゲノム-疾患関連のRDFデータ、並びにNANDO, DOID, MONDO, ORDOなどの疾患オントロジー及 びMP, HPなどの表現型オントロジーを統合.
- ・統合したRDFデータから、例えば、疾患名や国際疾病分類コードを検索キーワードとして、疾患研究に活用が期 待されるバイオリソースを横断的に探索すること可能.





















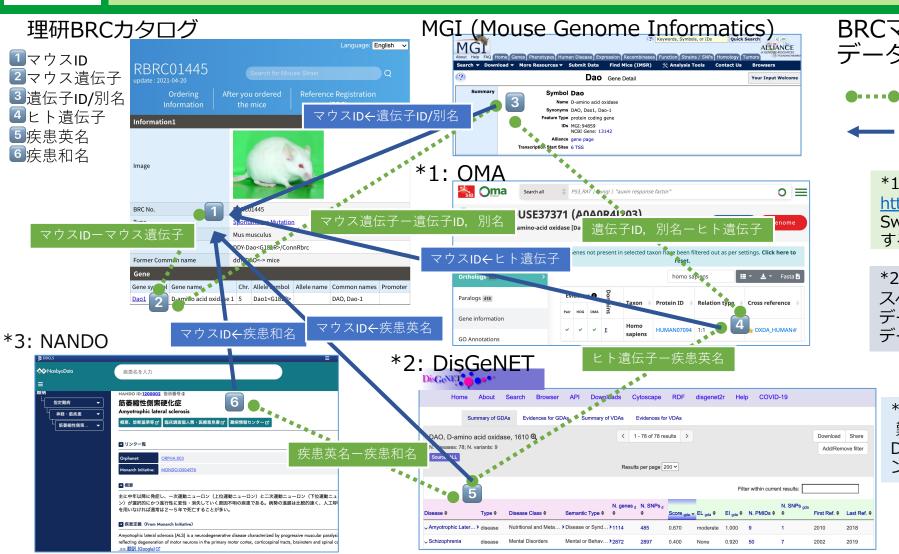




39

### セマンティックウェブ技術を用いたバイオリソース横断検索システム 2/7

#### トーゴーの日シンポジウム2021



BRCマウスカタログを中心とした外部 データベースの統合

●・・・・・
データベース間のリンク

◆ 各概念からマウス材料への検索リンク

\*1: OMA (Orthologous Matrix, <a href="https://omabrowser.org/">https://omabrowser.org/</a> )
Swiss Institute of Bioinformatics (SIB)が開発するオルソログデータベース

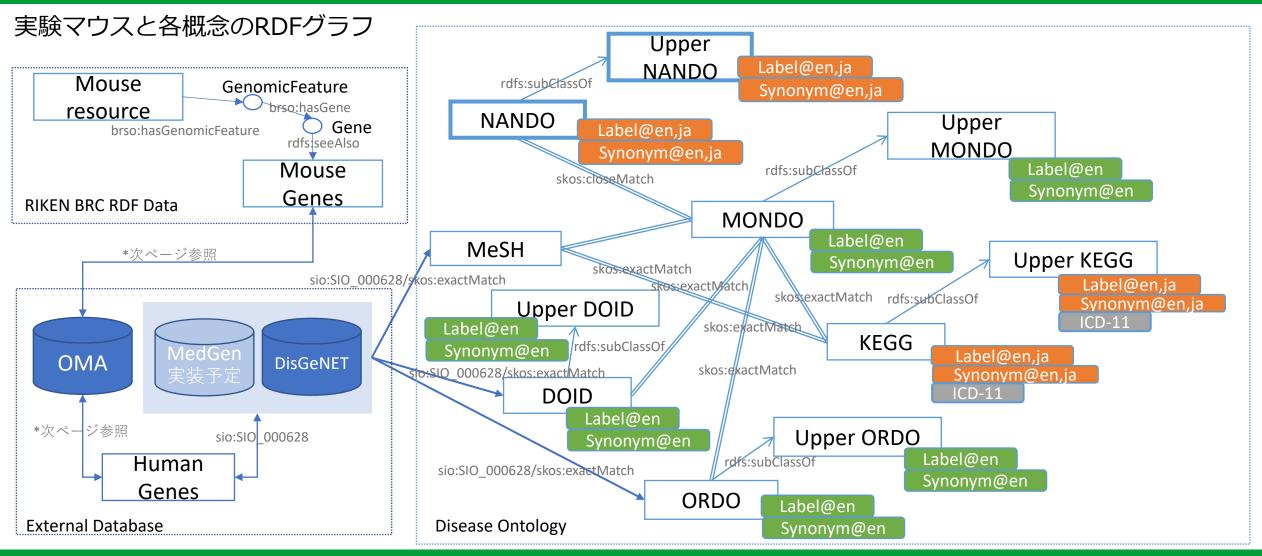
\*2: DisGeNET (https://www.disgenet.org/ ) スペインの医学研究機関が開発する文献や外部 データベースから抽出した疾患と遺伝子の関係 データベース.

\*3: NANDO (Nanbyo Disease Ontology, 難病オントロジー, https://nanbyodata.jp/) DBCLSが開発する指定難病, 小児慢性特定疾病のオントロジー. 2300件以上の日英の疾患名を収録



## セマンティックウェブ技術を用いたバイオリソース横断検索システム 3/7

#### トーゴーの日シンポジウム2021

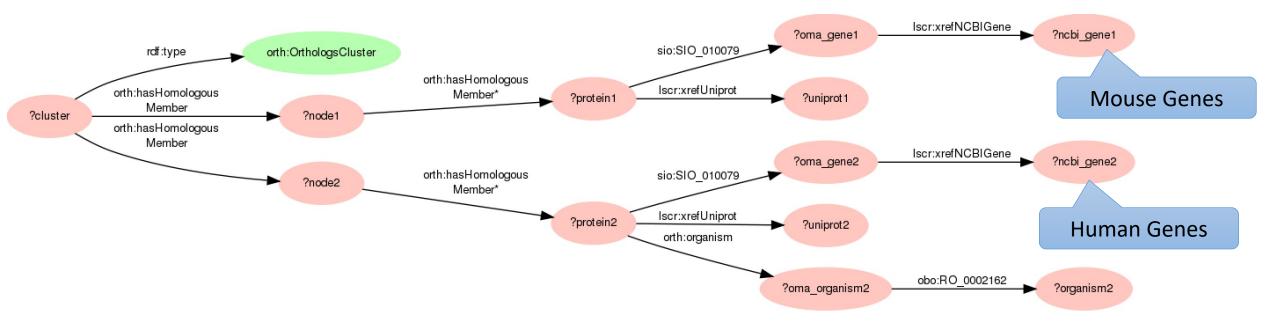




## セマンティックウェブ技術を用いたバイオリソース横断検索システム 4/7

#### トーゴーの日シンポジウム2021

#### OMA RDFグラフ



OMAを活用したオルソログ遺伝子の取得



RRIKEN BRC

### セマンティックウェブ技術を用いたバイオリソース横断検索システム 5/7

#### トーゴーの日シンポジウム2021

#### BRCカタログシステムの検索インターフェイス

- 1. SPARQL Endpoint (<a href="https://knowledge.brc.riken.jp/sparql">https://knowledge.brc.riken.jp/sparql</a>)
- 2. キーワード検索 (https://web.brc.riken.jp/)

#### キーワード検索(バイオリソース検索)の例

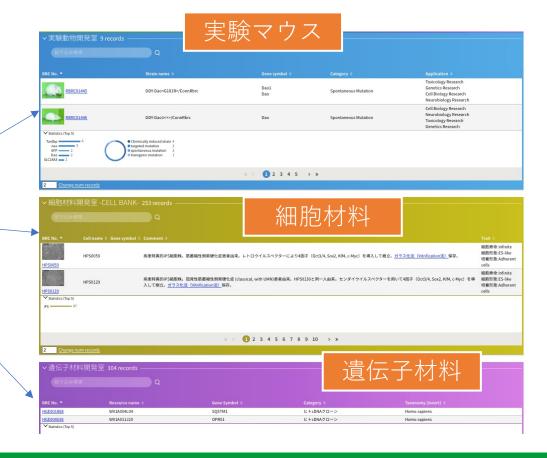
- ・疾患名,遺伝子シンボル,国際疾病分類コードなどで検索可能
- ・サジェスチョンリストで検索ワードの入力をサポート





#### 検索結果

・リソースタイプ別に検索結果を表示



Licensed under a Creative Commons 表示4.0国際ライセンス ©2021 櫛田達矢(理研バイオリソース研究センター)



# セマンティックウェブ技術を用いたバイオリソース横断検索システム 6/7

### BRCバイオリオースの統計量

	リソース総数	疾患情報(疾患名) を持つリソース数[%]	疾患種類(MONDO のラベルで算出)
実験マウス	7,321	1,680 [22.9%]	2,369
細胞株	7,176	3,628 [50.6%]	3,979
遺伝子材料	126,068	20,683 [16.4%]	4,015

	リソース総数	遺伝子数	植物種	Phenotype
植物材料	278,492	356,075	28	151

	リソース総数	Habitat (MEO, PDO)	微生物株カテゴリ
微生物株	18,752	15,767	<b>5 (</b> 酵母様藻類, <u>アーキア</u> , 細菌, 糸状菌, 酵母)



## セマンティックウェブ技術を用いたバイオリソース横断検索システム 7/7

#### トーゴーの日シンポジウム2021

#### 要旨

理化学研究所バイオリソース研究センター(BRC)は、マウス系統、植物種子、がん細胞株、DNAクローン及び微生物株などのバイオリソースを総合的に取り扱う世界最大級のリソースセンターである。当センターでは、リソース及びリソースが持つ情報を簡単な操作で正確に見つけることができるRDFベースの情報検索システムBRCカタログシステム(https://brc.riken.jp)の開発を進めている。このシステムでは、Ensembl、UniProtのほか、Orthologous Matrix (OMA)のオルソログ、DisGeNETの疾患-遺伝子間相互作用、KEGG Medicusのゲノム-疾患関連のRDFデータ、並びにNANDO、DOID、MONDO、ORDOなどの疾患オントロジー及びMP、HPなどの表現型オントロジーを組み入れている。この統合したRDFグラフから、例えば、疾患研究に活用が期待される遺伝子改変マウスの探索が可能である。本報告では、疾患名や国際疾病分類コードを使ったバイオリソース検索の方法などBRCカタログシステム活用事例やそのRDFデータ構造を紹介する。

#### 関連ポスター

疾患名を利用

・ポスター番号:19

「日本の難病に関する中心的基盤となる難病オントロジーNANDOと難病ポータルサイトNanbyoDataの正式リリース」

・ポスター番号:40

「BRSO:Biological Rresource Schema Ontology の利用と現状 †

・ポスター番号:43

「MoG+:生物医科学研究推進のためのマウスゲノム多型データベース」

RIKEN BRC RDF Dataに活用

BRCカタログシステムと 連携予定

