AJACS徳島

疾患・表現型データベース (Gendoo, OMIM, 他)

仲里 猛留

NAKAZATO, Takeru

@chalkless

情報・システム研究機構 データサイエンス共同利用基盤施設
ライフサイエンス統合データベースセンター

Database Center for Life Science (DBCLS),
Joint Support-Center for Data Science Research, Research Organization of Information and Systems (ROIS)

CC BY

撮影可

2019/6/5
@大塚製薬株式会社 徳島研究所

イベント

統合データベース講習会：
AJACS

受入れ機関募集

学会・展示会

トーゴーの日シンポジウム

BioHackathon

その他のイベント

関連情報

› 講師派遣

統合データベース講習会：AJACS徳島

https://events.biosciencedbc.jp/training/ajacs76

資料：https://github.com/AJACS-training/AJACS76

統合データベース講習会：AJACSは、生命科学系のデータベースやツールの使い方、データベースを統合する活動を紹介する講習会です。

今回の講習会では、生命科学系データベースのカタログ、横断検索、アーカイブ、ヒトデータに関するサービス・ツール等の紹介に加えて、多型データベース、疾患・表現型データベース、相互作用・パスウェイデータベース、ゲノムブラウザ・発現・局在関連データベース、タンパク質立体構造データベース、化合物データベースについてご紹介します。参加者全員がハンズオンでコンピュータを使いながらの講習です。

このセクションはtweet OK

ハッシュタグは #AJACS

データベース統合化における問題点

・似たようなデータベースがたくさんありすぎる
　→ 何を見ていいのかわからない
　　えてして全部見ないといけない…

・そもそもどう見たらいいかわからない

・データベース間でエントリ間の対応がとれない

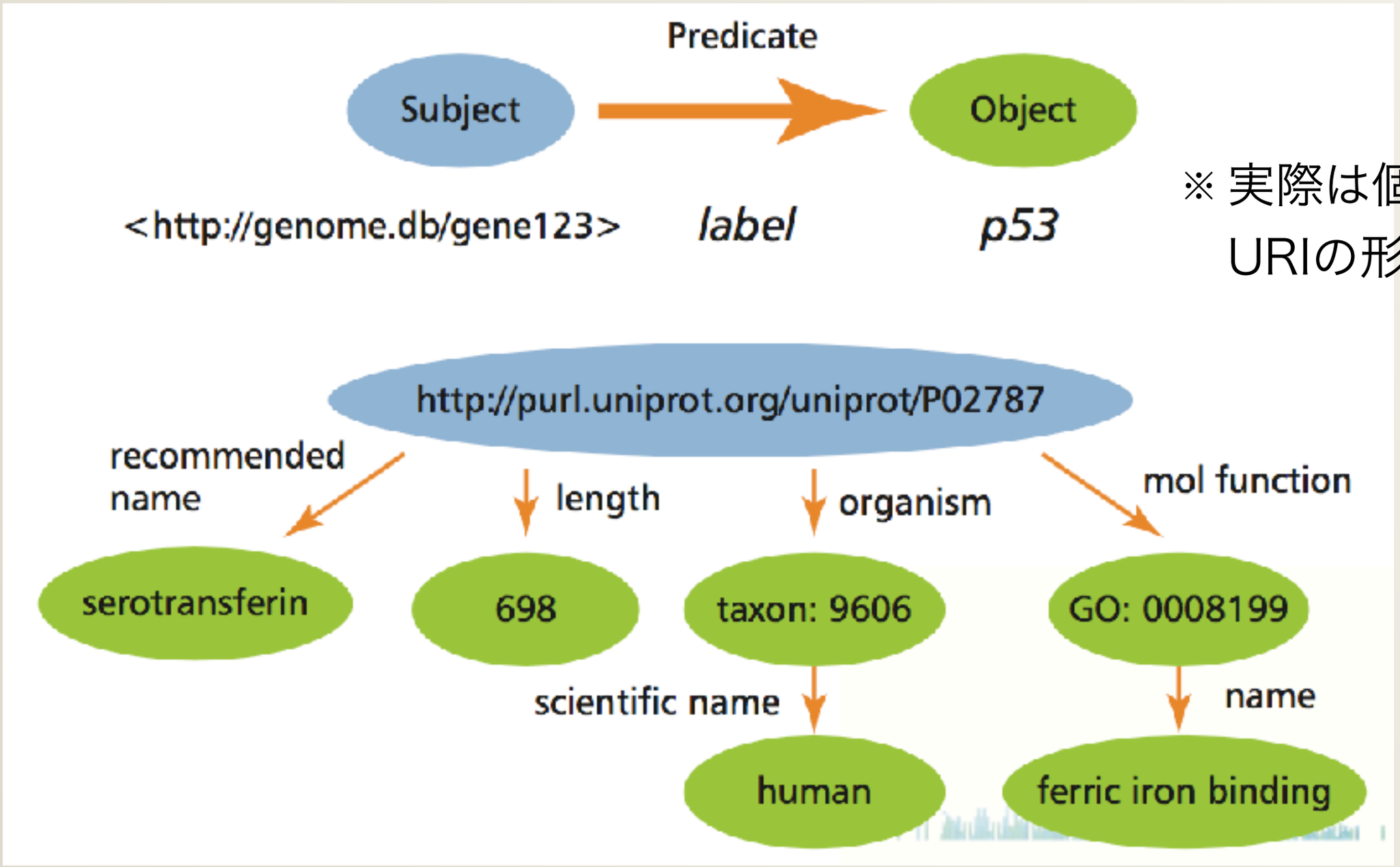
・記述内容がバラバラで同じデータベース内でも対応がとれない

・文章（自然言語）で書かれていてコンピューターで
　一括処理できない

統合化を支えるSemantic Web技術

RDF：Resource Description Framework

主語　　述語　　目的語



※ 実際は個々をURIの形で表現

そんなこと

言われましても

誰でもできるデータベース統合の考え方

別に Excel でデータを管理してもいい
（あとで関係性をバラしてRDFにできる）

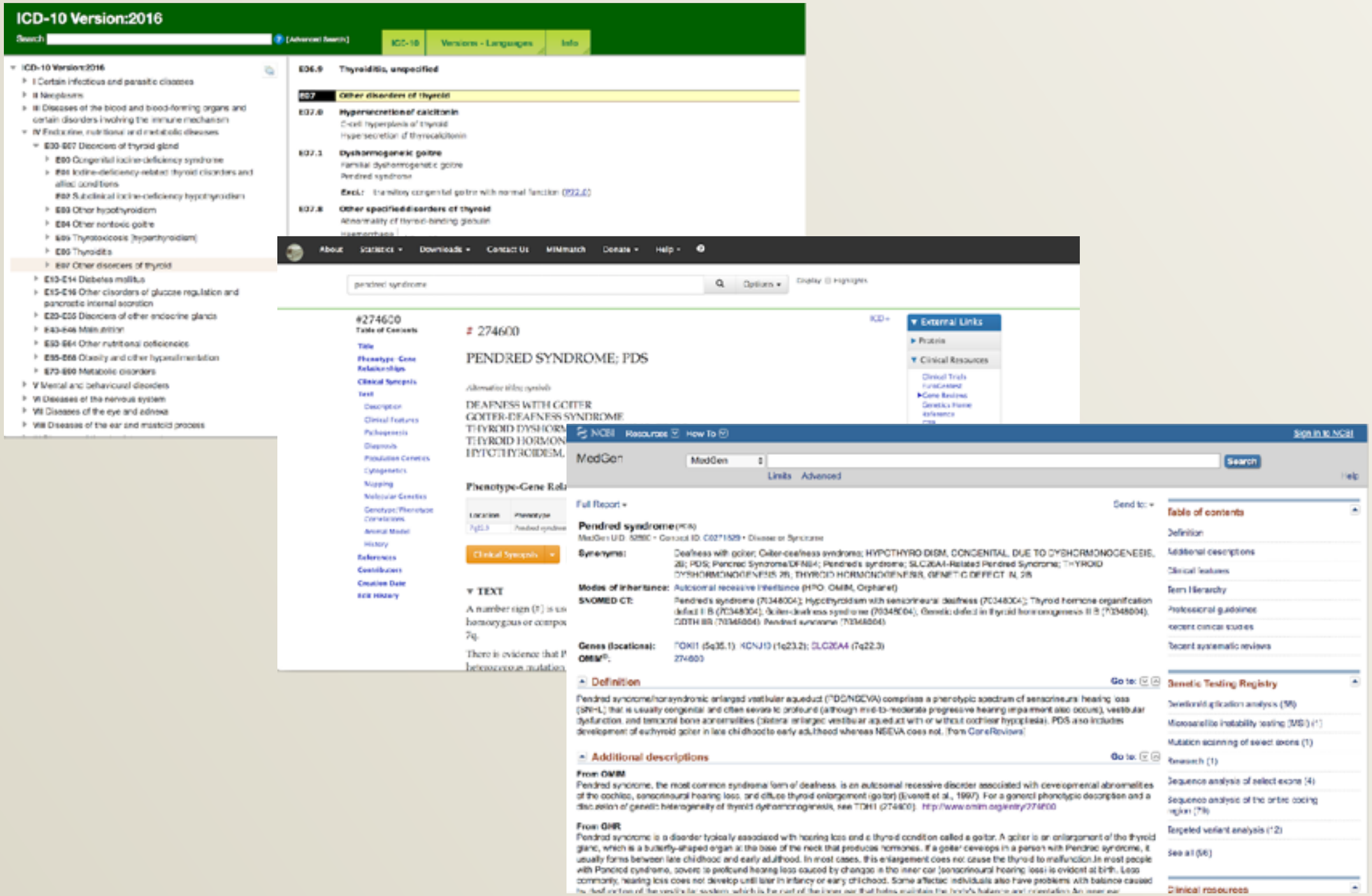
が、

同じことが書かれる列は同じラベルをつけておく
（そしてできれば誰かのマネをする）

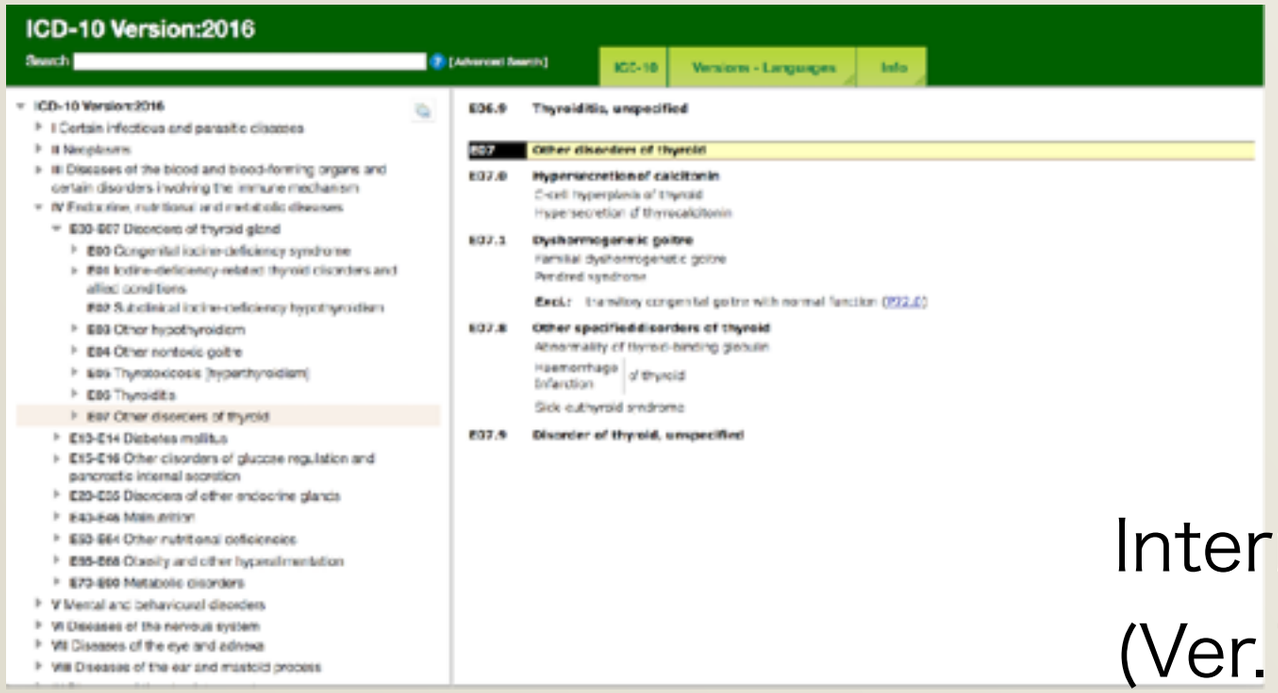
データ部分は既存の用語集から用語をもってくる
せめて同じデータ内容は同じ記述をする
※ 用語集＝オントロジー

これだけで統合化がだいぶ楽に

さまざまな疾患データベース



ICD-10



International Classification of Diseases

(Ver.10)

https://www.who.int/classifications/icd/icdonlineversions/en/

by WHO

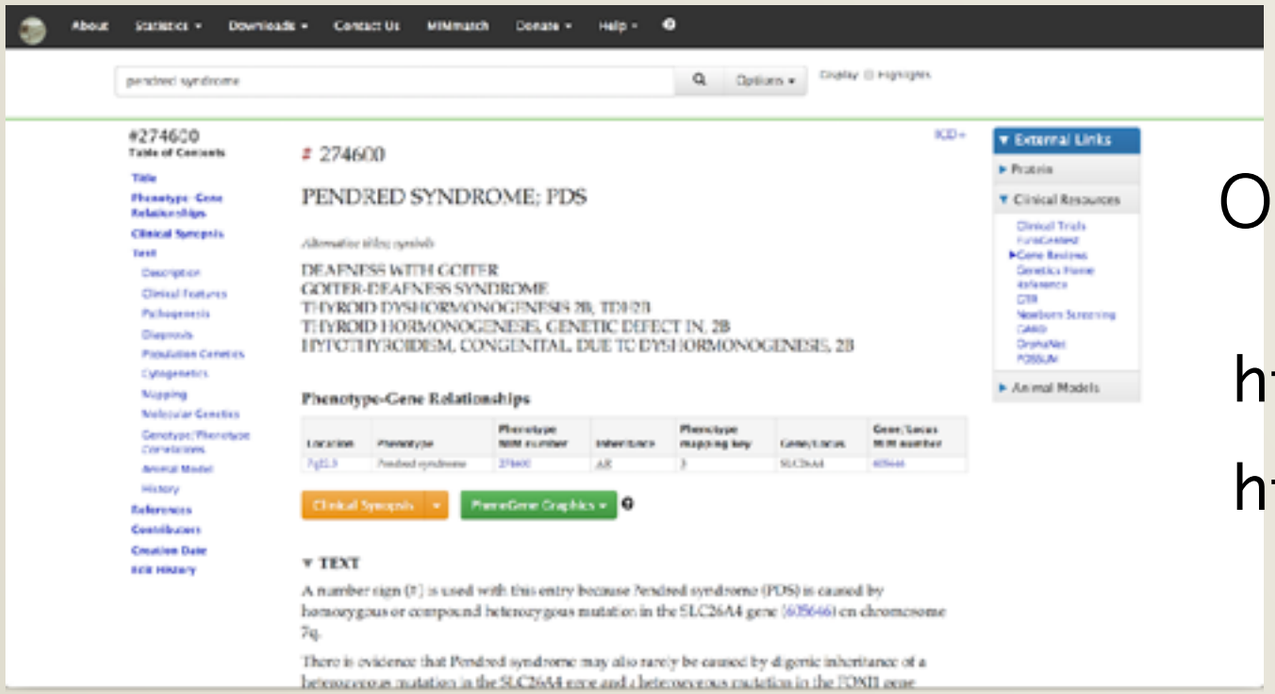
疾患名リスト

日本語化もされている

電子カルテ等でも病名の基準に

病名だけで、詳細な解説はない
ましてや関連遺伝子はわからない

OMIM



Online Mendelian Inheritance in Man

https://www.omim.org/

https://www.ncbi.nlm.nih.gov/omim/

Dr McKusickらによる詳細な解説

あくまで遺伝子による疾患なので
原因遺伝子により独自に
ナンバリングされているものも
例：NIDDM1/2/3

最近ライセンスがきびしくなった?

ヒト疾患と疾患関連遺伝子の
データベース

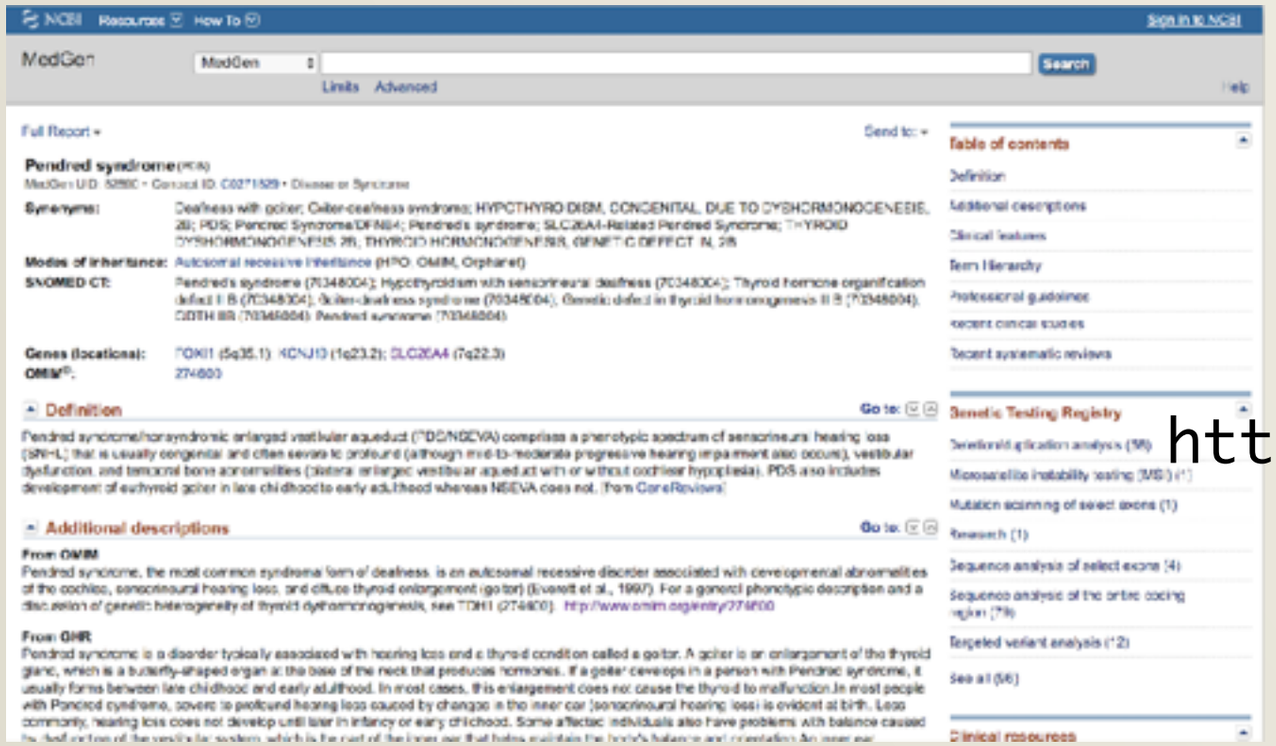
Johns Hopkins 大学

26256 entries

MedGen

一括処理に便利

閲覧にも便利?



https://www.ncbi.nlm.nih.gov/medgen/

疾患（表現型）と遺伝子の
統合的なデータベース

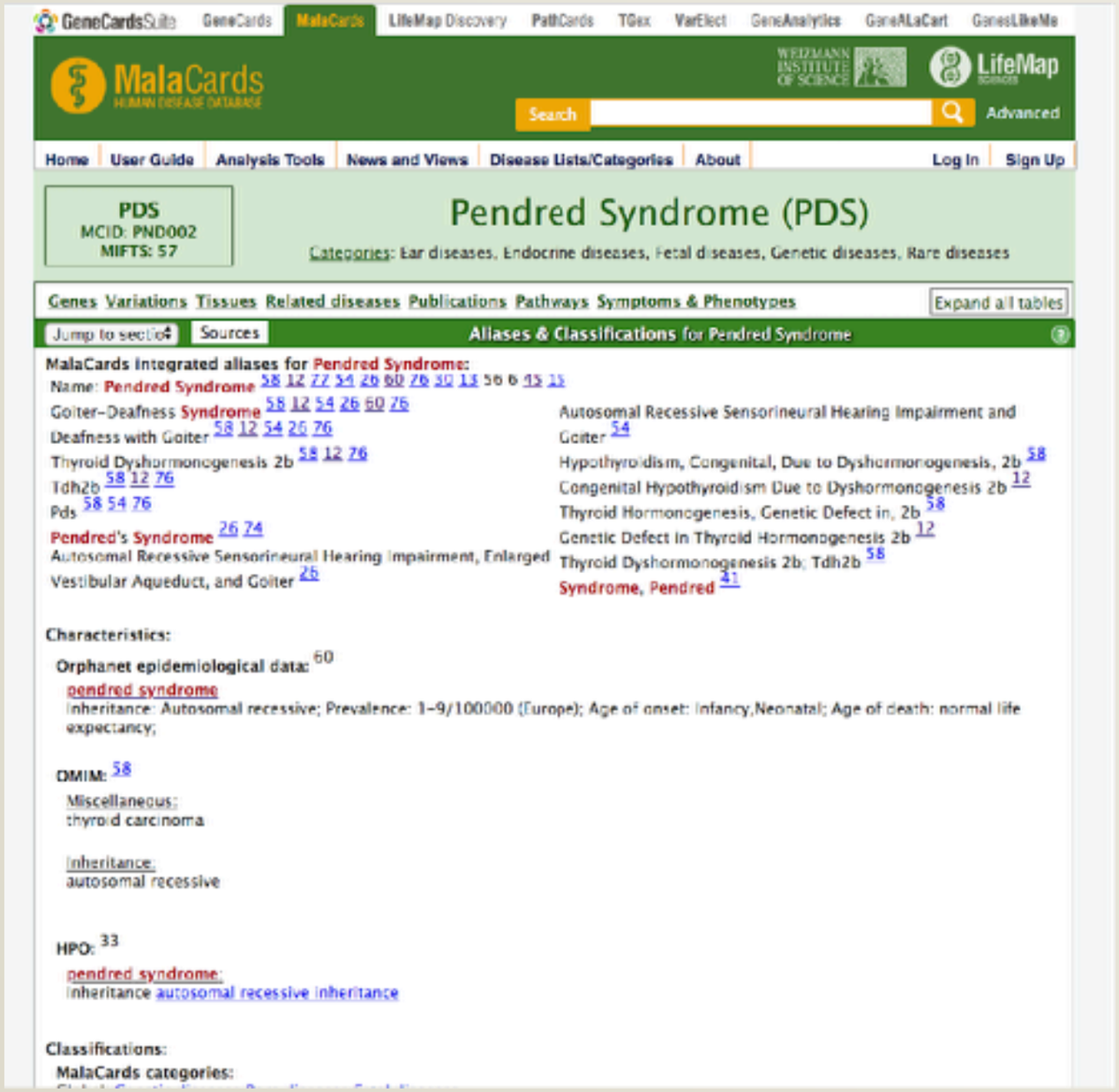
by NCBI

30937 entries

FTPサイトにデータが置かれている

他データベースのエントリとの
変換テーブルも作成されている
(OMIM, Gene など)

MalaCards



https://www.malacards.org/

疾患（表現型）と遺伝子の
統合的なデータベース

by Weizmann Institute of Science
(Israel)

19941 entries

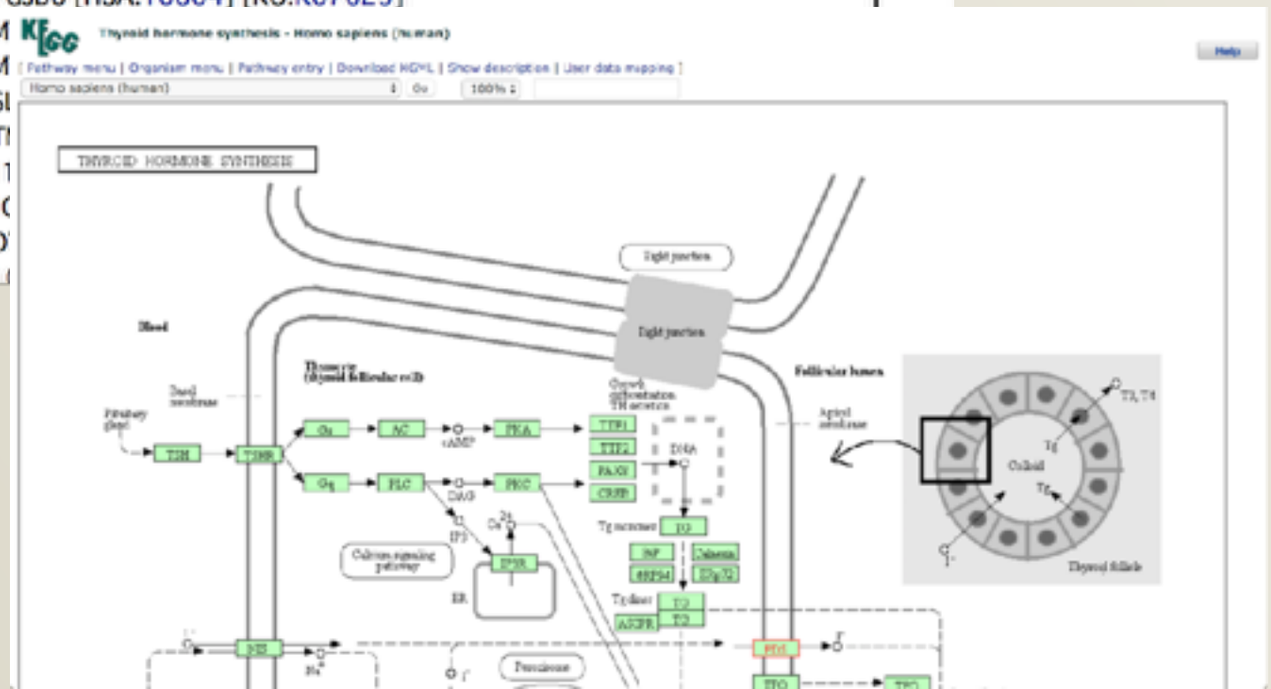
GeneCardのシリーズ

KEGG Disease

KEGG DISEASE: 遺伝性難聴 (常染色体劣性)

Help

エントリ	H00605
名称	遺伝性難聴 (常染色体劣性)
概要	遺伝性難聴は、症候型(様々な他の異常を伴う難聴)と非症候型に分けられる。非症候型が遺伝性症例の 70%, 症候型が 30% を占める。遺伝形式は、常染色体劣性遺伝形式が最もよく見られ (非症候型難聴症例の 75%-85%), 次いで常染色体優性遺伝形式 (12-13%) とX-連鎖性やミトコンドリア型 (2-3 %) が続く。常染色体劣性遺伝性難聴は、一般的に最も重症で、専ら蝸牛の欠陥(感音声難聴)に起因する。症候型ではほとんどの症例で 伝導性難聴 (外耳、中耳の発達異常)または混合型であるのと対照的である。
カテゴリ	神経系疾患
階層分類	ヒト疾患 [BR:jp08402] 神経系疾患 耳疾患 H00605　遺伝性難聴 (常染色体劣性) ICD-11 による疾患分類 [BR:jp08403] 10 Diseases of the ear or mastoid process Disorders with hearing impairment AB50　Congenital hearing impairment H00605　遺伝性難聴 (常染色体劣性) BRITE hierarchy
病因遺伝子	(DFNB1A) GJB2 [HSA:2706] [KO:K07621] (DFNB1A) GJB3 [HSA:2707] [KO:K07622] (DFNB1B) GJB6 [HSA:10804] [KO:K07625] (DFNB2) M (DFNB3) M (DFNB4) SI (DFNB6) TI (DFNB7/11 (DFNB8/10 (DFNB9) O (DFNB12)



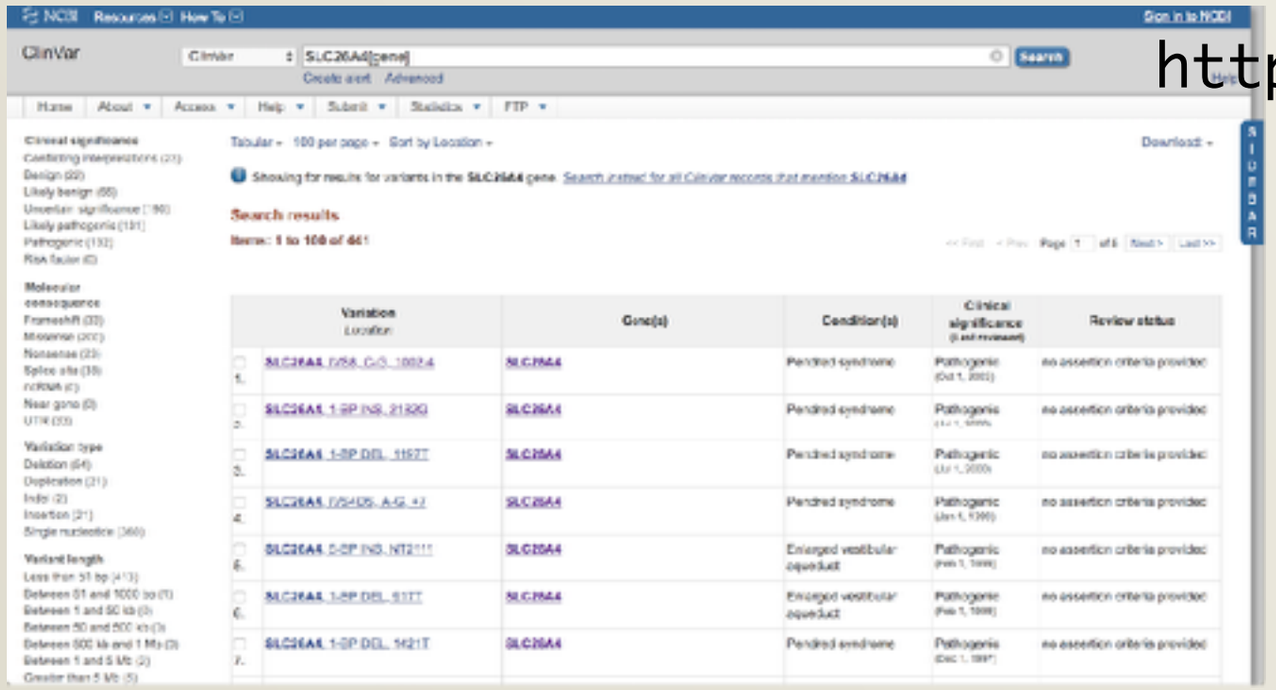
https://www.genome.jp/kegg/disease/disease_ja.html

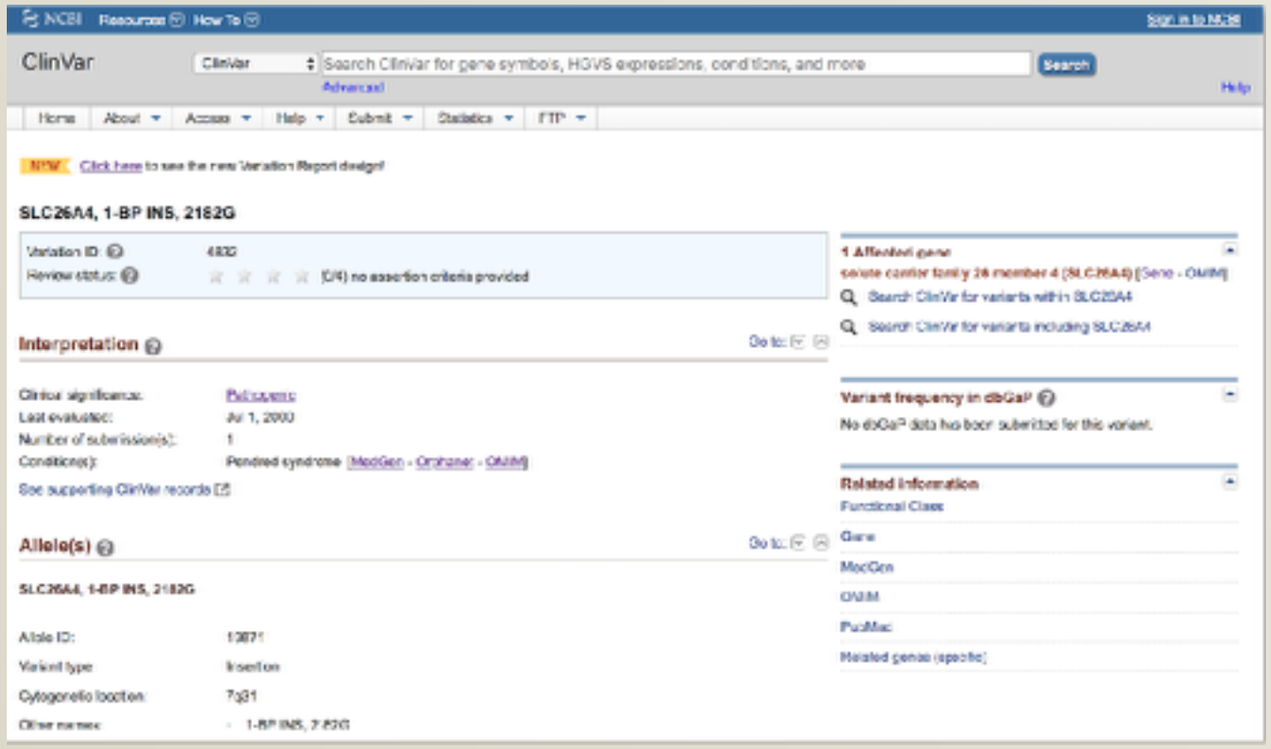
pathwayから派生した疾患のデータベース

（疾患 → 遺伝子 →pathway）

by 京都大・金久研究室

ClinVar





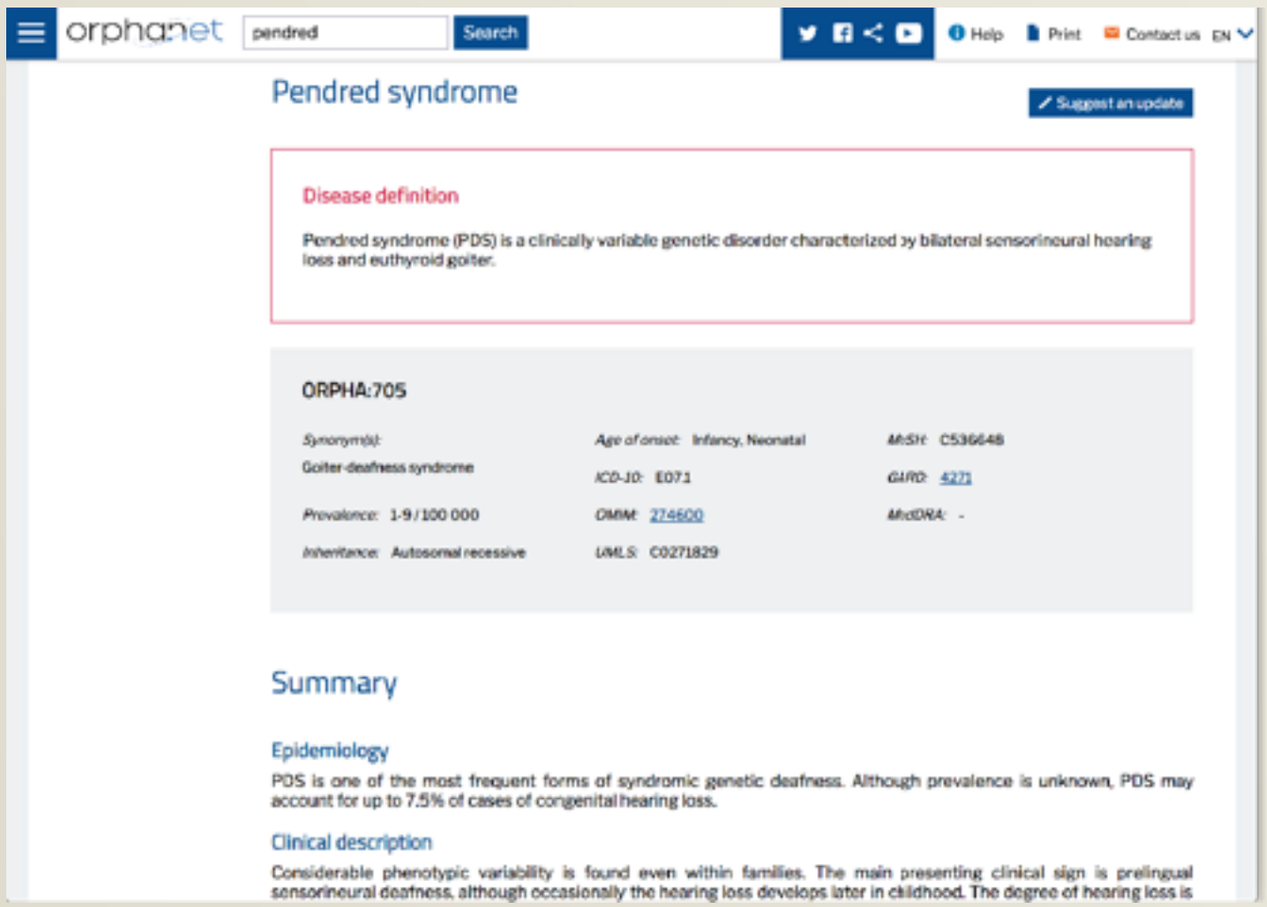
https://www.ncbi.nlm.nih.gov/clinvar/

ヒトvariationと疾患（表現型）のデータベース

by NCBI

509259 entries

Orphanet



https://www.orpha.net/

希少疾患とorphaned drugのデータベース

Orphanet コンソーシアム
（仏・Insermが中心）

5856 diseases and
3573 genes

PubCaseFinder

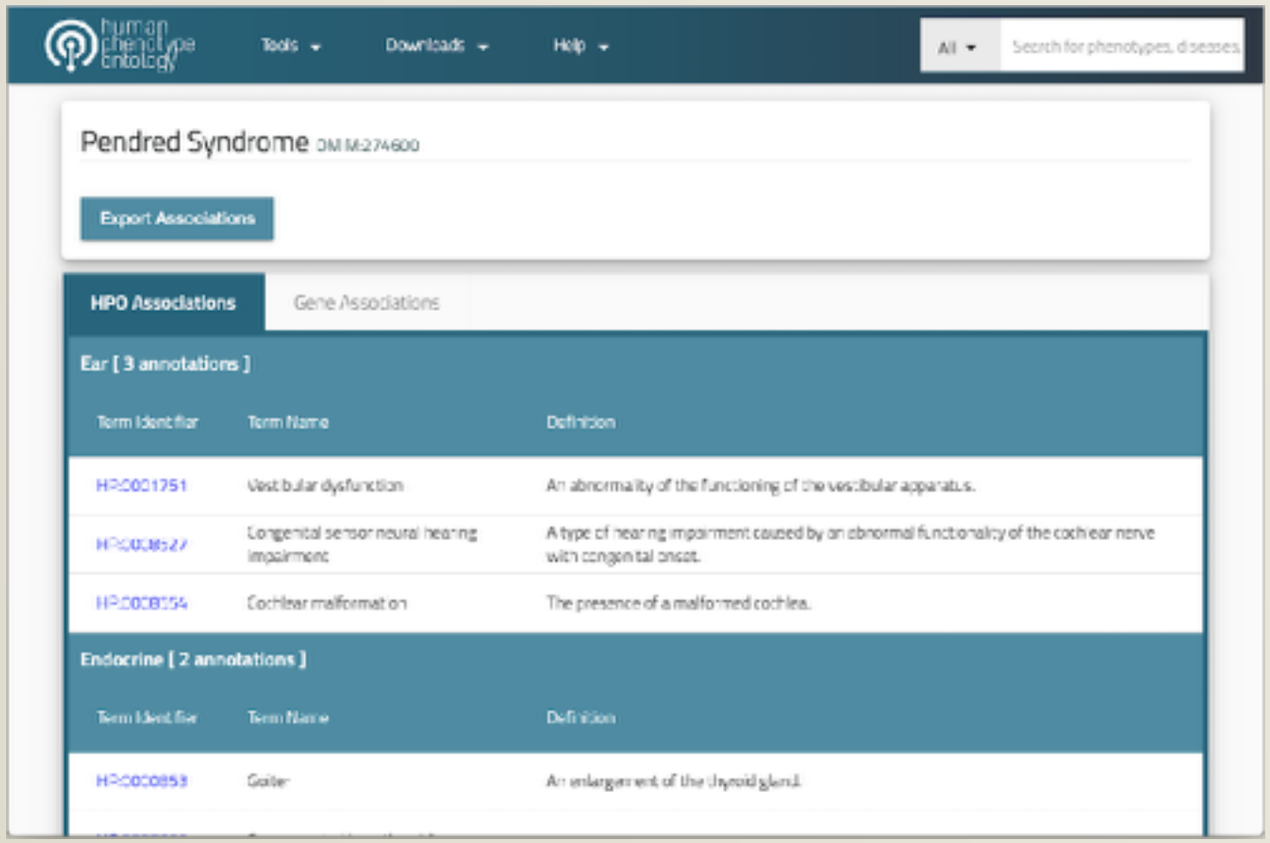
https://pubcasefinder.dbcls.jp/

症状から希少疾患を検索する
ウェブサービス

by DBCLS

4066 rare diseases,
4002 disease-gene associations,
1M+ case reports

HPO (Human Phenotype Ontology)

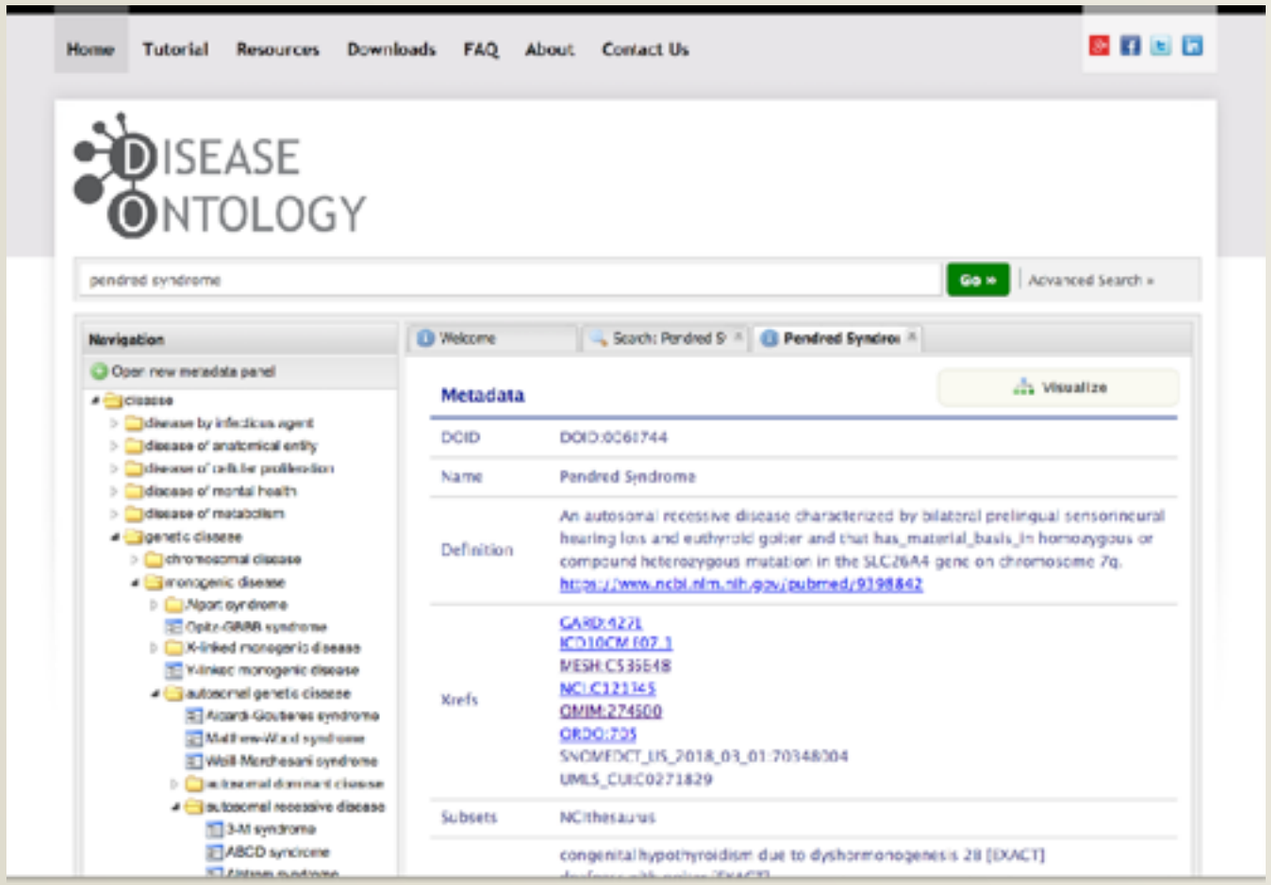


https://hpo.jax.org/app/

ヒト表現型のオントロジー

遺伝子や疾患に対して
対応づけられた情報も

DO (Disease Ontology)



http://disease-ontology.org/

ヒト疾患のオントロジー

疾患データベースの活用例：Gendoo

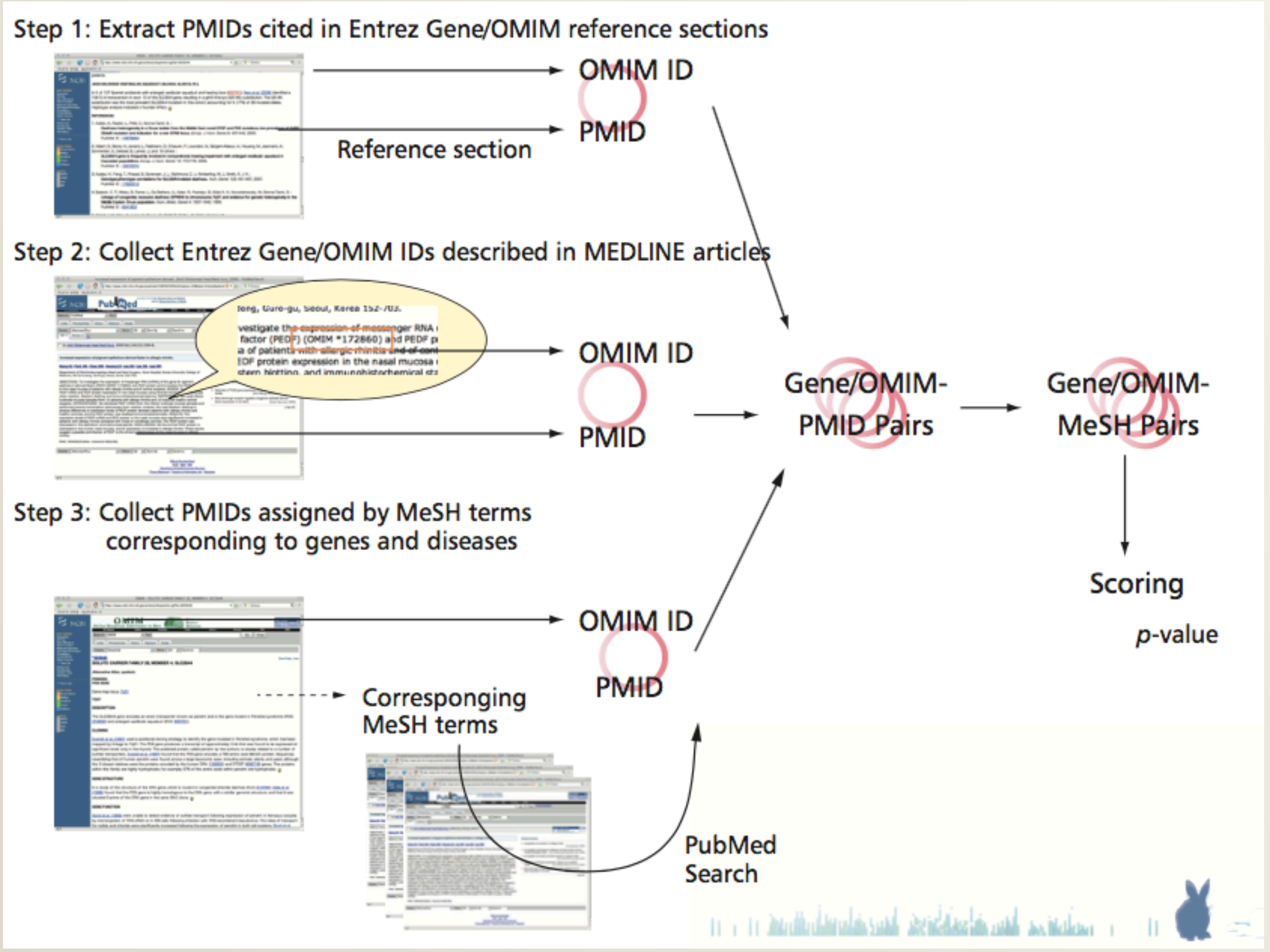
Disease

→ Tree View

	MeSH term	Japanese	Link	01	02
☐	Diabetes Mellitus, Type 1	糖尿病-1型	MeSH LSDB		
☐	Diabetes Mellitus, Type 2	糖尿病-2型	MeSH LSDB		
☐	Insulin Resistance	インスリン抵抗性	MeSH LSDB		
☐	Diabetes Mellitus	糖尿病	MeSH LSDB		
☐	Genetic Predisposition to Disease	遺伝的素因(疾患)	MeSH LSDB		
☐	Obesity	肥満	MeSH LSDB		
☐	Disease Susceptibility	疾病感受性	MeSH LSDB		
☐	Diabetes Mellitus, Experimental	糖尿病-実験的	MeSH LSDB		
☐	Hyperglycemia	高血糖症	MeSH LSDB		
☐	Prediabetic State	前糖尿病状態	MeSH LSDB		
☐	Syndrome	症候群	MeSH LSDB		
☐	Body Weight	体重	MeSH LSDB		
☐	Glucose Intolerance	耐糖能障害	MeSH LSDB		
☐	Autoimmune Diseases	自己免疫疾患	MeSH LSDB		
☐	Pancreatic Diseases	膵臓疾患	MeSH LSDB		

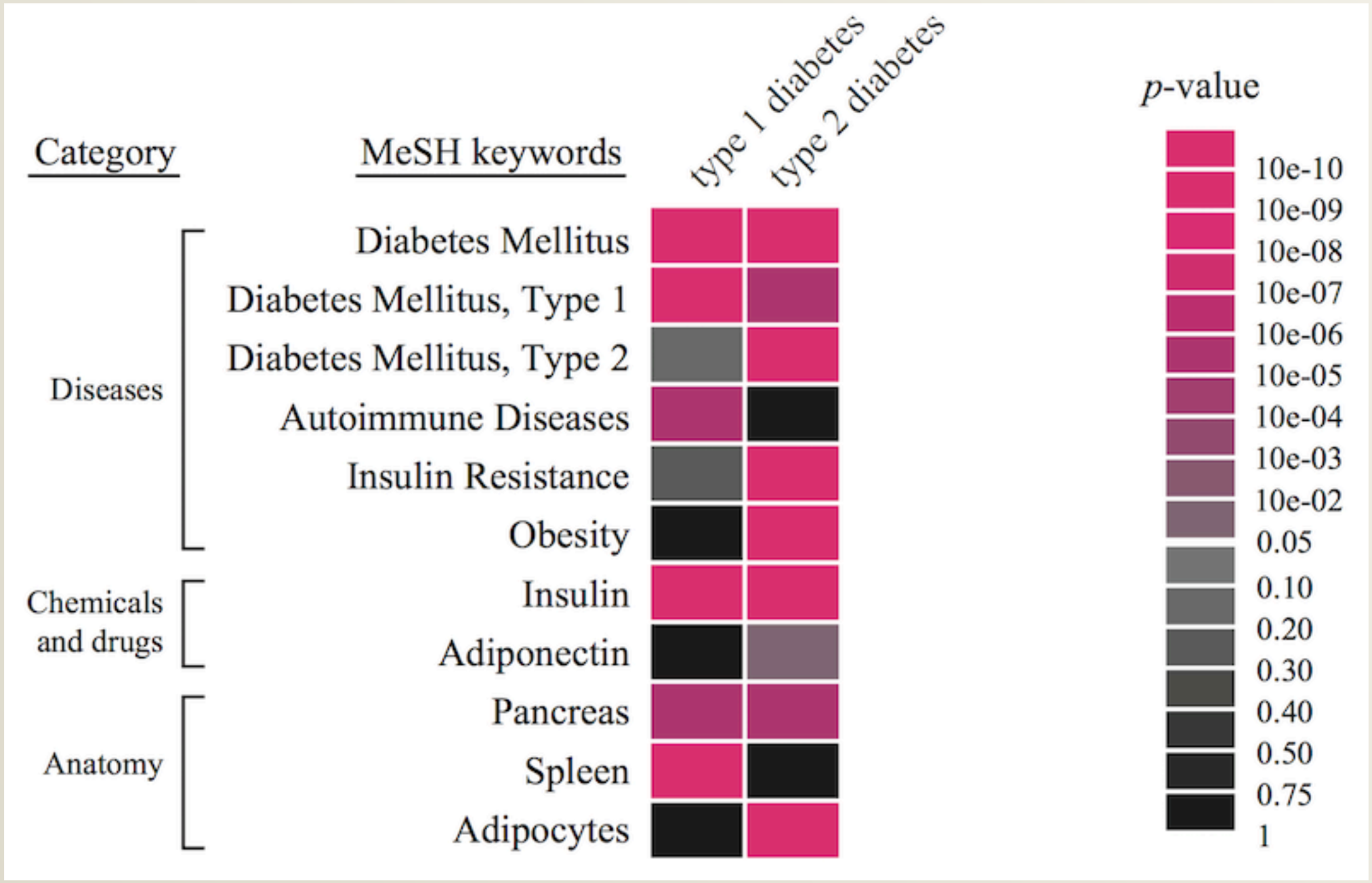
疾患をMeSHキーワードを用いてプロファイリング

データ作成の流れ


Step 1: Extract PMIDs cited in Entrez Gene/OMIM reference sections
OMIM ID
PMID
Reference section
Step 2: Collect Entrez Gene/OMIM IDs described in MEDLINE articles
OMIM ID
PMID
Gene/OMIM-PMID Pairs
Gene/OMIM-MeSH Pairs
Scoring
p-value
Step 3: Collect PMIDs assigned by MeSH terms corresponding to genes and diseases
OMIM ID
PMID
Corresponging MeSH terms
PubMed Search

疾患データの外部データとの連携例

例：1型/2型糖尿病のプロファイルの違い



別の例：疾患関連NGSデータの検索

Disease Type: ANY　Search

Total: 303 << first < prev **1** 2 3 4 5 6 7 8 9 10 next > last >> 10

Disease	疾患名	# of submission
Breast Neoplasms	乳房腫瘍	43
Prostatic Neoplasms	前立腺腫瘍	22
Disease Models, Animal	疾患モデル(動物)	21
Genetic Predisposition to Disease	遺伝的素因(疾患)	
Disease Progression	病勢悪化	
Translocation, Genetic	転座	
Cell Transformation, Neoplastic	腫瘍細胞形質転換	
Lung Neoplasms	肺腫瘍	
Staphylococcal Infections	ブドウ球菌感染症	
Malaria	マラリア	

Total: 303 << first < prev **1** 2 3 4 5 6 7

Total: 6 << first < prev **1** next > last >> 10

SRA ID	SRA Title	Disease	疾患名	PMID
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18987736
SRA026055	DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18987736
SRA026055	Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome	Leukemia, Myeloid, Acute	白血病-急性骨髄性	19657110
SRA009897	In-depth characterization of the microRNA transcriptome in a leukemia progression model	Leukemia, Myeloid, Acute	白血病-急性骨髄性	18849523
SRA029797	Exome Sequencing Identifies Somatic Mutations in Acute Monocytic Leukemia	Leukemia, Myeloid, Acute	白血病-急性骨髄性	21399634

Total: 6 << first < prev **1** next > last >> 10

参考リソース

参考図書・その1 ～ ウェブツール使い方集


2点すべてのイメージを見る

生命科学データベース・ウェブツール 図解と動画で使い方がわかる! 研究がはかどる定番18選 単行本（ソフトカバー） – 2018/12/4

坊農秀雅 (編集), 小野浩雅 (編集)

カスタマーレビューを書きませんか？

› その他 () の形式およびエディションを表示する

単行本（ソフトカバー）
￥ 3,024

￥ 5,362 より 4 中古品の出品
￥ 3,024 より 3 新品

5/25 土曜日 にお届けするには、今から**21 時間 4 分**以内にお急ぎ便を選択して注文を確定してください（Amazonプライム会員は無料 詳細を見る

まとめ買いで 最大10% ポイント還元

【2冊で最大4%、3冊以上で最大8%、10冊以上で最大10%】ポイント還元

2冊を購入する際クーポンコード**「2BOOKS」**を、3冊以上は**「MATOME」**を入力すると最大8~10%ポイント還元！今すぐチェック

シェアする

￥ 3,024
ポイント: 30pt (1%)
詳細はこちら

通常配送無料 詳細

残り13点（入荷予定あり） 在庫状況について
この商品は、Amazon.co.jp が販売、発送します。

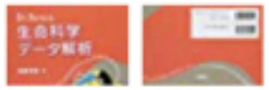
数量: 1

カートに入れる

今すぐ買う

時間を節約
1-Click注文を有効にすると、すぐに注文できます。
詳細はこちら

参考図書・その2 ～ 初心者向けバイオインフォ本



2点すべてのイメージを見る

Dr. Bonoの生命科学データ解析 単行本 – 2017/9/29

坊農秀雅 (著)

★★★★★ 3件のカスタマーレビュー

› その他 () の形式およびエディションを表示する

単行本
¥ 3,240

¥ 4,205 より 8 中古品の出品
¥ 3,240 より 4 新品

5/25 土曜日 にお届けするには、今から21 時間 4 分以内にお急ぎ便を選択して注文を確定してください（Amazonプライム会員は無料 詳細を見る

まとめ買いで 最大10% ポイント還元

【2冊で最大4%、3冊以上で最大8%、10冊以上で最大10%】ポイント還元

2冊を購入する際クーポンコード**「2BOOKS」**を、3冊以上は**「MATOME」**を入力すると最大8~10%ポイント還元！今すぐチェック

シェアする

¥ 3,240
ポイント: 96pt (3%)
詳細はこちら

通常配送無料 詳細

残り10点（入荷予定あり） 在庫状況について
この商品は、Amazon.co.jp が販売、発送します。

数量: 1

カートに入れる

今すぐ買う

時間を節約
1-Click注文を有効にすると、すぐに注文できます。
詳細はこちら