

仲里 猛留

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 (DBCLS)

# Motivation : 我々は疾患について詳しくは知らない

## 【新着論文レビューから抽出した疾患名の数々】

Duchenne型筋ジストロフィー

放射線誘導性腸線維症

パーキンソン病

T細胞性急性リンパ性白血病

急性骨髄性白血病

筋萎縮性側索硬化症

肝癌

骨髄異形成症候群

Hajdu-Cheney症候群

MIRAGE症候群

織毛病

Cushing症候群

家族性乳がん

数々の疾患研究が行われているが、  
(特に医学教育を受けていない) 我々には  
聞いたこともないような疾患が数々ある。

また、疾患についての情報は  
症状については多く見つかるが  
その分子的な機序 (や生命現象との関わり) は  
(特にバイオインフォマティクス解析に使える形では)  
得られにくい。

→ 疾患情報をバイオインフォマティクスで使える形で  
(しかも遺伝子の世界と同じ用語で) 表現できるようにしたい

# MeSH (Medical Subject Headings)

トーゴの日シンポジウム2020

特徴をプロファイリングするのにMeSHキーワードを用いた

MeSH (Medical Subject Headings)

PubMed収載の論文をindexingするための用語集

23,000語

15分野 (anatomy, taxonomy, chemical & drugs, disease, ...)

階層構造により用語を整理

NLM (Natl. Lib. Med) が管理

<https://www.ncbi.nlm.nih.gov/mesh>

The screenshot shows a PubMed article page. At the top, there is a search bar with the text 'Nakazato takenu' and a 'Search' button. Below the search bar, the article title 'Mechanism of acid adaptation of a fish living in a pH 3.5 lake' is displayed. The authors listed are Taku Hirata, Toyoji Kaneko, Toshihiro Ono, Takeru Nakazato, Norihisa Furukawa, Sanae Hasegawa, Shigeo Wakabayashi, Munekazu Shigekawa, Min-Hwang Chang, Michael F. Romero, and Shigehisa Hirose. The abstract text is visible, starting with 'Despite unfavorable conditions, a single species of fish, Osorezan dace, lives in an extremely acidic lake (pH 3.5) in Osorezan, Aomori, Japan. Physiological studies have established that this fish is able to prevent acidification of its plasma and loss of Na<sup>+</sup>. Here we show that these abilities are mainly attributable to the chloride cells of the gill, which are arranged in a follicular structure and contain high concentrations of Na<sup>+</sup>-K<sup>+</sup>-ATPase, carbonic anhydrase II, type 3 Na<sup>+</sup>-H<sup>+</sup> exchanger (NHE3), type 1 Na<sup>+</sup>-HCO<sub>3</sub><sup>-</sup> cotransporter, and aquaporin-3, all of which are upregulated on acidification. Immunohistochemistry established their chloride cell localization, with NHE3 at the apical surface and the others localized to the basolateral membrane. These results suggest a mechanism by which Osorezan dace adapts to acidic conditions.'

The screenshot shows the MeSH terms and substances for the article. The MeSH terms are listed in a hierarchical structure, starting with 'Adaptation, Physiological\*' and including 'Animals', 'Aquaporin 3', 'Carbonic Anhydrase II / genetics', 'Cloning, Molecular', 'Enzyme Induction', 'Fishes / physiology\*', 'Fresh Water / chemistry\*', 'Gills / enzymology', 'Hydrogen-ion Concentration', 'Japan', 'Kidney / enzymology', 'Oocytes', 'RNA, Messenger / genetics', 'RNA, Messenger / metabolism', 'Sodium / metabolism', 'Sodium-Bicarbonate Symporters / genetics', 'Sodium-Bicarbonate Symporters / metabolism', 'Sodium-Hydrogen Exchangers / genetics', 'Sodium-Hydrogen Exchangers / metabolism', 'Sodium-Potassium-Exchanging ATPase / genetics', 'Up-Regulation', and 'Xenopus laevis'. The substances listed are 'Aquaporin 3', 'RNA, Messenger', 'Sodium-Bicarbonate Symporters', 'Sodium-Hydrogen Exchangers', 'Aquaporin 3', 'Sodium', and 'Carbonic Anhydrase II'.

Licensed under a Creative Commons 表示4.0国際ライセンス

©2020仲里猛留 (ライフサイエンス統合データベースセンター)



Step 1: Extract PMIDs cited in Entrez Gene/OMIM reference sections



Reference section

OMIM ID

PMID

Step 2: Collect Entrez Gene/OMIM IDs described in MEDLINE articles



OMIM ID

PMID

Step 3: Collect PMIDs assigned by MeSH terms corresponding to genes and diseases



Corresponding MeSH terms

OMIM ID

PMID

PubMed Search

Gene/OMIM-  
PMID Pairs

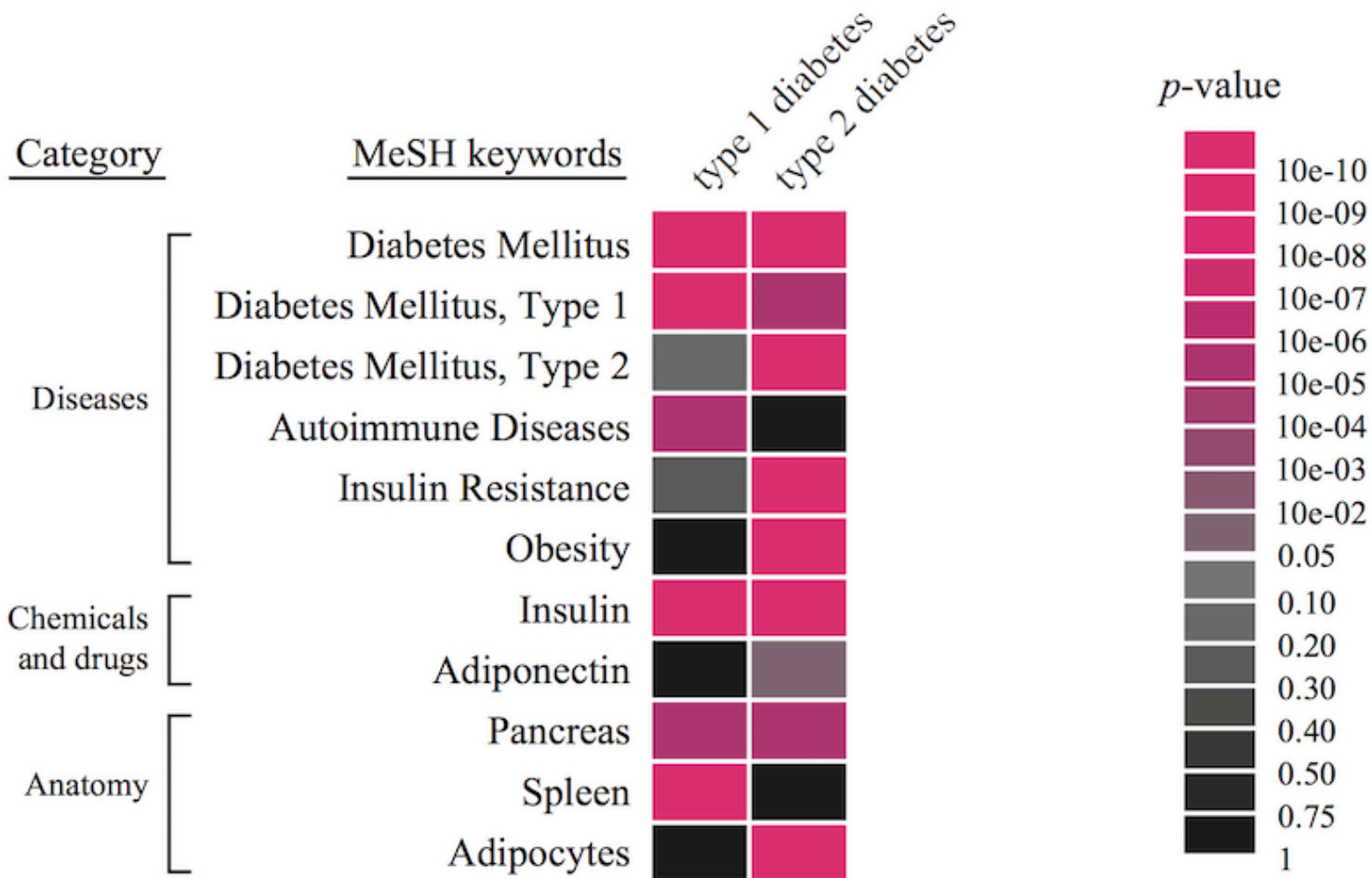
Gene/OMIM-  
MeSH Pairs

Scoring

p-value

MeSH自体は文献に付与されたキーワードなので各遺伝子に対して関連文献を収集し、そこから対応するMeSHを抽出することでGene-MeSHペアを作成している。

→ 以前、OMIMの疾患でプロファイリングも試みた



## 1型/2型糖尿病のプロファイル例

1型糖尿病は免疫系疾患なので自己免疫疾患や脾臓が、2型糖尿病は代謝系疾患なので肥満や脂肪細胞のスコアが高い。

ウェブサービスとしてGendooを公開  
(<http://gendoo.dbcls.jp/>)

ICD-10	電子カルテなどで使われる疾患名リスト。逆に疾患名リストしかない
OMIM	ヒトの疾患と遺伝子のリスト。階層構造になっていない。項目が少し特殊なものも。
MedGen	OMIMがNCBIからJohn Hopkins大に管理が戻ったので、NCBIがつくったデータベース。
ClinVar	ヒトvariationと疾患（表現型）のデータベース by NCBI
MONDO	さまざまな疾患関連のontologyをまとめたもの
HPO (Human Phenotype Ontology)	ヒトの表現型をまとめたデータベース
DO (Disease Ontology)	ヒト疾患のontology
KEGG Disease	pathwayから派生した疾患のデータベース（疾患 → 遺伝子 → pathway）
Orphanet	希少疾患と orphaned drugのデータベース

→ 他の疾患IDからも特徴プロファイルを確認できるように作業中

# エンリッチメント解析に向けて

トーゴの日シンポジウム2020

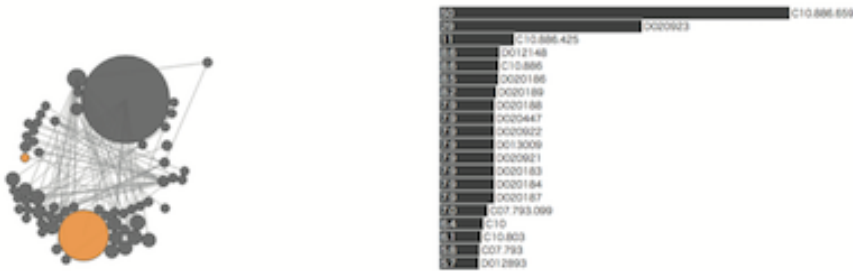
Gendoo 2

215,1780,6729,154810,2521,7691,10815,57147

MeSH Co-occurrence Rate Ranking



Pagerank - Disease Categories



既存のGendooでは、入力の遺伝子・疾患と個々のプロファイルを関連度により並べるにすぎなかった。これを実際に解析シーンで利用しやすくするべく（複数の遺伝子・疾患を入力に、など）エンリッチメント解析を行うためにウェブサイトを構築中である。