

トーゴの日シンポジウム2020

データベース整備とデータ統合

国立研究開発法人 医薬基盤・健康・栄養研究所
AI健康・医薬研究センター

大阪大学 蛋白質研究所



水口 賢司

大阪大学 蛋白質研究所
Institute for Protein Research, Osaka University



<https://mizuguchilab.org>

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

©2020水口 賢司(医薬基盤・健康・栄養研究所 AI健康・医薬研究センター／大阪大学 蛋白質研究所)



Sagace

創薬・疾患研究のための生命科学分野のデータベースの横断検索サービス

<https://sagace.nibiohn.go.jp/>



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

©2020水口 賢司(医薬基盤・健康・栄養研究所 AI健康・医薬研究センター/大阪大学 蛋白質研究所)



すべて

- 生物資源 (87)
- 特許 (6594)
- 実験プロトコル (8)
- 文献・用語集・辞書 (15933)
- その他のデータベース (28655)

研究関連の信頼出来るデータベースからの情報

- 植物 (319)
- 微生物 (319)
- 特定生物なし (18680)

すべて

- ゲノム・遺伝子 (24024)
- タンパク質 (1035)
- 糖鎖 (30)
- その他の生体分子 (38)
- 化合物 (57)
- パスウェイ・ネットワーク (1889)
- 細胞・組織 (895)
- 個体 (7)
- その他 (23487)

悪性リンパ腫の検索結果: 55645 hits 表示オプション: 同義語展開あり (デフォルト)

KEGG DRUG / Toralizumab (USAN/INN)

<http://www.genome.jp/> KEGG DRUG (Drugs viewed as perturbants to the molecular system)
 その他のデータベース | ヒト | 化合物
 DRUG: D06193 Entry D06193Drug Name Toralizumab (USAN/INN) Activity Treatment of antibody-mediateation], and B-cell malignancies [such as CLL/small lymphoma, follicular cell lymphoma grade I or II, marginal

Search Result

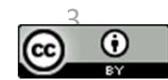
<http://alterna.cbrc.jp/> ASTRA(Alternative Splicing and Transcription Archives)
 その他のデータベース | ヒト, 動物 (ヒト以外), 植物, 微生物 | ゲノム・遺伝子
 ASTRA Home Statistics Human Mouse D. melanogaster C. elegans A. thaliana O. sativa Help Links Se3040355 9 e2:1 viewer Hs.94181 Homo sapiens T-cell lymphoma breakpoint associated target 1, mRNA (cDNA clone M

KW-0761 - 希少疾病用医薬品・希少疾病用医療機器

<http://www.nibio.go.jp/> 希少疾病用医薬品・希少疾病用医療機器
 その他のデータベース | ヒト | その他
 ID: (22薬)第232号
 医薬品 22 H22.8.11 (22薬)第232号 1、3※27 KW-0761 CCR4陽性の成人T細胞白血病リンパ腫※27 協和発酵キリン(株) 再発又は難治性のCCR4陽性の成人T細胞白血病リンパ腫※27 協和発酵キリン(株) H24.3.30※27 ポテリジオ点滴静

いむーの:bcl-2

<http://immuno.med.kobe-u.ac.jp/> 免疫組織データベースいむーの
 文献・用語集・辞書 | ヒト | 細胞・組織
 bcl-2岡山大学病理病態学教室 吉野 正メーカー DAKOおよびNOVO castra (をmixture) 動物種 マウス単クロ心を除くB細胞、T細胞など用途その他:bcl2遺伝子は濾胞性リンパ腫のt(14;18)(q32; q21)の18番染色体切断点近傍からクロー



オントロジー、ID

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

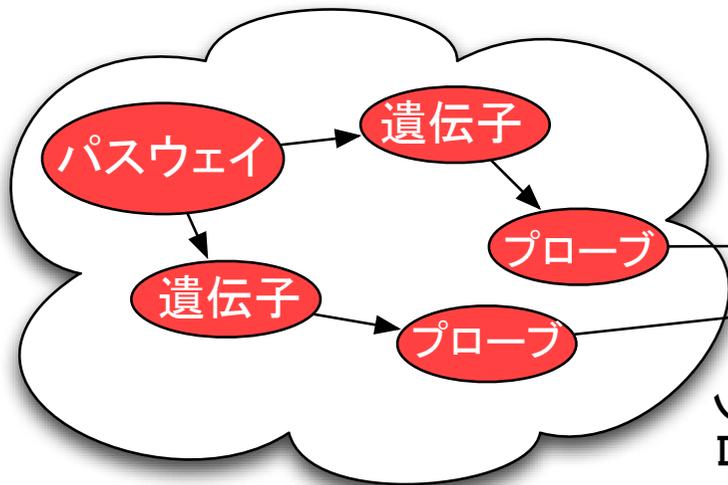
©2020水口 賢司(医薬基盤・健康・栄養研究所 AI健康・医薬研究センター／大阪大学 蛋白質研究所)



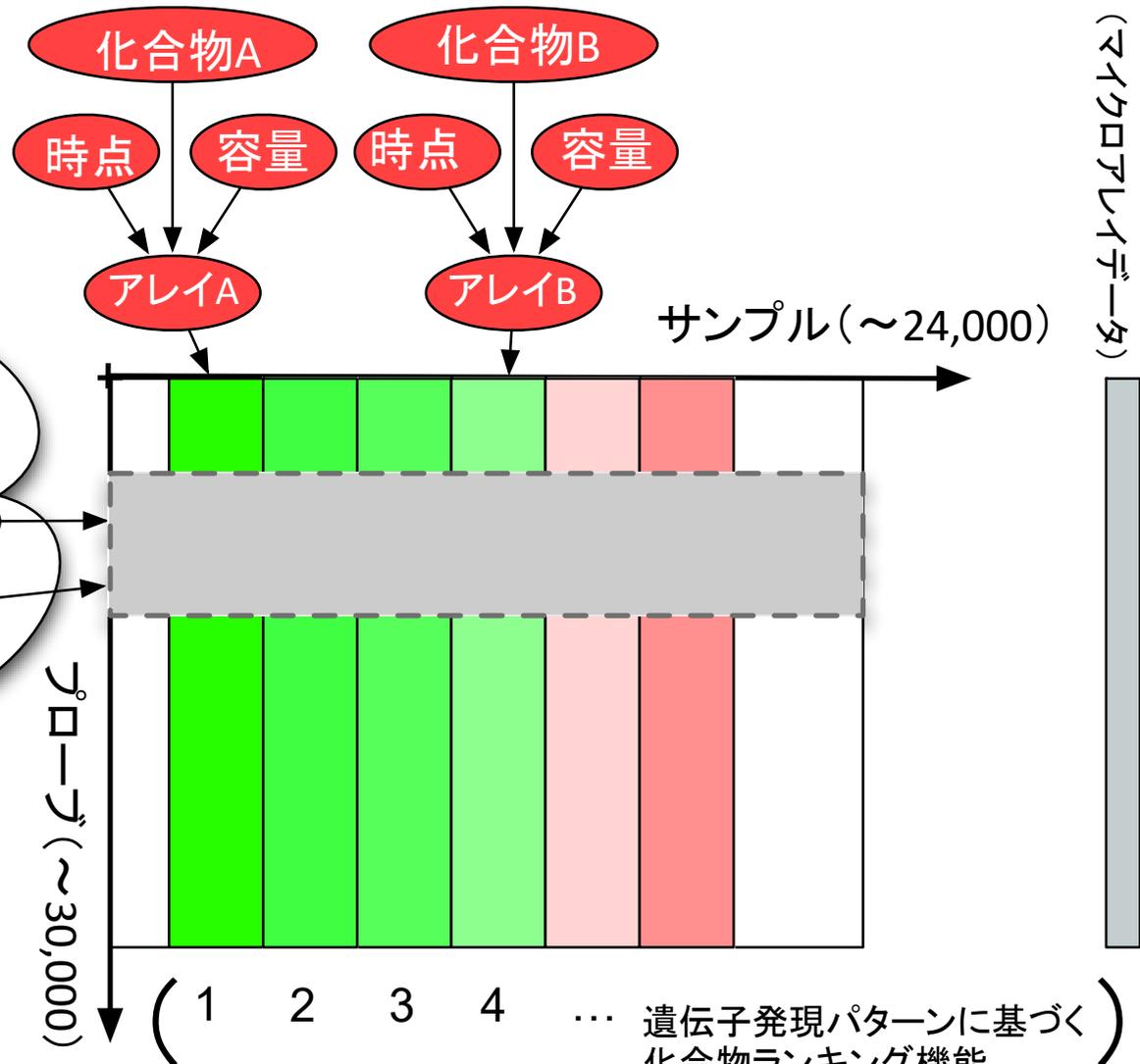
Toxygates

<http://toxygates.nibiohn.go.jp/>

インターネット上の公共データ
遺伝子, タンパク質, パスウェイ...



セマンティックウェブ技術を用いたリアルタイムの取得



プローブ (~30,000)

大規模データ行列

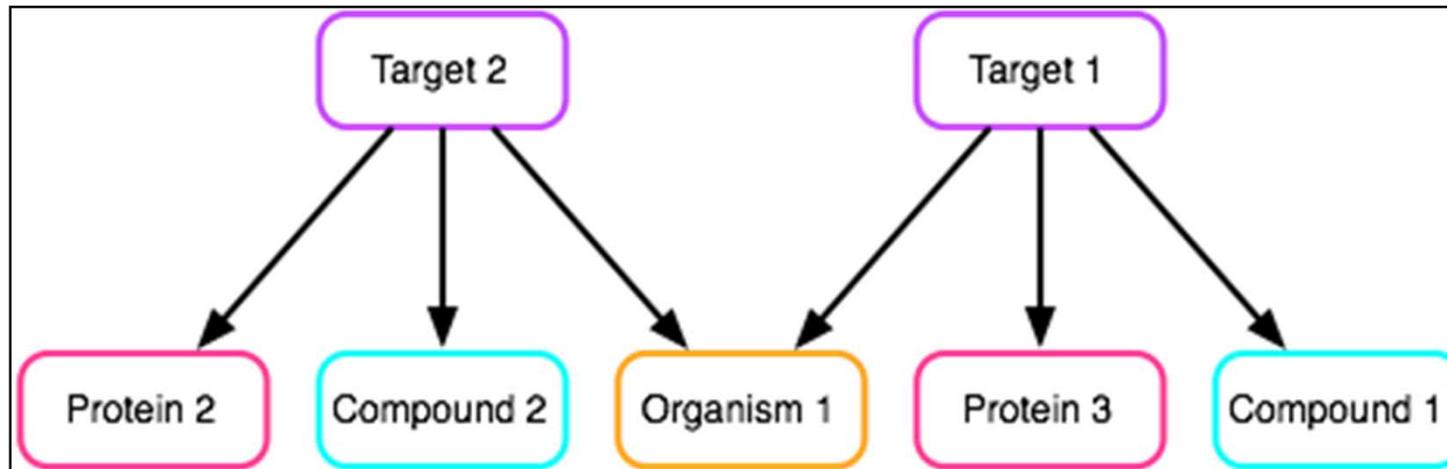
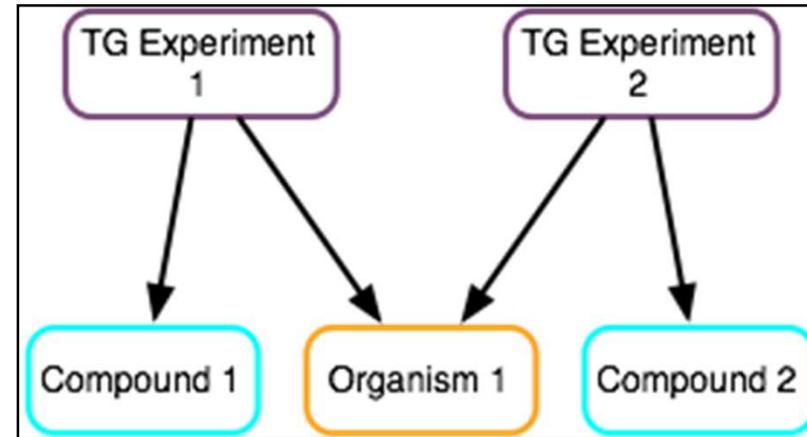
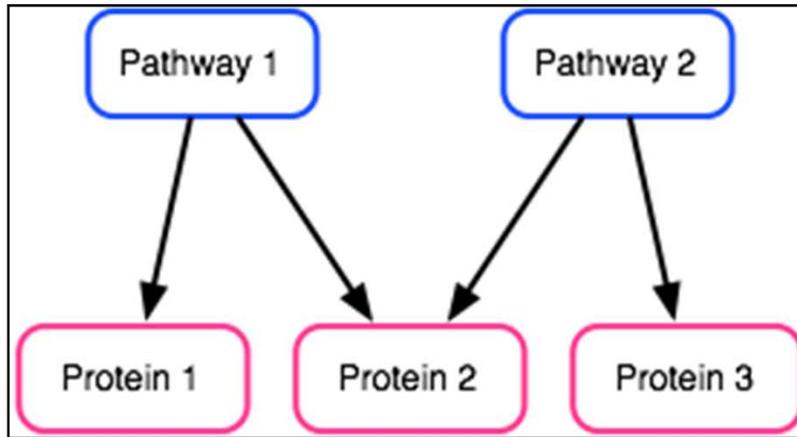
Nystrom et al., *Bioinformatics* (2013)

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

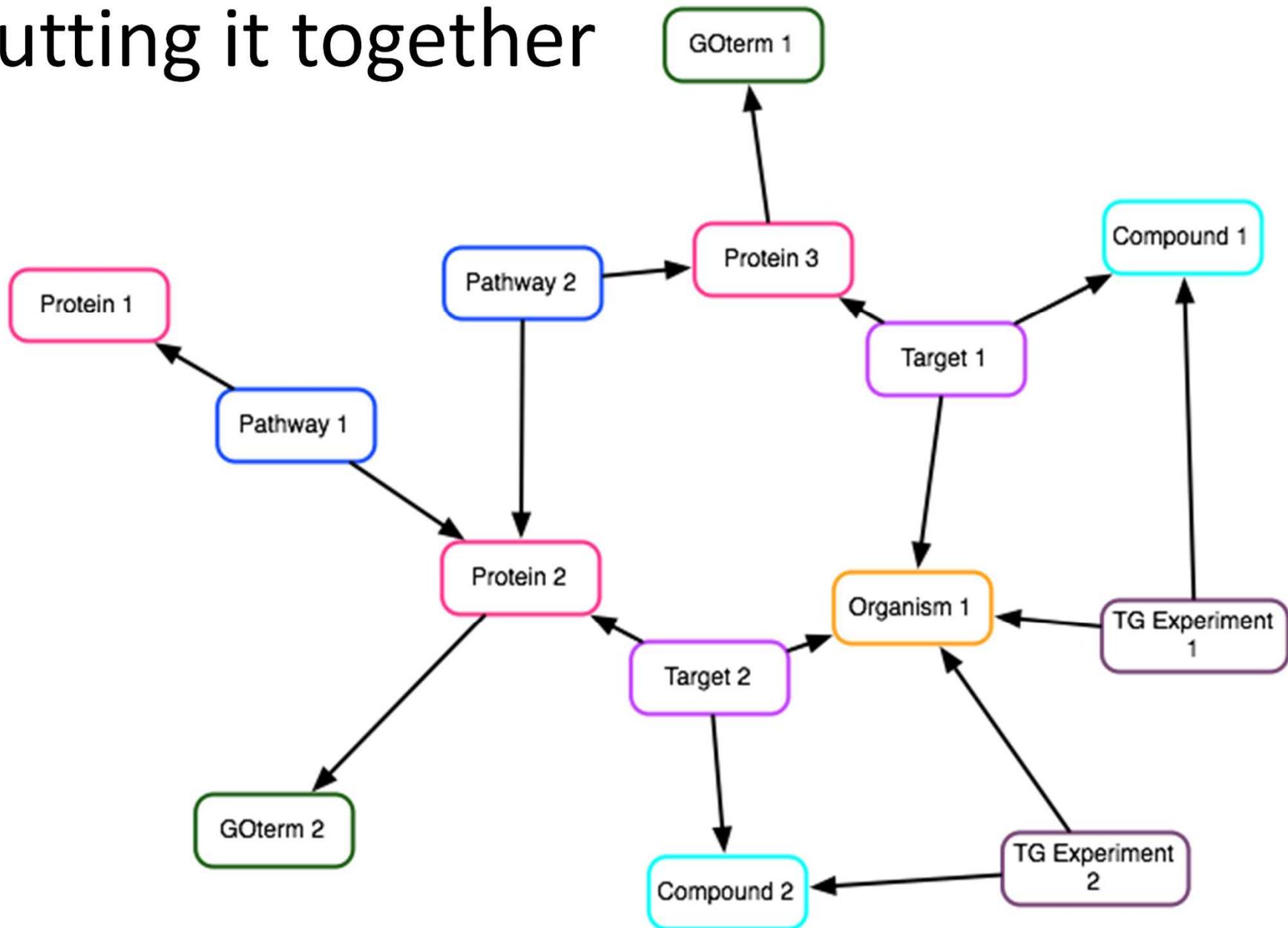
©2020水口 賢司 (医薬基盤・健康・栄養研究所 AI健康・医薬研究センター / 大阪大学 蛋白質研究所)



RDF datasets



Putting it together



DruMAP

(Drug Metabolism and pharmacokinetics Analysis Platform)

<https://drumap.nibiohn.go.jp/>

The screenshot shows the DruMAP website interface. At the top, there is a navigation bar with links for 'DruMAP', 'Home', 'Compound search', 'Activity search', 'New Prediction', and 'About'. Below the navigation bar, the main heading reads 'DruMAP ver.1.0' followed by the subtitle 'Drug Metabolism and pharmacokinetics Analysis Platform'. A warning message states: 'This application may not work on Internet Explorer or Microsoft Edge properly. Please use Google Chrome or Firefox.' Below this, a paragraph describes the platform's purpose: 'DruMAP consists of a database for DMPK parameters and a program that can predict some DMPK parameters based on the chemical structure of a novel compound. The DruMAP database contains data of DMPK parameters from curated public data and newly acquired experimental data obtained by unified conditions. It also contains predicted data using our prediction programs. Users can predict several DMPK parameters at once for novel compounds. Our platform provides a simple method to search and predict DMPK parameters and could contribute to the acceleration of new drug development. DruMAP has three functions.' Three blue buttons are displayed: 'Compound search', 'Activity search', and 'New prediction'. Below the buttons, a text line says 'Current version contains following chemical records.' This is followed by a table with two columns: the first column lists record types, and the second column, titled 'Number of records', shows the corresponding counts.

	Number of records
All registered compounds	30,628
Freebase compounds	27,237
Freebase compounds ignoring stereo structure	25,387

This research is supported by the Drug Discovery Support Promotion Project from Japan Agency for Medical Research and Development, AMED.

 Mizuguchi Laboratory

 NIBIOHN 国立研究開発法人医薬基盤・健康・栄養研究所
National Institutes of Biomedical Innovation, Health and Nutrition

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

©2020水口 賢司 (医薬基盤・健康・栄養研究所 AI健康・医薬研究センター/大阪大学 蛋白質研究所)



Data quality influences prediction accuracy

Public data

Meta data with various attributes

Name	Value	Unit	Type	Description
verapamil	1	mL.min-1.g-1	CL	Intrinsic clearance in mouse microsomes at 1 μ M
verapamil	8	mL.min-1.kg-1	CL	Intrinsic clearance in mouse microsomes by LC-MS/MS

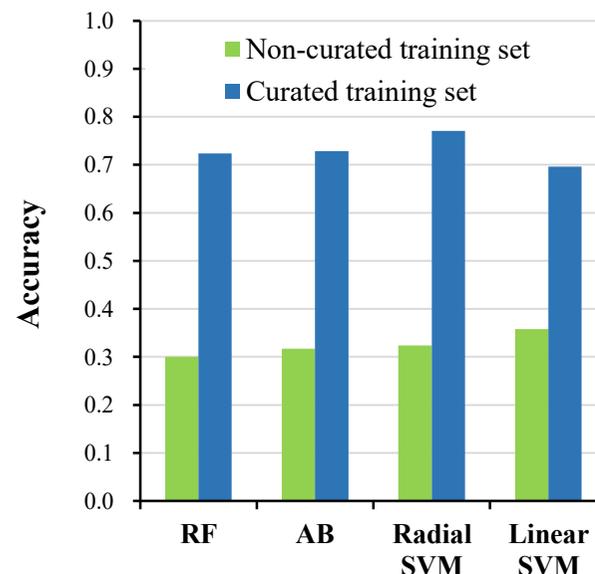
Name	Value	Type	Description
oxprenolol	0.2	LogP	Partition coefficient (logP)
propranolol	-4.68	LogP	Permeability coefficient through artificial membrane
propranolol	1.18	Log P	Partition coefficient (logD7.4)

Necessary to build a database with curated (high quality) information

Prediction model of metabolic stability (*CL_{int}*)

Esaki et al, *Mol. Inf.*, 2019

Dataset
Non-curated training: 6,999
Curated training: 4,567
Test: 586



インタビューフォーム構造化への取り組み

医薬品情報におけるデータ活用の現状と課題

大阪大学蛋白質研究所 医薬基盤・健康・栄養研究所 長尾知生子
医薬基盤・健康・栄養研究所 深川明子
京都大学大学院医学研究科 中津井雅彦・鎌田真由美
ライフサイエンス統合データベースセンター 川島秀一・片山俊明
理化学研究所 安倍理加



PMDA
医薬品情報検索
サイトで得られる
情報のほとんどは、
構造化されていない

学習用データとして
活用するためには
データクレンジングの
必要があり、
それが活用の壁となっている

検索結果一覧で表示する文書を選ぶ

- 添付文書
- 患者向医薬品ガイド/ワクチン接種を受ける人へのガイド
- インタビューフォーム
- 医薬品リスク管理計画 (RMP)
- 改訂指示反映履歴および根拠症例
- 審査報告書/再審査報告書/最適使用推進ガイドライン等
- 重篤副作用疾患別対応マニュアル
- くすりのしおり
- 緊急安全性情報/安全性速報
- 医薬品の適正使用等に関するお知らせ
- 厚生労働省発表資料 (医薬品関連)
- 医薬品に関する評価中のリスク等の情報
- 医薬品添付文書改訂相談に基づく添付文書改訂
- DSU (医薬品安全対策情報)
- PMDA医療安全情報
- 医療用医薬品問合せ先

長尾知生子さんより



RDF (Resource Description Framework) 化

- インタビューフォームIFの構造化をRDFを用いて行った。
- RDFは、W3C(World Wide Web Consortium)標準仕様。特定の企業やOSに依存していない。
- RDFモデルはグラフ構造であり、データの変更・追加などに柔軟に対応。
- Med2RDFなど膨大な公共の医薬・生命科学データベースとシームレスに統合。
- オントロジーで標準化することで、添付文書 (PI) とIFの対応も管理できる。

2019年12月改訂 (第15版)

日本標準品分類番号 87259

医薬品インタビューフォーム
日本病院薬剤師会のIF記載要領2018に準拠して作成

選択的β₂アドレナリン受容体作動性過活動膀胱治療剤
ミラベグロン錠
ベタニス錠25mg
ベタニス錠50mg
Betanis® Tablets 25mg・50mg

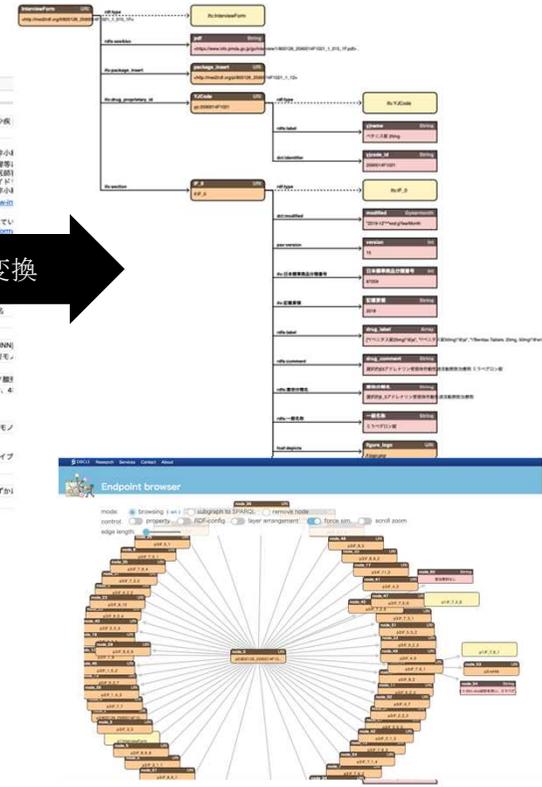
剤形	フィルムコーティング錠
製剤の規制区分	創薬、処方箋医薬品(注意-医師等の処方箋により使用すること)
規格・含量	ベタニス錠 25mg: 1錠中にミラベグロン 25mgを含有する。 ベタニス錠 50mg: 1錠中にミラベグロン 50mgを含有する。
一般名	和名: ミラベグロン(JAN) 洋名: Mirabegron(JAN)
製造販売承認年月日 薬価基準収載年月日 販売開始年月日	製造販売承認年月日: 2011年7月1日 薬価基準収載年月日: 2011年9月12日 販売開始年月日: 2011年9月16日
製造販売(輸入)・ 提携・販売会社名	製造販売: アステラス製薬株式会社
医薬情報担当者の連絡先	
問い合わせ窓口	アステラス製薬株式会社 メディア・カル・インフォメーションセンター TEL:0120-189-371 医療従事者向け情報サイト(Astellas Medical Net) https://amn.astellas.jp/

テンプレート

A	B	C	D	E	F	G	H	I	J	K	L	M	N
IF	タイトル	IF-PI	subject	predicate	object	object (記入例: betans)	object (記入例: optivo)						
47	1.4.4 保険薬用上の留意事項(適)	IF_1_4.4	rdfs:comment										
48	1.5 承認条件及び流通・使用	IF_1_5	rdfs:comment										
49	1.5.1 承認条件	PI_21	IF_1_5.1	rdfs:value									
50	1.5.2 流通・使用上の制限事項	PI_21	IF_1_5.2	rdfs:value									
51													
52	1.6 RMPの概要	PI_21	IF_1_6.2	rdfs:seeAlso									
53		PI_1	IF_1_6	rdfs:seeAlso									
54		PI_1	IF_1_6	rdfs:seeAlso									
55		PI_1	IF_1_6	rdfs:seeAlso									
56		PI_0.7	IF_2_1.1	rdfs:value		ベタニス錠 25mg, ベタニス錠 50mg							
57		PI_0.7	IF_2_1.2	rdfs:value		Betanis Tablets 25mg, Betanis Tablets 50mg							
58		PI_2	IF_2_1.3	rdfs:value		Beta 3 agonist より命名した。							
59	2.1.3 名称の由来	PI_1	IF_2_1.3	rdfs:value									
60	2.2 一般名	PI_2	IF_2_2	rdfs:value									
61	2.2.1 和名(命名)	PI_0.7, PI_11	IF_2_2.1	rdfs:value		ミラベグロン(JAN)							
62	2.2.2 洋名(命名)	PI_0.7, PI_11	IF_2_2.2	rdfs:value		Mirabegron(JAN), mirabegron(INN)							
63	2.2.3 ステム(命名)	PI_2	IF_2_2.3	rdfs:value		β3 アドレナリン受容体作動薬: -begon structure.png							
64	2.2.4 構造式又は示性式	PI_19	IF_2_2.4	rdfs:value									
65	2.4 サイテックス及びサイテ	PI_19	IF_2_4	rdfs:value									
66													
67													
68	2.5 化学(命名法)又は物質	PI_19	IF_2_5	rdfs:value		2-(2-Amino-1,3-thiazol-4-yl)-N-(4-(2-(2R)-2-hydroxy-2-phenyl							
69	2.6 慣用名、別名、略号、記	PI_19	IF_2_6	rdfs:value		白色〜微黄色の結晶又は粉末である。							
70	3 有効成分に関する項目		IF_3	rdfs:value									
71	3.1 物理化学的性質	PI_19	IF_3_1	rdfs:value									
72	3.1.1 外観・性状	PI_19	IF_3_1.1	rdfs:value									
73	3.1.2 溶解性	PI_19	IF_3_1.2	rdfs:value									
74		PI 19	IF 3 1 2	rdfs:table									

自動変換

11品目 (ベタニス、オプジーボ点滴静注、アイリーア、エレンゾ錠、プログラフ注射液、リアルダ錠、ルセフィ錠、ヘムライプ皮下注、ボナロン(経口、点滴)、セレキノ錠、ヒルテプソ点滴静注) について、PJ26の有志の方のご協力も得て、RDF化。



長尾知生子さんより



TargetMineの遺伝子・疾患関連性データ

PCSK9を例に

データソースによって、疾患・表現型の用語がバラバラ

61 Genetic disease associations GWAS catalog	Number of publications	Number of SNPs
ischemic cardiomyopathy - PCSK9		
LDL cholesterol change measurement - PCSK9		
angina pectoris - PCSK9	1	1
percutaneous transluminal coronary angioplasty - PCSK9	1	1
total cholesterol measurement - PCSK9		
coronary artery bypass - PCSK9		
blood metabolite measurement - PCSK9	1	1
low density lipoprotein cholesterol measurement - PCSK9	6	4
ClinVar		
Hypercholesterolemia, autosomal dominant, 3 - PCSK9	54	91
Hypocholesterolemia - PCSK9	8	4
Familial hypercholesterolemia - PCSK9		
Familial hypobetalipoproteinemia - PCSK9		
Low density lipoprotein cholesterol level quantitative trait locus 1 - PCSK9	11	4
Familial hypercholesterolemias - PCSK9	5	1
dbSNP-PubMed-MeSH		
Cognition - PCSK9		
Hyperlipoproteinemia Type II - PCSK9		
Alcohol Drinking - PCSK9	1	1
Smoking - PCSK9	2	2
Obesity - PCSK9		
Heart Diseases - PCSK9		
Hypercholesterolemia - PCSK9	12	14
Cardiovascular Diseases - PCSK9	3	5

total cholesterol measurement

low density lipoprotein cholesterol measurement

Hypercholesterolemia, autosomal dominant, 3

Familial hypercholesterolemia

Hypercholesterolemia

統合解析プラットフォーム MANTA の公開

Home

Example DB

Download

Documentation



Microbiota And pheNotype correlaTion Analysis platform

An integrative database and analysis platform that relates microbiome and phenotypic data

Example Database

Here is an example database which based on the MANTA platform. The data was collected from 20 healthy adult volunteers (21 – 41 years old, male) from Minami-
uonuma city, Niigata Prefecture, Japan. [➔](#)

The 16S rRNA gene amplicon sequencing for the fecal samples was performed in the National
Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN). Different parameters were
collected from the volunteers including physical and exercising measurements, blood profiles, and
lifestyle questionnaires.



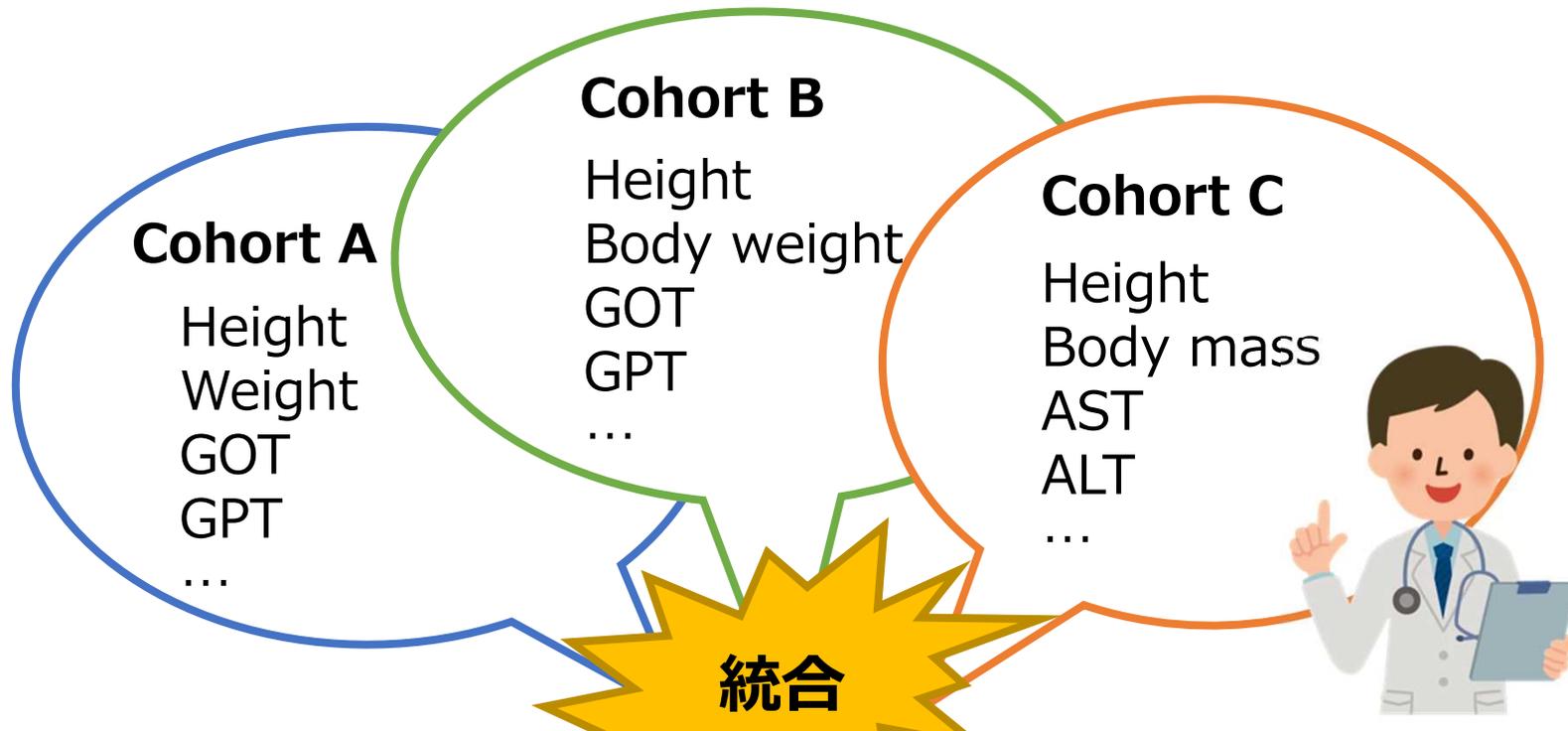
<https://mizuguchilab.org/manta/>

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

©2020水口 賢司 (医薬基盤・健康・栄養研究所 AI健康・医薬研究センター/大阪大学 蛋白質研究所)



コホートによって、表現型の用語がバラバラ



Cohort	Sample	Height	Weight	AST	ALT
A	sa01	170.5	65.2	47	56
B	sb01	165.0	58.6	52	69
C	sc01	168.5	64.0	19	13
...

用語の統合によって、比べられるようになった

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

©2020水口 賢司 (医薬基盤・健康・栄養研究所 AI健康・医薬研究センター/大阪大学 蛋白質研究所)



- インフォマ人材≠DB人材
- 多面的な課題（技術、制度、慣習…）
- データ公開、長期的な維持について各方面からの強い期待