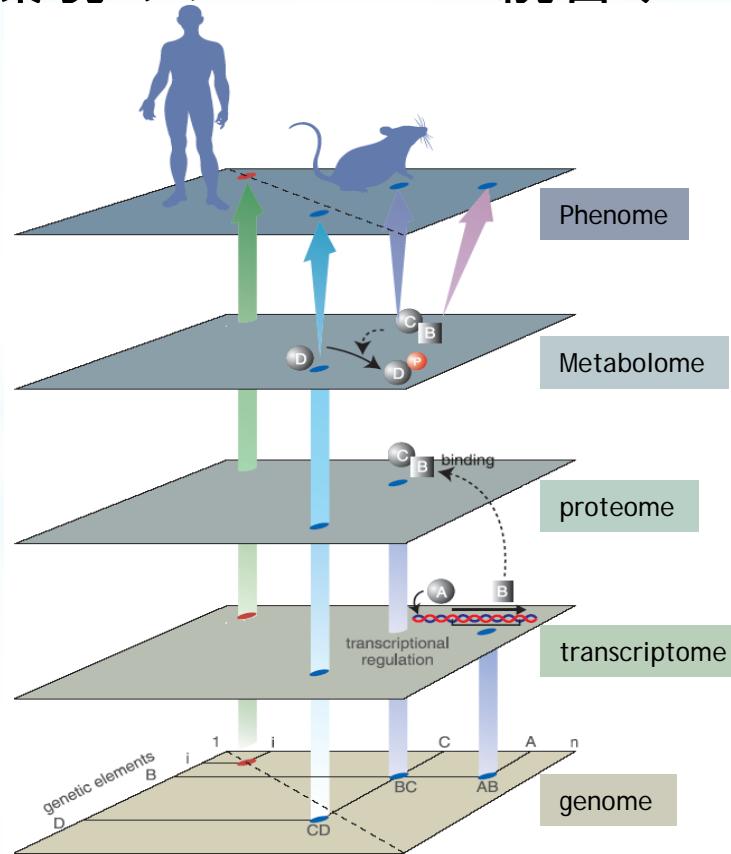


フェノタイプ情報の高度な活用による バイオリソースの付加価値創造

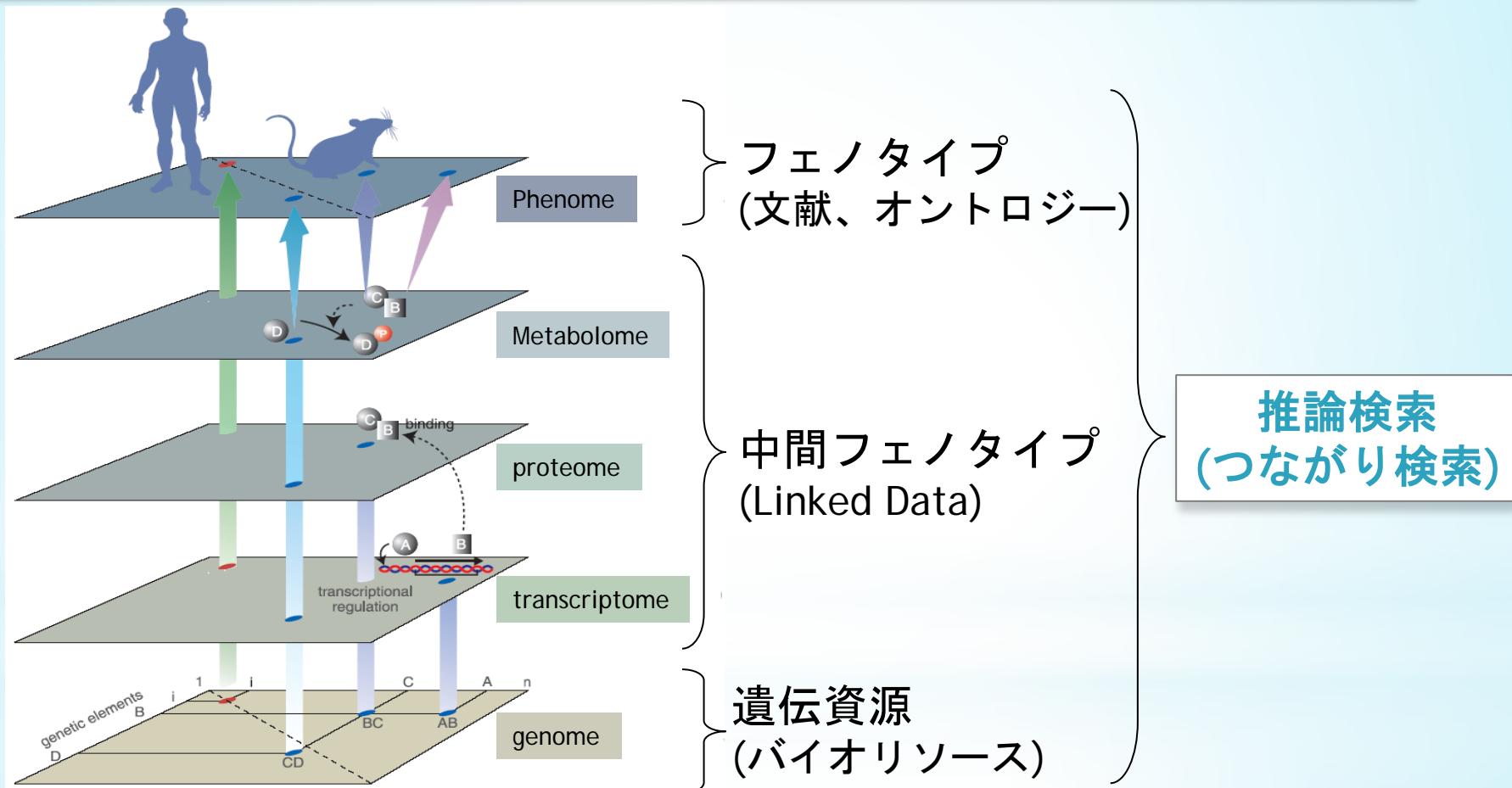
～ 生命と環境のフェノーム統合データベース ～



(独) 理化学研究所 豊田哲郎、榎屋啓志

フェノーム統合データベース

フェノタイプ情報からバイオリソースを検索できるようにする



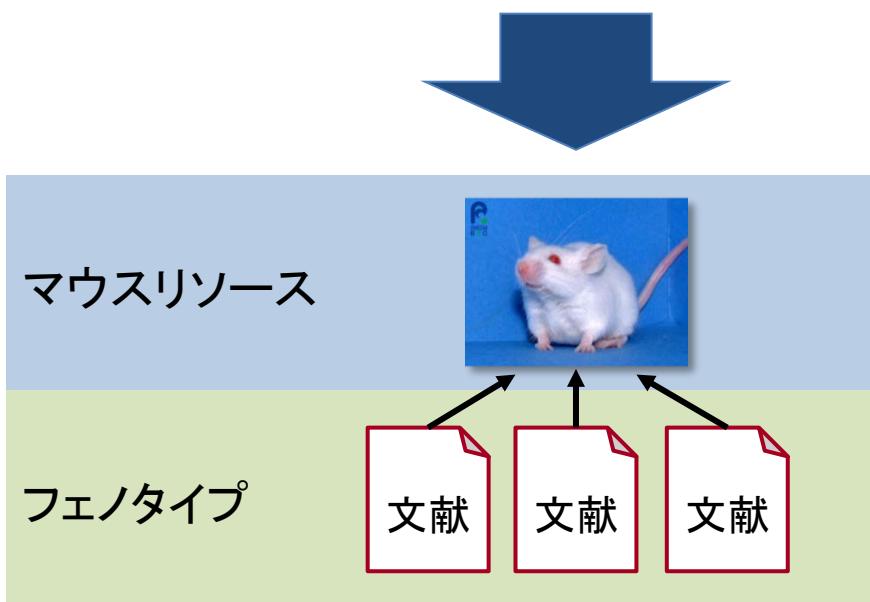
バイオリソースからフェノタイプ情報を検索できるようにする

フェノタイプ情報の蓄積はリソースの利用価値を高める



従来: リソース名を知らないと検索できない

リソース名で
検索

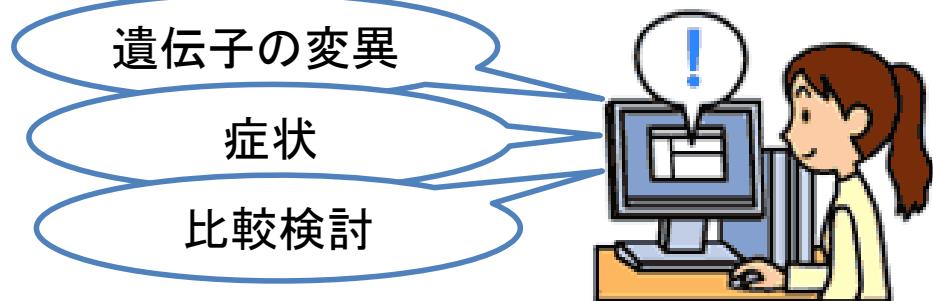


フェノーム統合データベース: フェノタイプ経由で研究にふさわしいリソースを探索、比較検討

遺伝子の変異

症状

比較検討



リソースの利用価値はフェノタイプ情報の質と量に比例

フェノタイプ情報からのリソース検索の実現

keyword 糖尿病

← キーワード「糖尿病」でマウスリソース検索

1. [BKS.Cg-Lepr<db> Nos3<tm1Unc>...](#)

← ヒットしたマウスリソースをランキング表示



1168 docs

← このマウスには 1,168 件の文献報告あり

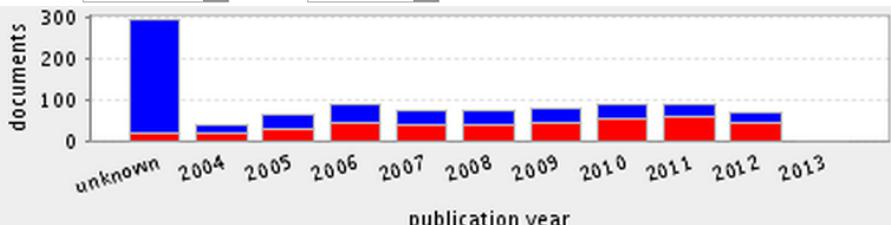
↳ 524 hits P value: 5.51E-497

← 1,168 件の文献中、524件に「糖尿病」がヒット

↳ 糖尿病 diabetes

↓ このマウスに関する文献の詳細表示

← このマウスの文献の年次推移



show digest order by fact download

Documents related to BKS.Cg-Lepr<db>...

- 2: In vivo evidence of enhanced di-methylation of histone tissue of diabetic db/db mice.

[PMID:21110946](#)

Biochemical and biophysical research communications 2011 Jan 7
AB- ...In this study, we examined whether the development of **diabetes** in upregulated genes. ...

- 3: Mitochondria are impaired in the adipocytes of type 2 d

[PMID:16501941](#)

Diabetologia 2006 Apr
AB- ...Interestingly, there was an increase in mitochondria and mitochondria treated with rosiglitazone, an agent that enhances insulin sensitivity. Taken in adipose tissue is correlated with the development of type 2 **diabetes**.

マウスリソース



フェノタイプ



利用者はキーワードにヒットした文献に紐づくマウスを一発で探せる



文献経由で研究者ランキング

keyword "nature neuroscience"

search

clear

←「nature neuroscience」で研究者リソース検索

1. Morgan Sheng

Morgan Sheng 47 docs P
↳ 4 hits P value: 3.22E-16
↳ "nature neuroscience"

←ヒットした研究者リソースを文献数でランキング表示

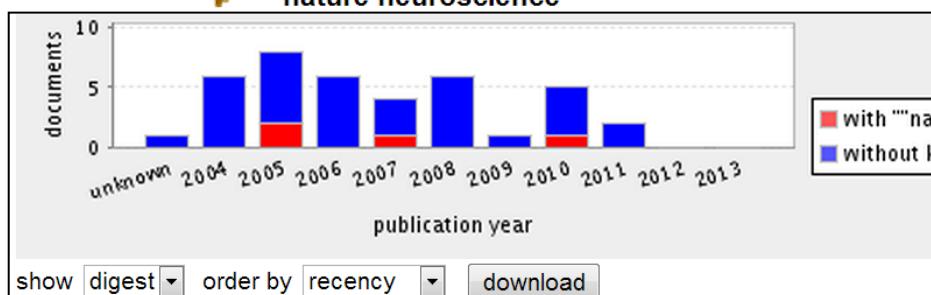
←この研究者には最近47件の文献報告あり

←47件の文献中、4件に「nat. neurosci.」がヒット

2. Michisuke Yuzaki

Michisuke Yuzaki 24 docs
↳ 3 hits P value: 3.22E-16
↳ "nature neuroscience"

↓この研究者が著者の文献の詳細表示と年次グラフ



Documents related to Morgan Sheng

1: Muscarinic receptors induce LTD of [NMDAR](#) EPSCs via AP2 and PSD-95.

PMID:20852624

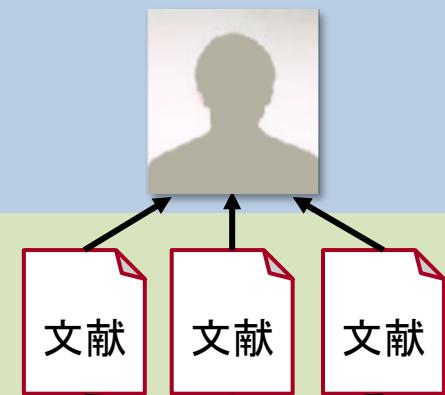
Nature neuroscience 2010 Oct

AB- ...We propose that [hippocalcin](#) binds to the SH3 region of PSD-95 at the membrane on sensing Ca²⁺; in doing so, it causes PSD-95 to dissociate from PSD-95 during dynamin-dependent endocytosis.

JO- [Nature neuroscience](#) ...

研究者リソース

フェノタイプ



利用者はキーワードに
ヒットした文献に紐づく
研究者を一発で探せる



「研究者のフェノタイプ=論文」という考え方

さらに高度な推論検索(つながり検索)

生薬の“アシュワガンダ”(*Withania somnifera*)から
マウスリソースへの“つながり検索”を実現した例

	Name	No. of docs	P value	Position
1. B6.Cg-Tg(APPswe,PSEN1dE9)85Dbo...	B6.Cg-Tg(APPswe,PSEN1dE9)85Dbo/Mmjjax Withania somnifera	271 docs	1 doc 8.52E-19	Score: 8.52E-50019
2. B6C3-Tg(APPswe,PSEN1dE9)85Dbo...	B6C3-Tg(APPswe,PSEN1dE9)85Dbo/Mmjjax Withania somnifera	273 docs	1 doc 8.52E-19	Score: 8.52E-50019
3. B6.Cg-Tg(PDGFB-APPsWInd)20Lms/...	B6.Cg-Tg(PDGFB-APPsWInd)20Lms/2Mmjjax Withania somnifera	193 docs	1 doc 8.52E-19	Score: 8.52E-50019
4. C3fBAnl.Cg-Cat<sub>b</sub>/AnlJ	C3fBAnl.Cg-Cat <sub>b</sub> /AnlJ Cat Withania somnifera	374 docs	Mm:2:103294006-103325317	Score: 2.91E-40100

「生薬成分が Catalase を増加させる」という報文を経由した間接ヒット

文献を経由したヒット例

生薬名(キーワード) → 文献のテキスト → マウスリソース

文献と遺伝子情報を経由したヒット例

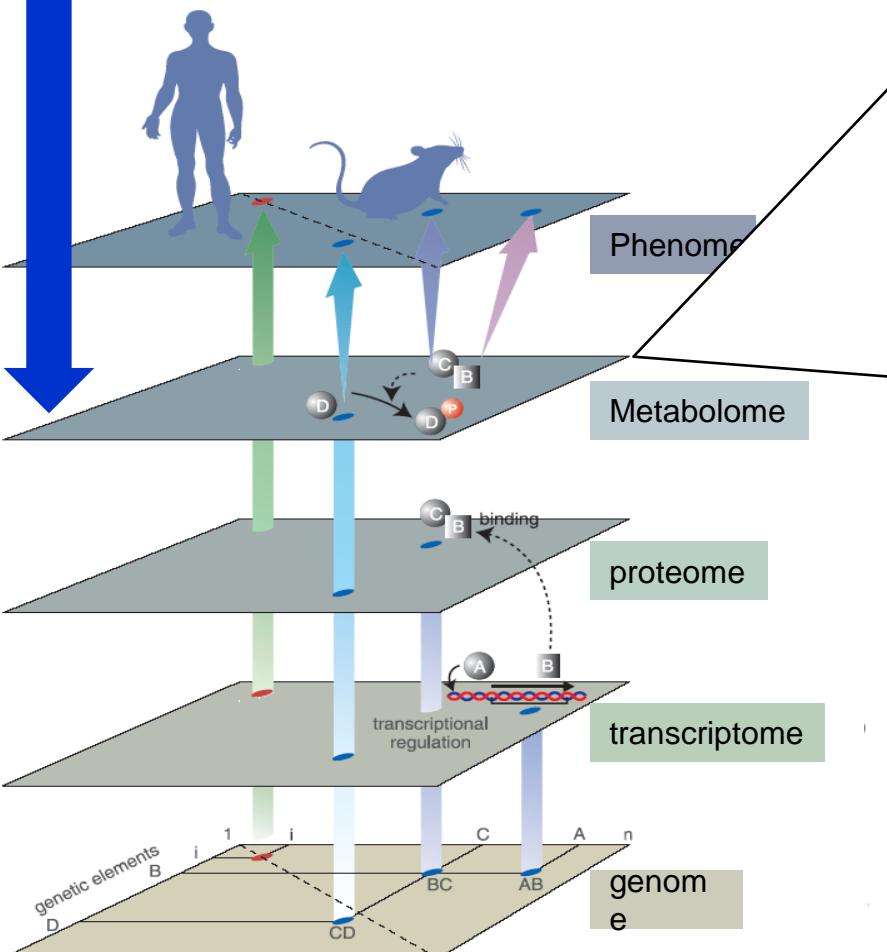
生薬名(キーワード) → 文献のテキスト → 変異遺伝子Catalase → マウスリソース

CC-BY Tetsuro Toyoda, October 5, 2013.

6

フェノーム情報は他のDBの価値も高める(バイオリソースだけではない)

キーワード「多動性障害」から KNApSAck(金谷先生のDB)へのつながり検索を実現した例



1. [Dopamine](#) Score: 4.29E-53211

Name	No. of docs	P value	Position
Dopamine	117384 docs		
多動性障害 ("hyperactivity disorder")	1236 docs	4.29E-3211	

2. [Amphetamine](#) Score: 3.11E-51088

Name	No. of docs	P value	Position
Amphetamine	23048 docs		
多動性障害 ("hyperactivity disorder")	423 docs	3.11E-1088	

3. [Serotonin](#) Score: 1.91E-51031

Name	No. of docs	P value	Position
Serotonin	119321 docs		
多動性障害 ("hyperactivity disorder")	401 docs	1.91E-1031	

4. [L-Noradrenaline](#) Score: 2.43E-50812

Name	No. of docs	P value	Position
L-Noradrenaline	96919 docs		
多動性障害 ("hyperactivity disorder")	316 docs	2.43E-812	

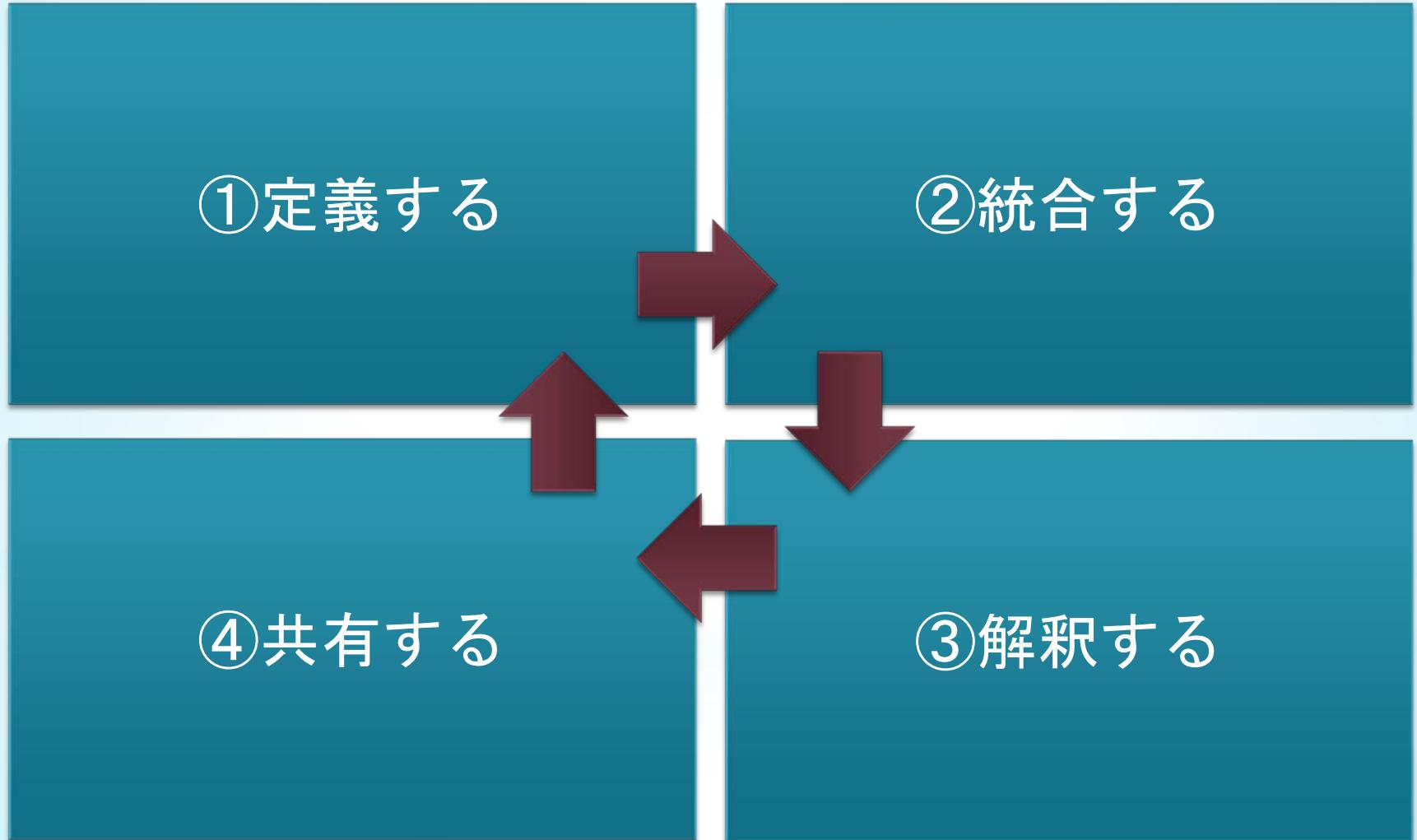
5. [Nicotine](#) Score: 4.07E-50496

Name	No. of docs	P value	Position
Nicotine	30770 docs		
多動性障害 ("hyperactivity disorder")	193 docs	4.07E-496	

6. [Cocaine](#) Score: 1.74E-50452

Name	No. of docs	P value	Position
Cocaine	31422 docs		
多動性障害 ("hyperactivity disorder")	176 docs	1.74E-452	

フェノーム統合プラットフォーム



①フェノタイプを定義する

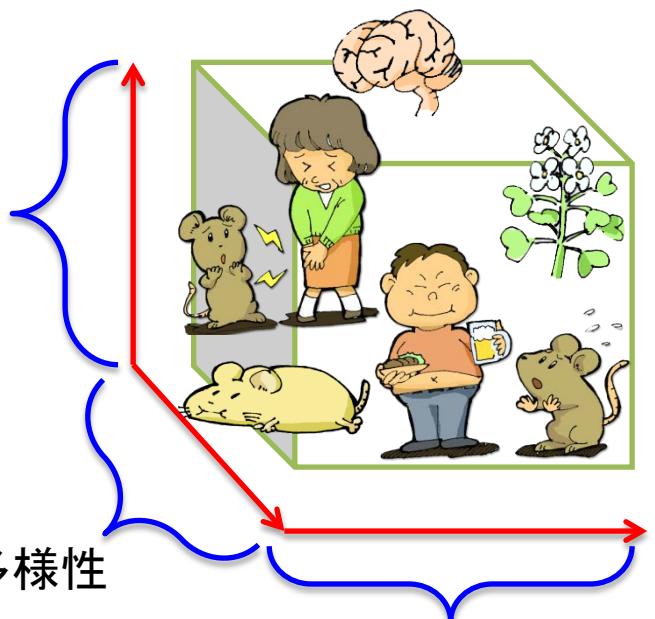
フェノタイプの記述対象は
極めて自由度が高い！

生命の階層性

- 個体レベル
- 臓器レベル
- 細胞レベル

遺伝的要因の多様性

- 生物種
- エピゲノム
- 遺伝子変異



環境要因の多様性

- 栄養状態
- 薬剤投与
- ストレス

従来の問題点：

- 文章等で表現されていたため、語彙のゆらぎの影響が大きく、フェノタイプの網羅的検索が不可能だった。

改善ポイント：

- 辞書=オントロジー化により、網羅的検索を可能にする。

期待される効果：

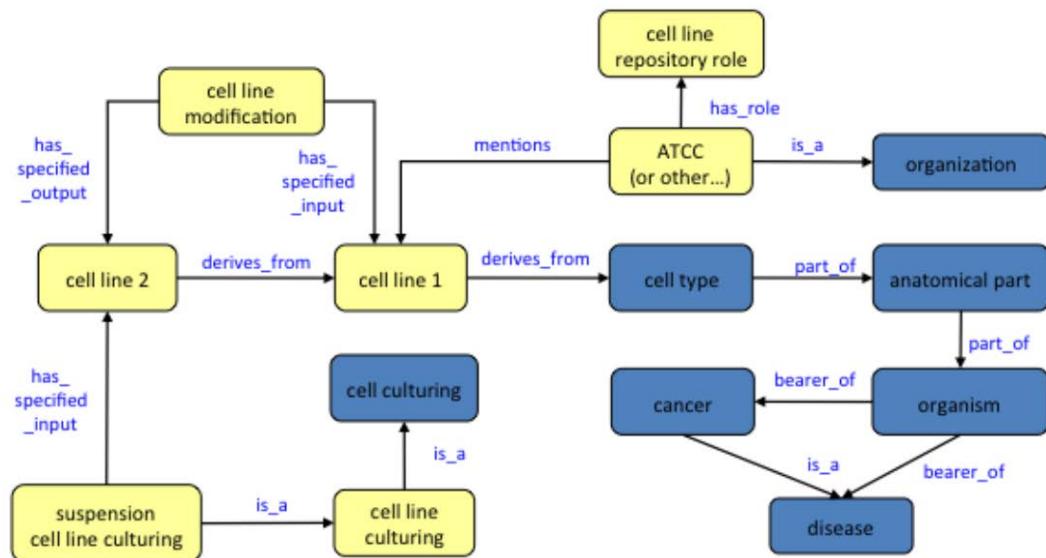
- **標準化：**
ゆらぎのないフェノタイプ表現で、検索漏れを防止
- **統合化：**
複数データベースで表現を統一し、ワンストップショッピング検索を実現
- **体系化：**
フェノタイプの類似性で比較・体系化
(マウスの異常とヒト疾患の対応づけ)

フェノタイプの語彙を標準化して世界的な統合化へ

定義オントロジー国際連携

Cell Line Ontology (CLO) Consortium

- OWL/RDF スキーム標準化の推進
- 世界規模でのフェデレーション型情報統合
- 他分野データとの統合／連携推進



Sarntivijai et al, Journal of Biomedical Semantics, (2013: submitted)

これにより、全世界的な培養細胞情報統合を目指す

フェノタイプの表現方法(スキーマ)を検索しやすく統一

部位と特性で、リソース情報を整理

異常部位の内訳(MAオントロジー分類)

神経系の異常: 10系統

骨格系: 36系統

長骨: 16

長骨以外: 5

中軸系: 12

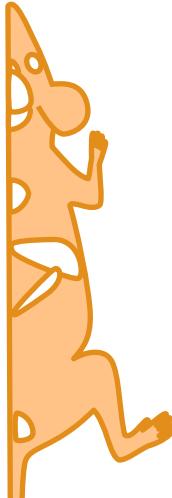
関節: 3

表皮系: 87系統

結合組織系: 17系統

血液、リンパ系: 102系統

尿: 6系統



長骨異常の内訳(PATOオントロジー分類)

形態: 35

形態以外: 1

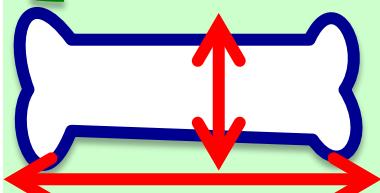
サイズ: 20

形状: 15

長さ短縮: 13

径の短縮: 6

形成不全: 1



長骨の長軸方向成長
コントロール変異: 13
系統、太さコントロール
変異: 6系統

統一スキーマでの表現例

フェノタイプ

何の?

長骨

どんな
形質が?

長さ

どうなった?

短縮

柔軟なリソース検索の実現

「骨」に異常のあるマウスは?

異常部位でフィルター

表現型アノテーション	系統	mammalian phenotype ontology	測定部位	測定対象の形質	測定結果
phenotype of M100451	M100451	長骨骨端の形態異常	long bone epiphysis	morphology	abnormal
phenotype of M100856	M100856	長骨骨端軟骨板増殖帯の異常	long bone epiphyseal plate proliferative zone	physical object quality	abnormal
phenotype of B6.129S-Osmr<tm1Mtan>	B6.129S-Osmr<tm1Mtan>	骨髓球系前駆細胞の形態異常	common myeloid progenitor	morphology	abnormal
phenotype of M100856	M100856	下肢帶骨の形態異常	pelvis	abnormal	abnormal
phenotype of M101679	M101679	頭部の短縮		diameter	decreased diameter
phenotype of M100906	M100906	頭部の短縮		diameter	decreased diameter
phenotype of M101532	M101532	頭部の短縮	cranium	diameter	decreased diameter
phenotype of PIC/Nem-Pdn/+	PIC/Nem-Pdn/+	骨癒合症	bone	attachment quality	attached to
phenotype of M100576	M100576	頭蓋骨の短縮	cranium	height	decreased height
phenotype of M100117	M100117	半球状の頭蓋	cranium	curvature	domed

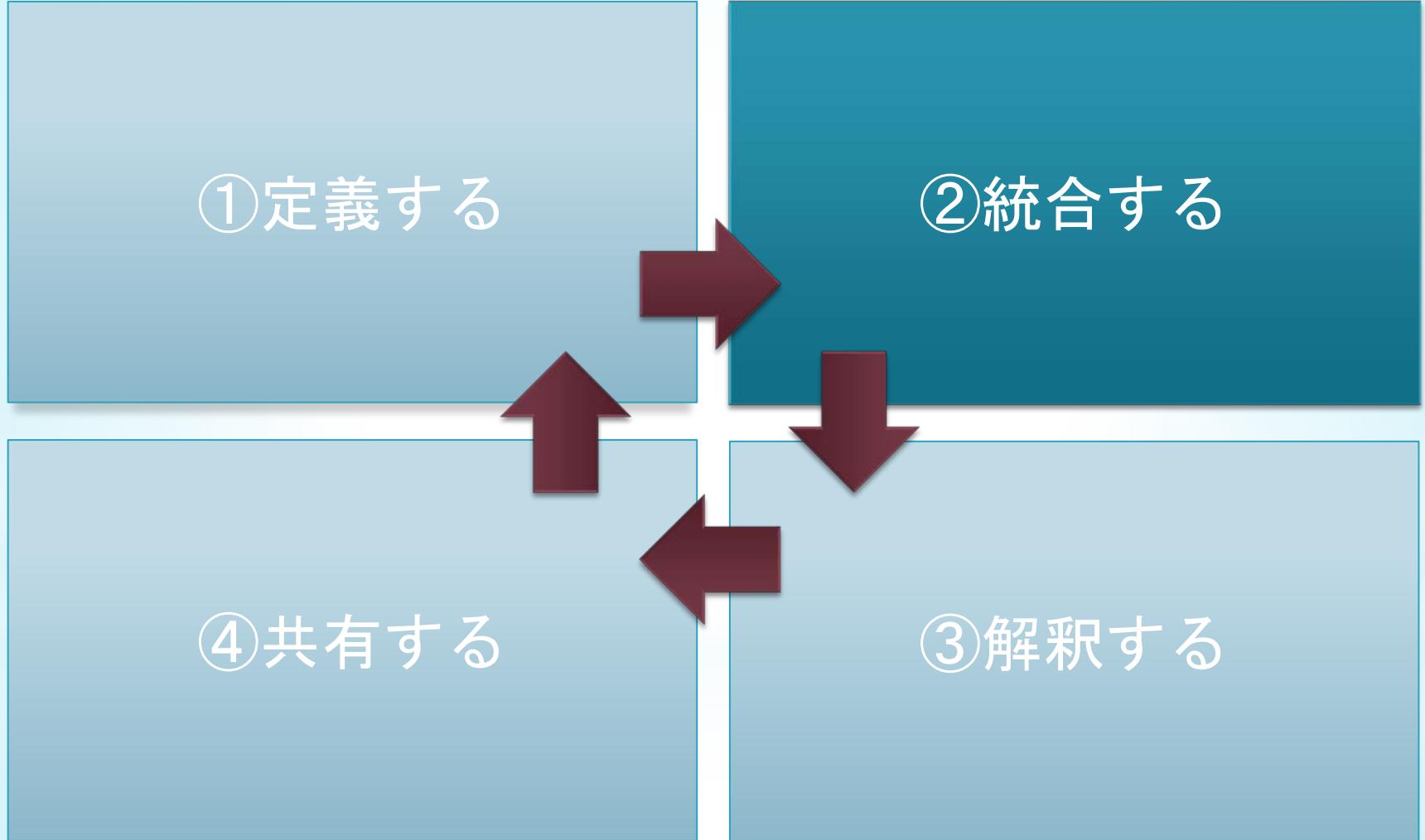
マウス系統のリスト

スキーマ統一の国際連携で可能になったこと

表現型という『複雑で自由度の高い』情報に対して…

- ゆらぎのないフェノタイプ表現で、検索漏れを防止
=>「長骨の長軸方向の成長が阻害されるKOマウス」等を、語彙表現のゆらぎを超えて検索する事が可能になった。
- 複数データベースで表現を統一し、ワンストップショップ検索の実現へ
=>現在、黒川G(微生物)、大浪G(時空間解析)、DBCLSと連携中
- フェノタイプの類似性で比較・体系化
=>「お勧めマウスリソース」提案機能
=>疾患モデルマウス検索(つながり検索)
- 国際的なデータベースの相互利用の実現へ
=>Mouse Genome Informatics (MGI: ジャクソン研)
PhenomeNet (ケンブリッジ大)

フェノタイプ情報を分かりやすく統合



定義したスキーマに基づいて フェノタイプデータを生物種横断的に収集・統合

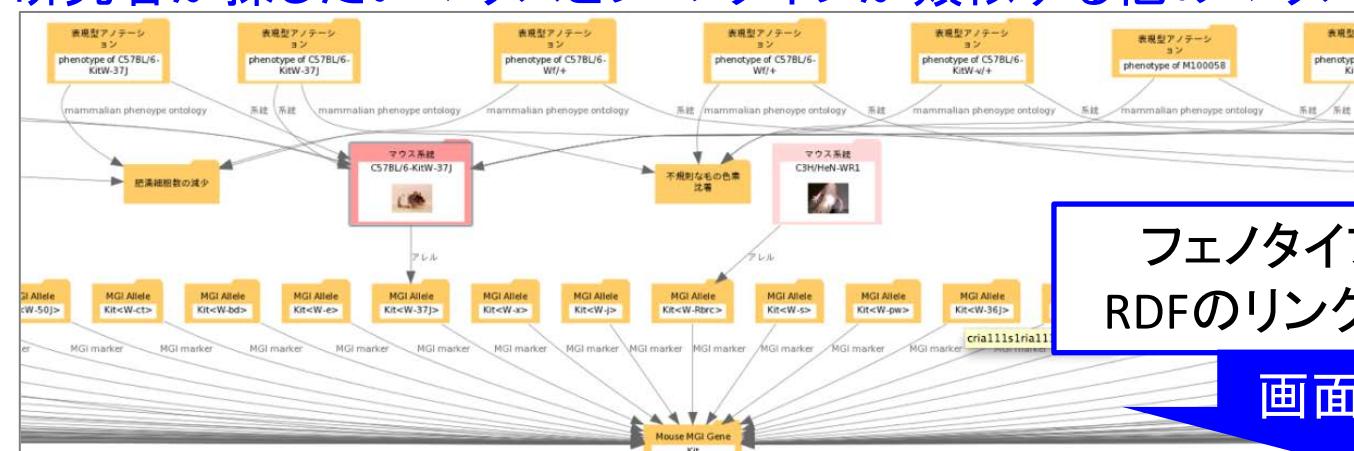
表現型データ汎用RDFスキーム(抜粋)
(OBOコンソーシアム提案の改良型)



生物横断的に特性情報を格納できる汎用スキーム(各生物の特性情報 is_a 表現型データ)

フェノタイプの類似性を自動探索して「お勧めマウス」を提示

研究者が探したいマウスとフェノタイプが類似する他のマウスを分かりやすく提示



研究者に新たな
気づきを与える

フェノタイプに関する
RDFのリンクを自動探索

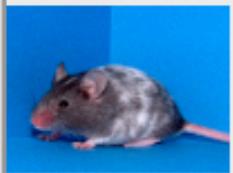
画面生成

同じ表現型を示すマウス系統 (56)

(各系統の示す表現型の一一致度が高い順番に提示)

7 個のmammalian phenotype ontologyが一致

C57BL/6-Wf/+



共通の mammalian phenotype...

- 肥満細胞数の減少
- 薄い毛色
- 腹部の白斑
- 大球性貧血
- 赤血球数の減少

4 個のmammalian phenotype ontologyが一致

C57BL/6-Wsh/Wsh



共通の mammalian phenotype...

- 肥満細胞数の減少
- 被毛色異常
- 赤血球数の減少
- 可変性体部斑点

3 個のmammalian phenotype ontologyが一致

C57BL/6-KitW-37J



共通の mammalian phenotype...

- 肥満細胞数の減少
- 薄い毛色
- 白色スポット

C3.Cg-Kitl<Sl-pan>...



共通の mammalian phenotype...

- 薄い毛色
- 大球性貧血
- 白色スポット

類似度高

類似度低

フェノタイプ推薦機能でバイオリソースの付加価値が向上

マウス・フェノタイプ統合の国際コンソーシアム

アジアマウス変異リソース連合 (AMMRA)

 Asian Mouse Mutagenesis Resource Association
AMMRA

[Home](#)
[About AMMRA](#)
[Greetings from Chair](#)
[Meetings and Reports](#)
[Organization](#)
[AMMRA Resource DB \(in preparation\)](#)
[Links](#)



Asian Mouse Mutagenesis Resource Association



News :
Information of upcoming workshop:
[Genetically Modified Animal Care and Refinement Workshop](#)
Date : 12 November 2013 (Tuesday)
Venue : Academia Sinica, Taipei
Registration : 19 August 2013 ~ 21 October 2013 (Please note that only [online registration](#) is accepted.)
Fee: USD60

Institute members of AMMRA :
[Biological Resource Center, Singapore](#) (founding member)
[National Laboratory Animal Center, National Applied Research Laboratories, Taipei](#) (founding member)
[Model Animal Research Center, Nanjing Univ. Nanjing](#) (founding member)
[Shanghai Institute of Biological Sciences, Shanghai](#) (founding member)
[Nanfang Center for Model Organisms, Shanghai](#) (founding member)
[Institute of Laboratory Animal Sciences of Peking Union Medical College, Beijing](#) (founding member)
[Peking University-BLARC, Beijing](#) (founding member),
[Biomedical Mouse Resource Center \(BMRC\), KRIE, Daejeon](#) (founding member)

中国のグループと連携し、
アジア統合リソースデータベース目指す

国際マウス表現型解析コンソーシアム (IMPC)

 HOME ABOUT IMPC SEARCH NEWS AND EVENTS CONTACT US MY IMPC REGISTER

Welcome to the International Mouse Phenotyping Consortium

SEARCH IMPC DATA
Enter your favorite gene in the search box and register your interest or type a standard protocol name and retrieve the parameters measured in IMPReSS.



The International Mouse Phenotype Consortium
Comprises a group of major mouse genomics research groups that have joined together to tackle the challenge of developing an encyclopedic resource of mouse phenotypic data.

BROWSE GENE LIST
Find out the latest status of your gene of interest

REGISTER FOR GENE STATUS UPDATES
Register to be kept up to date in the latest progress of your gene of interest

PHENOTYPE PROTOCOLS (IMPReSS)
View Phenotype Procedures and Parameters held in the IMPReSS Database

TRACK MOUSE PRODUCTION (IMITS)
Follow microinjection progress and mouse production

Project Status



All Projects
Project started
Microinjection in progress
Genotype Confirmed Mice

網羅的表現型データのRDF化を
理研が主導(来年公開予定)

表現型解析結果 (IMPC)
C57BL/6N-D1Pas1tm1(KOMP)Vlcg>D4K0Crc

Basic information: MGI Allele	D1Pas1tm1(KOMP)Vlcg>D4K0Crc	関連情報を見る	関連情報検索			
Basic information: BRC ID	RBR0205432					
Open Field		パラメーター	p値	統計的有意差	雌の検査結果	雄の検査結果
Open Field	IMPC_OFD_005_001_05_DT	Distance travelled	5.76790E-01	normal	Male result of Distance travelled	Female result of Distance travelled
Open Field	IMPC_OFD_005_001_10_DT	Distance travelled	5.38244E-01	normal	Male result of Distance travelled	Female result of Distance travelled
Open Field	IMPC_OFD_005_001_15_DT	Distance travelled	6.45193E-01	normal	Male result of Distance travelled	Female result of Distance travelled
Open Field	IMPC_OFD_005_001_20_DT	Distance travelled	7.66731E-01	normal	Male result of Distance travelled	Female result of Distance travelled
Open Field	IMPC_OFD_007_001_WA_RT	Whole arena resting time	3.39672E-01	normal	Male result of Whole arena resting time	Female result of Whole arena resting time
Open Field	IMPC_OFD_007_001_WA_RM	Whole arena permanence	1.00000E+00	normal	Male result of Whole arena permanence	Female result of Whole arena permanence
Open Field	IMPC_OFD_007_001_WA_AVG	Average speed	7.48251E-01	normal	Male result of Whole arena average speed	Female result of Whole arena average speed
Open Field	IMPC_OFD_007_001_PERI_RM	Periphery travelled	7.93119E-01	normal	Male result of Periphery distance travelled	Female result of Periphery distance travelled
Open Field	IMPC_OFD_007_001_PERI_AVG	Periphery time	8.08771E-01	normal	Male result of Periphery resting time	Female result of Periphery resting time
Open Field	IMPC_OFD_007_001_PERI_RM	Periphery permanence time	8.16525E-01	normal	Male result of Periphery permanence time	Female result of Periphery permanence time
Open Field	IMPC_OFD_007_001_CTR_RM	Center speed	8.48350E-01	normal	Male result of Periphery average speed	Female result of Periphery average speed
Open Field	IMPC_OFD_007_001_CTR_AVG	Center travelled	7.92953E-01	normal	Male result of Center distance travelled	Female result of Center distance travelled
Open Field	IMPC_OFD_007_001_CTR_RM	Center resting time	7.29193E-01	normal	Male result of Center resting time	Female result of Center resting time
Open Field	IMPC_OFD_007_001_CTR_AVG	Center permanence time	8.16525E-01	normal	Male result of Center permanence time	Female result of Center permanence time
Open Field	IMPC_OFD_017_001_C_A5	Center average speed	8.97311E-01	normal	Male result of Center average speed	Female result of Center average speed
Open Field	IMPC_OFD_018_001_L_CE	Latency to center entry	7.95401E-01	normal	Male result of Latency to center entry	Female result of Latency to center entry
Open Field	IMPC_OFD_019_N_CE	Number of center entries	8.38610E-01	normal	Male result of Number of center entries	Female result of Number of center entries
Open Field	IMPC_OFD_020_001_DT_T	Distance travelled - total	7.48251E-01	normal	Male result of Distance travelled - total	Female result of Distance travelled - total
Open Field	IMPC_OFD_022_001_C1_T	Percentage center time	8.16525E-01	normal	Male result of Percentage center time	Female result of Percentage center time

設計スキーマに
従ったデータ作成

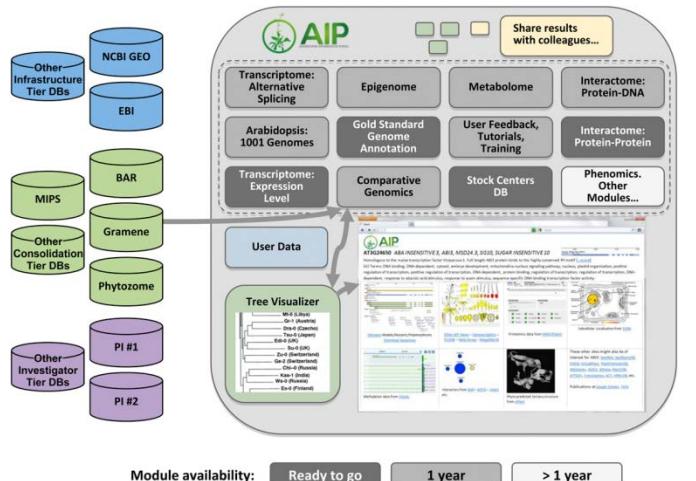
植物フェノタイプ統合の国際コンソーシアム

Arabidopsis Information Portal (AIP)



- シロイヌナズナの国際的コンソーシアム
- 国際的な分業体制でシロイヌナズナの統合データベースを目指す
- 理研はフェノームで連携

参画

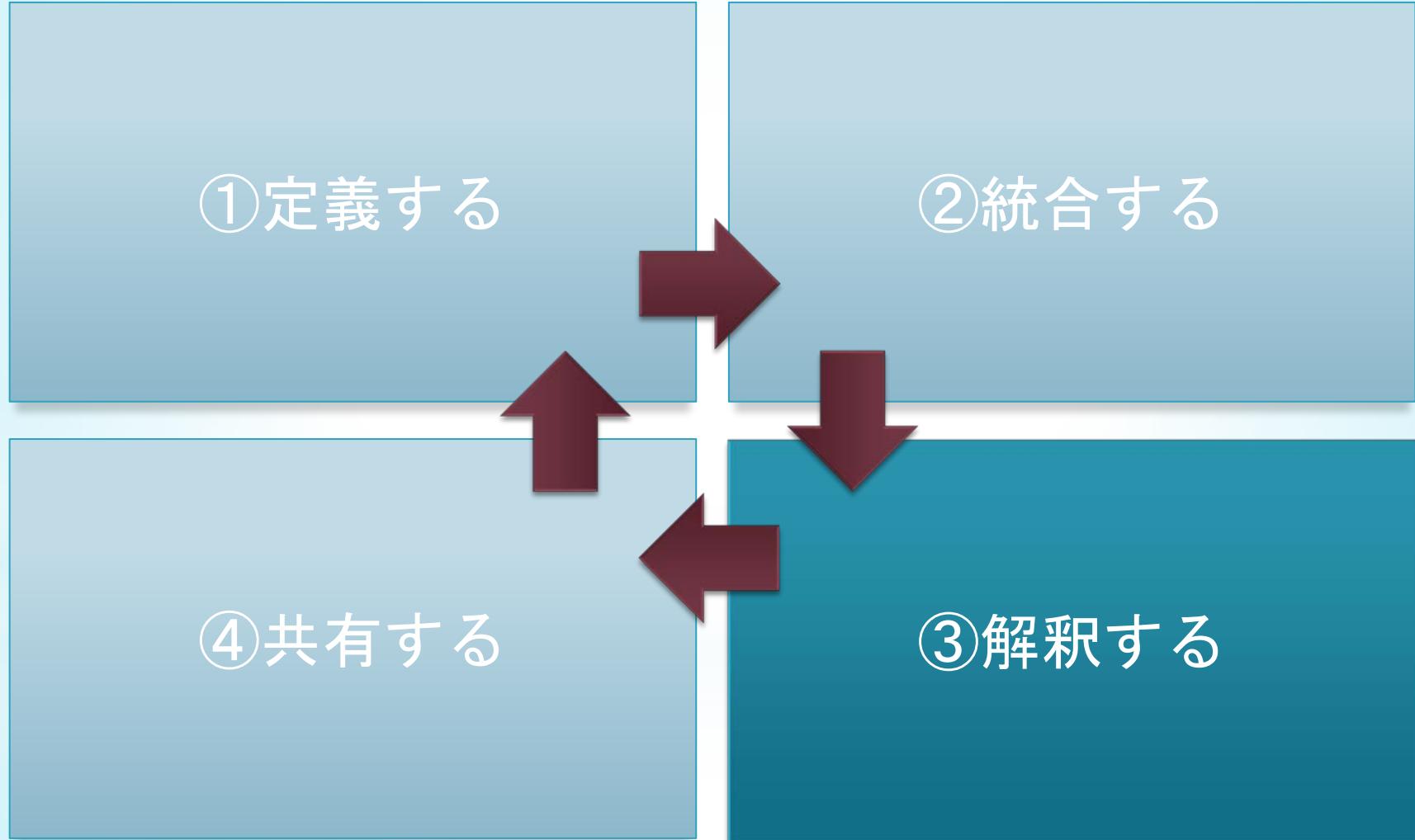


本事業のフェノームデータベース



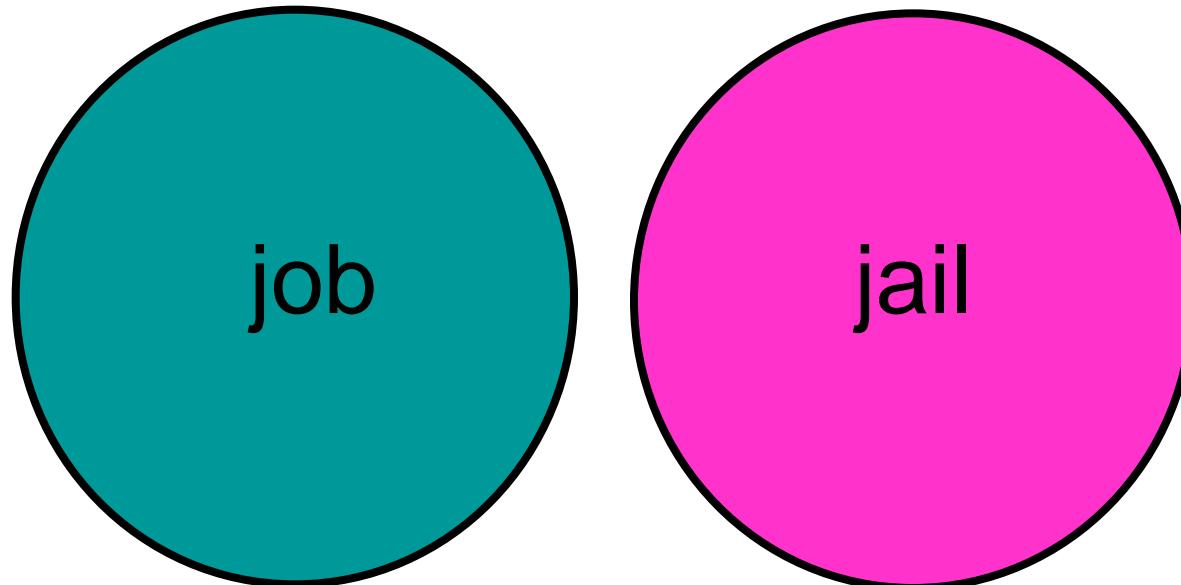
多様なシロイヌナズナ変異体を開発し、研究コミュニティに提供

フェノタイプ情報で利用者のデータ解釈を支援



通常のデータベース検索の限界

“My job is a jail” を解釈できない！



Boolean では $\text{job} \neq \text{jail}$

脳における Open-endedness

連想力で “My job is a jail” を解釈

job

jail

拘束感かな？

脳は概念の枠を容易に広げて解釈する

→ Linked Data のつながりで解釈を支援

つながり検索エンジン(PosMed)の適用

Linked Dataを使った推論検索(脳型データベース)

Makita et al. *Nucleic Acids Res.* (2013)

閲覧用文書ファイル
~1000万件

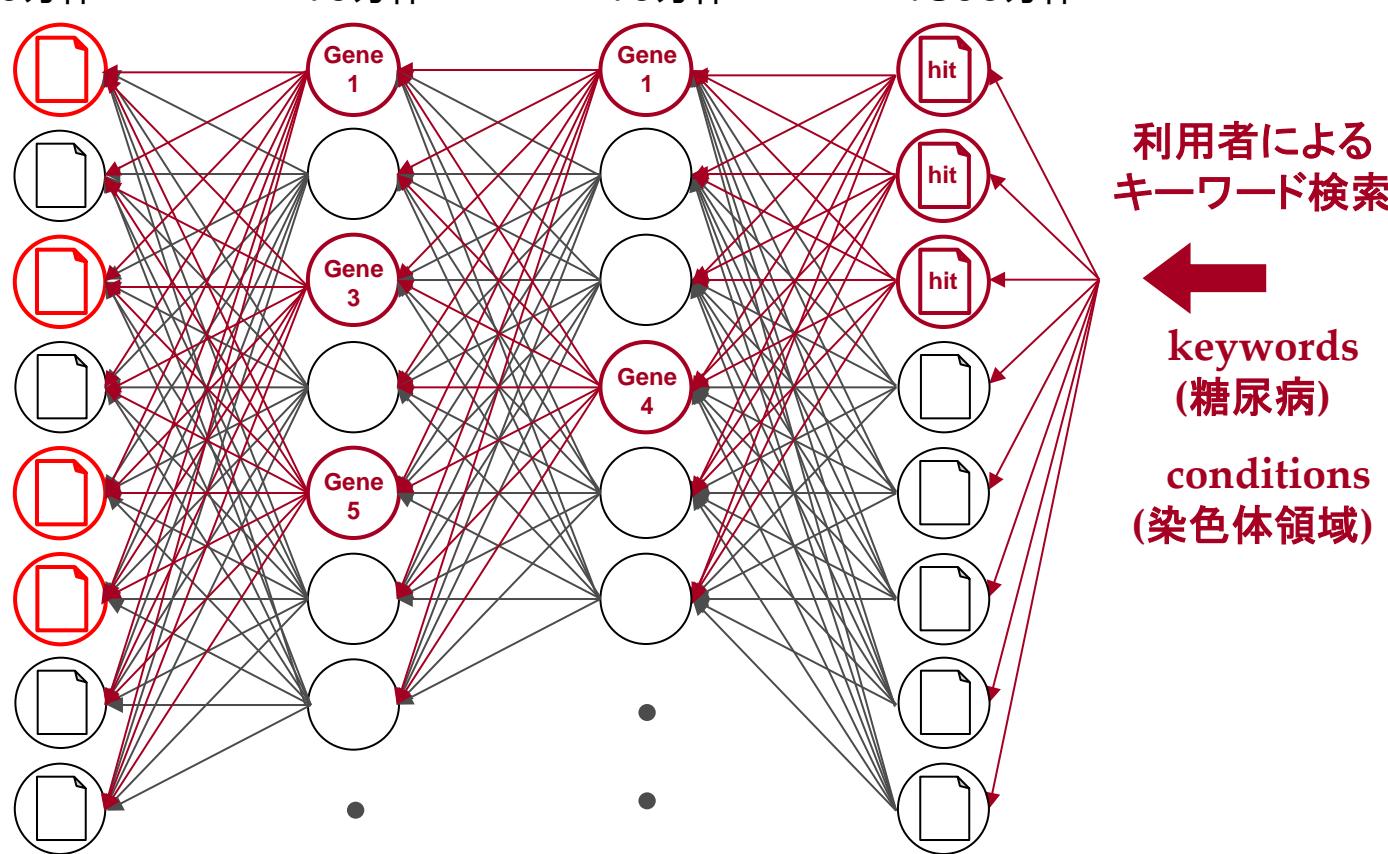
高速検索

概念ノード
~10万件

概念ノード
~10万件

文献等の文書
~1800万件

検索結果



文献と遺伝子の正確な対応付けルールを構築 → 正確なテキストマイニング

数万個の集計表を高速に計算する技術を適用

すべての Gene (X=1,2,,,数万) に対して、
三次元集計表を瞬時に作成する技術

	検索ワードを含む	検索ワードを含まない
Gene Xを含む	A個のドキュメント	B個のドキュメント
Gene Xを含まない	C個のドキュメント	D個のドキュメント

プロパティ(ドキュメントの種類)による
分類(Medline, OMIM, PPI,
Bio resources 等)

この表に基づいて、統計量の算出や
フィッシャーの正確検定などの
統計的推論を行うことができる

遺伝子名や
概念ノード等
(~10万件)

文献等の文書ファイル
(~千万件)

利用者による
フェノタイプ検索
(検索ワード)

キュレーターが対応づけた
遺伝子と文書の対応関係

すべての計算が1秒程度で終わる → 高速ウェブ検索の実現

つながり検索でバイオリソースのフェノタイプを推論

search bioresource condition bioresource name species mouse keyword diabetes GWAS bioresource name search clear recent 4 years 4 show 20

Hit resources

download rank list set as target check all

- 1. C57BL/6J
└ Myd88
- 2. IGIF/IL-18 KO
└ Il18
- 3. C3H/HeJ
└ Cntrnap1
- 4. C3HeB/FeJ
└ Asgr1
- 5. C57BL/6JScfAnu-Cd4^{thb}/A...
└ Cd4
- 6. B6:129-Tnfrsf1a^{tm1Mak}
└ Tnfrsf1a
- 7. C.BA-Tg(CD2-IL5)C2...
└ Cd2
- 8. B6.CBACa-Tg(CD2-II5)...
└ Cd2
- 9. CBA/Ca-Tg(CD2-II5)C...
└ Cd2
- 10. CBA/Ca-Tg(CD2-II5)C...
└ Cd2
- 11. C3.SW-H2/Sn...
└ Cntrnap1
- 12. C.Cg-Myd88^{tm1Aki}Tg(CD2-IL...
└ Myd88
- 13. NOD.Cg-Tnfrsf1a^{tm1Blt}Fas^{lpr}
└ Tnfrsf1a
- 14. C57BR/cdJ
└ Cd4
- 15. SHM/Nem-Re/+;shm/+
└ Cntrnap1
- 16. C57BL/6JAnu-Cd4^{thb}/A...
└ Cd4
- 17. RKO-1

Links from the resource to the keyword

BALB/cByJ Acads diabetes GWAS

Mouse bioresource Provenance Mouse Locus keyword

About the resource

Acads 182 show graph

Symbol Acads
Name acyl-Coenzyme A dehydrogenase, short chain
Other aliases SCAD, Bcd-1, Bcd1, Hdlq8

Documents about the resource

all (2/182) Mouse bioresource (0/3) Gene ontology (0/16) MEDLINE (1/125) Mammalian phenotype ontology (0/13) Mouse protein-protein interaction (0/0) Mouse gene (0/1) OMIM (1/13) Pathway information (0/11) other databases

show related entities

keyword diabetes GWAS year all Find

documents publication year

with "diabetes GWAS"
without keyword

show digest order by fact download

1-20 of 182 Page 1

Documents related to Acads

2: The minor C-allele of rs2014355 in **ACADS** is associated with reduced insulin release following an oral glucose load.

PMID:21211036
BMC medical genetics 2011
AB-

A genome-wide association study (GWAS) using metabolite concentrations as proxies for enzymatic activity, suggested that two variants: rs2014355 in the gene encoding short-chain acyl-coenzyme A dehydrogenase (ACADS) and rs11161510 in the gene encoding medium-chain acyl-coenzyme A dehydrogenase (ACADM) impair fatty acid ?-oxidation.

Chronic exposure to fatty acids due to an impaired ?-oxidation may down-regulate the glucose-stimulated insulin release and result in an increased risk of type 2 diabetes (T2D). ... The C-allele was not associated with T2D in the case-control analysis (OR 1.07, 95% CI 0.96-1.18, P = 0.21). rs11161510 of ACADM did not associate with any indices of glucose-stimulated insulin

キーワード
「糖尿病, GWAS」で
バイオリソースを検索



GWASで糖尿病原因遺伝子とされている
Acads遺伝子と文献共起のある、
マウスリソースが提示された。

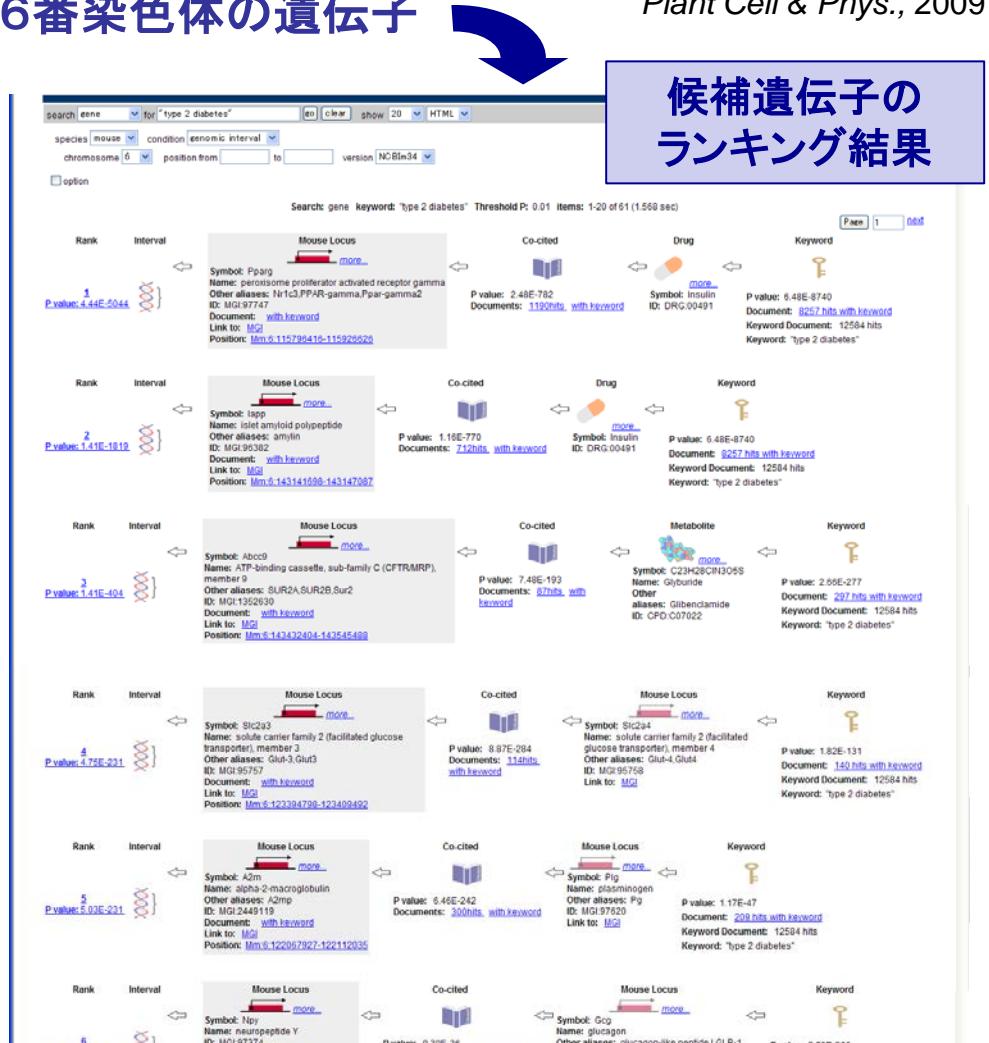
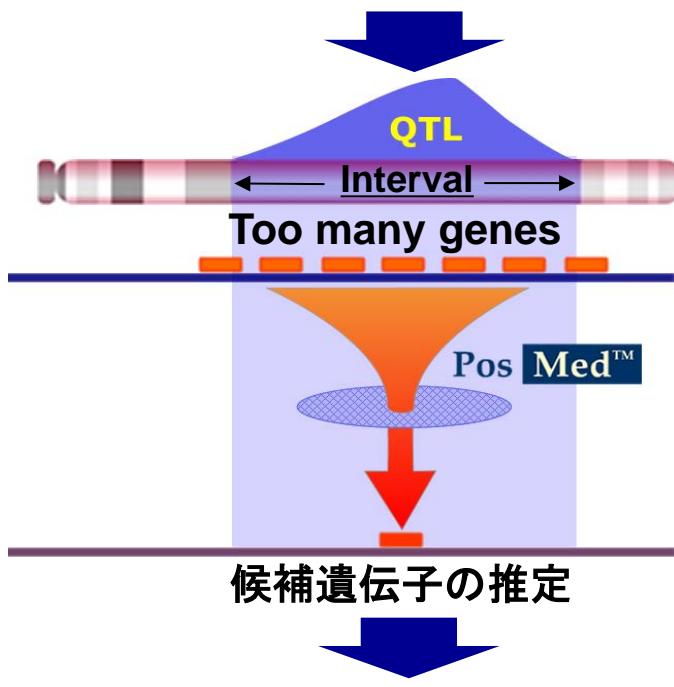
PosMed: Positional Medline (つながり検索のシステム)



複合検索条件 { キーワード: type 2 diabetes
ゲノム条件: 6番染色体の遺伝子

ポジショナルクローニングでの候補遺伝子選び

交配実験によるラフマッピング
(~10Mbp wide interval)



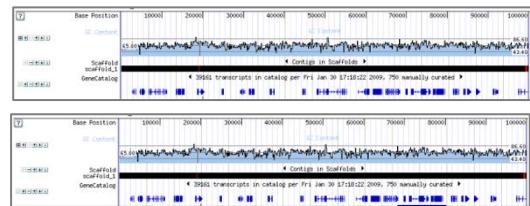
ENUマウスマутタジェネシスでは、約65のマウス系統の原因遺伝子の探索成功に貢献
→ NGSで見つかってくるヒト細胞の変異部位の機能解釈に応用できるインターフェースを提供

PosMedの用途例：がん細胞の原因変異遺伝子の絞り込み

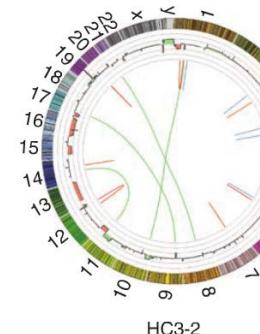
①NGSデータの比較による変異部位の同定

正常細胞

がん
細胞

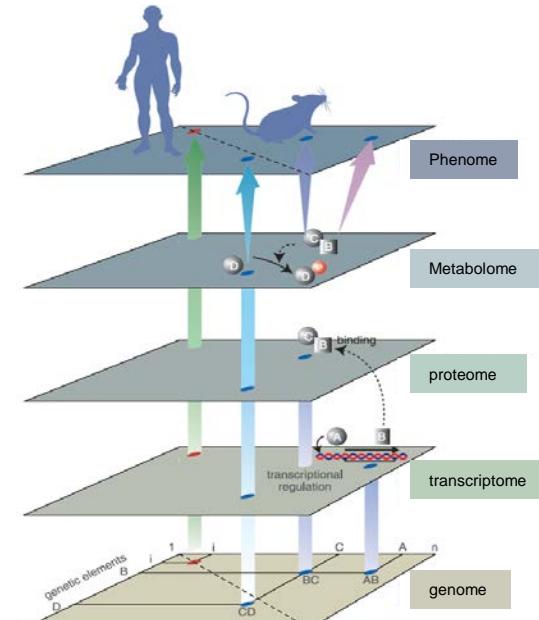


②変異のある遺伝子のリストの作成(数百～数千)



Gene1
Gene2
▪
▪
▪

変異遺伝子リスト



フェノーム統合DB



③変異遺伝子とフェノタイプ情報を推論的に結び付けて提示(原因遺伝子変異の絞り込み)

Gene1

4. IL1A , interleukin 1, alpha	Human Locus	Mouse Locus	Co-citation	Mouse Locus	Keyword
Interval show graph	Human Locus show graph	Mouse Locus show graph	Co-citation show graph	Mouse Locus show graph	Keyword show graph

Gene2

5. TGFA , transforming growth factor, alpha	Human Locus	Mouse Locus	Co-citation	Mouse Locus	Keyword
Interval show graph	Human Locus show graph	Mouse Locus show graph	Co-citation show graph	Mouse Locus show graph	Keyword show graph

フェノタイプ：
Cancer

PosMed でつながり検索した多数のデータベース群 (キーワード例：“induced Pluripotent Stem cell”)

PosMed | induced | biosparql.org/PosMed/search

Search Categories Document Set Mouse Gene マウス遺伝子

Gene MEDLINE 3847 hits Pou5f1, POU domain, class 5, transcr... 515 hits (2897 docs)
OMIM 87 hits Sox2, SRY-box containing gene 2 316 hits (1470 docs)
Biosource Mouse bioresource (complete) 12 hits Klf4, Kruppel-like factor 4 (gut) 150 hits (698 docs)

Disease Search more Nanog, Nanog homeobox 114 hits (644 docs)
Metabolite Lif, leukemia inhibitory factor 96 hits (3770 docs)

Drug Fgf2, fibroblast growth factor 2 62 hits (16078 docs)
Researcher Nes, nestin 59 hits (3136 docs)
Document set Stat3, signal transducer and activat... 45 hits (8923 docs)
Dictionary Ctnnb1, catenin (cadherin associated) 44 hits (12126 docs)
Database Tg(CAG-EGFP)1Osb, transgene insertio... 42 hits (12220 docs)

Search more

ユニークユーザ数 毎月約 12,000 人

Metabolite Patulin 175 hits (85834 docs)
Dopamine 54 hits (117384 docs)
D-Glucose 46 hits (353172 docs)
ATP 36 hits (170061 docs)
Cytosine 32 hits (23305 docs)
alpha-6-Deoxyoxotetraacycline 28 hits (10683 docs)
Factor I 25 hits (15330 docs)
Indole-3-acetonitrile 24 hits (35799 docs)
AMP 23 hits (112779 docs)
Integrin 19 hits (27548 docs)
Search more

Drug Tretinoïn 109 hits (18756 docs)
Dopamine, C8H11NO2 62 hits (117586 docs)
Dexamethasone 33 hits (54156 docs)
Doxycycline, C22H24N2O8 28 hits (10626 docs)
Interferon 20 hits (70508 docs)
Isoproterenol, C11H17NO3 17 hits (34770 docs)
Cyclophosphamide 14 hits (54882 docs)
Valproic Acid 13 hits (11116 docs)
Adriamycin, C27H29NO11 12 hits (48062 docs)
Verapamil, C27H38N2O4 9 hits (22867 docs)
Search more

KNAPSAck KEGG Drug (予定)

Disease Isolation 125 hits (176023 docs)
Alzheimer disease 32 hits (83115 docs)
Cardiomyopathy 23 hits (45563 docs)
Schizophrenia 21 hits (96985 docs)
DIAMOND-BLACKFAN ANEMIA 17 hits (23693 docs)
Acute myocardial infarction 16 hits (41798 docs)
Immunoglobulin 16 hits (252906 docs)
Osteoporosis 16 hits (56630 docs)
Neuroblastoma 15 hits (30843 docs)
Ganglion 15 hits (51975 docs)
Search more

ICD10 研究者リソース

Mouse Biosource Shinya Yamanaka 39 hits (50 docs)
Kazutoshi Takahashi 27 hits (35 docs)
Keisuke Okita 26 hits (29 docs)
Hideyuki Okano 15 hits (122 docs)
Yohei Okada 12 hits (27 docs)
Hirofumi Suemori 11 hits (37 docs)
Norio Nakatsuji 10 hits (36 docs)
Masayo Takahashi 10 hits (43 docs)
Yukio Nakamura 8 hits (34 docs)
Fumitaka Osakada 7 hits (21 docs)
Search more

マウスリソース Human Gene ヒト遺伝子

Mouse Gene Human Gene

Gene MRGPRX2, MAS-related GPR_member X2 1 hit (2 docs)
ESRRB, estrogen-related receptor bet... 1 hit (2 docs)
Disease SHH, sonic hedgehog 1 hit (2 docs)
SIX1, SIX homeobox 1 1 hit (3 docs)
TET2, tet oncogene family member 2 1 hit (2 docs)
PDX1, pancreatic and duodenal homeob... 1 hit (2 docs)
MYT1L, myelin transcription factor 1 1 hit (4 docs)
FBXO15, F-box protein 15 1 hit (2 docs)

Gene Human Gene

Document Set Human Gene

Mouse Gene Human Gene

Drug Human Gene

Researcher Human Gene

Dictionary Human Gene

Database Human Gene

About PosMed 研究者リソース

User's guide

RIKEN Mouse Biosource Search

マウスリソース

マウス遺伝子

KEGG Drug (予定)

研究者リソース

ヒト遺伝子

<http://biosparql.org/PosMed> で公開中

BIO Lod PosMed

biolod.org/PosMed/

PosMed Positional Medline

keyword

Tutorial in 2013 [[English](#)] [Sample trial \(search "diabetes" related mouse bioresource of RIKEN Bio Resource Center\)](#)

Search Categories

- Gene
- Bioresource
- Phenotype
- Disease
- Metabolite
- Drug
- Researcher
- Document set
- Database

PosMed: “Positional MEDLINE” assists your positional-cloning studies

Since whole genome sequences were first elucidated, knowledge-based ranking of candidate genes has become one of the most important bioinformatics tasks in the forward-genetics and positional-cloning approaches to identify phenotype-responsible gene mutations.

This task requires creating a form of artificial intelligence that can solve a genetic researcher's problems by learning computationally a vast amount of information accumulated in documents and published data.

We have developed a system named "PosMed," an artificial intelligence that guides you to the key information waiting to be discovered in the sea of data.

Please check [our tutorial](#) for details.

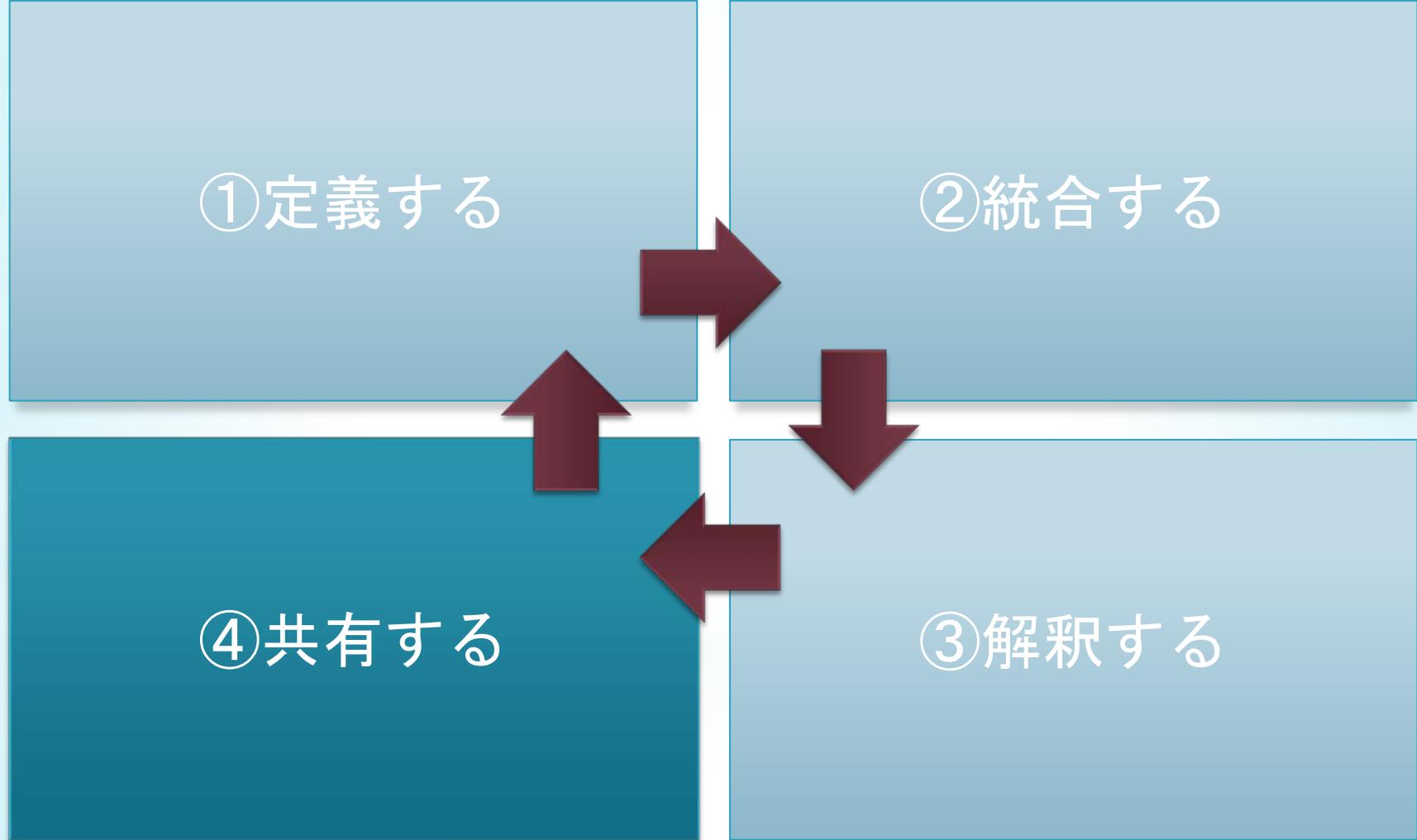
PosMed: inferential search of disease-related genes and bioresources

Bioresource search provides not only normal direct keyword search but also inference searches using various biomedical databases, including OMIM, MEDLINE, ICD-10. These indirect search functions are unique and not provided at any other bioresource databases.

RIKEN Mouse Bioresource Search

月間ユニークユーザー 12,000 人

フェノタイプ情報を共有する



成果共有(マウスリソースフェノーム)



C3H/HeN-WR1 オリジナルサイト

BRC ID	RBRC00111
正式系統名	C3H/HeN-KitW-Rbrc 1/Rbrc
由来生物の分類群	Mus musculus 関連情報を見る 関連情報を検索
系統カテゴリー	Spontaneous Mutation 関連情報を見る 関連情報を検索
アレル	Kit<W-Rbrc+ 関連情報を見る 関連情報を検索
遺伝的背景(系統)	C3H/HeN 関連情報を見る 関連情報を検索
寄託者	RIKEN BRC
系統の保存状態	Frozen embryos 関連情報を見る 関連情報を検索
提供までの期間	凍結卵より作出した子孫を2~4ヶ月以内に提供可能、 凍結胚を1ヶ月以内に提供可能 関連情報を見る
説明(日本語)	頭部腹部に白斑ができる。ホモは胎生期致死。
供与元	発注する

ディレクトリ

Basic Formal Ontology 1.1.1
CHEBI Chemical entities of biological interest
遺伝学オントロジー(GXO) Lite版
理研哺乳類統合データベース
RIKENBASE
マウス系統 (理研バイオリソースセンター)
Mouse strain
理研バイオリソース

マウス系統

遺伝的背景検査結果
Genotyping PCR protocol
表現型アノテーション
理研疾患モデルマウス
その他の疾患モデルマウス
理研疾患モデルマウス
生習慣病モデルマウス
理研疾患モデルマウス
精神・神経疾患モデルマウス
理研疾患モデルマウス
理研疾患モデルマウス
骨系疾患モデルマウス
理研疾患モデルマウス
感覚疾患モデルマウス
感覚疾患モデルマウス
理研疾患モデルマウス
血流疾患モデルマウス
理研疾患モデルマウス

理研ENU誘発変異マウスライン(BRC)

細胞株
ヒトiPS細胞
細胞株 (理研バイオリソースセンター)
ヒトES細胞
細胞株 (理研バイオリソースセンター)
動物iPS細胞
細胞株 (理研バイオリソースセンター)
動物ES細胞及び生殖細胞由来多能性幹細胞
細胞株 (理研バイオリソースセンター)
The Gene-trap Mouse ES cell clone
細胞株 (理研バイオリソースセンター)
ヒト及び動物由来の培養細胞株
細胞株 (理研バイオリソースセンター)
日本人由来不死化細胞株
細胞株 (理研バイオリソースセンター)
研究用ヒト間葉系幹細胞
細胞株 (理研バイオリソースセンター)
後藤コレクション細胞

マウス系統数:5,217
総インスタンス数:11,098
総トリプル数:76,076

リソース情報

フェノタイプ情報

各マウスリソースのフェノタイプ
を結ぶRDFグラフ

フェノタイプが類似するマウス
「おすすめ」機能

RDFやテーブル形式でデータを一括ダウンロード可能

成果共有(細胞リソースフェノーム)

CiRA00026 [オリジナルサイト]

[ページ一覧を見る]

種類	ヒトiPS細胞 (HPS) 関連情報を見る 関連情報を検索
細胞が由来した組織	skin
樹立者	Inoue, Haruhisa
Approver	http://www2.brc.riken.jp/lab/cell/address.cgi?cno=HPS0294
BRC ID	HPS0294
カテゴリ	iPS細胞, 分化能のある細胞
由来生物の分類群	Homo sapiens 関連情報を見る 関連情報を検索
由来生物（一般名）	human 関連情報を見る 関連情報を検索
説明（日本語）	疾患特異的iPS細胞株。筋萎縮性側索硬化症（TDP-43遺伝子変異）患者由来。HPS0292、HPS0293と同一人由来。エピソーマルスペクターにより6因子（Oct3/4, Sox2, Klf4, L-Myc, Lin28, shp53）を導入。提供依頼書式（C-0041、C-0056、C-0003、C-0058：必要な場合のみ）。
説明（英語）	Disease specific iPS cell line derived from a patient: Amyotrophic lateral sclerosis. HPS0293 and HPS0293 are derived from the same patient. Order Form (C-0042, C-0057, C-0007).
提供者	Nakahata, Tatsutoshi
供与元	発注する
臓器提供者あるいは個体	[person]HPS0292,HPS0293,HPS0294 関連情報を隠す 関連情報を検索 発育段階（臓器提供者、個体の） 60 yearsold 性別（臓器提供者、個体の） female 関連情報を検索 地域（臓器提供者、個体の） 日本 関連情報を見る 関連情報を検索 人種（臓器提供者） Mongoloid disease of donor ALS筋萎縮性側索硬化症（TDP-43遺伝子変異） "臓器提供者あるいは個体" が [person]HPS0292,HPS... の細胞株 CIRA00026, CIRA00025, CIRA00024 関連情報を見る 関連情報を検索
"BRC ID" が CiRA00026 の 培養条件	HPS0294_culture condition 関連情報を隠す 関連情報を検索 BRC ID CIRA00026 関連情報を検索 培養最適温度 37°C 二酸化炭素濃度 (%) .05 継代方法 0.25% trypsin+0.1% collagenase type IV+20% KSR+1mM CaCl2+PBS(-), 0.1% gelatin coated dish 関連情報を見る 関連情報を検索 継代頻度 subculture : once/4-6days. medium change : every day 関連情報を見る 関連情報を検索 フィーダー細胞条件 SNL 76/7 (X-rays:5000R or MMC) 1.5-2.5x10 ⁶ cells/100mm dish 関連情報を見る 関連情報を検索 培養液 Primate ES Cell Medium+4ng/ml human basic FGF 関連情報を見る 関連情報を検索 継代密度 1:3-5 split
"BRC ID" が CIRA00026 の 凍結保存条件	HPS0294_cryopreservation condition 関連情報を見る 関連情報を検索

細胞数: 3,676
総インスタンス数: 12,206
総トリプル数: 62,787

**iPS細胞のドナー情報
(個人情報は無し、
疾患情報は有り)**

**細胞特性、
培養環境情報**

- iPS細胞を含む、細胞特性／環境情報を共有
- Cell Line Ontology (CLO) を通じたグローバルな情報共有へ

成果共有(先端計測フェノーム)

This image shows a screenshot of the RIKENbase search interface. The main search bar at the top contains the query "FANTOM5の情報". Below the search bar, the results are displayed under the heading "共同研究者者の情報" (Information about joint researchers). The results include details about a researcher named Yukio NAKAMURA, such as their position (Postdoctoral Fellow), supervisor (Yukio NAKAMURA), and affiliation (RIKEN Center for Genomic Medicine). The interface also features a sidebar with various filters and links related to the database.



■ 細胞リソースデータベースと FANTOMプロジェクトデータベース のRDFによる統合化

- 全1,350系統の内、理研BRC等の細胞リソース配布機関に関連する582細胞系統(ヒト562、マウス20)
 - CAGE、転写因子、モチーフ、共発現等の計測フェノーム情報とともに統合作業中

成果共有(微生物リソースフェノーム)

画像



JCM Strains (理研 微生物系)
JCM 1432 [オリジナルサイト]

[ページ一覧を見る]

BRIC ID	JCM 1432	
来歴	<--ATCC 29900 ,<-- BCRC 14526 ,<-- CCM 6168 ,<-- CECT 4322 ,<-- CIP 106966 ,<-- DSM 1552 ,<-- NCTC 11211	
他機関の菌株番号	<input checked="" type="checkbox"/> 関連情報を隠す	
	他機関の菌株番号	機関名
ATCC 29900	American Type Culture Collection	
BCRC 14526	Bioresources Collection and Research Center, Food...	
CCM 6168	Czech Collection of Microorganisms, Masaryk Unive...	
CECT 4322	Colección Española de Cultivos Tipo, Universida...	
CIP 106966	Pasteur Institute Collection, Biological Resource...	
DSM 1552	DSMZ-Deutsche Sammlung von Mikroorganismen und Ze...	
NCTC 11211	National Collection of Type Cultures, Central Pub...	

学名 Clostridium spiroforme 関連情報を見る 関連情報を検索

供給元（所属） http://www.jcm.riken.jp/cgi-bin/jcm/jcm_number?JCM=1432

分離に関する説明 Human feces

至適生育温度 37°C

培養培地 関連情報を隠す

培養培地	培地名	機関
GRMD=13	BL AGAR (GLUCOSE BLOOD LIVER AGAR)	微生物材料開発室
GRMD=14	EG MEDIUM	微生物材料開発室

分類基準 Type strain

塩基配列番号 16S rRNA gene (X73441)

メタ情報

ID URN:RIKEN:cria381s1ria381u1432i
URI <http://scinets.org/item/cria381s1ria381u1432i>
帰属ページ http://www.jcm.riken.jp/cgi-bin/jcm/jcm_number?JCM=1432
最終更新日 2013-09-20

画像

PosMed Related Information in Other Databases

Drug	Metabolite	Document Set	Mouse Gene
• Folic Acid 1 hit • Folate, C19H19N7O6 1 hit • Thiamin, C12H17N4OS 1 hit	• Thiamine 1 hit • Folic acid 1 hit • Biotin 1 hit	• MEDLINE 1 hit	• mt-Rnr2, mitochondrially... 1 hit

13,732株
総トリプル数194,389
(今年度開始)

黒川Gの Microbial Culture Collection Vocabulary (MCCV) を用いたデータ記述

分離源、生育温度、培地等の、特性／環境情報

黒川G、DBCLSと連携してデータ記述をすすめている

成果共有(植物リソースフェノーム)

オントロジーによるシロイヌナズナのフェノタイプデータの統合

理研の旧植物センターのデータ統合

- CSRSで開発したシロイヌナズナ変異体の表現型情報を標準化・統合

Arabidopsis Phenome Database

IN Mutants Literature Curation Data List Contact Us

表現型(シロイヌナズナ変異体フェノーム情報)

petiole:decreased length

特性(PATO)

器官・組織(PO)

"観察された表現型"がpetiole:decreased lengthのシロイヌナズナ変異体観察情報

[ページ一覧を見る]

decreased length [関連情報を探す]

petiole [関連情報を探す]

標準化された表現型
「短い葉柄」

注目した表現型を示すバイオリソース(研究に役立つ変異体)のリスト

(1-650-1: "Leaves - short petiole" (Ds tagging line))

F13631_obs_15 (FOX hunting),

F13631_obs_16 (FOX hunting),

Z089635_obs_1 (Activation tagging line),

F29745_obs_3 (FOX hunting),

12-3613-1: "Leaves - short petiole" (Ds tagging line),

F29399_obs_8 (FOX hunting),

F03114_obs_6 (FOX hunting),

Z083211_obs_5 (Activation tagging line),

Z082802_obs_2 (Activation tagging line),

Z082802_obs_1 (Activation tagging line),

Z039935_obs_7 (Activation tagging line),

13-2505-1: "Leaves - short petiole" (Ds tagging line),

13-5315-1: "Leaves - short petiole" (Ds tagging line),

F13631_obs_17 (FOX hunting)

[関連情報を見る] [関連情報を探す]



表現型の原記載(バイオリソース開発者による)

世界のフェノタイプ文献の収集と整理

- 学術論文を一報づつ注意深く読解し、シロイヌナズナの表現型情報を収集・標準化・統合

Delayed flowering under long day length regimen

実験条件: under environment of long day length regimen

~の過程において: delayed flowering

品質(PATO): "表現型の変異"が delayed flowering under long day length regimen の現象 - 遺伝子

結果: delayed flowering under long day length regimen for AT1G26830, delayed flowering under long day length regimen for AT1G78370, delayed flowering under long day length regimen for AT3G04610, delayed flowering under long day length regimen for AT4G04865, delayed flowering under long day length regimen for AT4G31120, delayed flowering under long day length regimen for AT5G62000

phenotype-gene (Plant Phenome)

delayed flowering under long day length regimen for AT1G26830

annotation: delayed flowering under long day length regimen

has_mutation: AT1G26830

outcomePaper: Dieterle Monika et al. 2005 Feb. Plant J. 41(3):386-99

journal: The Plant Journal

Abstract: Culin proteins, which belong to multigenic families in all eukaryotes, associate with other proteins to form ubiquitin protein ligases (E3s) that target substrates for proteolysis by the 26S proteasome. In this study, we analyzed the function of the Culin genes of a plant, *Arabidopsis thaliana*, in contrast to fission yeast and mice. The genome of the model plant *Arabidopsis thaliana* contains two related CUL3 genes, called CUL3A and CUL3B. We found that CUL3A is ubiquitously expressed in plants and is able to interact with the ring-finger protein RBX1. A genomic search revealed the existence of at least 76 BTB-domain proteins in *Arabidopsis* belonging to 11 major families. Yeast two-hybrid experiments indicate that representative members of certain families are able to physically interact with both CUL3A and CUL3B, suggesting that CUL3A and CUL3B E3 protein complexes with other BTB-domain proteins. In order to determine the function of CUL3A, we used a reverse genetic approach. The cul3a null mutant flowers slightly later than the control plants. Furthermore, this mutant exhibits a reduced sensitivity of the inhibition of hypocotyl growth in far-red light and mis-expresses COP1. The viability of the mutant plants suggests functional redundancy between the two CUL3 genes in *Arabidopsis*.

表現型
「長日条件下で開花が遅い」

表現型－遺伝子の関係性の情報

文献

- 変異株フェノタイプ観察情報 14,631件
- 文献からのフェノタイプ情報 824件

Biophenome Linked Open Databases

<http://biolod.org>

The screenshot shows the BioLOD website homepage. At the top, there is a navigation bar with links for Home, BioLODについて, データベースリスト, チュートリアル, ユースケース, リンク, and ニュース. Below the navigation bar is a search bar with the placeholder "Enter keyword...". To the right of the search bar is a "Search" button. The main content area has a blue header bar with the text "BioLOD (Biophenome Linked Open Databases)". Below this, there is a section titled "BioLOD data statistics" containing a table with the following data:

	Databases	Classes	Instances	Literature
Open Data	139	715	9,366,352	272,163
Not Open Data	66	228	286,429	0
Total (Browsable)	205	943	9,652,781	272,163

as of January 2013.

Below the statistics, there is a section titled "Integrated databases in various field" with the following list:

- マウス系統 - マウス系統（理研バイオリソースセンター） (5115 Instances)
- 細胞株 - 細胞株（理研バイオリソースセンター） (3611 Instances)
- シロイヌナズナ変異株観察情報 - シロイヌナズナ変異体フェノーム情報 (14629 Instances)
- 表現型の変異 - 植物フェノーム (487 Instances)

At the bottom of the page, there is a "Facebookもチェック" button and a "Bio LOB.org" logo.

基盤普及活動実績

震災後の情報基盤提供の依頼を受け、分野を越えて日本のオープンデータ普及を支援
様々な分野から700を超えるデータセットのRDF公開を支援し、その関連活動が評価される

受賞歴（7件）

2011年12月	文部科学省科学技術政策研究所NISTEP “ナイスステップな研究者” 受賞
2012年3月	Linked Open Data Challenge Japan 2011 アプリケーション部門最優秀賞
2013年2月	International Open Data Hackathon Tokyo 優勝
2013年3月	Linked Open Data Challenge Japan 2012 未来ヘルスケア賞
2013年4月	International Space Apps Challenge Tokyo 2013 AWSアーキテクト賞
2013年4月	International Space Apps Challenge Tokyo 2013 samurai fab ヨコハマものづくり工房賞
2013年9月	Earth Communication Award ハッカソン東京 準優勝



2011年から、LODチャレンジのコンテストを開催し、日本のオープンデータ推進に大きく貢献

コンテスト開催（4件）

2011年	Linked Open Data Challenge Japan 2011
2012年	GenoCon2（合成生物学におけるゲノム設計の国際コンテスト）
2012年	Linked Open Data Challenge Japan 2012
2013年	Linked Open Data Challenge Japan 2013



LOD Challenge2013



外部セミナー講演（31件）

2011年	LODチャレンジデー開催（全4回、東京・大阪）
2012年5月	Linked Data 勉強会にて講演（東京）
2012年9月	NASA Ames Research Center にてセミナー講演（California）
2012年9月	New College of Florida にてセミナー講演（Florida）
2012年	LODチャレンジデー開催（全9回、東京・大阪・名古屋・鯖江・奈良・函館・横浜）
2013年	LODチャレンジデー開催（全4回、東京・石川・名古屋）
2013年9月	東京医科歯科大学大学院の特別セミナーにて講演（和光）



※ その他、直接関係のない外部組織においても本基盤を用いたセミナーが行われた実績有(例: 第1回岐阜オープンデータハッカソン)

まとめと将来展望

- フェノタイプの膨大な文献とセマンティックウェブを統合する「つながり検索」を実現
- バイオリソースに蓄積する知見を高度に活用できる
- バイオリソースに限らず幅広いリソースの成果情報を統合
- 細胞リソースに関するフェノタイプ情報の需要が今後見込まれる(細胞DBのユニークユーザ数は月間2,000人程度)

