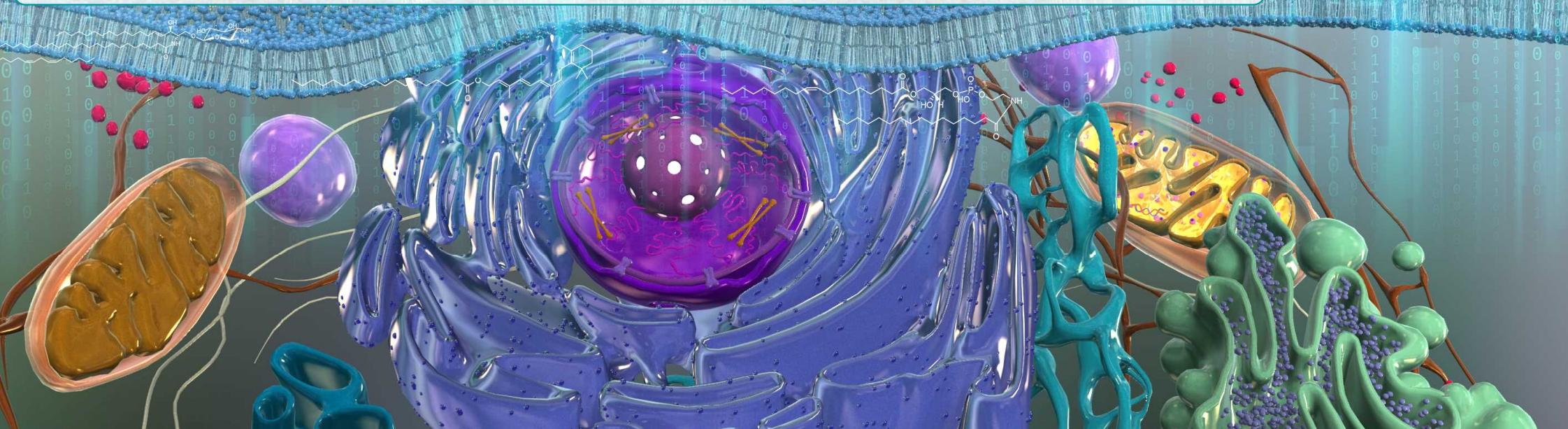
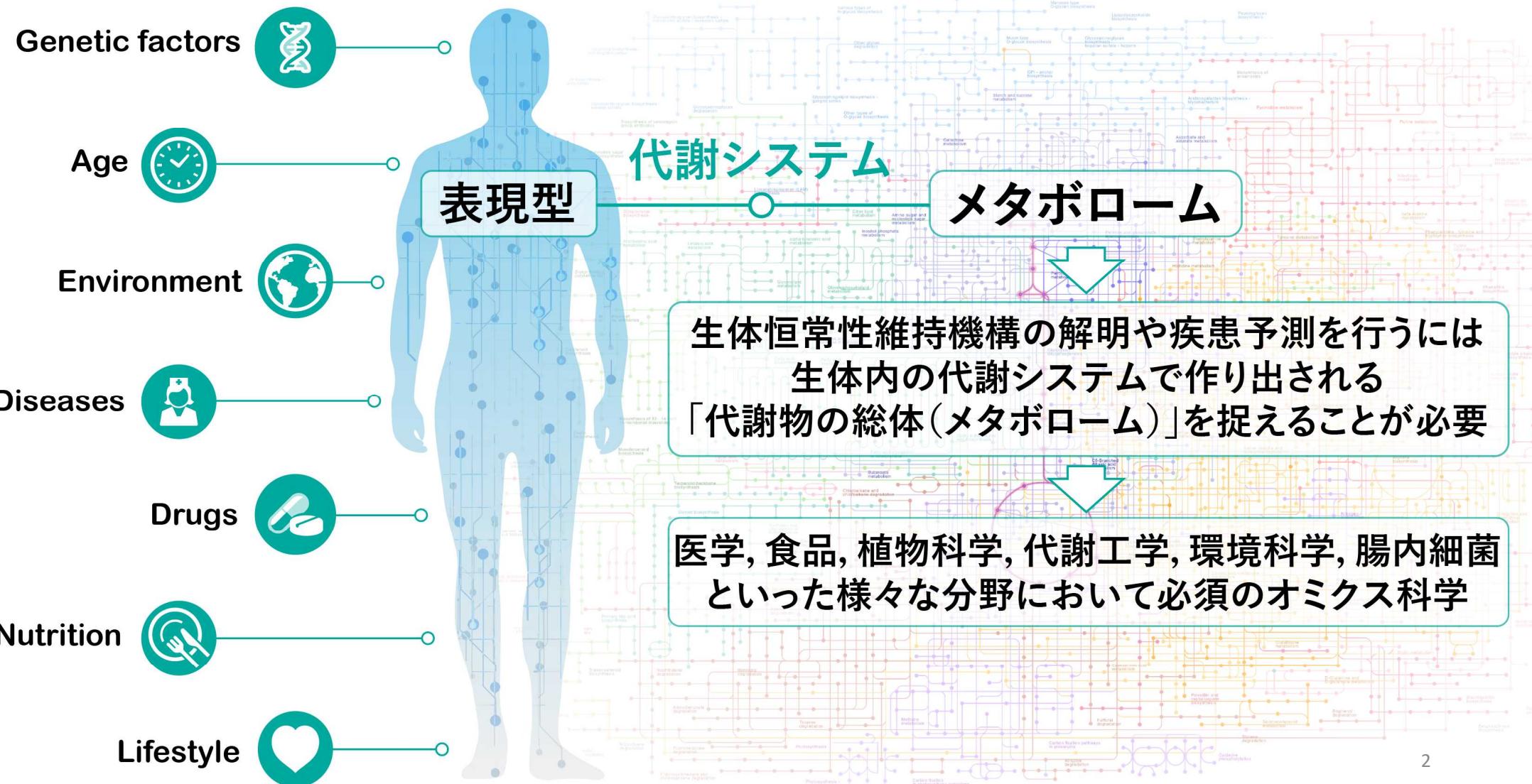


## 化合物データベース（代謝DBを中心に、KEGGやHMDBなど）

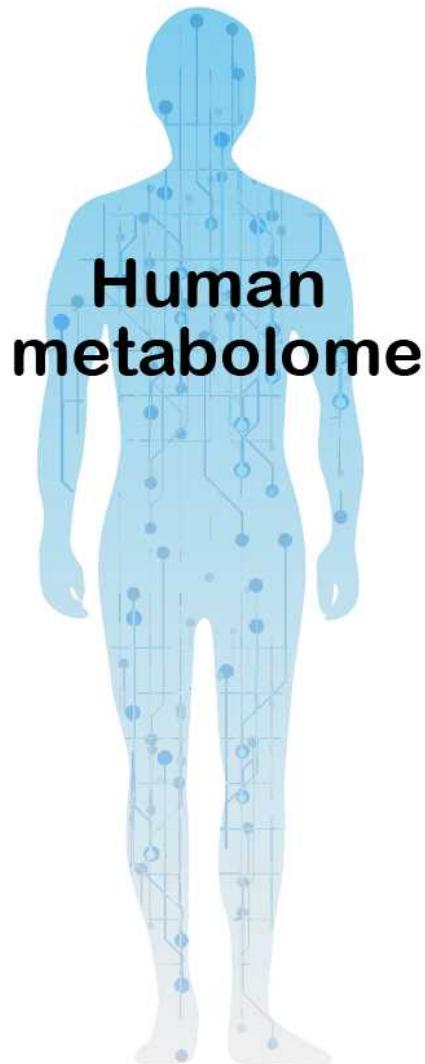
津川 裕司（つがわひろし）



# システムバイオロジーにおけるメタボローム解析の重要性



# メタボロームの多様性



Natural products:  
>>250,000 molecules



<http://www.bdsnatural.com/>

Compounds in foods:  
>>15,000 molecules

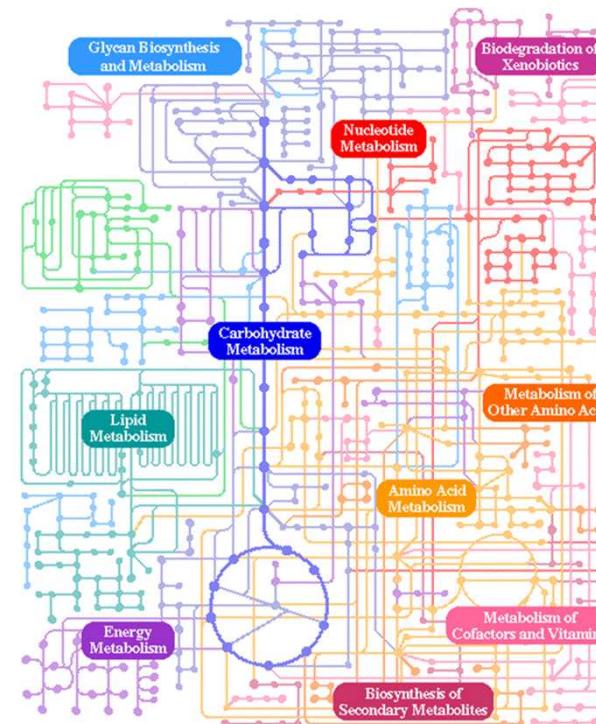


Various new metabolites  
from microbiome

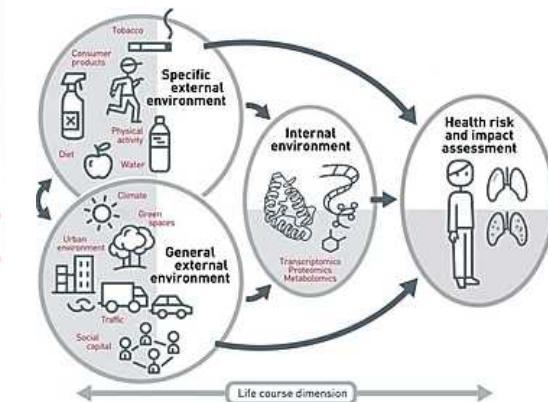


<http://snalime.com/health/gut-flora/>

Endogenous metabolites  
in HMDB: 3040



Human exposome  
> Million metabolites



# メタボロームデータを出す側と使う側

## 出す側

- 分析化学
- 質量分析
- イオン
- マススペクトル
- 質量分析インフォマティクス
- アノテーション
- データ標準化

## 使う側

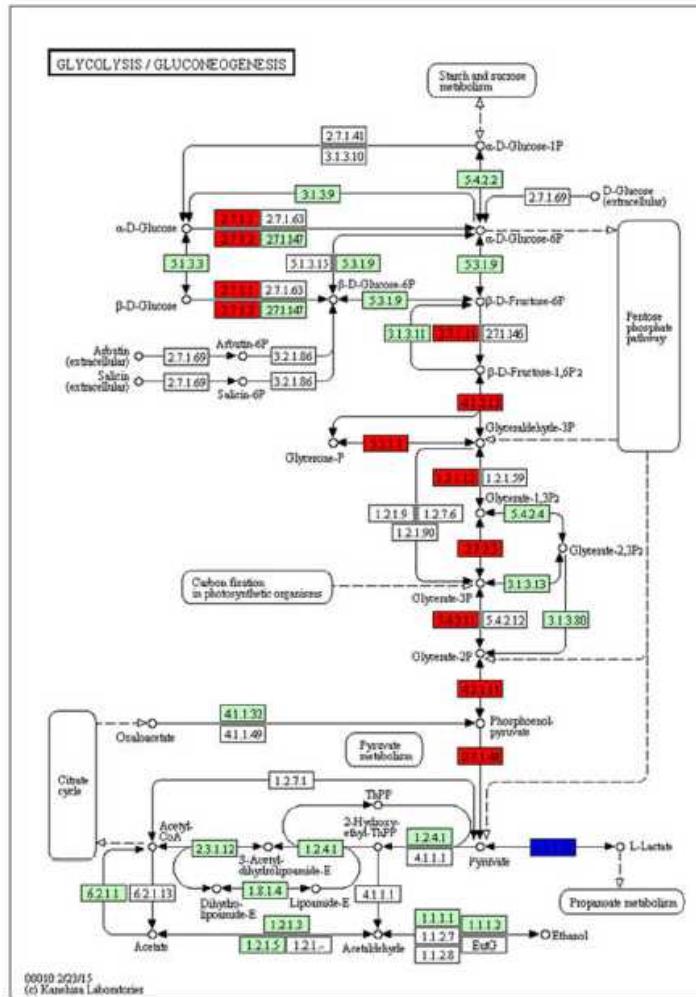
- データの特徴理解
- 化合物IDの取得
- データの正規化
- 統計・多変量解析
- データベース検索
- 代謝マップ投影
- 適切な数理モデルの選択
- オミクスモジュールへ統合

# メタボロームデータを使う側が求めるアウトプット

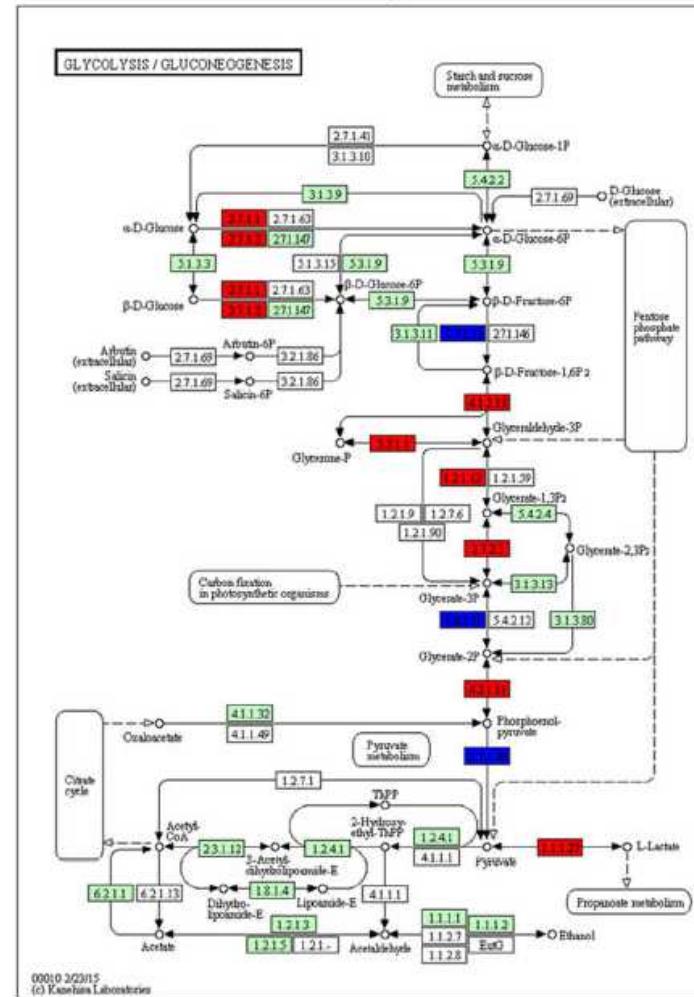
## 代謝物アノテーションのメタデータ・定量値が整理されてるテーブル

# KEGG MapperとかKEGG Scapeとか

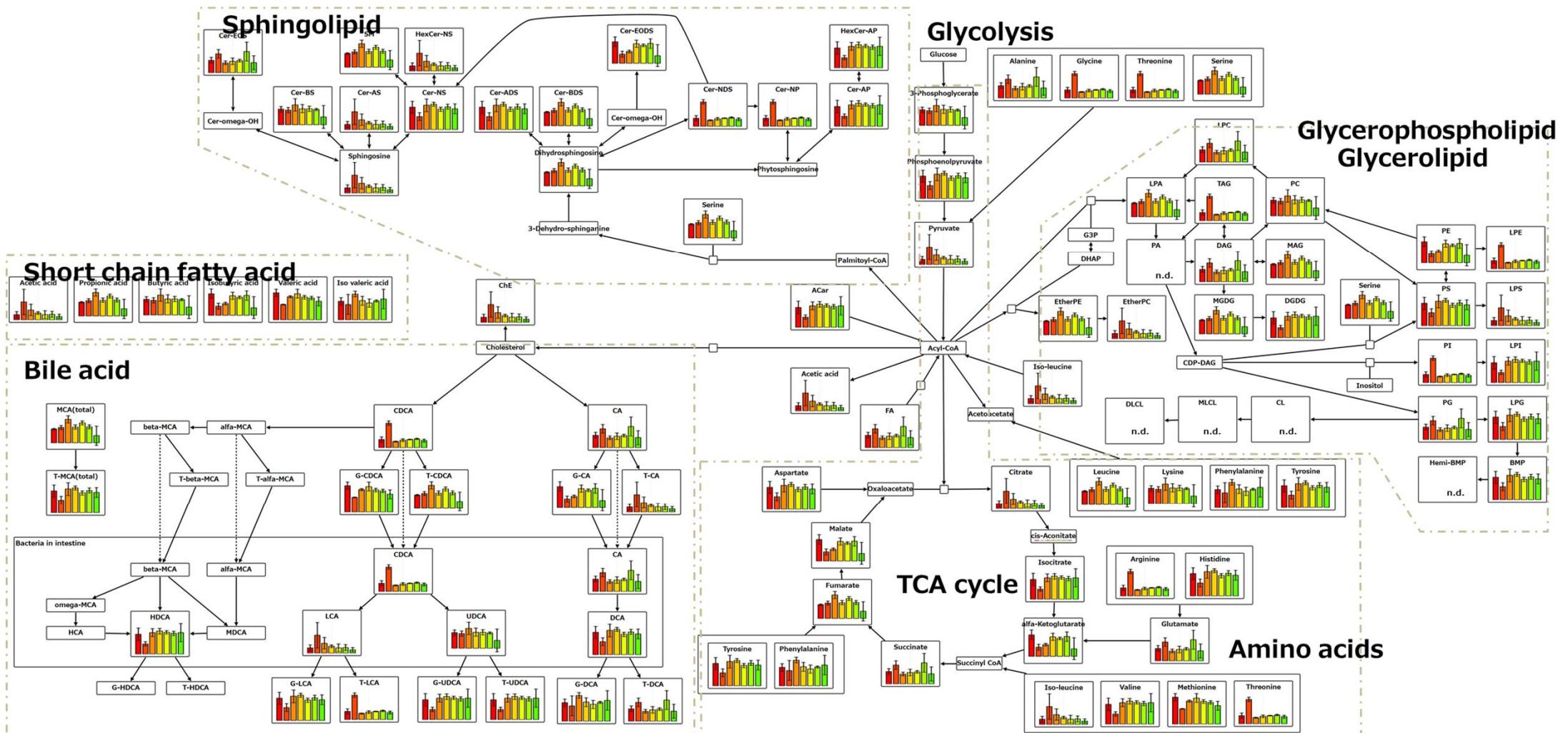
## Normal vs cancer



## Normal vs paracancer

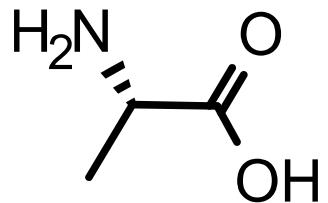


# PathVisioやVANTEDを使った代謝マップ投影



# 親水性メタボロミクスではInChIKeyが世界標準

Alanine



Synonyms: L-alanine, 56-41-7, H-Ala-OH, (S)-Alanine

IUPAC name: (2S)-2-aminopropanoic acid

SMILES: C[C@H](C(=O)O)N

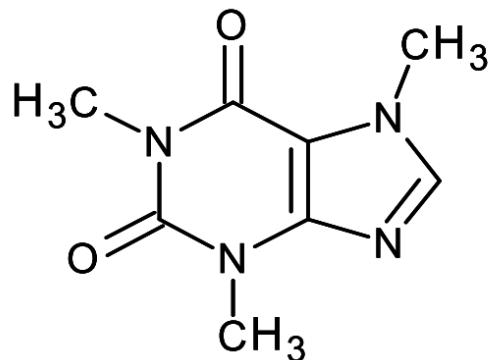
Formula: C<sub>3</sub>H<sub>7</sub>NO<sub>2</sub>

InChI: 1S/C3H7NO2/c1-2(4)3(5)6/h2H,4H2,1H3,(H,5,6)/t2-/m0/s1

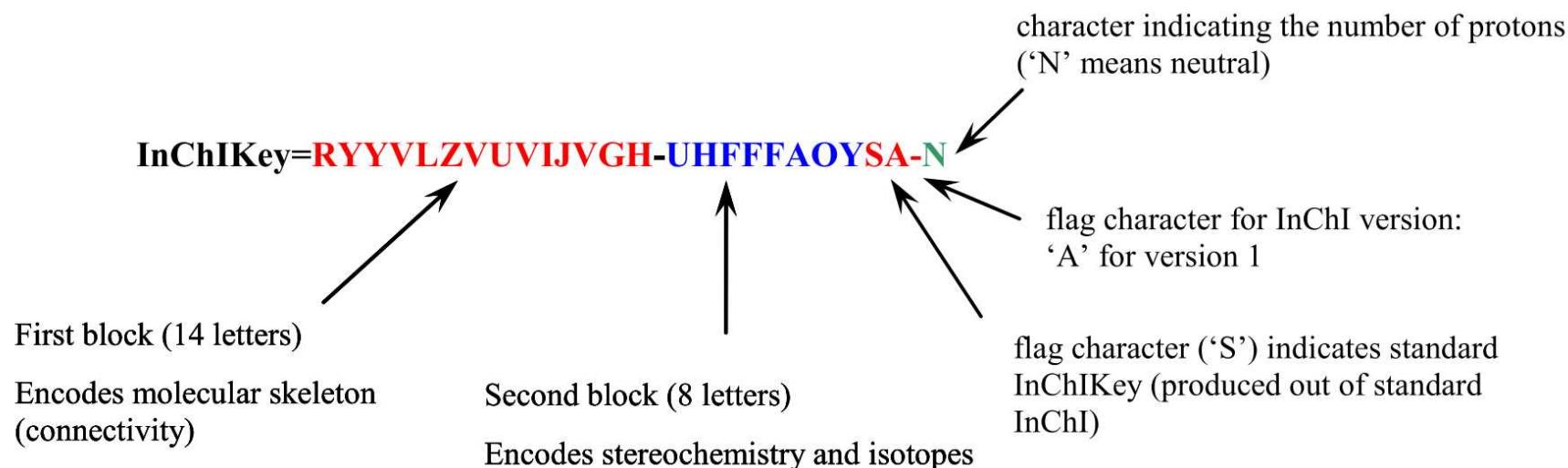
InChIKey: QNAYBMKLOCPYGJ-REOHCLBHSA-N

Ontology: Alanine and derivatives

# 親水性メタボロミクスではInChIKeyが世界標準



InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3 (caffeine)



# KEGG mapツールを使いたければKEGG IDが必要

Alanine

InChIKey: QNAYBMKLOCPYHGJ-REOHCLBHSA-N

The screenshot shows the 'Batch Conversion' page of the CTS service. The URL in the address bar is `cts.fiehnlab.ucdavis.edu/batch`. The main heading is 'Batch Conversion'. A text input field contains the InChIKey: `QNAYBMKLOCPYHGJ-REOHCLBHSA-N`. To the right of this is a dropdown menu set to 'InChIKey'. Below the input field is a 'Convert' button. Another dropdown menu is set to 'KEGG'. A file upload section shows 'Choose File' and 'No file chosen'. Below this, a search result table has one row: 'InChIKey' column shows 'QNAYBMKLOCPYHGJ-REOHCLBHSA-N'; 'KEGG' column shows 'C00041'. At the bottom, there are 'Download Style' (set to 'Table'), 'File Type' (set to 'CSV'), and a checked checkbox for 'Top Hit Only'. A blue 'Download' button is at the bottom right.

# KEGG mapツールを使いたければKEGG IDが必要

## Alanine

InChIKey: QNAYBMKLOCPYGJ-REOHCLBHSA-N

The screenshot shows two windows. On the left is the 'PubChem Identifier Exchange Service' interface. In the 'Input ID List' section, 'alanine' is entered into the text input field under 'Synonyms'. The 'Output IDs' dropdown is set to 'KEGG'. A large blue arrow points from this window to a right-hand window titled '3376337956341249430.txt - Notepad'. This notepad window displays a list of KEGG IDs for Alanine, including C01401, C00041, D00012, C00133, C01401, C00041, D00012, C00133, C01401, C00041, D00012, C00133, C01401, C00041, D00012, C00133.

KEGG ID
C01401
C00041
D00012
C00133
C01401
C00041
D00012
C00133
C01401
C00041
D00012
C00133

PubChem identifier exchange service

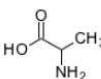
# KEGG mapツールを使いたければKEGG IDが必要

## Alanine

InChIKey: QNAYBMKLOCPYGJ-REOHCLBHSA-N

```
3376337956341249430.txt - Notepad
File Edit Format View Help
alanine C01401
alanine C00041
alanine D00012
alanine C00133
Ln 1, Col 1 100% Unix (LF) UTF-8
```



KEGG COMPOUND: C01401	
Entry	C01401 Compound
Name	Alanine; 2-Aminopropionic acid; 2-Aminopropanoic acid
Formula	C3H7NO2
Exact mass	89.0477
Mol weight	89.0932
Structure	 C01401
	[Mol file] [KCF file] [DB search]
Reaction	R03542 R05840
Pathway	map00460 Cyanoamino acid metabolism map01100 Metabolic pathways
Enzyme	3.5.5.1
Brite	Risk category of Japanese OTC drugs [BR:br08312] Third-class OTC drugs Inorganic and organic chemicals L-Alanine C01401 Alanine [BRITE hierarchy]
Other DBs	CAS: 302-72-7 PubChem: 4590 ChEBI: 16449 CHEMBL: CHEMBL12198 3DMET: B04793 NIKKAJI: J44.439A
KCF data	[Show]

KEGG COMPOUND: C00041	
Entry	C00041 Compound
Name	L-Alanine; L-2-Aminopropionic acid; L-alpha-Alanine
Formula	C3H7NO2
Exact mass	89.0477
Mol weight	89.0932
Structure	 C00041
	[Mol file] [KCF file] [DB search]
Remark	Same as: D00012
Reaction	R00258 R00369 R00396 R00397 R00398 R00400 R00401 R00453 R00576 R00585 R00692 R00863 R00907 R00987 R01215 R01261 R01712 R02050 R02271 R02293 R02668 R02970 R03001 R03038 R03193 R03210 R03502 R03599 R03854 R03936 R04112 R04152 R04187 R05652 R07460 R08197 R08714 R08872 R09048 R09254 R10124 R10178 R10180 R10992 R11038 R11064 R11528 R11583 R12343 R12417 R12419 R12622
Pathway	map00250 Alanine, aspartate and glutamate metabolism map00270 Cysteine and methionine metabolism map00430 Taurine and hypotaurine metabolism map00450 Selenocompound metabolism map00473 D-Alanine metabolism map00710 Carbon fixation in photosynthetic organisms map00970 Aminoacyl-tRNA biosynthesis map00998 Biosynthesis of various secondary metabolites - part 2 map01060 Biosynthesis of plant secondary metabolites map01100 Metabolic pathways map01110 Biosynthesis of secondary metabolites map01120 Microbial metabolism in diverse environments map01200 Carbon metabolism map01230 Biosynthesis of amino acids map01502 Vancomycin resistance map02010 ABC transporters map04122 Sulfur relay system map04974 Protein digestion and absorption map04978 Mineral absorption map05230 Central carbon metabolism in cancer

# KEGG IDを揃えておくと簡単なPathway mappingが可能

[https://www.genome.jp/kegg/tool/map\\_pathway2.html](https://www.genome.jp/kegg/tool/map_pathway2.html)

The image shows two side-by-side Notepad windows. The left window is titled 'kegg\_mapper\_colors.txt' and contains a list of KEGG IDs followed by their color codes. The right window is titled 'pathway\_rno04910\_rno...' and contains a similar list. Both files use the same color mapping scheme.

KEGG ID	Color
C00002	green
C00003	green
C00004	green
C00005	green
C00006	green
C00008	green
C00010	green
C00015	green
C00016	black
C00019	green
C00020	green
C00021	blue
C00024	green
C00025	green
C00025	red
C00026	red
C00029	green
C00035	green
C00037	green
C00041	red
C00042	green
C00043	green
C00044	green
C00047	green
C00049	blue
C00049	green
C00051	green
C00055	green
C00062	green
C00063	green
C00064	green
C00065	green
C00073	green
C00074	red
C00075	green
C00077	red
C00078	green
C00079	green
C00082	green
C00083	green
C00085	blue
C00086	green
C00091	red
C00092	blue
C00093	green
C00096	green
C00097	black
C00099	green
C00103	green
C00105	green
C00111	green
C00112	green

The screenshot shows the KEGG Mapper interface. At the top, there's a logo and the title 'KEGG Mapper – Search&Color Pathway'. On the left, a sidebar lists various tools: About KEGG Mapper, Reconstruct Pathway (and Brite, Module), Search Pathway (and Brite, Module, Network, Disease), Search&Color Pathway (and Brite, Module), Color Pathway, Join Brite, Convert ID, Annotate Sequence, BlastKOALA, Map Taxonomy, and KEGG. The main area has sections for 'Target databases', 'Search mode', and 'Enter objects one per line followed by bgcolor, fgcolor:'. A scrollable text area contains the color-coded KEGG ID list from the Notepad windows. Below this, there are options for file upload, changing default bgcolor, including aliases, using uncolored diagrams, searching for pathways containing all objects, and executing the search.

# KEGG IDを揃えておくと簡単なPathway mappingが可能

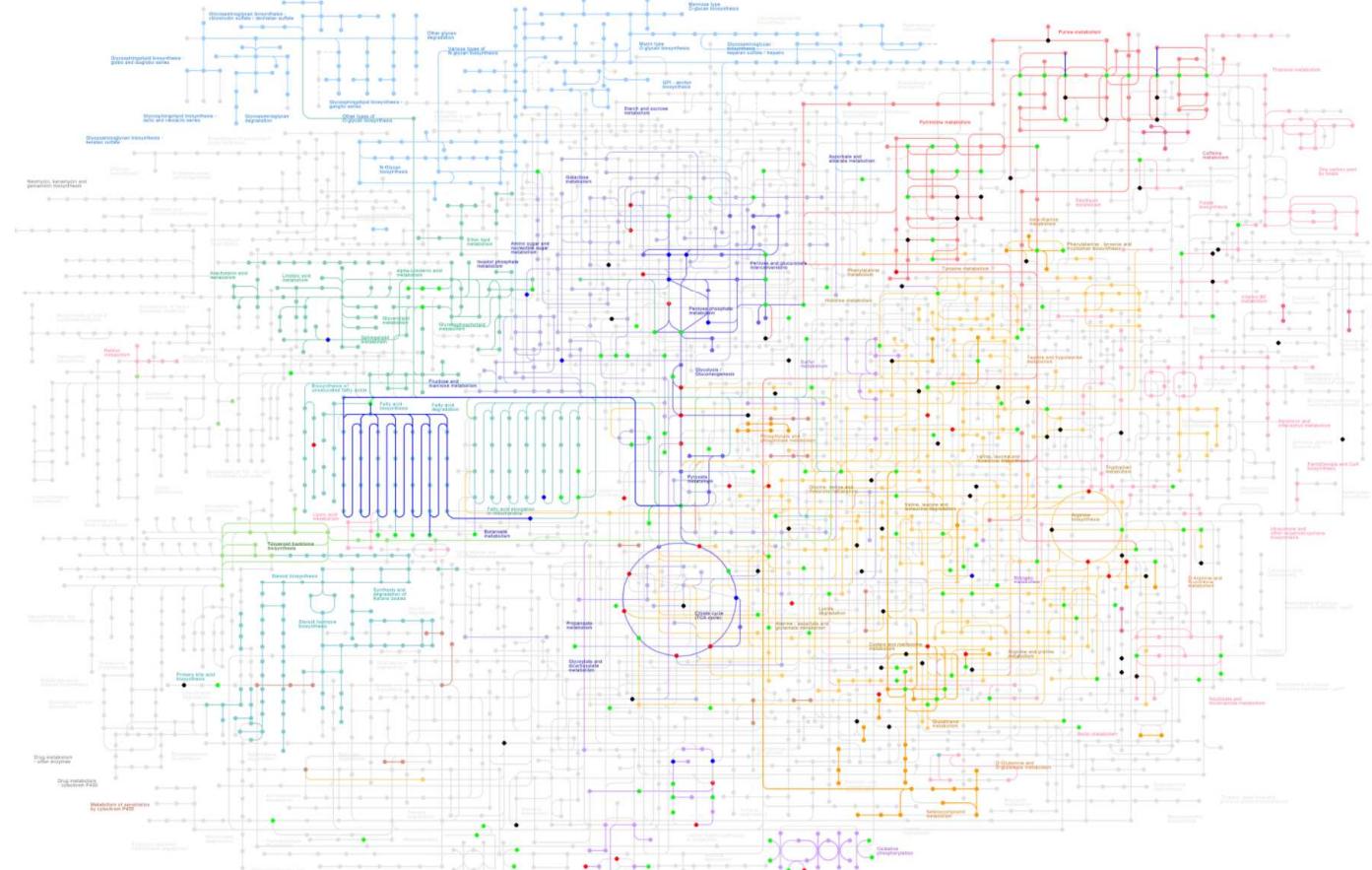
[https://www.genome.jp/kegg/tool/map\\_pathway2.html](https://www.genome.jp/kegg/tool/map_pathway2.html)

The image shows two side-by-side Notepad windows. The left window is titled 'kegg\_mapper\_colors.txt - Notepad' and contains a list of KEGG IDs and their corresponding colors. The right window is titled 'pathway\_rno04910\_rno...' and contains a list of KEGG IDs and their corresponding RNO numbers.

KEGG ID	Color
C0002	green
C0003	green
C0004	green
C0005	green
C0006	green
C0008	green
C0010	green
C0015	green
C0016	black
C0019	green
C0020	green
C0021	blue
C0024	green
C0025	green
C0025	red
C0026	red
C0029	green
C0035	green
C0037	green
C0041	red
C0042	green
C0043	green
C0044	green
C0047	green
C0049	blue
C0049	green
C0051	green
C0055	green
C0062	green
C0063	green
C0064	green
C0065	green
C0073	green
C0074	red
C0075	green
C0077	red
C0078	green
C0079	green
C0082	green
C0083	green
C0085	blue
C0086	green
C0091	red
C0092	blue
C0093	green
C0096	green
C0097	black
C0099	green
C0103	green
C0105	green
C0111	green
C0112	green

KEGG ID	RNO number
rno:24697	gray
rno:84006	blue
rno:293508	red
rno:116636	green
rno:24244	red
rno:24362	gray
rno:25623	gray
rno:287871	red
rno:24242	red
rno:29304	red
rno:25233	green
rno:50663	red
rno:170851	red
rno:24679	gray
rno:78968	gray
rno:686098	red
rno:24855	green
rno:24185	red
rno:84027	red
rno:29516	blue
rno:29699	blue
rno:50689	red
rno:83840	red
rno:501563	gray
rno:81745	gray
rno:362282	gray
rno:361377	gray
rno:60581	green
rno:678739	blue
rno:60445	red
rno:56718	gray
rno:25636	red
rno:114486	green
rno:50671	blue
rno:29376	green
rno:313845	red
rno:24703	green
rno:78975	red
rno:25467	gray
rno:361042	gray
rno:83803	green
rno:64639	red
rno:64363	blue



# HMDBは代謝物の情報が豊富に集積されている

HMDB Browse Search Downloads About Contact Us

Search metabolites Q Search

TMIC The Metabolomics Innovation Centre Your source for quantitative metabolomics technologies and bioinformatics.

Showing metabocard for D-2-Hydroxyglutaric acid (HMDB0000606)

Jump To Section: Identification Taxonomy Ontology Physical properties Spectra Biological properties Concentrations Links References XML

enzymes (3) Show 3 proteins Show Metabolites with Similar Structures

Record Information

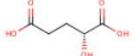
Version	4.0
Status	Detected and Quantified
Creation Date	2005-11-16 15:48:42 UTC
Update Date	2020-10-09 20:59:51 UTC
HMDB ID	HMDB0000606
Secondary Accession Numbers	<ul style="list-style-type: none"><li>HMDB0002323</li><li>HMDB00606</li><li>HMDB02323</li></ul>

Metabolite Identification

Common Name	D-2-Hydroxyglutaric acid
Description	In humans, D-2-hydroxyglutaric acid is formed by a hydroxyacid-oxoacid transhydrogenase whereas in bacteria it is formed by a 2-hydroxyglutarate synthase. D-2-Hydroxyglutaric acid is also formed via the normal activity of hydroxyacid-oxoacid transhydrogenase during conversion of 4-hydroxybutyrate to succinate semialdehyde. The compound can be converted to alpha-ketoglutaric acid through the action of a 2-hydroxyglutarate dehydrogenase (EC 1.1.99.2). In humans, there are two such enzymes (D2HGDH and L2HGDH). Both the D and the L stereoisomers of hydroxyglutaric acid are found in body fluids. D-2-Hydroxyglutaric acid is a biochemical hallmark of the inherited neurometabolic disorder D-2-hydroxyglutaric aciduria (OMIM: 600721) and the genetic disorder glutaric aciduria II. D-2-Hydroxyglutaric aciduria (caused by loss of D2HGDH or gain of function of LHD) is rare, with symptoms including cancer, macrocephaly, cardiomyopathy, mental retardation, hypotonia, and cortical blindness. An elevated urine level of D-2-hydroxyglutaric acid has been reported in patients with spondyloenchondrodysplasia (OMIM: 271550). D-2-Hydroxyglutaric acid can be converted to alpha-ketoglutaric acid through the action of 2-hydroxyglutarate dehydrogenase (D2HGDH). Additionally, the enzyme D-3-phosphoglycerate dehydrogenase (PHGDH) can catalyze the NADH-dependent reduction of alpha-ketoglutarate (AKG) to D-2-hydroxyglutarate (D-2HG). Nyhan et al. (1995) described 3 female patients, 2 of them sibs, who were found to have excess accumulation of D-2-hydroxyglutaric acid in the urine. The phenotype was quite variable, even among the sibs, but included mental retardation, macrocephaly with cerebral atrophy, hypotonia, seizures, and involuntary movements. One of the patients developed severe intermittent vomiting and was given a

Read more...

Structure



Chemical structure of D-2-Hydroxyglutaric acid: A four-carbon chain with a carboxyl group at one end and a hydroxyl group at the second carbon from the carboxyl group. The hydroxyl group is shown with a wedge bond, indicating its stereochemistry.

MOL SDF PDB SMILES InChI

Nucleic Acid Res. 46, D608-D617 (2018)

# HMDBは代謝物の情報が豊富に集積されている

Normal Concentrations								
Biospecimen	Status	Value	Age	Sex	Condition	Reference	Details	
Blood	Detected and Quantified	0.280-0.930 uM	Infant (0-1 year old)	Male	Normal	8981317		
Blood	Detected and Quantified	0.3-0.9 uM	Infant (0-1 year old)	Both	Normal	22391998		
Blood	Detected and Quantified	0.7(0.3-0.9) uM	Infant (0-1 year old)	Female	Normal	7609436		
Blood	Detected and Quantified	0.7 +/- 0.2 uM	Adult (>18 years old)	Both	Normal	8134166		
<a href="#">Show more...</a>								
Abnormal Concentrations								
Biospecimen	Status	Value	Age	Sex	Condition	Reference	Details	
Blood	Detected and Quantified	217 uM	Infant (0-1 year old)	Male	D-2-hydroxyglutaric aciduria	8981317		
Blood	Detected and Quantified	2.6(2.2-3.2) uM	Infant (0-1 year old)	Both	Combined D,L-2-hydroxyglutaric aciduria	22391998		
Blood	Detected and Quantified	366(99-757) uM	Infant (0-1 year old)	Female	D-2-Hydroxyglutaric aciduria II	7609436		
Blood	Detected and Quantified	68(26-123) uM	Infant (0-1 year old)	Female	D-2-Hydroxyglutaric aciduria I	7609436		
<a href="#">Show more...</a>								
Associated Disorders and Diseases								
Disease References	<b>D-2-hydroxyglutaric aciduria</b>							
	1. Gibson KM, ten Brink HJ, Schor DS, Kok RM, Bootsma AH, Hoffmann GF, Jakobs C: Stable-isotope dilution analysis of D- and L-2-hydroxyglutaric acid: application to the detection and prenatal diagnosis of D- and L-2-hydroxyglutaric acidemias. <i>Pediatr Res.</i> 1993 Sep;34(3):277-80. [PubMed:8134166]							
	2. Chalmers RA, Lawson AM, Watts RW, Tavill AS, Kamerling JP, Hey E, Ogilvie D: D-2-hydroxyglutaric aciduria: case report and biochemical studies. <i>J Inherit Metab Dis.</i> 1980;3(1):11-5. [PubMed:6774165]							
	3. Geerts Y, Renier WO, Bakkeren J, de Jong J: 2-Hydroxyglutaric aciduria: a case report on an infant with the D-isomeric form with review of the literature. <i>J Neurol Sci.</i> 1996 Nov;143(1-2):166-9. [PubMed:8981317]							
	4. G.Frauenfeld-Egger, Friedrich K. Trefz (2017). MetaGene: Metabolic & Genetic Information Center (MIC: <a href="http://www.metagene.de">http://www.metagene.de</a> ). METAGENE consortium.							
D, L-2-hydroxyglutaric aciduria								
	1. Kranendijk M, Struys EA, Salomons GS, Van der Knaap MS, Jakobs C: Progress in understanding 2-hydroxyglutaric acidurias. <i>J Inherit Metab Dis.</i> 2012 Jul;35(4):571-87. doi: 10.1007/s10545-012-9462-5. Epub 2012 Mar 6. [PubMed:22391998]							
	2. Muntau AC, Roschinger W, Merkenschlager A, van der Knaap MS, Jakobs C, Duran M, Hoffmann GF, Roscher AA: Combined D-2- and L-2-hydroxyglutaric aciduria with neonatal onset encephalopathy: a third biochemical variant of 2-hydroxyglutaric aciduria? <i>Neuropediatrics.</i> 2000 Jun;31(3):137-40. doi: 10.1055/s-2000-7497. [PubMed:10963100]							
	3. G.Frauenfeld-Egger, Friedrich K. Trefz (2017). MetaGene: Metabolic & Genetic Information Center (MIC: <a href="http://www.metagene.de">http://www.metagene.de</a> ). METAGENE consortium.							
Spondyloenchondroplasia								
	1. Talkhani IS, Saklatvala J, Dwyer J: D-2-hydroxyglutaric aciduria in association with spondyloenchondromatosis. <i>Skeletal Radiol.</i> 2000 May;29(5):289-92. [PubMed:10883451]							

# MetaboAnalystのウェブサイトに行くと意外に何でもできる

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Click a module to proceed, or scroll down for more details:

The homepage features a central circular hub containing various analytical modules, each represented by a colored circle with a white border. The modules include: Statistical Analysis (blue), Enrichment Analysis (green), Pathway Analysis (green), MS Peaks to Pathways (cyan), MS Spectral Processing (cyan), Joint Pathway Analysis (green), Network Explorer (green), Other Utilities (grey), Biomarker Meta-analysis (blue with diagonal stripes), Power Analysis (blue), Time-series / Two-factor (blue), Biomarker Analysis (blue), and Other Utilities (grey). To the right of the hub is a legend with the following items:

- Display R command history
- Targeted or untargeted metabolomics
- Targeted or annotated metabolomics
- Untargeted metabolomics
- Multiple metabolomics data
- Integrating other omics

On the left side of the page is a vertical sidebar with the following links:

- [Home](#)
- [User Stats](#)
- [Overview](#)
- [Data Formats](#)
- [FAQs](#)
- [Tutorials](#)
- [MetaboAnalystR](#)
- [Contact](#)
- [Resources](#)
- [APIs](#)
- [Update History](#)
- [About](#)

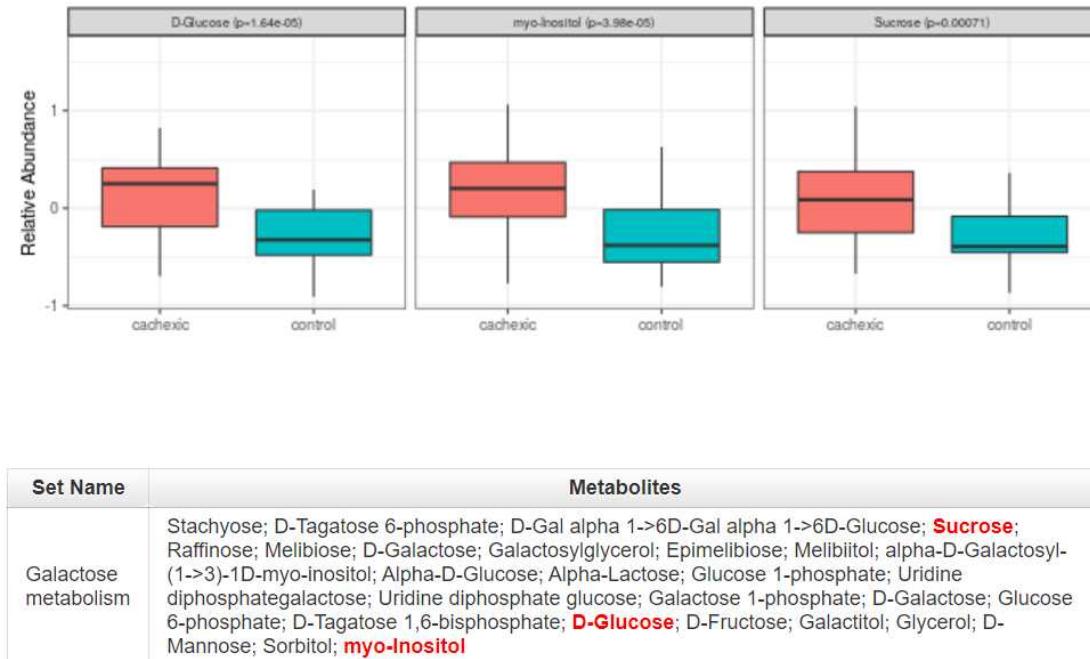
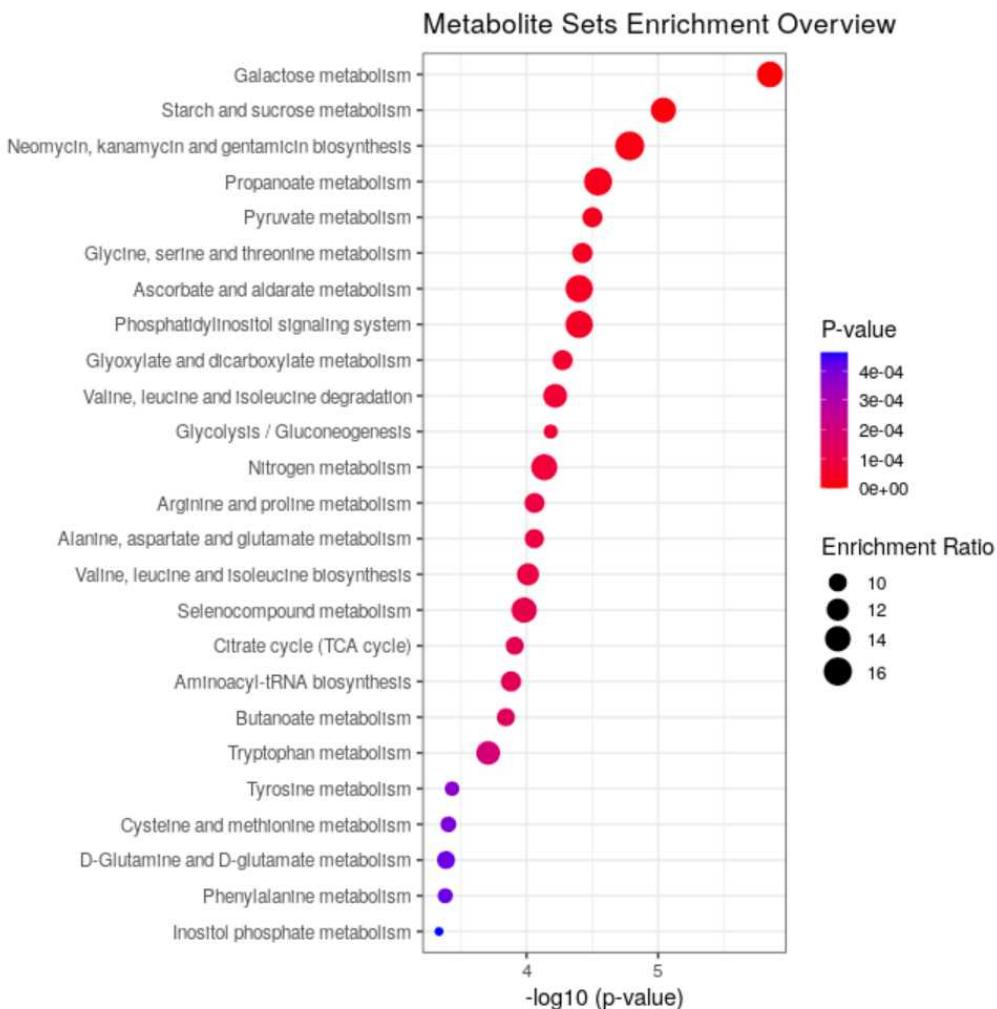
At the bottom of the sidebar are two logos:

- GenomeCanada
- GenomeQuébec

# MetaboAnalystのウェブサイトに行くと意外に何でもできる

*Current Protocols in Bioinformatics* 68, e86 (2019)

# MetaboAnalystにおけるEnrichment Analysis



Set Name	Metabolites
Galactose metabolism	Stachyose; D-Tagatose 6-phosphate; D-Gal alpha 1->6D-Gal alpha 1->6D-Glucose; <b>Sucrose</b> ; Raffinose; Melibiose; D-Galactose; Galactosylglycerol; Epimelibiose; Melibitol; alpha-D-Galactosyl-(1->3)-1D-myo-inositol; Alpha-D-Glucose; Alpha-Lactose; Glucose 1-phosphate; Uridine diphosphategalactose; Uridine diphosphate glucose; Galactose 1-phosphate; D-Galactose; Glucose 6-phosphate; D-Tagatose 1,6-bisphosphate; <b>D-Glucose</b> ; D-Fructose; Galactitol; Glycerol; D-Mannose; Sorbitol; <b>myo-Inositol</b>

# MetaboAnalystにおける統合解析



MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data



Upload

▶ Integrative Analysis  
Download  
Exit

Please upload a gene list and a metabolite list below.

Gene List

Metabolite List

Gene list with optional fold changes

```
#Official logFC
AASS -0.139042168
ACAA2 1.401267672
ACADL -2.608712824
ACADM -0.876538515
ACADS 0.150535255
ACADSB -1.637743607
ACHE 2.567118372
ACSM1 -2.348501729
ACTA2 -0.282176735
ACTB 1.559623747
ACTC1 -1.690352151
ADCY1 2.916857724
ADH1A -0.87610472
AGL -0.399133917
AGTR1 -1.078340189
AKR1A1 2.178398898
AKR1B1 -1.077265882
AKT2 0.2815620216
```

ID Type: Official Gene Symbol

Compound list with optional fold changes

```
#KEGG logFC
C0006 0.512160717
C0024 0.351757155
C0026 -2.669056963
C0029 0.379186578
C0031 1.669222153
C0047 -2.492289379
C0049 2.963835134
C0062 -2.558919927
C0064 1.77810046
C0072 0.632536475
C0077 -2.09045808
C0084 0.347392968
C0089 -1.460843412
C0097 3.046798674
C0101 -1.495004303
C0109 0.476718643
C0111 -2.672997377
C02117 0.572115007
```

ID Type: KEGG ID

Specify organism:

Homo sapiens (human)

Try our example data

Submit

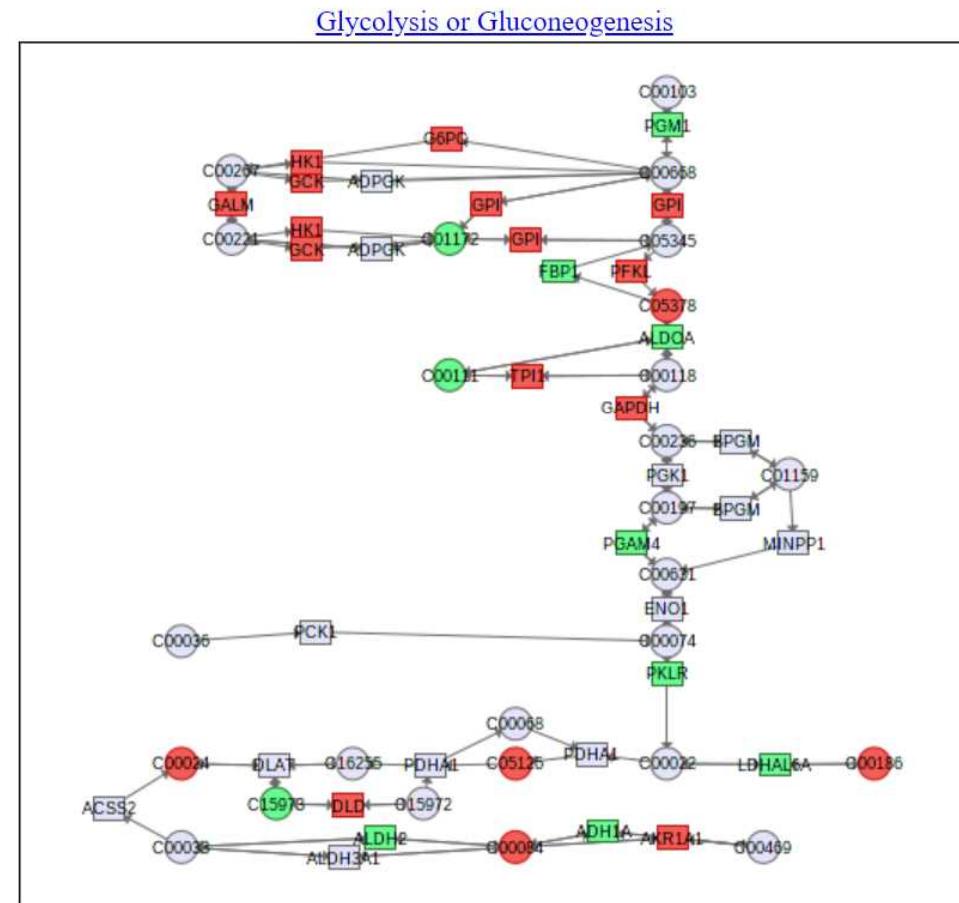
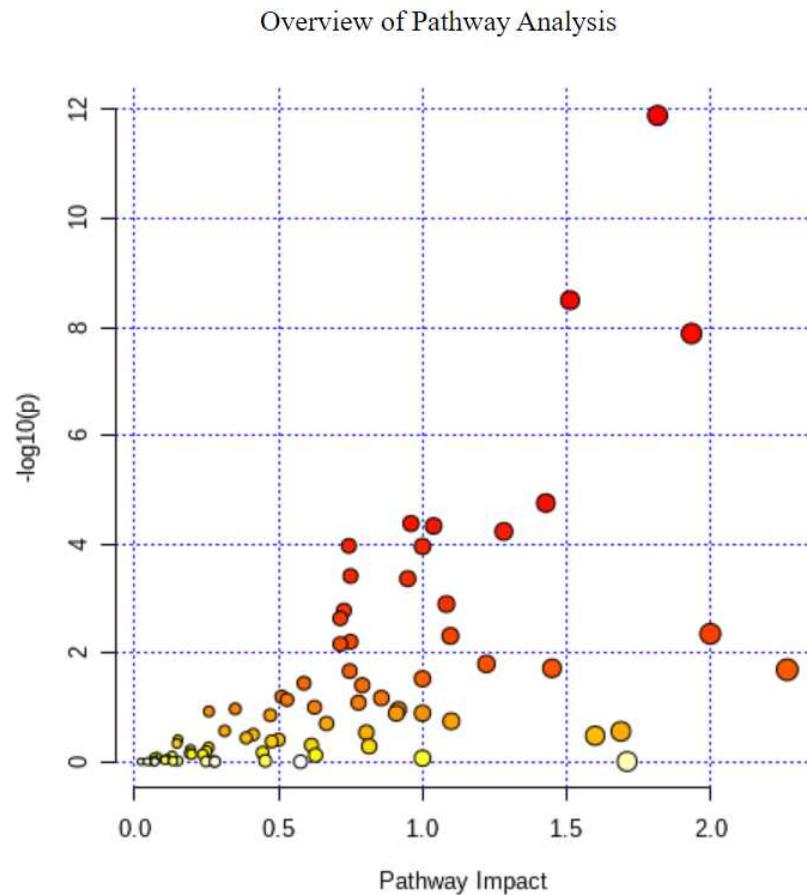
R Command History

Keep collapsed

Save

```
1. mSet<-InitDataObjects("conc", "msetqea", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "roww", "disc");
3. mSet<-CrossReferencing(mSet, "name");
4. mSet<-CreateMappingResultTable(mSet)
5. mSet<-SanityCheckData(mSet)
6. mSet<-ReplaceMin(mSet);
7. mSet<-PreparePrenormData(mSet)
8. mSet<-Normalization(mSet, "NULL", "LogNorm", "MeanCenter", ratio=FALSE, ratioNum=20)
9. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)
10. mSet<-PlotSampleNormSummary(mSet, "snorm_0_", "png", 72, width=NA)
11. mSet<-SetMetabolomeFilter(mSet, F);
12. mSet<-SetCurrentMsetLib(mSet, "kegg_pathway", 2);
13. mSet<-CalculateGlobalTestScore(mSet)
14. mSet<-PlotQEA.Overview(mSet, "qea_0_", "net", "png", 72, width=NA)
15. mSet<-PlotEnrichDotPlot(mSet, "qea", "qea_dot_0_", "png", 72, width=NA)
16. mSet<-PlotQEA.MetSet(mSet, "Galactose metabolism", "png", 72, width=NA)
17. mSet<-PlotQEA.MetSet(mSet, "Galactose metabolism", "png", 72, width=NA)
18. mSet<-PlotQEA.MetSet(mSet, "Galactose metabolism", "png", 72, width=NA)
19. mSet<-PreparePrenormData(mSet)
20. mSet<-SanityCheckData(mSet)
21. mSet<-CreateMappingResultTable(mSet)
```

# MetaboAnalystにおける統合解析



# MetaboAnalystにおける統合解析

Click the corresponding **Pathway Name** to view its graphical presentation; click **Match Status** to view the pathway members (with matched ones highlighted).

Pathway Name	Match Status	p	-log(p)	Holm p	FDR	Impact	Link
<a href="#">Glycolysis or Gluconeogenesis</a>	<a href="#">28/61</a>	1.2805E-12	11.893	1.0756E-10	1.0756E-10	1.8167	<a href="#">KEGG</a>
<a href="#">Fructose and mannose metabolism</a>	<a href="#">19/40</a>	3.1764E-9	8.4981	2.6364E-7	1.3341E-7	1.5128	<a href="#">KEGG</a>
<a href="#">Pentose phosphate pathway</a>	<a href="#">20/47</a>	1.2988E-8	7.8865	1.065E-6	3.6367E-7	1.9348	<a href="#">KEGG</a>
<a href="#">Starch and sucrose metabolism</a>	<a href="#">15/43</a>	1.7621E-5	4.754	0.0014273	3.7004E-4	1.4286	<a href="#">KEGG</a>
<a href="#">Galactose metabolism</a>	<a href="#">16/51</a>	4.1966E-5	4.3771	0.0033573	6.5122E-4	0.96	<a href="#">KEGG</a>
<a href="#">Arginine biosynthesis</a>	<a href="#">11/27</a>	4.6516E-5	4.3324	0.0036748	6.5122E-4	1.0385	<a href="#">KEGG</a>
<a href="#">Retinol metabolism</a>	<a href="#">15/47</a>	5.8685E-5	4.2315	0.0045774	7.0422E-4	1.2826	<a href="#">KEGG</a>
<a href="#">beta-Alanine metabolism</a>	<a href="#">14/44</a>	1.082E-4	3.9658	0.0083311	0.0010314	0.74419	<a href="#">KEGG</a>
<a href="#">Pantothenate and CoA biosynthesis</a>	<a href="#">12/34</a>	1.1051E-4	3.9566	0.0083989	0.0010314	1.0	<a href="#">KEGG</a>
<a href="#">Lysine degradation</a>	<a href="#">14/49</a>	3.8705E-4	3.4122	0.029029	0.0032513	0.75	<a href="#">KEGG</a>
<a href="#">Amino sugar and nucleotide sugar metabolism</a>	<a href="#">19/79</a>	4.3262E-4	3.3639	0.032014	0.0033036	0.94872	<a href="#">KEGG</a>
<a href="#">Ascorbate and aldarate metabolism</a>	<a href="#">6/13</a>	0.001268	2.8969	0.092562	0.0088758	1.0833	<a href="#">KEGG</a>
<a href="#">Glutathione metabolism</a>	<a href="#">14/56</a>	0.0016645	2.7787	0.11984	0.010755	0.72727	<a href="#">KEGG</a>
<a href="#">Butanoate metabolism</a>	<a href="#">9/29</a>	0.0023213	2.6343	0.16481	0.013928	0.71429	<a href="#">KEGG</a>
<a href="#">Neomycin, kanamycin and gentamicin biosynthesis</a>	<a href="#">3/4</a>	0.0044608	2.3506	0.31225	0.02498	2.0	<a href="#">KEGG</a>
<a href="#">Pentose and glucuronate interconversions</a>	<a href="#">9/32</a>	0.0048735	2.3122	0.33627	0.025586	1.0968	<a href="#">KEGG</a>
<a href="#">Pyruvate metabolism</a>	<a href="#">11/45</a>	0.0062287	2.2056	0.42355	0.030777	0.75	<a href="#">KEGG</a>
<a href="#">Arginine and proline metabolism</a>	<a href="#">16/78</a>	0.0068754	2.1627	0.46065	0.032085	0.71429	<a href="#">KEGG</a>
<a href="#">D-Glutamine and D-glutamate metabolism</a>	<a href="#">4/10</a>	0.015946	1.7973	1.0	0.0705	1.2222	<a href="#">KEGG</a>
<a href="#">Biotin metabolism</a>	<a href="#">6/21</a>	0.019191	1.7169	1.0	0.080601	1.45	<a href="#">KEGG</a>

# MetaboAnalystにおける統合解析

Click the corresponding **Pathway Name** to view its graphical presentation; click **Match Status** to view the details.

[Download Results Table](#)
1 2 3

Pathway Name	Match Status	
Glycolysis or Gluconeogenesis	28/61	1.2805
Fructose and mannose metabolism	19/40	3.1764
Pentose phosphate pathway	20/47	1.2988
Starch and sucrose metabolism	15/43	1.7621
Galactose metabolism	16/51	4.1966
Arginine biosynthesis	11/27	4.6516
Retinol metabolism	15/47	5.8685
beta-Alanine metabolism	14/44	1.082E-05
Pantothenate and CoA biosynthesis	12/34	1.1051
Lysine degradation	14/49	3.8705
Amino sugar and nucleotide sugar metabolism	19/79	4.3262
Ascorbate and aldarate metabolism	6/13	0.0012
Glutathione metabolism	14/56	0.0016
Butanoate metabolism	9/29	0.0023
Neomycin, kanamycin and gentamicin biosynthesis	3/4	0.0044
Pentose and glucuronate interconversions	9/32	0.0048
Pyruvate metabolism	11/45	0.0062
Arginine and proline metabolism	16/78	0.0068
D-Glutamine and D-glutamate metabolism	4/10	0.0159
Biotin metabolism	6/21	0.0191

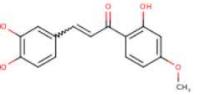
Matched Features

Pathway	Members
Glycolysis or Gluconeogenesis	<b>Acetaldehyde</b> ; Ethanol; Thiamin diphosphate; Pyruvate; <b>2-(alpha-Hydroxyethyl)thiamine diphosphate</b> ; Enzyme N6-(lipoyl)lysine; <b>Acetyl-CoA</b> ; <b>Enzyme N6-(dihydrolipoyl)lysine</b> ; <b>(S)-Lactate</b> ; Phosphoenolpyruvate; 2-Phospho-D-glycerate; D-Glyceraldehyde 3-phosphate; <b>beta-D-Fructose 1,6-bisphosphate</b> ; beta-D-Fructose 6-phosphate; alpha-D-Glucose 6-phosphate; D-Glucose 1-phosphate; <b>beta-D-Glucose 6-phosphate</b> ; beta-D-Glucose; alpha-D-Glucose; 3-Phospho-D-glycerate; Oxaloacetate; 2,3-Bisphospho-D-glycerate; 3-Phospho-D-glyceroyl phosphate; Acetate; [Dihydrolipoilysine-residue acetyltransferase] S-acetyl dihydrolipoilysine; <b>Glycerone phosphate</b> ; <b>ALDH2, ALDH-E2, ALDHI, ALDM...</b> ; ALDH3A1, ALDH3, ALDHIII...; <b>AKR1A1, ALDR1, ALR, ARM, DD3, HEL-S-6, ADH1A, ADH1...</b> ; PDHA1, PDHA, PDHAD, PDHCE1A, PHE1A...; DLAT, DLTA, E2, PDC-E2, PDCE2; <b>LDHAL6A, LDH6A...</b> ; <b>PKLR, PK1, PKL, PKRL, RPK...</b> ; ENO1, ENO1L1, HEL-S-17, MPB1, NNE, PPH...; <b>PGAM4, PGAM-B, PGAM1, PGAM3, dJ1000K24.1...</b> ; <b>GAPDH, G3PD, GAPD, HEL-S-162eP...</b> ; <b>TPI1, HEL-S-49, TIM, TPI, TPID</b> ; ALDOA, ALDA, GSD12, HEL-S-87p...; PFKL, ATP-PFK, PFK-B, PFK-L...; FBP1, FBP...; GPI, AMF, GNPI, NLK, PGI, PHI, SA-36, SA36; PGM1, CDG1T, GSD14...; HK1, HK, HK1-ta, HK1-tb, HK1-tc, HKD, HKI, HMSN, HXK1, RP79, hexokinase...; GCK, FGQTL3, GK, GLK, HHF3, HK4, HKIV, HXKP, LGLK, MODY2; GALM, BLOCK25, GLAT, HEL-S-63p, IBD1; G6PC, G6PC1, G6PT, G6Pase, GSD1, GSD1a...; DLD, DLDD, DLDH, E3, GCSL, LAD, PHE3; PGK1, HEL-S-68p, MIG10, PGKA...; PCK1, PCKDC, PEPCK-C, PEPCK1, PEPCKC...; BPGM, DPGM, ECYT8; ACSS2, ACAS2, ACECS, ACS, ACSA, dJ1161H23.1...; ADPGK, 2610017G09Rik, ADP-GK; MINPP1, HIPER1, MINPP2, MIPP

# 化合物のオントロジー化

ClassyFire   Browse▼   Classify   About ClassyFire   Contact   Downloads   Help▼

Structure Information



Compound Identification

SMILES

COC1=CC(O)=C(C=C1)C(=O)C=CC1=CC(O)=C(O)C=C1

InChIKey

InChIKey=ILUVGTQOZKYHCS-UHFFFAOYSA-N

Formula

C<sub>16</sub>H<sub>14</sub>O<sub>5</sub>

Mass

286.283

Export to:

JSON   SDF   CSV

Taxonomic Classification

Taxonomy Tree

- Kingdom [Organic compounds](#)
- Superclass [Phenylpropanoids and polyketides](#)
- Class [Linear, 1,3-diarylpropanoids](#)
- Subclass [Chalcones and dihydrochalcones](#)
- Level 5 [2'-Hydroxychalcones](#)

# Chemical similarity based enrichment analysis

## Input file structure

A	B	C	D	E	F
Compound Name	InChiKeys	Pubchem ID	SMILES	pvalue	foldchange
1 SM (d42:2) B	DACOGJMBLYZYDH-GXJPFDISA-N	52931217	CCCCCCCCCCCCCCCCCCCCCC(=O)N[C@H]1CC	1.32E-13	0.303587195
3 N-Tetracosenoyl-4-spiro[5H-chromene-5,3'-cyclohexene]-1,2-dione	WKZHECFHXLTOLI-QYKFWSDSSA-N	44260126	CCCCCCCCCC/C=C/[C@H](CC1=CC=CC=C1)OC(=O)COP(=O)(O)C	3.71E-13	0.270951732
4 LPC (16:0)	ASWBNNHCZGQVJV-UHFFFAOYSA-N	86554	CCCCCCCCCC(=O)OCC(COP(=O)(O)C)OC[C@H]1COF	1.50E-12	0.363195589
5 PC (18:1/16:0)	WTJKGGKOPKCXLL-VYOBOKEXSA-N	5497103	CCCCCCCCCC(=O)OC[C@H]1COF	5.38E-12	0.369242722

Structure of the input file. See here an [example data file](#) you can use as template.

## The input file must have 6 columns, in this order:

- Column 1 = Compound Name
- Column 2 = InChiKeys
- Column 3= Pubchem ID
- Column 4 = SMILES
- Column 5 = pvalue
- Column 6= foldchange

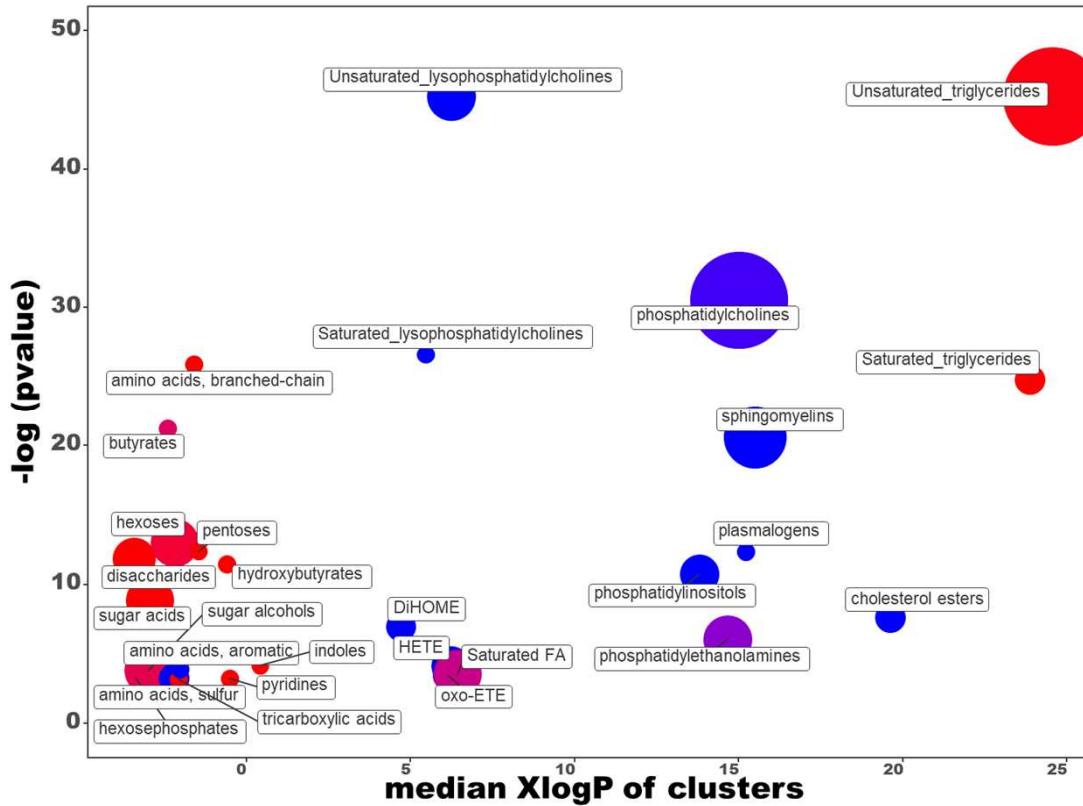
You can obtain PubChem CIDs and InChI keys from [the Chemical Translation Service](#). You can obtain SMILES from the [PubChem Identifier Exchanger tool](#).

You can also use this excel file for getting identifiers and SMILES codes for your compounds. [ChemRICH Metabolite Identifier File](#) .

## Input file must satisfy these conditions

- No duplicate names OR CIDs
- No Missing names, SMILES, pvalue or fold-change
- Minimum one class with at least three compounds
- Minimum one class shall be significantly enriched

# Chemical similarity based enrichment analysis



A	B	C	D	E	F	G	H	
Cluster name	Cluster	p-value	FDR	Key compound	Altered	Increases	Decreases	In
1 Unsaturated_lysophosphatidylcholines	9	2.2E-20	5.4E-19	LPC (16:1)	9	0	9	
2 Unsaturated_triglycerides	46	2.2E-20	5.4E-19	TG (53:4)	29	28	1	
3 phosphatidylcholines	44	5.5E-14	9E-13	PC (18:1/16:0)	26	1	25	
4 Saturated_lysophosphatidylcholines	3	2.8E-12	3.4E-11	LPC (16:0)	3	0	3	
5 amino acids, branched-chain	3	5.6E-12	5.5E-11	isoleucine	3	3	0	
6 Saturated_triglycerides	4	1.7E-11	1.4E-10	TG (48:0)	4	4	0	
7 butyrates	3	6.2E-10	4.4E-09	2-hydroxy-2-methylbutanoic acid	3	2	1	
8 sphingomyelins	16	1.1E-09	7E-09	SM (d42:2) B	12	0	12	
9 hexoses	9	2.2E-06	0.000012	1,5-anhydroglucitol	7	6	1	
10 pentoses	3	4.2E-06	0.000019	ribose	3	3	0	
11 plasmalogens	2	4.3E-06	0.000019	plasmalogenLDF (38:6)	2	0	2	

A	B	C	D	E	F	G	H
Compound.Name	InChiKey	Pubche	SMILES	pvalue	foldcha	CID	Cluster
165 PC 32:0	KILNVBDSD	452110	CCCCCCCC	1.7E-06	0.47	452110	36
203 PC (34:2)	JLPULHDH	5287971	CCCCCCCC	0.01	0.9	5287971	2
212 PC (18:1/16:0)	WTJKGGK	5497103	CCCCCCCC	5.4E-12	0.37	5497103	2
222 PC (32:1)	QIBZFHLFI	6443788	CCCCCCCC	2.3E-10	0.2	6443788	2
226 SM (d16:1/20:0)	LKQLRGM	6453725	CCCCCCCC	0.65	1.1	6453725	3
271 PC (36:2)	SNKAWJB	10350317	CCCCCCCC	0.79	1.1	10350317	2
273 PC (36:4) A	IIZPYDJI	10747814	CCCCCCCC	3.3E-07	0.63	10747814	2
300 PC (38:4) A	PSVRFUPC	16219824	CCCCCCCC	0.0005	0.69	16219824	2
304 PC (34:0)	PZNPLUBH	24778686	CCCCCCCC	6.4E-08	0.44	24778686	36
305 PC (32:2)	GPWHCU	24778764	CCCCCC/C	6.9E-07	0.38	24778764	2
306 PC (36:5) C	DYDDZDM	24778771	CCCCCC/C	0.48	1.1	24778771	2
307 PC (36:3) A	BXRLDROZ	24778937	CCCCCCCC	0.98	1	24778937	2
308 PC (40:7)	BPUROMF	24778982	CCCCC/C=	9.7E-07	0.32	24778982	2
309 PC (38:2)	KXXLFCAP	24779263	CCCCCCCC	0.000011	0.5	24779263	2
334 PC (35:2)	ZSKWZJYU	52922491	CCCCCCCC	0.000026	1.5	52922491	2
335 PC (39:6)	QMCWOG	52922637	CCCCC/C=	0.25	1.1	52922637	2
336 PC 33:2	SBNDHGB	52922715	CCCCCCCC	0.31	0.88	52922715	2
337 PC (37:2)	MCZUABD	52922735	CCCCCCCC	0.00011	0.52	52922735	2
338 PC (36:4)	NKQPOV	52922783	CCCCCCCC	2E-07	0.56	52922783	2
339 PC (36:6)	SPWBDEZ	52922847	CCCCC/C	0.000034	0.42	52922847	2

# メタボロームデータを出す側と使う側

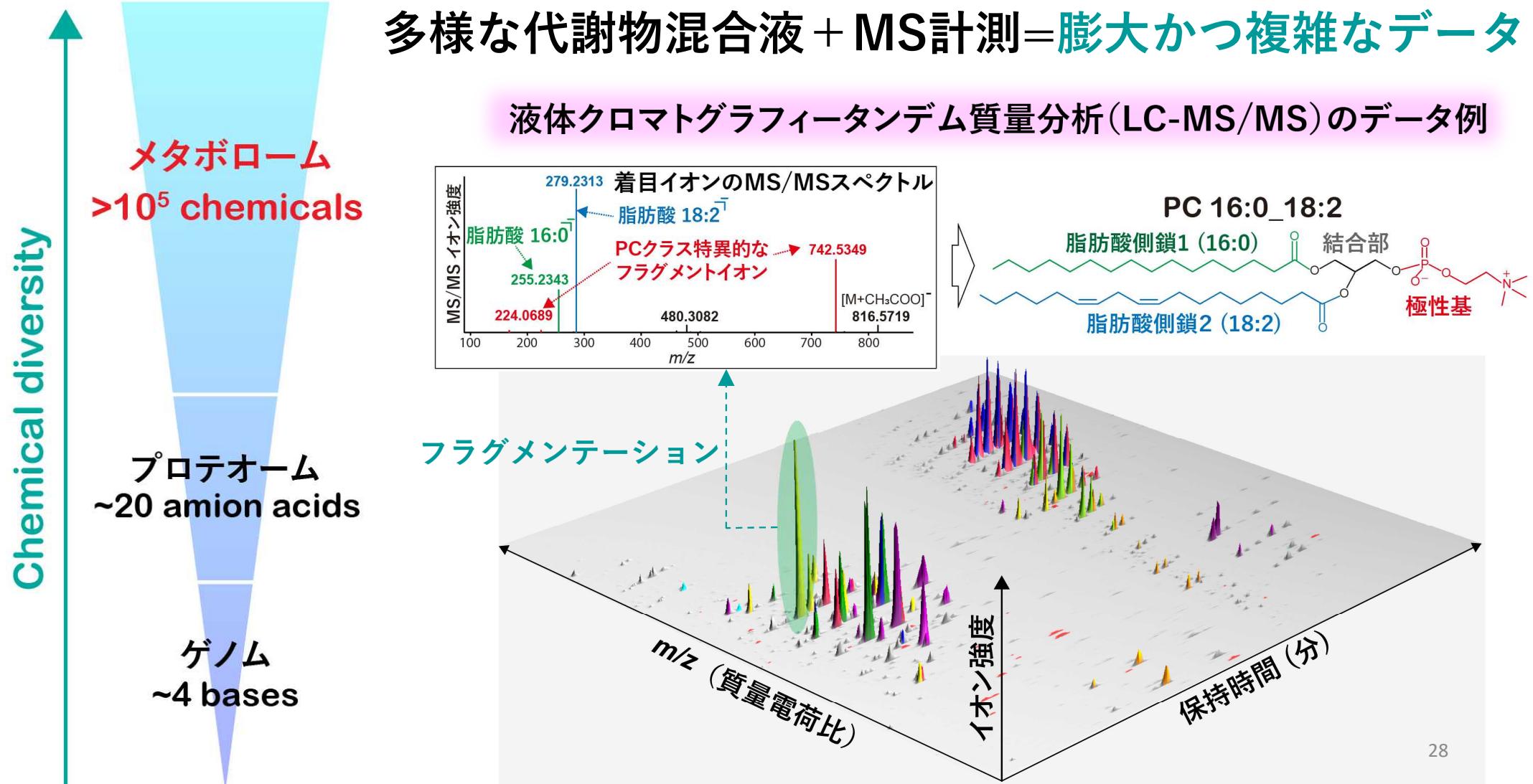
## 出す側

- 分析化学
- 質量分析
- イオン
- マススペクトル
- 質量分析インフォマティクス
- アノテーション
- データ標準化

## 使う側

- データの特徴理解
- 化合物IDの取得
- データの正規化
- 統計・多変量解析
- データベース検索
- 代謝マップ投影
- 適切な数理モデルの選択
- オミクスモジュールへ統合

# メタボロームの多様性とデータ解析の複雑さ

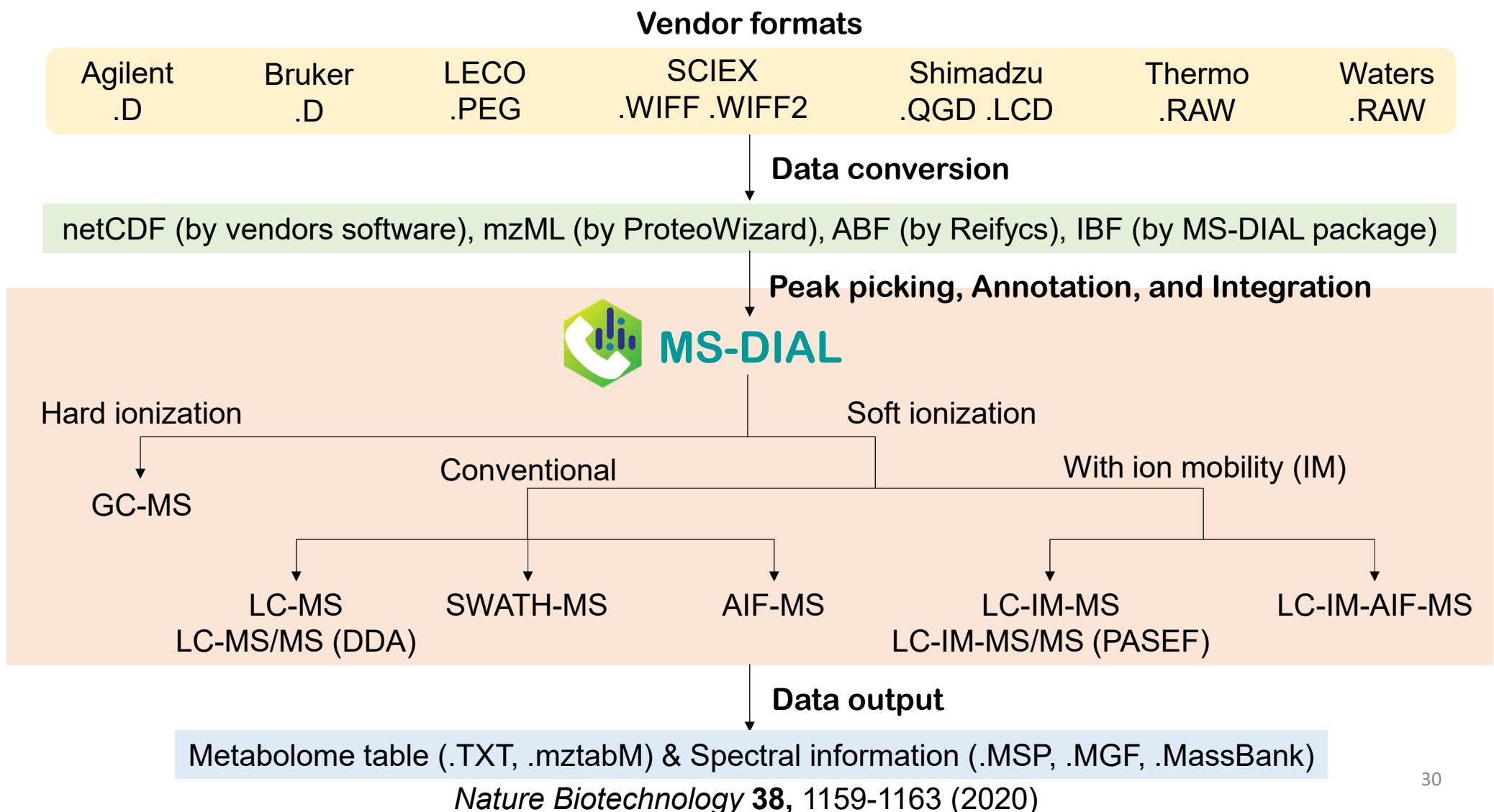


# 質量分析インフォマティクス研究の開拓

「質量」情報を「化合物」情報へ変換し「生命システム」情報へ



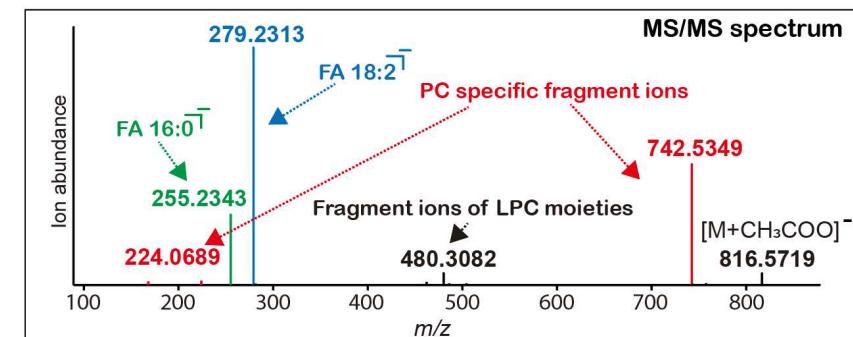
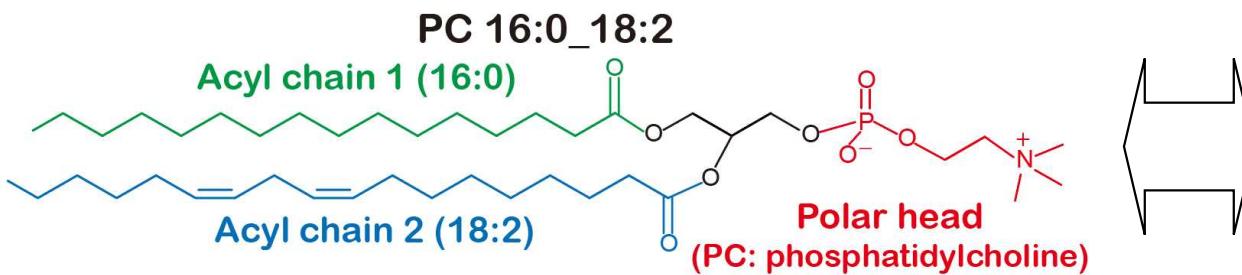
# Summary of MS-DIAL environment



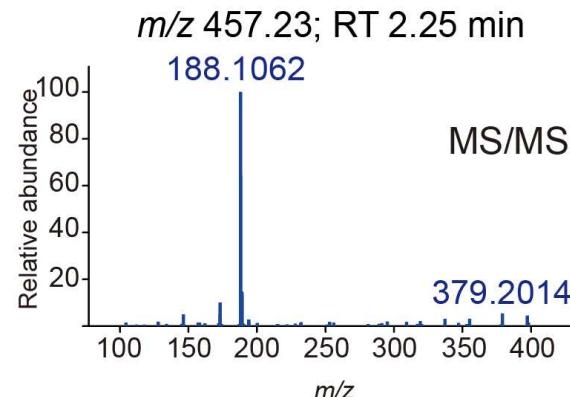
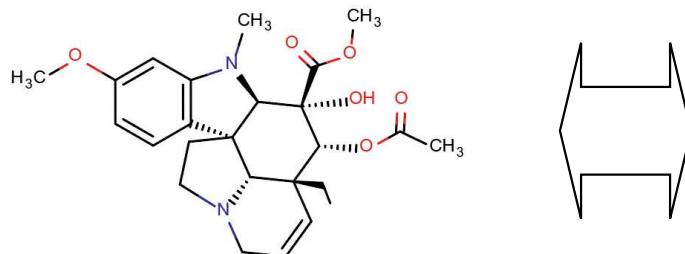
# 天然物化合物のアノテーションは難しい

プロテオミクス：およそ20種のアミノ酸で構成される

リピドミクス：アシル基の多様性はあれど開裂様式は一定



## 親水性メタボローム・天然物：標準品化合物が必須



# マススペクトルデータベースは有効に使いましょう

The screenshot shows a web browser window for the MoNA (MassBank of North America) website. The URL in the address bar is `mona.fiehnlab.ucdavis.edu/downloads`. The page title is "MoNA - MassBank of North America". The main content area is titled "Downloads" and lists various predefined queries with their counts and download options.

**Downloads**

A set of commonly referenced predefined queries. Clicking the name of the query will display the associated spectra in the query browser. Each query is also available to download in either the MoNA internal JSON format or as NIST MS Search compatible MSP files.

Display Hidden Downloads

<a href="#">All Spectra (670,148 spectra)</a>	<a href="#"> Download</a>
<a href="#">In-Silico Spectra (490,087 spectra)</a>	<a href="#"> Download</a>
<a href="#">Experimental Spectra (180,061 spectra)</a>	<a href="#"> Download</a>
<a href="#">GC-MS Spectra (18,886 spectra)</a>	<a href="#"> Download</a>
<a href="#">LC-MS Spectra (142,909 spectra)</a>	<a href="#"> Download</a>
<a href="#">LC-MS/MS Spectra (135,076 spectra)</a>	<a href="#"> Download</a>
<a href="#">LC-MS/MS Positive Mode (91,824 spectra)</a>	<a href="#"> Download</a>
<a href="#">LC-MS/MS Negative Mode (42,470 spectra)</a>	<a href="#"> Download</a>
<hr/>	
<b>Libraries</b>	
<a href="#">MassBank (72,439 spectra)</a>	<a href="#"> Download</a>
<a href="#">CASMI 2016 (622 spectra)</a>	<a href="#"> Download</a>
<a href="#">CASMI 2012 (26 spectra)</a>	<a href="#"> Download</a>
<a href="#">ReSpect (6,374 spectra)</a>	<a href="#"> Download</a>
<a href="#">HMDB (7,415 spectra)</a>	<a href="#"> Download</a>
<a href="#">GNPS (23,801 spectra)</a>	<a href="#"> Download</a>
<a href="#">LipidBlast (485,796 spectra)</a>	<a href="#"> Download</a>
<a href="#">FAHFA (4,290 spectra)</a>	<a href="#"> Download</a>

# マススペクトルデータベースは有効に使いましょう

J. Mass Spectrom. Soc. Jpn.

Vol. 65, No. 5, 2017

特集：メタボロミクス

COMMENTARY

## メタボロミクスにおける化合物同定 —生体内低分子代謝物の構造推定ガイド—

The Guide for Metabolite Annotation/Identification  
in Untargeted Metabolomics

津川 裕司<sup>1,2</sup>・早川 英介<sup>3</sup>・三浦 大典<sup>4,\*</sup>

Hiroshi TSUGAWA,<sup>1,2</sup> Eisuke HAYAKAWA,<sup>3</sup> and Daisuke MIURA<sup>4,\*</sup>

<sup>1</sup> 理化学研究所環境資源科学研究センター RIKEN Center for Sustainable Resource Science

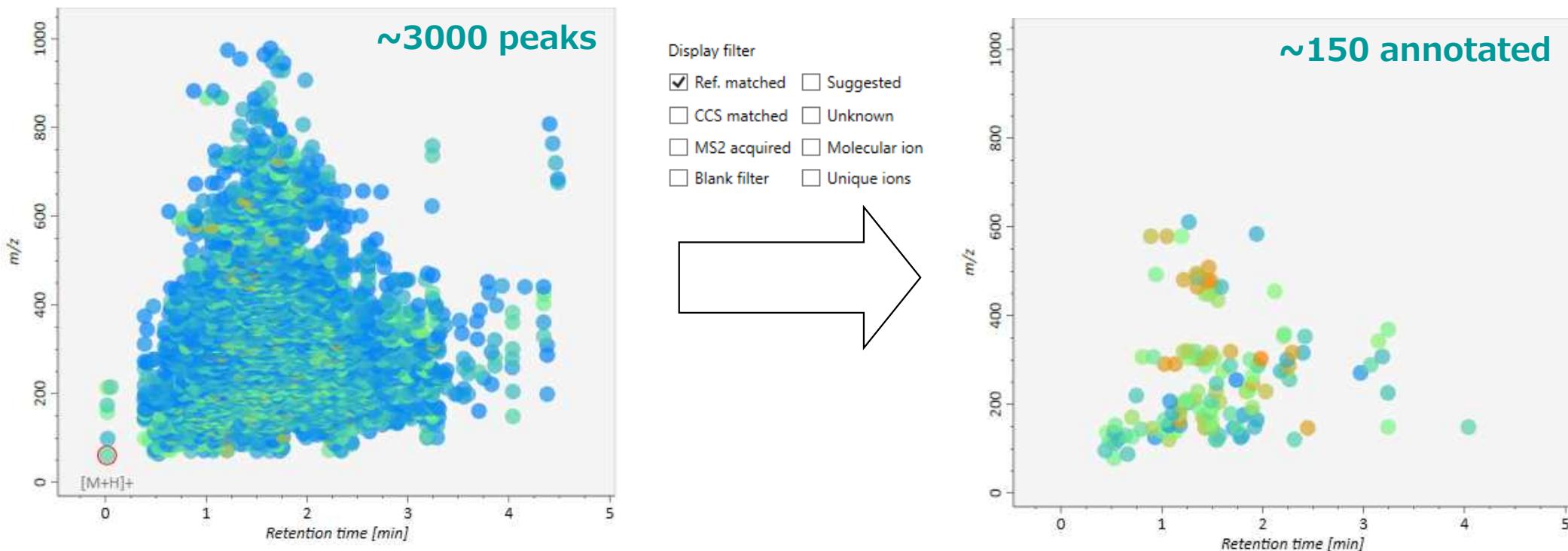
<sup>2</sup> 理化学研究所統合生命医科学研究センター RIKEN Center for Integrative Medical Sciences

<sup>3</sup> 沖縄科学技術大学院大学進化神経生物学ユニット Evolutionary Neurobiology Unit, Okinawa Institute of Science and Technology

<sup>4</sup> 九州大学農学研究院生命機能科学部門 Department of Bioscience and Biotechnology, Kyushu University

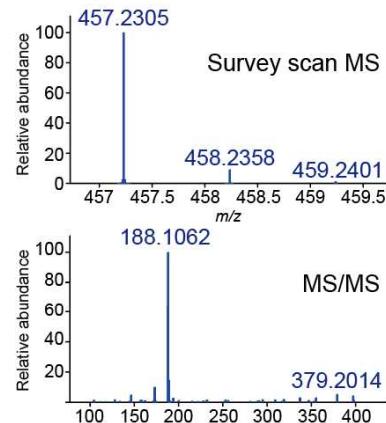
# 標準品によってアノテーションできる数は限られる

ワインのLC-MS/MS分析データ

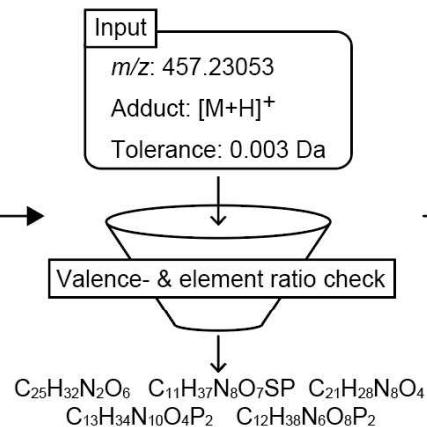


# In silico fragmentationによる構造予測

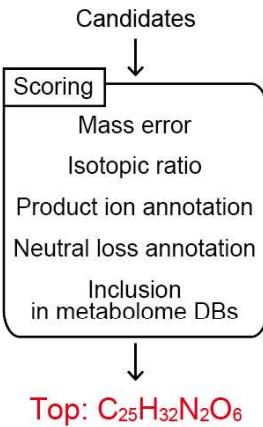
## a. Get MS & MS/MS spectrum



## b. Formula generation



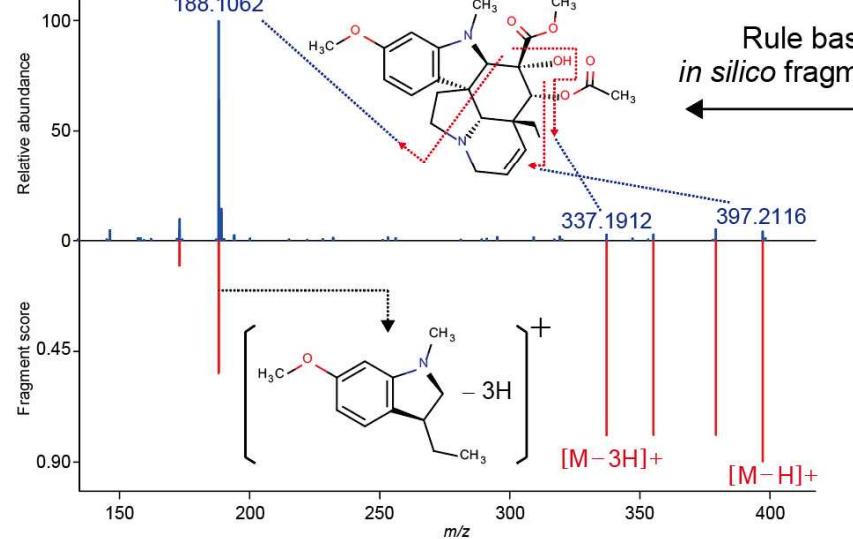
## c. Formula ranking



## Metabolome databases

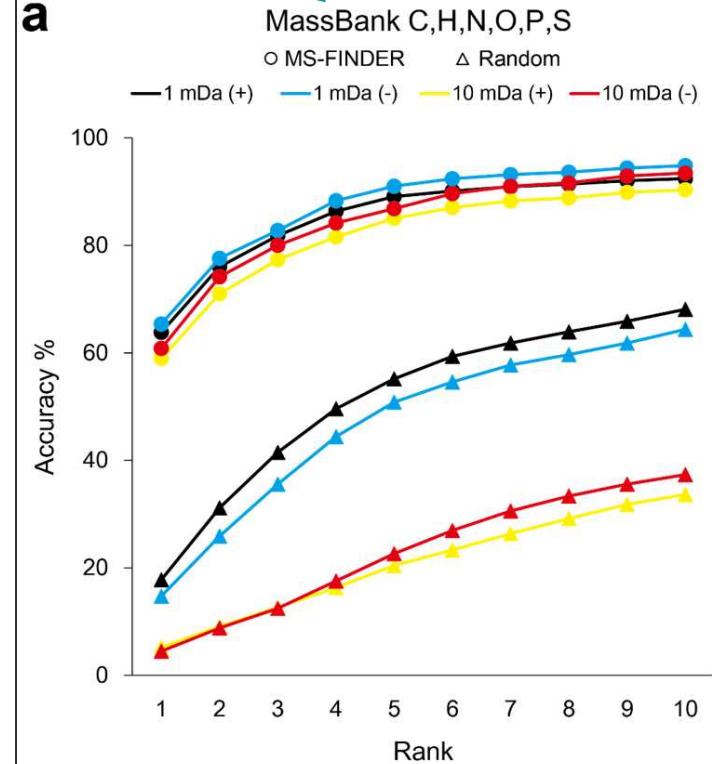
HMDB  
YMDB  
SMPDB  
UNPD  
ChEBI  
PlantCyc  
PubChem (Biopathway)  
DrugBank  
BMDB  
KNAPSAck  
FooDB  
ECMDB  
T3DB

## e. Structure ranking



## 3000種の既知構造とMS/MSのセット

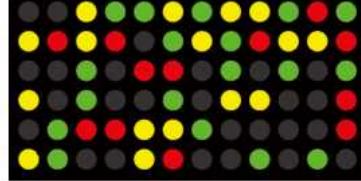
a



# フラグメントセットを用いた代謝物クラスの推定

a

GSEA  
mRNA expression profile



Applied to GO-gene set

GO 0019368

**Fatty acid elongation**

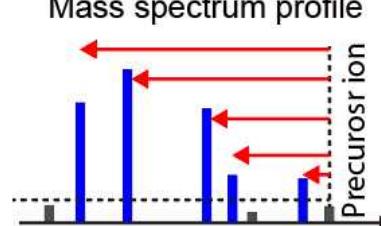
**Elovl1 Elovl2 Elovl3 Elovl4**

**Elovl5 Elovl6 Elovl7**

(Bold: significant genes)

p-value = 0.0032

FSEA  
Mass spectrum profile



Applied to CO-fragment set

CO 0003533

**Phosphatidylcholine**

H<sub>2</sub>O; **Phosphocholine**;

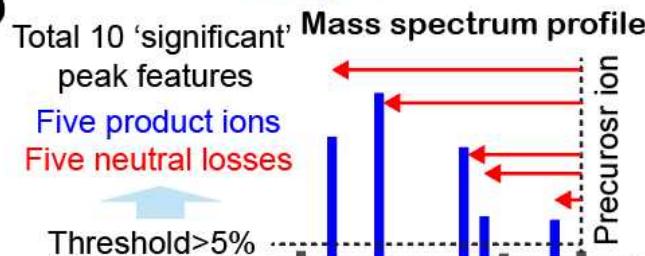
**Choline**; **Fatty acids**; LysoPC

(Bold: detected fragments)

Fisher's exact test  
p-value = 0.0056

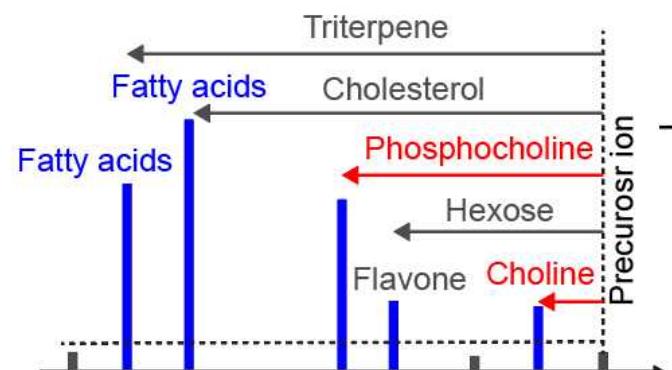
b

Step 1



Step 2

Formula assignment followed by fragment ontology assignment (now 459 ontology registered)



Blue ontology: true positive assignment for product ions  
Red ontology: true positive assignment for neutral losses  
Gray ontology: false positive assignment

Step 4

Calculate p-value by Fisher's exact test

p-value = 0.0056

Step 3

Cross-tabulation table

Metabolite class

Phosphatidylcholine

Required fragments

H<sub>2</sub>O; **Phosphocholine**;

**Choline**; **Fatty acids**; LysoPC

MS/MS spectrum

	Significant	N.S.
Included	3	2
Not included	4	450

Required fragments

Calculated by (1) registered ontologies (n = 459),  
(2) <5% peak features,  
or (3) reversed MS/MS spectrum

# メタボロームデータを出す側と使う側

## 出す側

- 分析化学
- 質量分析
- イオン
- マススペクトル
- 質量分析インフォマティクス
- アノテーション
- データ標準化

## 使う側

- データの特徴理解
- 化合物IDの取得
- データの正規化
- 統計・多変量解析
- データベース検索
- 代謝マップ投影
- 適切な数理モデルの選択
- オミクスモジュールへ統合

# 参考Webサイト：メタボロミクス若手会(ESI友の会)

ESI友の会：メ  
タボロミクス若  
手会

## メタボロミクスの発展を目指して

トップページ

【ご挨拶】

ESI友の会は、日本のメタボロミクス研究を盛り上げよう！もっと色々な人たちにメタボロミクスの技術を使えるようになってもらって、日本の科学を発展させよう！！という目的で、メタボロミクス研究者の中から有志が集まって結成されたコミュニティです。「ESI」という名前がついていますが、集まっている人たちは本当に様々です。扱っているサンプルでいえば微生物・食品・動植物、分離装置でいえばガスクロマトグラフィー・液体クロマトグラフィー・キャピラリー電気泳動・超臨界流体、検出装置でいえば、四重極型・飛行時間型・フーリエ変換型・タンデム型(QqQ・Q-TOF)・Orbitrap。さらにデータ処理や統計学を専門としている人がいます。「メタボロミクスをはじめるなら今！！」というくらい、メタボロミクス研究はいま多くの研究分野から注目を集めています。従来、メタボロミクスの技術は難しく一筋縄ではいかないから手を出しにくい...と言われてきました。しかしその時代はもう終わりつつあります！ESI友の会で集まった人たちは、それぞれ目的に応じたメタボロミクスプロトコールを開発し、実用的なシステムを構築してきました。そして今回、それぞれの技術を持ち寄り、「**メタボロミクスのプロトコール集**」を作成致しました。これを自由にダウンロードしていただき、日本の研究者の皆様方の一助になることをESI友の会は目指しています。もちろん、これからも様々な意見を取り入れ、改定を重ねていき、より良いメタボロミクスプロトコール集を作成・更新していくたいと思っています。乱筆ではありますが、これから日本の科学の発展を願いメタボロミクス研究の立場から日本科学に貢献できるよう日々精進して参ります。

ESI友の会一同

ESI友の会 プロトコール集執筆者一覧

九州大学 和泉自泰 (Izumi Yoshihiro) :幹事

理化学研究所 津川裕司 (Tsugawa Hiroshi) :幹事