

○細田正恵¹、山田一作²、塩田正明¹、高橋悠志¹、新町大輔³、小野多美子¹、土屋伸一郎²、松原正陽²、木村直貴²、瀬野瑛⁴、藤田晶大¹、金進東⁵、岡谷千晶³、久野敦³、藤田典昭³、安形清彦³、梶裕之³、成松久³、奥田修二郎⁴、木下聖子¹

1. 創価大学糖鎖生命システム融合研究所 (GaSIC)、2. 公益財団法人野口研究所、3. 産業技術総合研究所 (AIST)、4. 新潟大学医学部メディカル AIセンター、5. 情報・システム研究機構ライフサイエンス統合データベースセンター (DBCLS)

<https://glycosmos.org>

GlyCosmos Portalは糖鎖構造情報をはじめ、糖鎖に関わる遺伝子、タンパク質、脂質、疾患、パスウェイ情報をセマンティックウェブ技術を活用し、様々なデータベースと統合したウェブポータルである。

また日本糖質学会により公式認定されたポータルサイトとして糖鎖生物学者への利便性の向上を図っている。

ここでは主に疾患情報の拡充、外部サイトの相互リンク、開発・公開中のリポジトリについて紹介する。



Data Resources 疾患情報の拡充

Diseases

ACGG-DB

Glyco-Disease Genes Database (GDGDB)

GDGDB is a database of glycan-related diseases and their responsible genes.

— ACGG-DB

GlycoNAVI

GlycoNAVI-Diseases

GlycoNAVI-Diseases is dataset of glycoforms related to diseases. This is the content of GlycoNAVI.

— GlycoNAVI

DisGeNET

DisGeNET

DisGeNET is a knowledge management platform that integrates and standardizes the data of genes and mutations related to diseases from multiple sources. It covers entire range of human disease.

Alliance of Genome Resources

Alliance of Genome Resources

Alliance of Genome Resources is a gene dataset, which contains curated genome information essential for the understanding of the genetics and genomic basis of human biology, health and disease. All data is retrieved from six major model organism databases and the Gene Ontology including FlyBase, Mouse Genome Database (MGD), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), WormBase, Zebrafish Information Network (ZFIN), and the GO Consortium. In GlyCosmos, only glycan-related disease information has been extracted from Alliance of Genome Resources.

Alliance of Genome Resources

Alliance of Genome Resources is a gene dataset, which contains curated genome information essential for the understanding of the genetics and genomic basis of human biology, health and disease. All data is retrieved from six major model organism databases and the Gene Ontology including FlyBase, Mouse Genome Database (MGD), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), WormBase, Zebrafish Information Network (ZFIN), and the GO Consortium. In GlyCosmos, only glycan-related disease information has been extracted from Alliance of Genome Resources.

Gene Symbol	NCBI Gene ID	Species Name	Alliance of Genome Resources ID	FlyGlycoDB	Association Type	Dterm Name	With Orthologs	Evidence Code Name	Reference
A4GNT	51146	Homo sapiens	HGNC:17968	No data	is_marker_for	pancreatic cancer	No data	expression pattern evidence used in manual assertion	PMID:16441422
A4gnt	333424	Mus musculus	MG1:2143261	No data	biomarker_via_orthology	pancreatic cancer	HGNC:17968	evidence used in automatic assertion	MG1:6194238
Abo	80908	Mus musculus	MG1:2135738	No data	implicated_via_orthology	COVID-19 acute lymphoblastic leukemia asthma	HGNC:79	evidence used in automatic assertion	MG1:6194238

DisGeNET List

DisGeNET is a knowledge management platform that integrates and standardizes the data of genes and mutations related to diseases from multiple sources. It covers entire range of human disease.

Disease ID	Disease name	Gene symbol	Gene ID	Gene name	Protein
C0005684	Malignant neoplasm of urinary bladder	ABO	28	ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase	P16442
C0005695	Bladder Neoplasm	ABO	28	ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase	P16442
C0007182	Malignant tumor of colon	B3GALT5	10317	beta-1,3-galactosyltransferase 5	Q9Y2C3
C0007847	Malignant tumor of cervix	H3S5T2	9956	heparan sulfate-glucosamine 3-sulfotransferase 2	Q9Y278
C0017636	Glioblastoma	TUSC3	7991	tumor suppressor candidate 3	Q13454
C0017661	IGA Glomerulonephritis	C1GALT1C1	29871	C1GALT1 specific chaperone 1	Q96EU7
C0024790	Paroxysmal nocturnal hemoglobinuria	PIGA	5277	phosphatidylinositol glycan anchor biosynthesis class A	P37287
C0026697	Mucopolysaccharidosis I	NEU1	4758	neuraminidase 1	Q99519
C0026785	Mucopolysaccharidosis II	ID5	3423	iduronate 2-sulfatase	P22384
C0027851	Myocardial Infarction	NEEL3	55247	nei like DNA glycosylase 3	Q8TAT5

DisGeNETとAlliance of Genome Resourcesから、疾患と、関連する遺伝子の情報を取得し、GlyCosmosの糖鎖関連遺伝子に絞ってリストを作成した。

DisGeNETからは46件、Alliance of Genome Resourcesからは、575件の疾患情報を得ることができた。

今後、これらの糖鎖関連遺伝子が関与する疾患情報をまとめた、GlyCosmos Diseaseページを作成していく。



Data Resources 外部データベースとの相互リンク

GlyCosmos Pathways



GlyCosmos Pathways

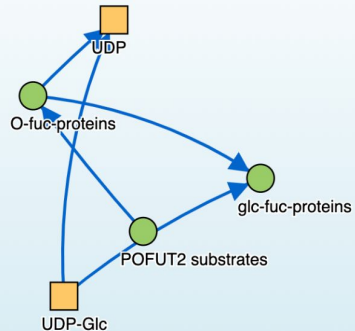
Species: Homo sapiens

Pathway Name: O-glycosylation of TSR domain-containing proteins(Link to Reactome)

External Links:

PubChem: (R-HSA-5173214)

Click on a node in the pathway to display its details. Gly



PubChem O-glycosylation of TSR domain-containing proteins (Pathway)

6.2 GlycoPathways

O-glycosylation of TSR domain-containing proteins

GlyCosmos Glycoscience Portal

Source: GlyCosmos Glycoscience Portal

URL: <https://glycosmos.org/pathways/show?id=R-HSA-5173214>

Description: The GlyCosmos Portal is a Web portal aiming to integrate the glycosciences with the life sciences. It consists of Standards, Repositories and Data Resources, providing information about genes, proteins, lipids, pathways and diseases related to glycans.

License Note: The GlyCosmos is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

License URL: <https://glycosmos.org/homes/license>

GlyCosmos Glycoproteins エントリーページ

Erythropoietin

Entry of glycoprotein extracted from UniProt

Summary

Erythropoietin

UniProt ID: P01588

Gene Symbol: EPO

Organism:

Scientific Name: Homo sapiens

External Links

- PubChem : P01588
- GlyConnect : 140
- GlyGen : P01588
- UniCarbKB : P01588

PubChem Erythropoietin (Protein)

1.1.5 GlyCosmos Protein

P01588

GlyCosmos Glycoscience Portal

GlyCosmos Pathways、GlycoproteinsはPubChemの該当ページと相互リンクしており、効率的にデータを調べることができるようになった。



Data Resources 生物種による検索

GlyCosmos Organisms

List of all species in GlyCosmos Resources.

Show 10 entries

To search, enter a keyword in the text box and press Enter (return).

Search across columns

Organism	Taxon ID	Gene	Structure	Protein	Lectin	Pathway
Arabidopsis thaliana	3702	460	No data	1562	212	No data
Homo sapiens	9606	460	No data	1562	212	No data
Xenopus laevis	8355	346	No data	415	10	No data
Mus musculus	10090	319	189	3856	178	1695
Danio rerio	7955	307	175	484	6	1648
Rattus norvegicus	10116	265	No data	2099	104	1677
Bos taurus	9913	251	No data	1131	40	1324
Xenopus tropicalis	8364	229	No data	190	2	881
Pongo abelii	9601	228	No data	323	12	No data
Sus scrofa	9823	225	No data	463	23	1590

Showing 1 to 10 of 21935 entries

Download the displayed table (.tsv) Download all (.tsv)

Previous 1 2 3 4 5 ... 2194 Next

Line breaks are removed from cells.

Homo sapiens

エントリーページ

Summary

Taxonomy ID: 9606
Common Name: Human

Contents

- Summary
- List

Glycogenes Glycans Glycoproteins Lectins

カテゴリ別の検索結果を表示することができます。

GlyCosmos Glycogenes

Show 10 entries

To search, enter a keyword in the text box and press Enter (return).

Show hidden columns

Gene Symbol	Gene ID	KEGG	Species
A (ABO)	28	hsa:28	Homo sapiens
AGL, GDE	178	hsa:178	Homo sapiens
AMY1, AMY1A	276	hsa:276	Homo sapiens
AMY1, AMY1B	277	hsa:277	Homo sapiens
AMY1, AMY1C	278	hsa:278	Homo sapiens
AMY2, AMY2A, PA	279	hsa:279	Homo sapiens
AMY2, AMY2B, AMY3, HXA	280	hsa:280	Homo sapiens
CHI3, CHIT, CHIT1, CHITD	1118	hsa:1118	Homo sapiens
CTB, CTBS	1486	hsa:1486	Homo sapiens

GlyCosmos Organism

生物種のリストとなっており、GlyCosmos内のGlycogenes、Glycans、Glycoproteins、Lectins、Pathwaysの各カテゴリに対して、生物種別の件数を一覧として表示している。

件数をクリックすることで、カテゴリ別のリストが表示され、生物種名をクリックすると、エントリーページに移動し、こちらのページからもカテゴリ別のリストを確認できる。



Data Resources 横断検索

検索例: GlyTouCan ID “G40702WU”

Cross Search
Search across GlyCosmos Resources.

All

Example search words ▾

Search results for G40702WU

Genes/Proteins/Lipids

- GlyCosmos GlycoGenes 0
- GlyCosmos Glycoproteins 23
- GlyCosmos Lectins 0
- GlyCosmos Glycolipids 0
- GlyCosmos Glycans 1

Pathways/Diseases

- GlyCosmos Pathways 0

検索結果が1件の場合は、直接エントリーページの結果に飛ぶことができる。

G40702WU
Entry of glycan extracted from GlyTouCan

Summary

G40702WU

[Open LiteMol Viewer](#)

IUPAC: Man(α1-2)Man(α1-2)Man(α1-3)[Man(α1-3)[Man(α1-2)Man(α1-6)]Man(α1-6)Man(b1-4)GlcNAc(b1-4)GlcNAc(b1-4)

Species:
○ Homo sapiens

Monoisotopic mass: 1720.59

Subsumption: Linkage defined saccharide

[GNOme Browser](#)

GlyCosmos全体のデータを一括で検索することが可能

検索ワードの例

UniProt ID: 糖タンパク質、レクチン

GlyTouCan : 糖タンパク質、糖鎖構造

Gene Symbol: 糖鎖関連遺伝子、糖タンパク質

その他、フリーワードでも検索が可能

GlyCosmos Glycoproteins [?]

A list of glycoproteins extracted from UniProt and annotated with glycosylation. From the entry page, you can check various information related to glycoproteins, such as glycosylation site, sequence information, three-dimensional structure, and pathway.

Database Name	Last Update
UniProt	December 9, 2020

Search results for G40702WU

Show 10 entries

To search, enter a keyword in the text box and press Enter (return).

[Show hidden columns](#)

Protein Name ▾	UniProt ID	Gene Symbol	Organism	No. of Glycosylation Sites	No. of GlyTouCan IDs	Disease Name	Disease Aliases
Search	Search	Search	Search			Search	Search
Acetylcholine receptor subunit alpha	P02710	CHRNA1	Tetronarce californica	1	66	No data	No data
Acetylcholine receptor subunit beta	P02712	CHRB1	Tetronarce californica	1	66	No data	No data
Acetylcholine receptor subunit delta	P02718	chrnd	Tetronarce californica	3	66	No data	No data
Acetylcholine receptor subunit gamma	P02714	CHRG	Tetronarce californica	1	66	No data	No data
Alpha-amylase	P17654	AMY1.1	Oryza sativa subsp. japonica	1	15	No data	No data
Alpha-amylase inhibitor 1	P02873	LLP	Phaseolus vulgaris	3	14	No data	No data
Alpha-glucosidase	P56526	aglA	Aspergillus niger	24	11	No data	No data
Beta-fructofuranosidase, insoluble isoenzyme 1	P26792	INV1	Daucus carota	5	10	No data	No data

検索結果が複数の場合は、リストページに飛ぶ



Repositories



GlyTouCan

国際糖鎖構造リポジトリ。
糖鎖構造に固有のアクセッション番号を付与することができる。
登録された糖鎖構造は、GlyCosmos Portal にも反映され、閲覧することができる。



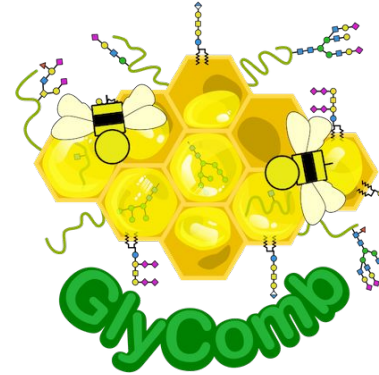
GlycoPOST

グライコミクス・グライコプロテオミクスのための質量分析データリポジトリ。
MIRAGE のガイドラインに従い、グライコプロテオミクス関連実験から得られた生データと実験に関するメタデータを登録することができる。



UniCarb-DR

質量分析から同定された糖鎖構造データのためのリポジトリ。
MIRAGE のガイドラインに従い、グライコミクス MS/MS スペクトルを登録することができる。



GlyComb

糖タンパク質、糖脂質といった複合糖質に対してアクセッション番号を割り当てるリポジトリ。

目下、開発中であり、まずは糖ペプチドを登録・処理する機能を公開予定である。



GlyComb

GlyComb

Data Registration Request Form

1 Upload 2 Confirmation 3 Completion

SELECT INPUT METHOD

Input from clipboard

Input from TSV file

Example Data

Paste your data to the text area below:

```
FTFTSHTPGEHQICLHSNSTK 18 G00681RS
FTFTSHTPGEHQICLHSNSTK 18 G11617NC
FTFTSHTPGEHQICLHSNSTK 18 G19243MH
FTFTSHTPGEHQICLHSNSTK 18 G28182LP
FTFTSHTPGEHQICLHSNSTK 18 G39448FX
FTFTSHTPGEHQICLHSNSTK 18 G52182RP
FTFTSHTPGEHQICLHSNSTK 18 G54675NM
FTFTSHTPGEHQICLHSNSTK 18 G58609GV
FTFTSHTPGEHQICLHSNSTK 18 G61514UE
FTFTSHTPGEHQICLHSNSTK 18 G92463DS
```

Submit

糖ペプチドを表すデータを入力
確認画面へ

GlyComb

Data Registration Request Form

1 Upload 2 Confirmation 3 Completion

No error found.

SEQUENCE 1

User input	Amino acid sequence	Glycosylation site	GlyYouCan ID
✓ FTFTSHTPGEHQICLHSNSTK 18 G00681RS	FTFTSHTPGEHQICLHSNSTK	18	G00681RS
			3x ☐
			4x ☐
			?? ☐
✓ FTFTSHTPGEHQICLHSNSTK 18 G11617NC	FTFTSHTPGEHQICLHSNSTK	18	G11617NC
			4x ☐
			4x ☐
			?? ☐
			?? ☐

...

✓ FTFTSHTPGEHQICLHSNSTK 18 G92463DS	FTFTSHTPGEHQICLHSNSTK	18	G92463DS
			3x ☐
			4x ☐
			?? ☐

Submit Back

受付番号が発行される。

GlyComb

Data Registration Request Form

1 Upload 2 Confirmation 3 Completion

✓ All your requests were accepted successfully.

Your receipt numbers are:

- [f438c9f8-540e-4c85-b499-e684d2499a0f](#)

バッチ処理により、アクセッション番号が発行
される。

データの閲覧ページは開発中。



謝辞

本研究は、科学技術振興機構(JST)、バイオサイエンスデータベースセンター(NBDC)ライフサイエンスデータベース統合推進事業「統合化推進プログラム」の助成を受けています。