GO, KEGG и все-все-все

Анастасия Жарикова, Юлия Алешина

14 марта 2025 – ФББ МГУ

ДАНО

- Набор интересных белков или генов

ДАНО

- Набор интересных белков или генов
- Откуда такой набор можно взять?

ДАНО

- Набор интересных белков или генов
- Откуда такой набор можно взять?

- Экспрессионный профиль
- Дифференциальная экспрессия
- Масспектрометрия
- Данные из статей
- Любимый набор
- …

СПИСОК ГЕНОВ

- PDHA2
- PDHA1
- TPI1
- AKR1A1
- G6PC1
- G6PC2
- GCK
- G6PC3
- GAPDHS

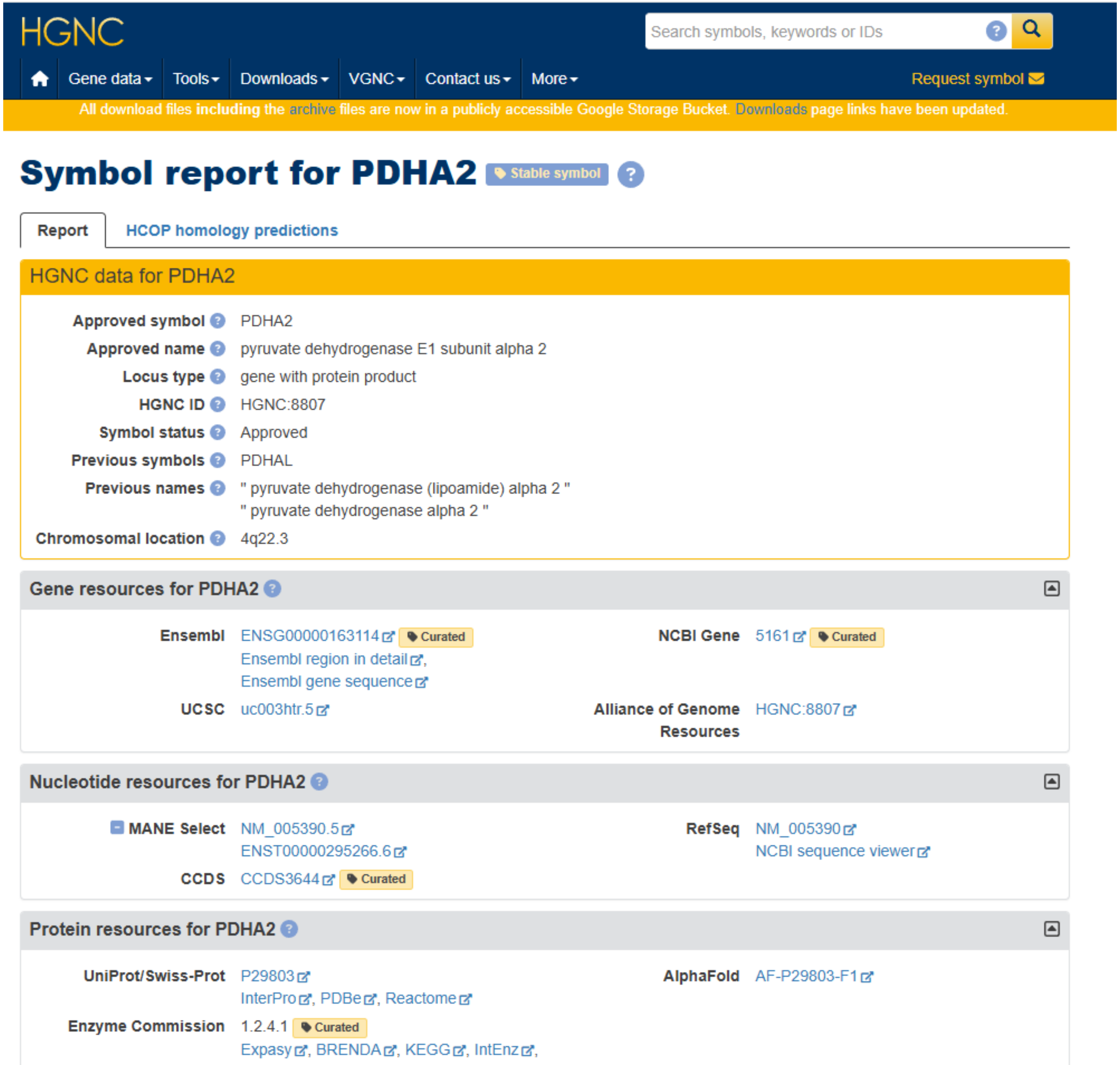
НАЗВАНИЯ ГЕНОВ

- Для некоторых организмов есть специальные комитеты, которые утверждают названия генов
- Названия генов человека утверждает **HUGO Gene Nomenclature Committee**

https://www.genenames.org/

- Для большинства видов найти ген можно в базе **NCBI Gene**

https://www.ncbi.nlm.nih.gov/gene



All download files including the archive files are now in a publicly accessible Google Storage Bucket. Downloads page links have been updated.

Symbol report for PDHA2

HGNC data for PDHA2

Approved symbol	PDHA2
Approved name	pyruvate dehydrogenase E1 subunit alpha 2
Locus type	gene with protein product
HGNC ID	HGNC:8807
Symbol status	Approved
Previous symbols	PDHAL
Previous names	" pyruvate dehydrogenase (lipoamide) alpha 2 " " pyruvate dehydrogenase alpha 2 "
Chromosomal location	4q22.3

Gene resources for PDHA2

Ensembl	ENSG00000163114 Curated Ensembl region in detail, Ensembl gene sequence	**NCBI Gene**	5161 Curated
UCSC	uc003htr.5	**Alliance of Genome Resources**	HGNC:8807

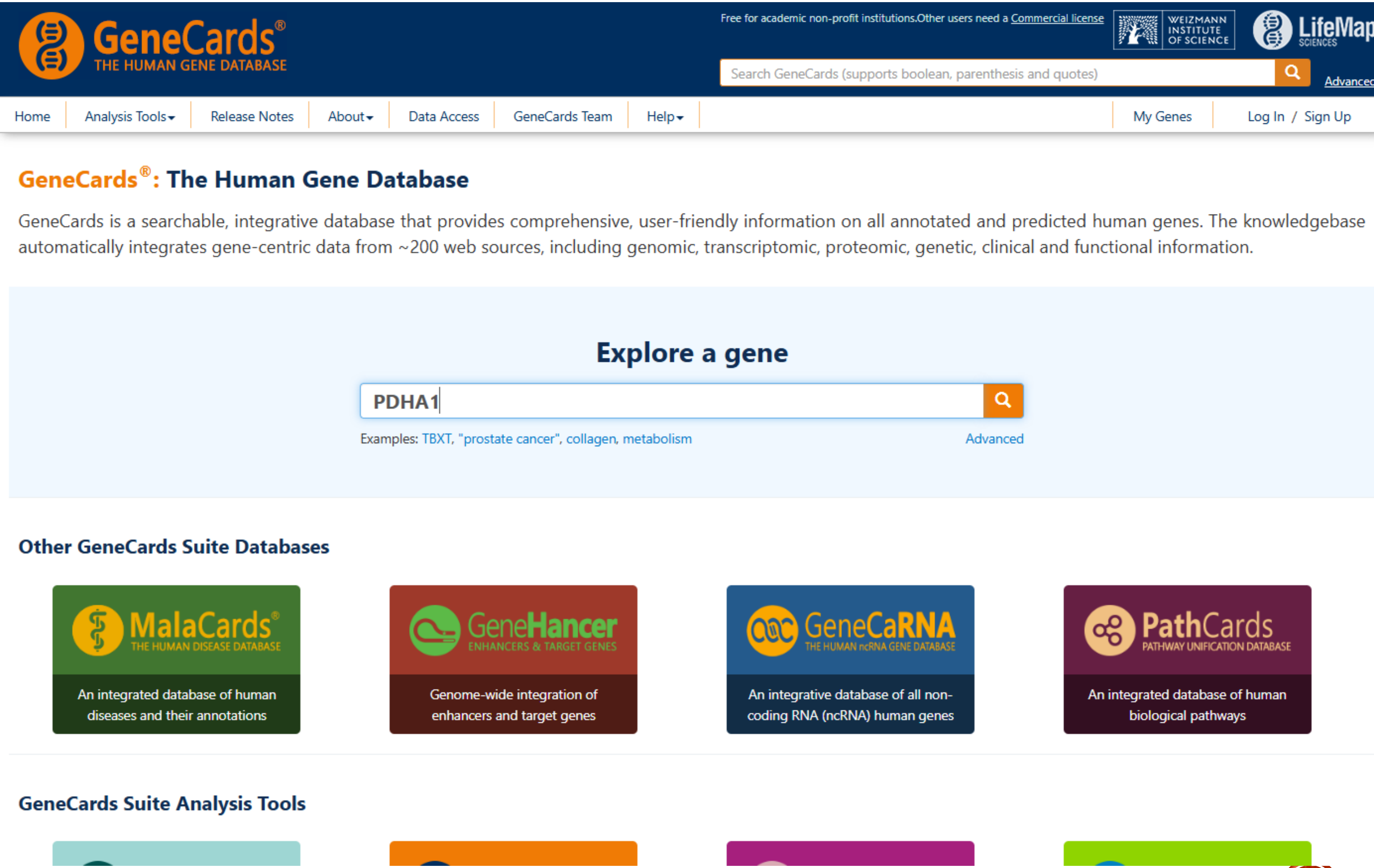
Nucleotide resources for PDHA2

MANE Select	NM_005390.5 ENST00000295266.6	**RefSeq**	NM_005390 NCBI sequence viewer
CCDS	CCDS3644 Curated		

Protein resources for PDHA2

UniProt/Swiss-Prot	P29803 InterPro, PDBe, Reactome	**AlphaFold**	AF-P29803-F1
Enzyme Commission	1.2.4.1 Curated Expasy, BRENDA, KEGG, IntEnz,		

https://www.genenames.org/



GeneCards®: The Human Gene Database

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. The knowledgebase automatically integrates gene-centric data from ~200 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information.

Explore a gene

PDHA1

Examples: TBXT, "prostate cancer", collagen, metabolism

Other GeneCards Suite Databases

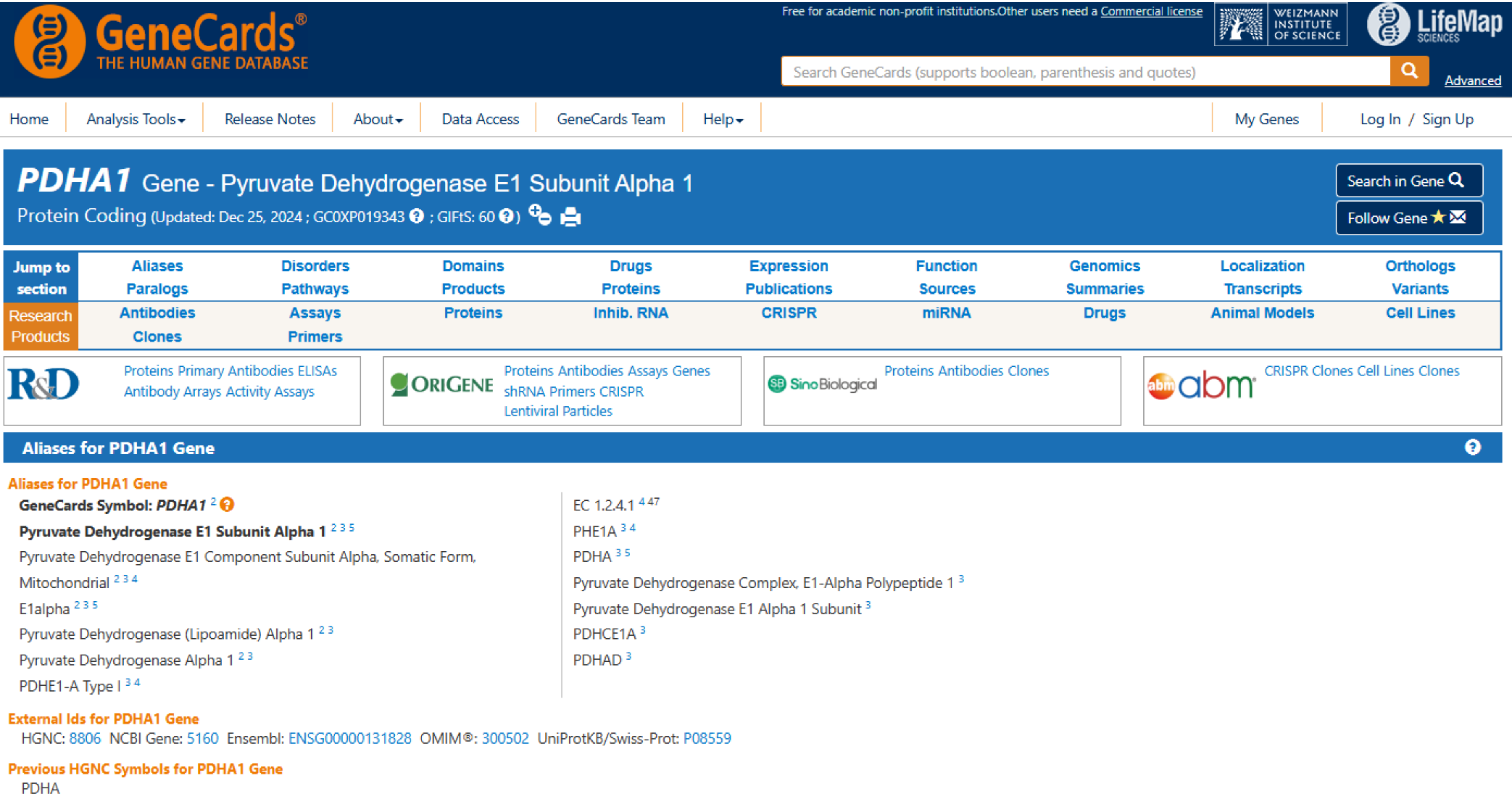
An integrated database of human diseases and their annotations

Genome-wide integration of enhancers and target genes

An integrative database of all non-coding RNA (ncRNA) human genes

An integrated database of human biological pathways

GeneCards Suite Analysis Tools



PDHA1 **Gene - Pyruvate Dehydrogenase E1 Subunit Alpha 1**

Protein Coding (Updated: Dec 25, 2024 ; GC0XP019343 ; GIFtS: 60)

Aliases for PDHA1 Gene

Aliases for PDHA1 Gene

GeneCards Symbol: *PDHA1* [2]

Pyruvate Dehydrogenase E1 Subunit Alpha 1 [2,3,5]

Pyruvate Dehydrogenase E1 Component Subunit Alpha, Somatic Form, Mitochondrial [2,3,4]

E1alpha [2,3,5]

Pyruvate Dehydrogenase (Lipoamide) Alpha 1 [2,3]

Pyruvate Dehydrogenase Alpha 1 [2,3]

PDHE1-A Type I [3,4]

EC 1.2.4.1 [4,47]

PHE1A [3,4]

PDHA [3,5]

Pyruvate Dehydrogenase Complex, E1-Alpha Polypeptide 1 [3]

Pyruvate Dehydrogenase E1 Alpha 1 Subunit [3]

PDHCE1A [3]

PDHAD [3]

External Ids for PDHA1 Gene

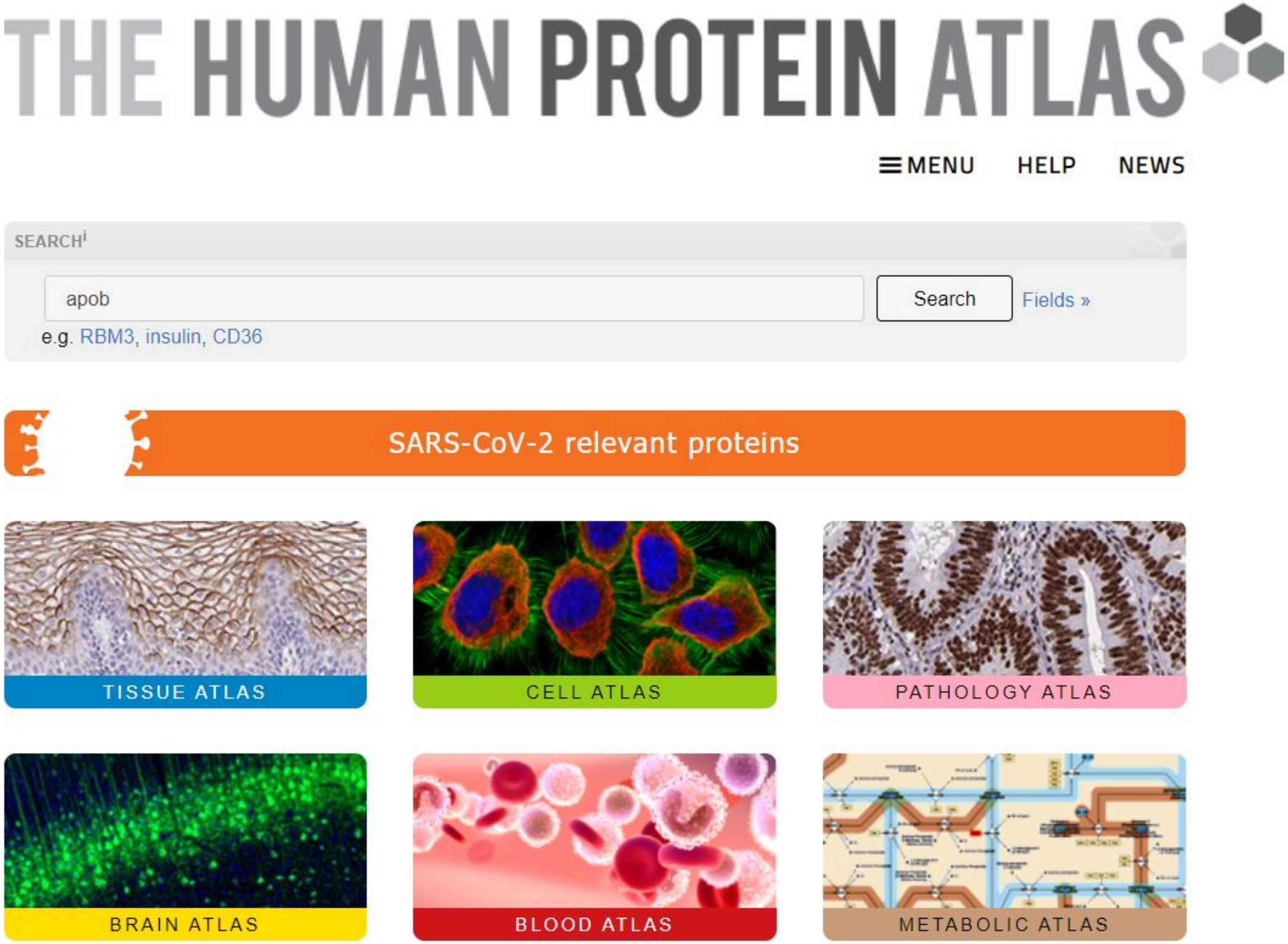
HGNC: 8806 NCBI Gene: 5160 Ensembl: ENSG00000131828 OMIM®: 300502 UniProtKB/Swiss-Prot: P08559

Previous HGNC Symbols for PDHA1 Gene

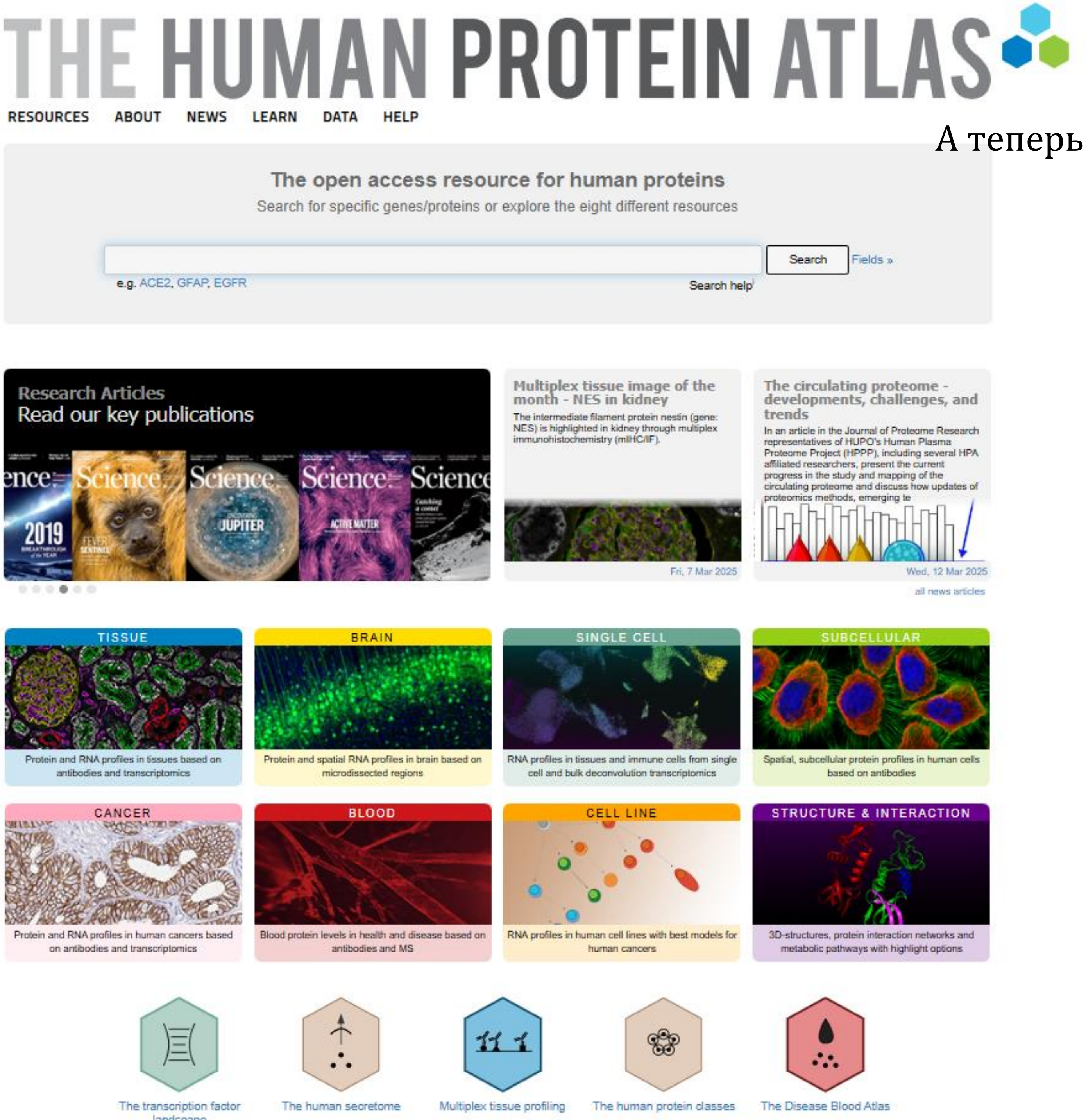
PDHA

https://www.proteinatlas.org/

Было 3 года назад



А теперь…



The open access resource for human proteins

Search for specific genes/proteins or explore the eight different resources

e.g. ACE2, GFAP, EGFR

Research Articles
Read our key publications

Multiplex tissue image of the month - NES in kidney

The intermediate filament protein nestin (gene: NES) is highlighted in kidney through multiplex immunohistochemistry (mIHC/IF).

Fri, 7 Mar 2025

The circulating proteome - developments, challenges, and trends

In an article in the Journal of Proteome Research representatives of HUPO's Human Plasma Proteome Project (HPPP), including several HPA affiliated researchers, present the current progress in the study and mapping of the circulating proteome and discuss how updates of proteomics methods, emerging te

Wed, 12 Mar 2025

all news articles

TISSUE
Protein and RNA profiles in tissues based on antibodies and transcriptomics

BRAIN
Protein and spatial RNA profiles in brain based on microdissected regions

SINGLE CELL
RNA profiles in tissues and immune cells from single cell and bulk deconvolution transcriptomics

SUBCELLULAR
Spatial, subcellular protein profiles in human cells based on antibodies

CANCER
Protein and RNA profiles in human cancers based on antibodies and transcriptomics

BLOOD
Blood protein levels in health and disease based on antibodies and MS

CELL LINE
RNA profiles in human cell lines with best models for human cancers

STRUCTURE & INTERACTION
3D-structures, protein interaction networks and metabolic pathways with highlight options

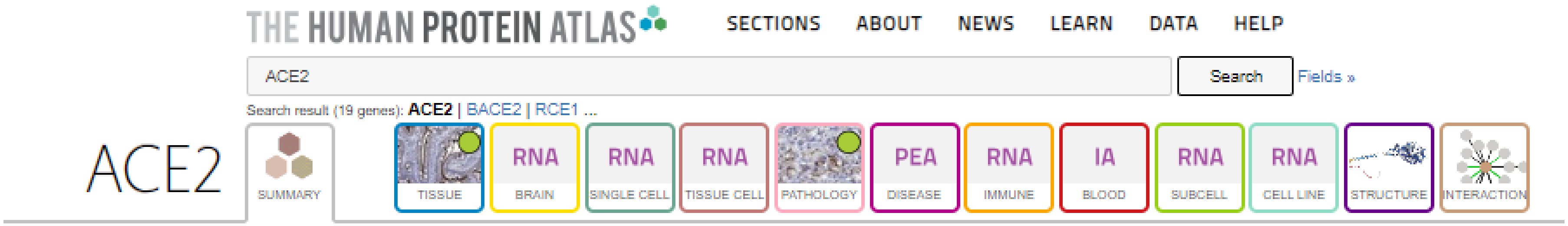
The transcription factor

The human secretome

Multiplex tissue profiling

The human protein classes

The Disease Blood Atlas

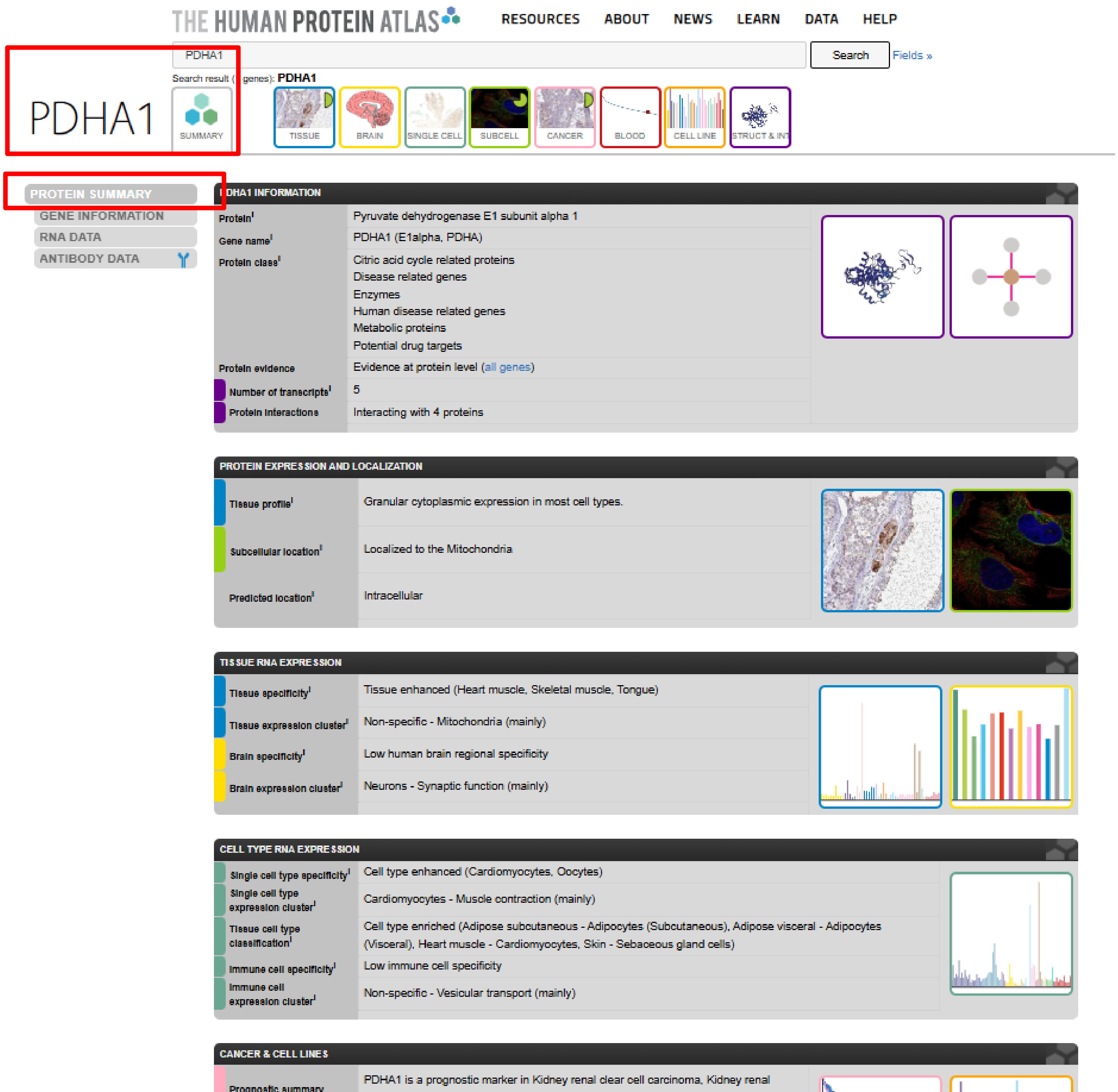


ACE2

Search result (19 genes): **ACE2** | BACE2 | RCE1 ...

ACE2

Ваш ID может быть специфичен для какой-либо ткани, может присутствовать везде или нигде не быть детектирован



Search result (genes): **PDHA1**

PDHA1

DHA1 INFORMATION

Protein	Pyruvate dehydrogenase E1 subunit alpha 1
Gene name	PDHA1 (E1alpha, PDHA)
Protein class	Citric acid cycle related proteins Disease related genes Enzymes Human disease related genes Metabolic proteins Potential drug targets
Protein evidence	Evidence at protein level (all genes)
Number of transcripts	5
Protein interactions	Interacting with 4 proteins

PROTEIN EXPRESSION AND LOCALIZATION

Tissue profile	Granular cytoplasmic expression in most cell types.
Subcellular location	Localized to the Mitochondria
Predicted location	Intracellular

TISSUE RNA EXPRESSION

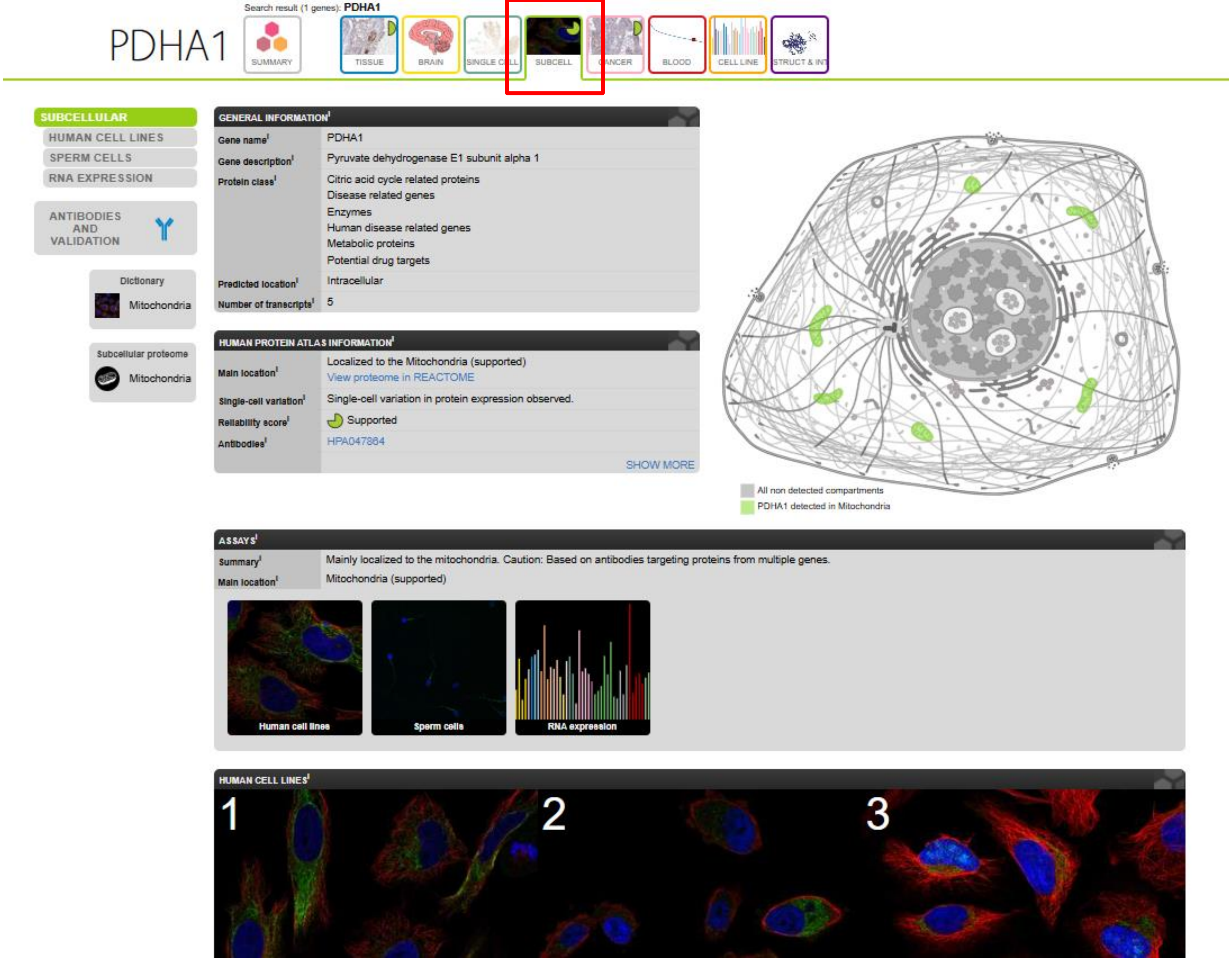
Tissue specificity	Tissue enhanced (Heart muscle, Skeletal muscle, Tongue)
Tissue expression cluster	Non-specific - Mitochondria (mainly)
Brain specificity	Low human brain regional specificity
Brain expression cluster	Neurons - Synaptic function (mainly)

CELL TYPE RNA EXPRESSION

Single cell type specificity	Cell type enhanced (Cardiomyocytes, Oocytes)
Single cell type expression cluster	Cardiomyocytes - Muscle contraction (mainly)
Tissue cell type classification	Cell type enriched (Adipose subcutaneous - Adipocytes (Subcutaneous), Adipose visceral - Adipocytes (Visceral), Heart muscle - Cardiomyocytes, Skin - Sebaceous gland cells)
Immune cell specificity	Low immune cell specificity
Immune cell expression cluster	Non-specific - Vesicular transport (mainly)

CANCER & CELL LINES

Prognostic summary	PDHA1 is a prognostic marker in Kidney renal clear cell carcinoma, Kidney renal



GENERAL INFORMATION

Gene name	PDHA1
Gene description	Pyruvate dehydrogenase E1 subunit alpha 1
Protein class	Citric acid cycle related proteins Disease related genes Enzymes Human disease related genes Metabolic proteins Potential drug targets
Predicted location	Intracellular
Number of transcripts	5

HUMAN PROTEIN ATLAS INFORMATION

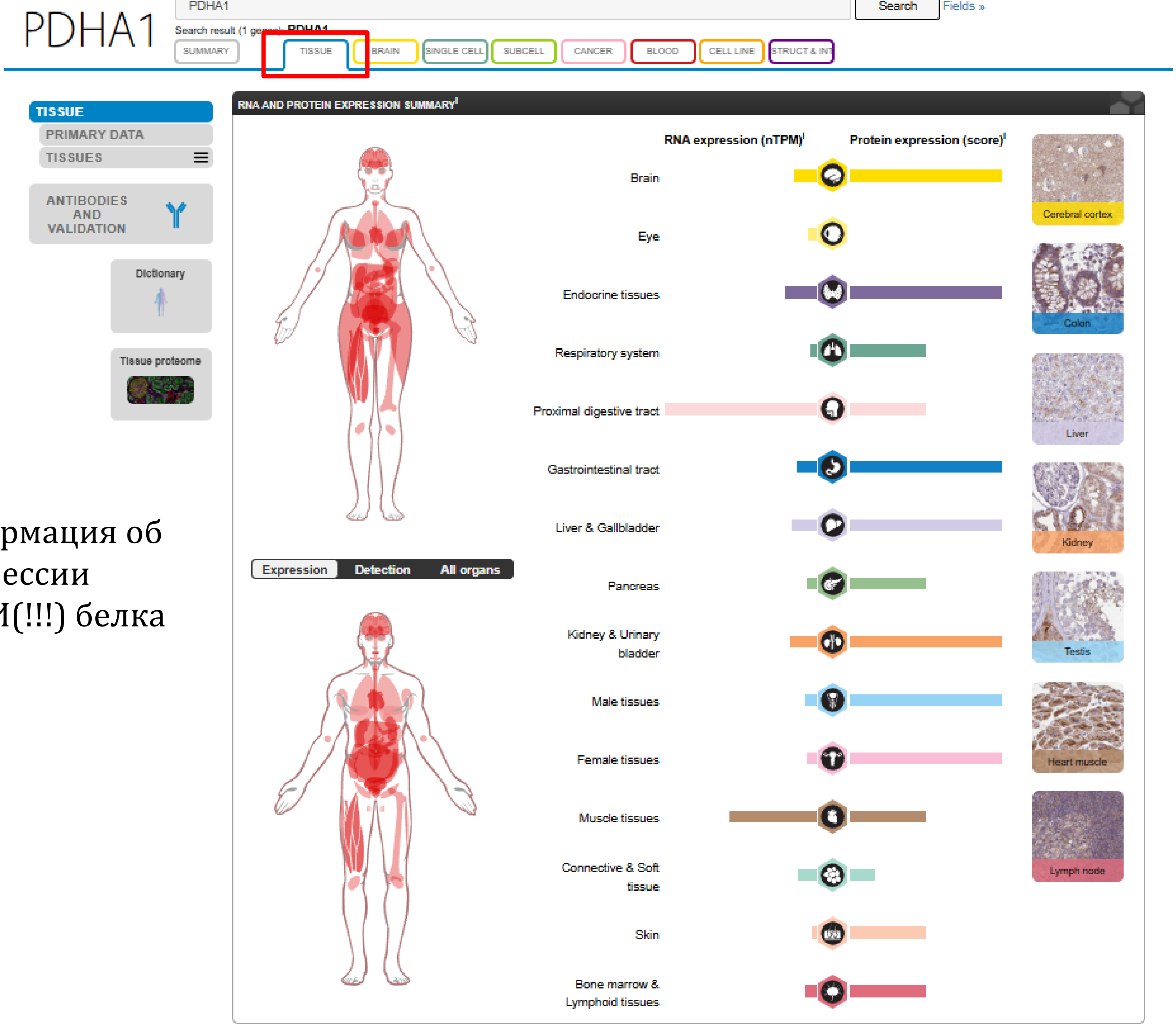
Main location	Localized to the Mitochondria (supported) View proteome in REACTOME
Single-cell variation	Single-cell variation in protein expression observed.
Reliability score	Supported
Antibodies	HPA047864

SHOW MORE

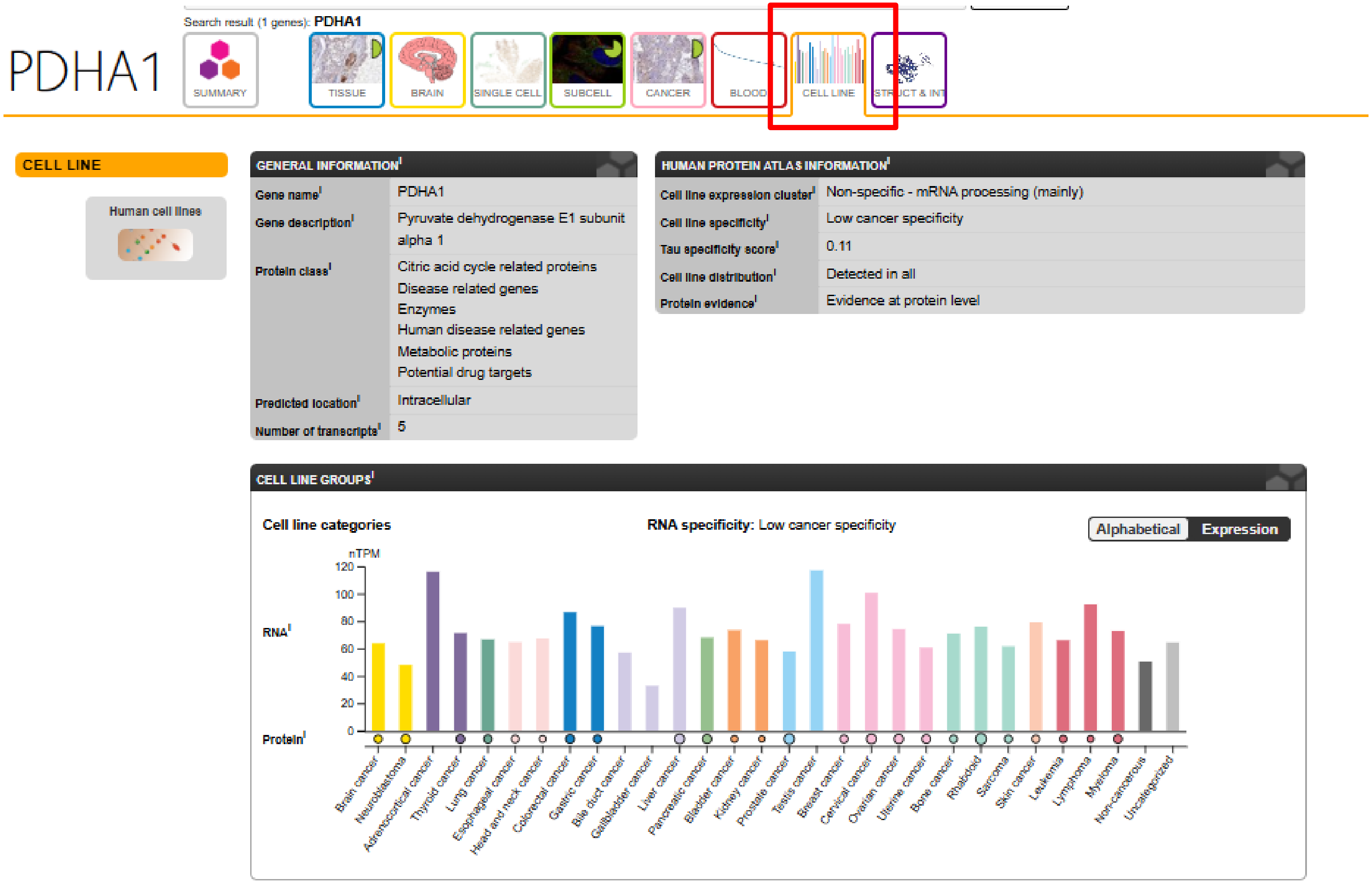
ASSAYS

Summary	Mainly localized to the mitochondria. Caution: Based on antibodies targeting proteins from multiple genes.
Main location	Mitochondria (supported)

HUMAN CELL LINES



Информация об экспрессии РНК И(!!!) белка



GENERAL INFORMATION

Gene name	PDHA1
Gene description	Pyruvate dehydrogenase E1 subunit alpha 1
Protein class	Citric acid cycle related proteins Disease related genes Enzymes Human disease related genes Metabolic proteins Potential drug targets
Predicted location	Intracellular
Number of transcripts	5

HUMAN PROTEIN ATLAS INFORMATION

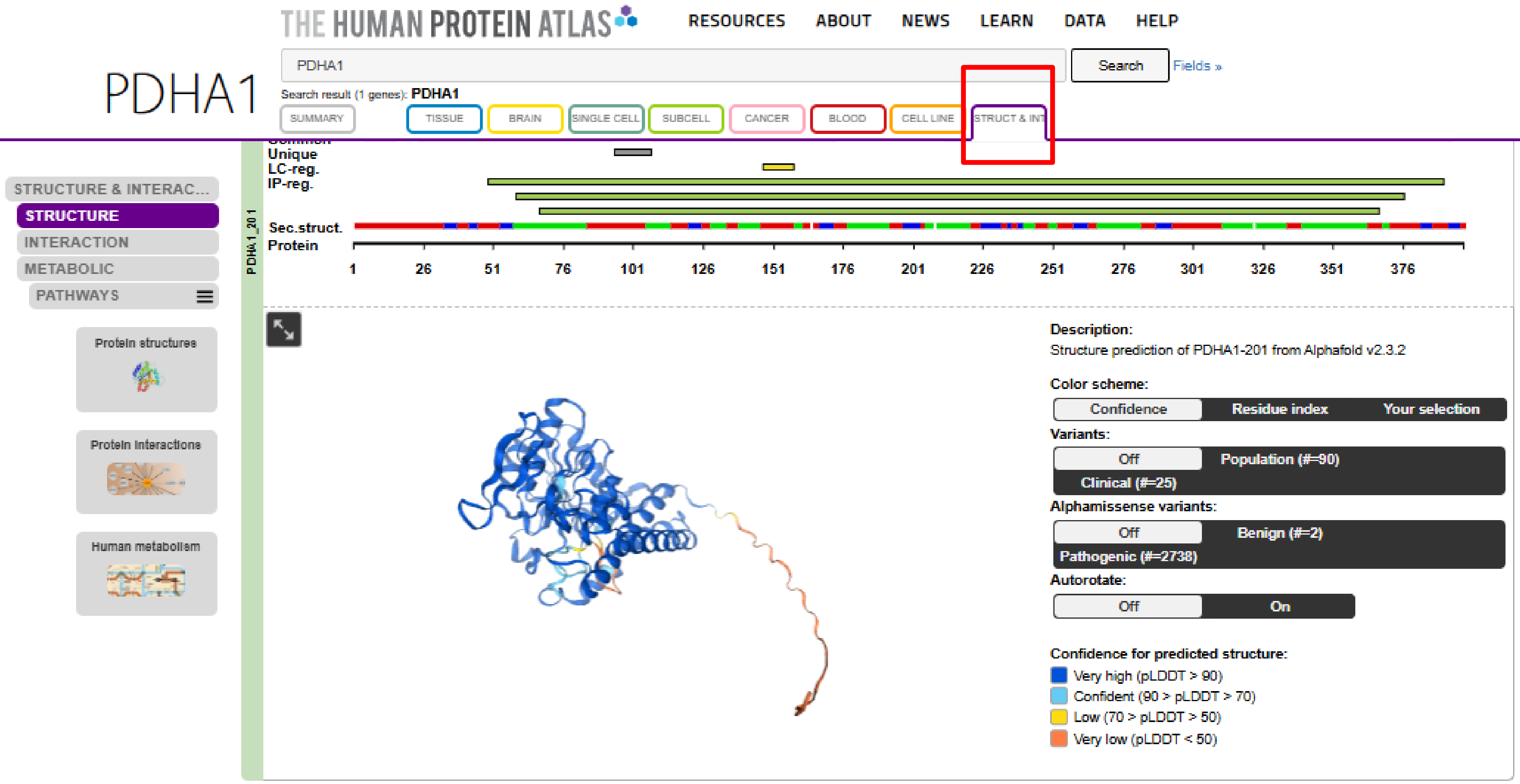
Cell line expression cluster	Non-specific - mRNA processing (mainly)
Cell line specificity	Low cancer specificity
Tau specificity score	0.11
Cell line distribution	Detected in all
Protein evidence	Evidence at protein level

CELL LINE GROUPS

Cell line categories

RNA specificity: Low cancer specificity

PDHA1



Search result (1 genes): **PDHA1**

Description:
Structure prediction of PDHA1-201 from Alphafold v2.3.2

Color scheme: Confidence | Residue index | Your selection

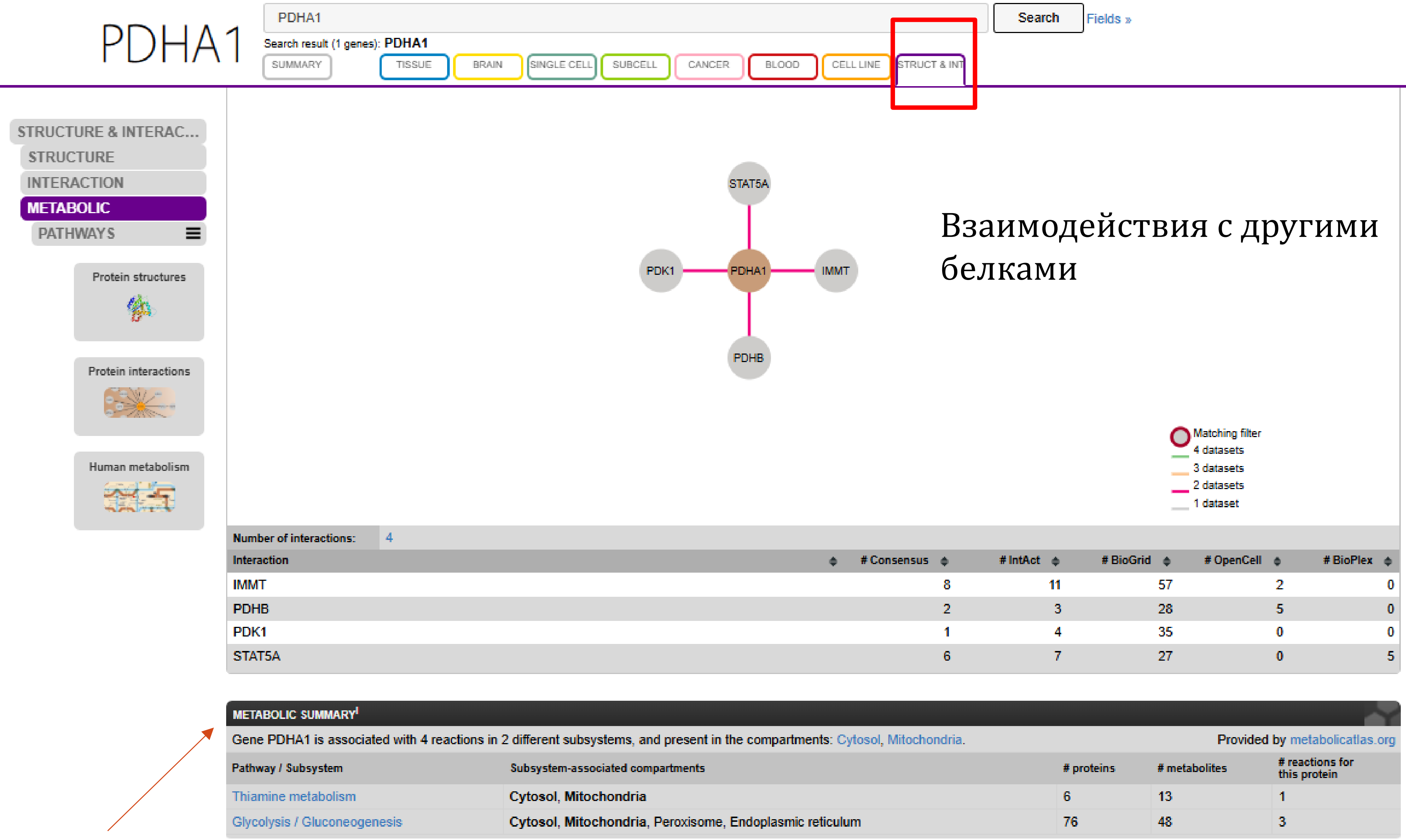
Variants: Off | Population (#=90) | Clinical (#=25)

Alphamissense variants: Off | Benign (#=2) | Pathogenic (#=2738)

Autorotate: Off | On

Confidence for predicted structure:

- 🟦 Very high (pLDDT > 90)
- 🟦 Confident (90 > pLDDT > 70)
- 🟨 Low (70 > pLDDT > 50)
- 🟧 Very low (pLDDT < 50)



Search result (1 genes): **PDHA1**

Взаимодействия с другими белками

Number of interactions: 4

Interaction	# Consensus	# IntAct	# BioGrid	# OpenCell	# BioPlex
IMMT	8	11	57	2	0
PDHB	2	3	28	5	0
PDK1	1	4	35	0	0
STAT5A	6	7	27	0	5

METABOLIC SUMMARY[i]

Gene PDHA1 is associated with 4 reactions in 2 different subsystems, and present in the compartments: Cytosol, Mitochondria.

Provided by metabolicatlas.org

Pathway / Subsystem	Subsystem-associated compartments	# proteins	# metabolites	# reactions for this protein
Thiamine metabolism	**Cytosol, Mitochondria**	6	13	1
Glycolysis / Gluconeogenesis	**Cytosol, Mitochondria**, Peroxisome, Endoplasmic reticulum	76	48	3

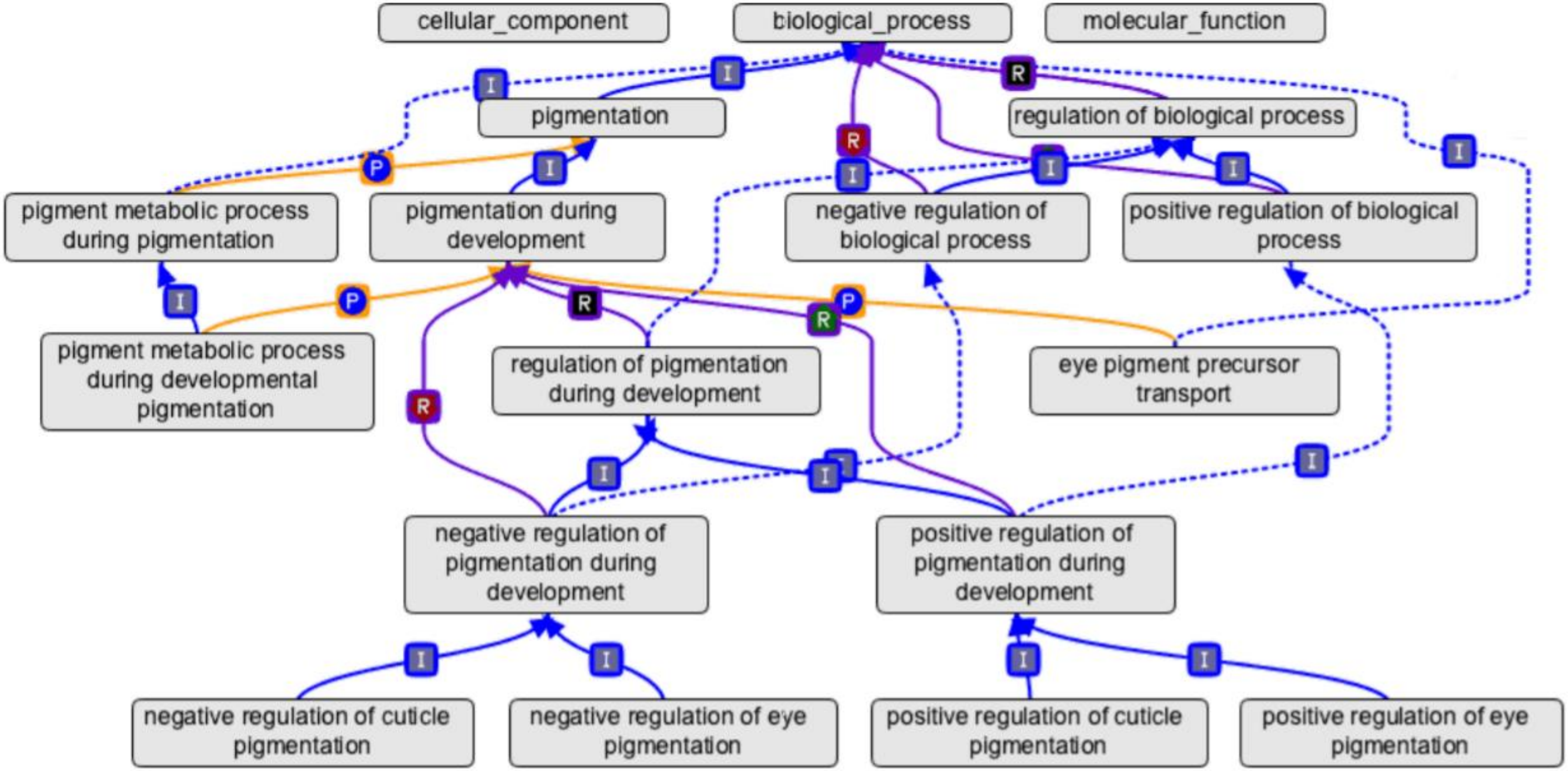
Метаболические пути

Что дальше делать?

GENE ONTOLOGY – GO

- http://geneontology.org/

- База данных GO представляет собой граф биологических терминов, соединенных различными отношениями



cellular_component

biological_process

molecular_function

pigmentation

regulation of biological process

pigment metabolic process during pigmentation

pigmentation during development

negative regulation of biological process

positive regulation of biological process

pigment metabolic process during developmental pigmentation

regulation of pigmentation during development

eye pigment precursor transport

negative regulation of pigmentation during development

positive regulation of pigmentation during development

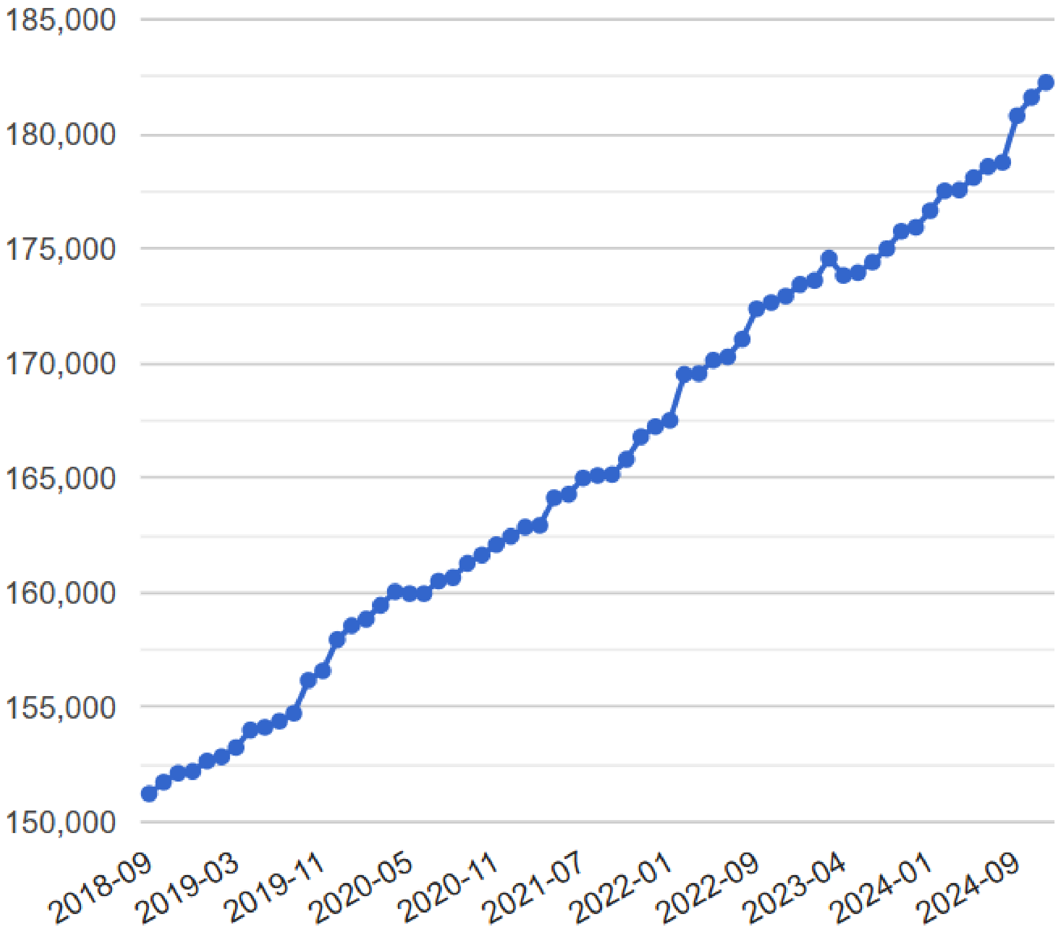
negative regulation of cuticle pigmentation

negative regulation of eye pigmentation

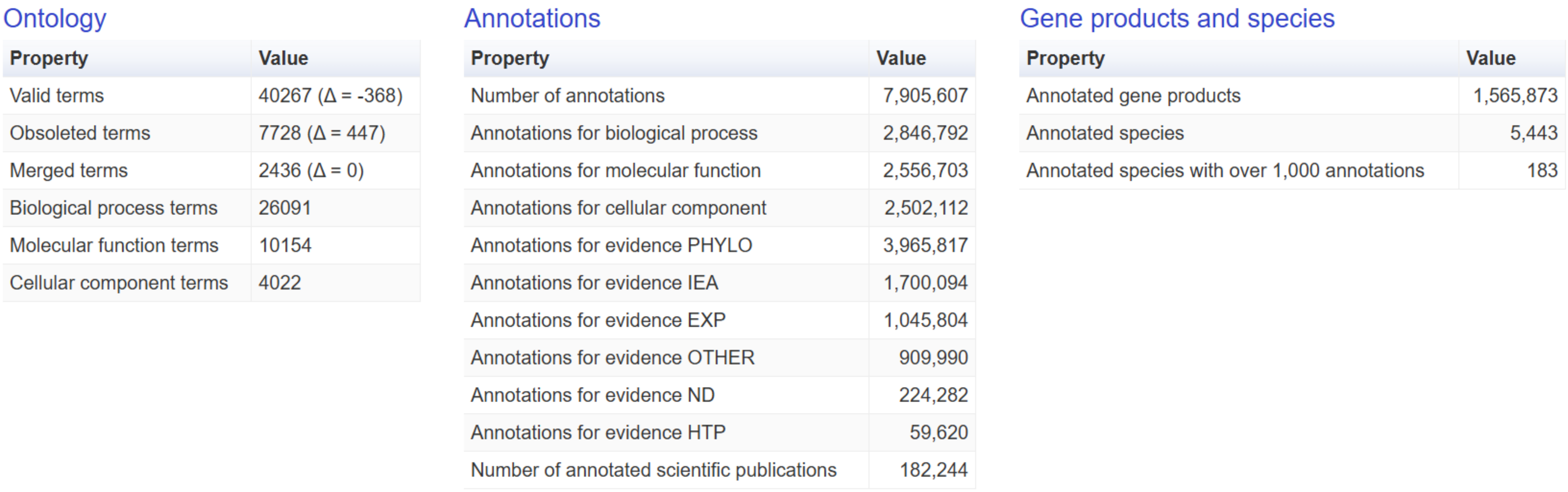
positive regulation of cuticle pigmentation

positive regulation of eye pigmentation

КОЛИЧЕСТВО ПУБЛИКАЦИЙ, В КОТОРЫХ ИСПОЛЬЗУЕТСЯ GO



release 2025-02-06



Ontology

Property	Value
Valid terms	40267 (Δ = -368)
Obsoleted terms	7728 (Δ = 447)
Merged terms	2436 (Δ = 0)
Biological process terms	26091
Molecular function terms	10154
Cellular component terms	4022

Annotations

Property	Value
Number of annotations	7,905,607
Annotations for biological process	2,846,792
Annotations for molecular function	2,556,703
Annotations for cellular component	2,502,112
Annotations for evidence PHYLO	3,965,817
Annotations for evidence IEA	1,700,094
Annotations for evidence EXP	1,045,804
Annotations for evidence OTHER	909,990
Annotations for evidence ND	224,282
Annotations for evidence HTP	59,620
Number of annotated scientific publications	182,244

Gene products and species

Property	Value
Annotated gene products	1,565,873
Annotated species	5,443
Annotated species with over 1,000 annotations	183

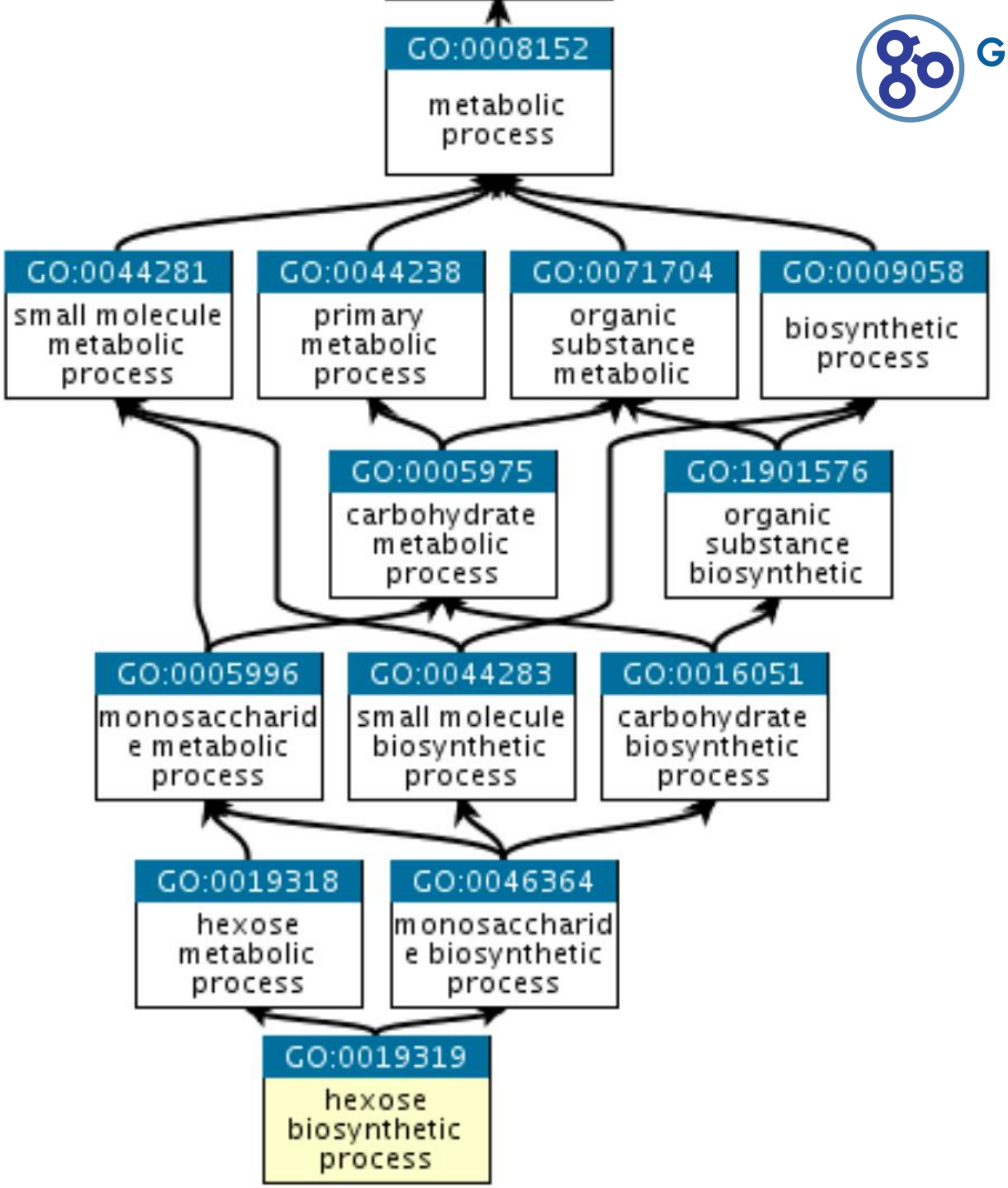
Аннотации – это утверждения, связывающие конкретный генный продукт с конкретным онтологическим термином, основанные на фактических данных.

Молекулярные функции (англ. molecular function) — специфическая активность генного продукта на молекулярном уровне, например, связывание углеводов или АТФазная активность.

Биологические процессы (англ. biological process) — сложные явления, необходимые для жизнедеятельности организмов и происходящие благодаря осуществлению последовательности молекулярных функций, например, митоз или биосинтез пуринов.

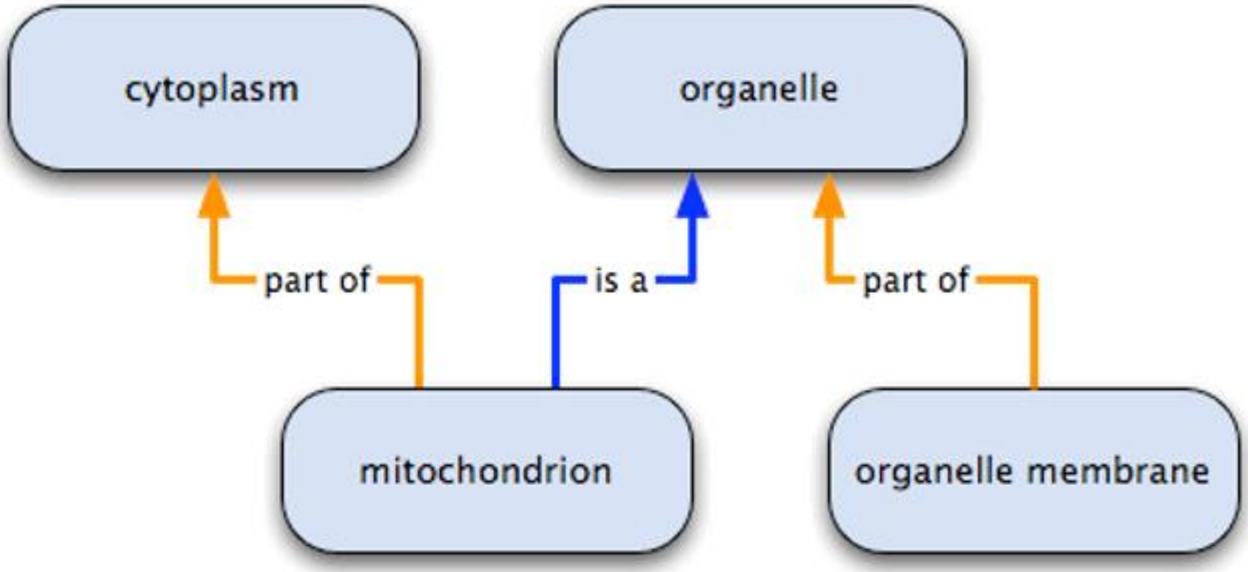
Клеточные компоненты (англ. cellular component) — части клетки или внеклеточного пространства, где осуществляется функция генного продукта, например, ядро или рибосома.

GENEONTOLOGY
Unifying Biology



Термин GO (GO term) – узел графа – может иметь любое количество связей с любыми другими узлами

Отношения между узлами – ребра графа бывают разного типа

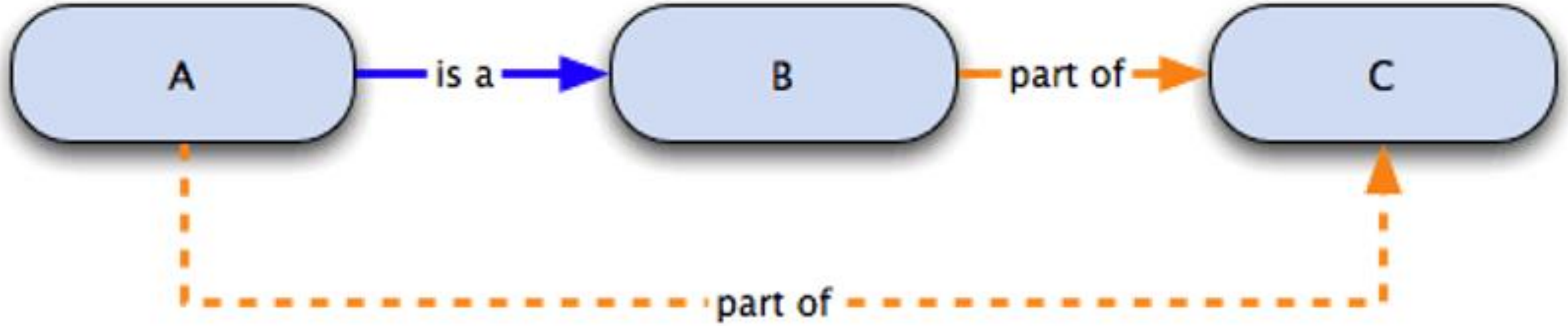


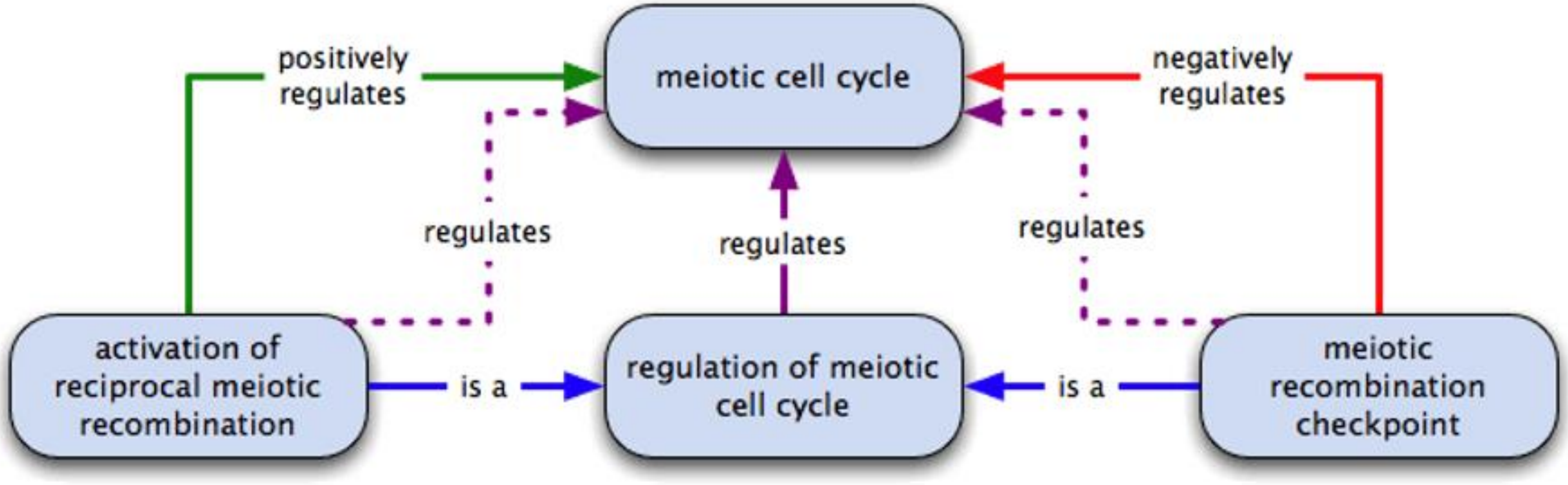
A *is a* B

B *part of* C

Значит, A *part of* C

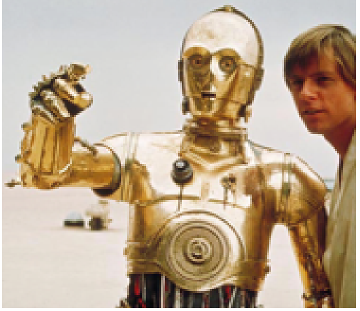
→	Указанные непосредственно
→	
⇢	Логически вытекающее




meiotic cell cycle
positively regulates
negatively regulates
regulates
regulates
regulates
activation of reciprocal meiotic recombination
is a
regulation of meiotic cell cycle
is a
meiotic recombination checkpoint

Типы достоверности

Inferred from Electronic Annotation (**IEA**)



Inferred from Sequence or structural Similarity (**ISS**)
- Inferred from Sequence Orthology (**ISO**)
- Inferred from Sequence Alignment (**ISA**)
- Inferred from Sequence Model (**ISM**)

Inferred from Experiment (**EXP**)
- Inferred from Direct Assay (**IDA**)
- Inferred from Physical Interaction (**IPI**)
- Inferred from Mutant Phenotype (**IMP**)
- Inferred from Genetic Interaction (**IGI**)
- Inferred from Expression Pattern (**IEP**)

Traceable Author Statement (**TAS**)
Non-traceable Author Statement (**NAS**)

🟥 Плохо
🟨
🟩 Хорошо

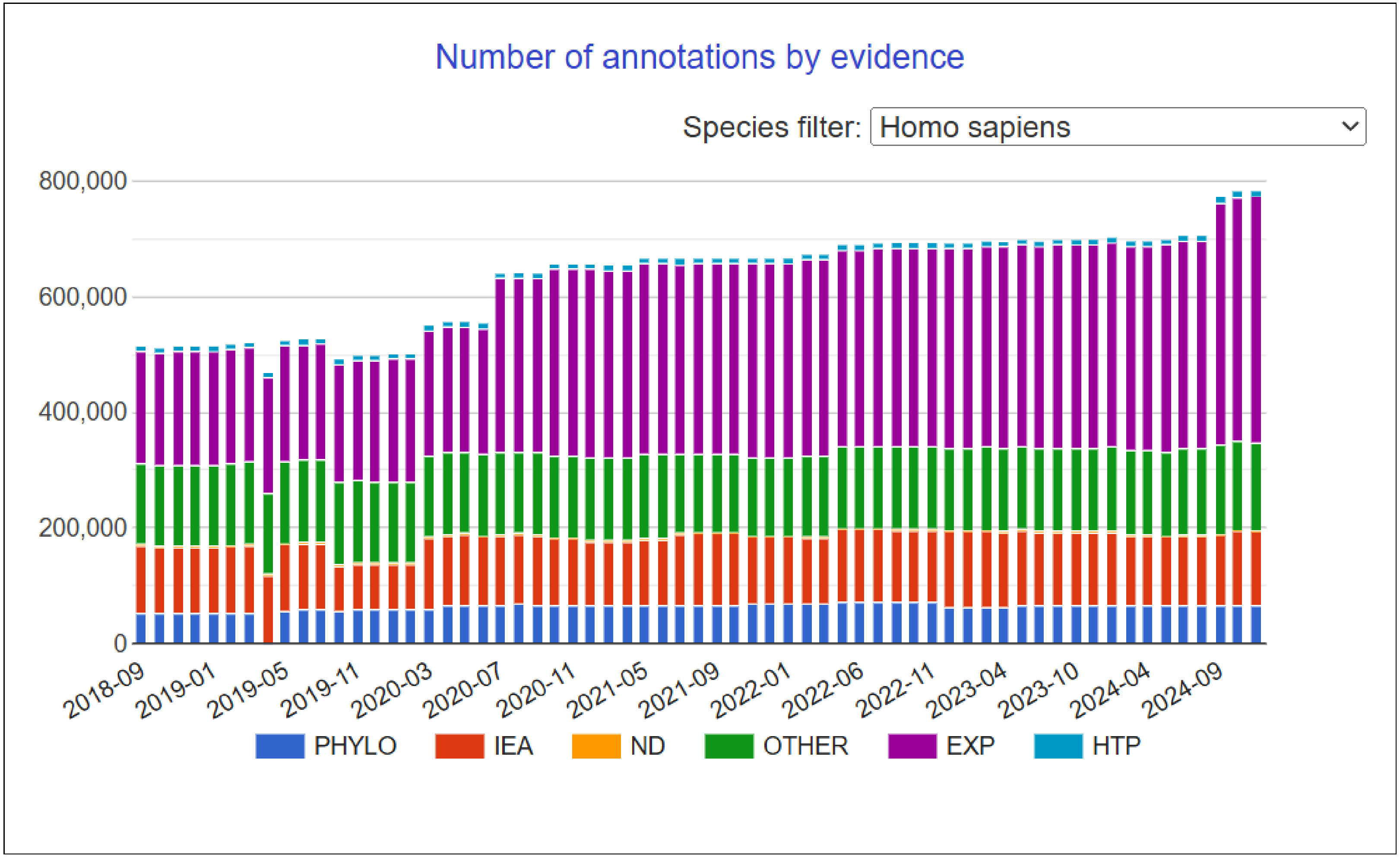
http://geneontology.org/docs/guide-go-evidence-codes/

ТИПЫ ДОСТОВЕРНОСТИ

- Экспериментальное подтверждение - Inferred from Experiment (EXP)
  - Inferred from Direct Assay (IDA)
  - Inferred from Physical Interaction (IPI)
  - Inferred from Mutant Phenotype (IMP)
  - Inferred from Genetic Interaction (IGI)
  - Inferred from Expression Pattern (IEP)
- На основе филогении:
  - Например, Inferred from Biological aspect of Ancestor (IBA)
- На основе анализа последовательностей или др:
  - Inferred from Sequence or structural Similarity (ISS)
  - Inferred from Sequence Orthology (ISO)
  - Inferred from Sequence Alignment (ISA)
  - Inferred from Sequence Model (ISM)
- Типы утверждений по наличию ссылки на статью:
  - Traceable Author Statement (TAS)
  - Non-traceable Author Statement (NAS)
- Inferred from Electronic Annotation (IEA)

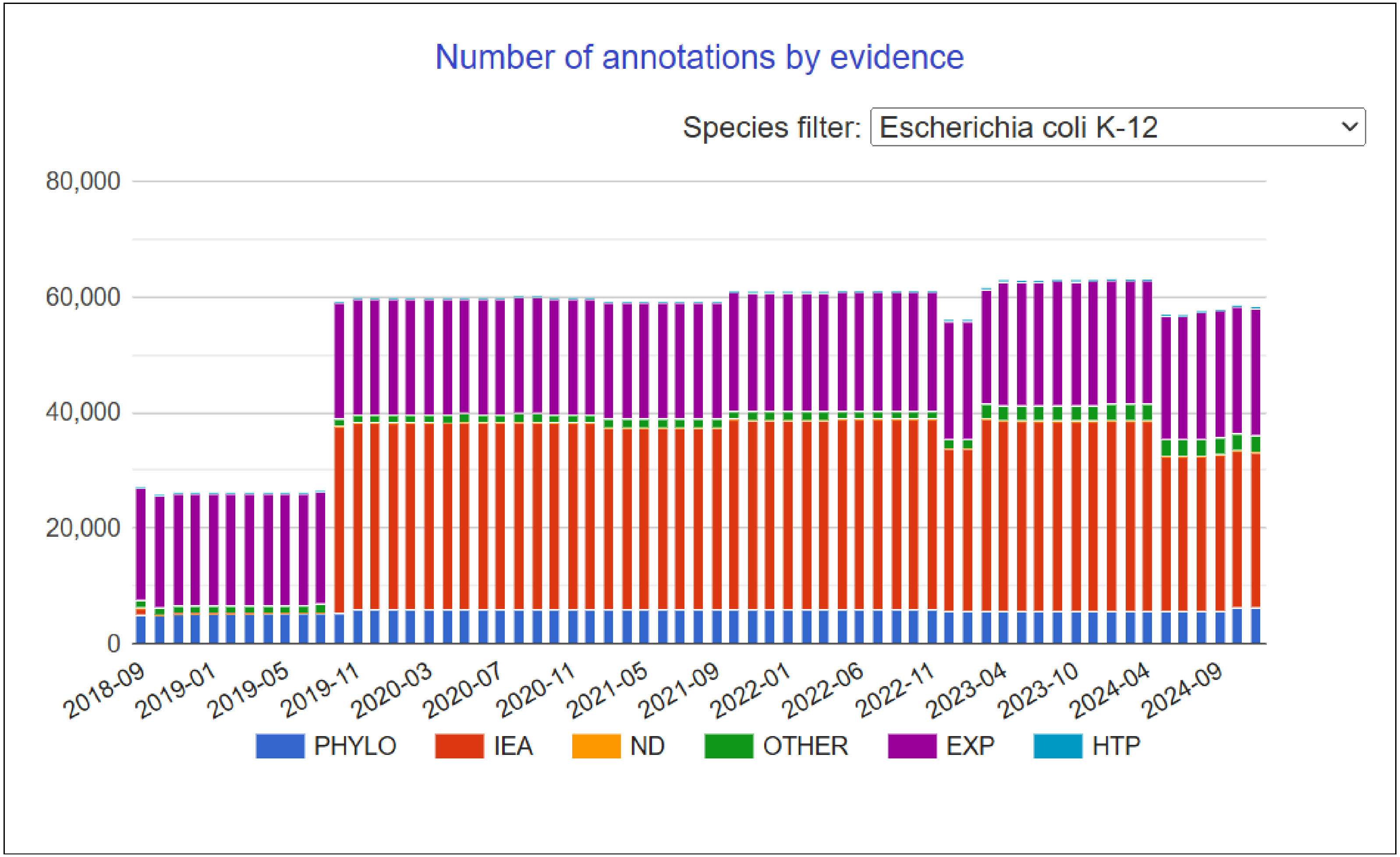
http://geneontology.org/docs/guide-go-evidence-codes/

☐	Gene/product	Gene/product name	Annotation qualifier	GO class (direct)	Annotation extension	Contributor	Organism	Evidence	Evidence with	PANTHER family	Type	Isofor
☐	Fcer1g	Fc fragment of IgE receptor Ig		positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
☐	Fcer1g	Fc fragment of IgE receptor Ig		positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
☐	Fcer1g	Fc fragment of IgE receptor Ig		positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:10572	high affinity immunoglobulin epsilon receptor gamma-subunit pthr16803	gene	
☐	Fcgr2a	Fc fragment of IgG receptor IIa		positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:736451	immunoglobulin fc receptor pthr11481	gene	
☐	Fcgr2a	Fc fragment of IgG receptor IIa		positive regulation of type IIa hypersensitivity		RGD	Rattus norvegicus	ISO	RGD:736451	immunoglobulin fc receptor pthr11481	gene	
☐	Fcgr2a	Fc fragment of IgG receptor		regulation of antibody-		RGD	Rattus norvegicus	IMP		immunoglobulin fc receptor pthr11481	gene	



Number of annotations by evidence

Species filter: Homo sapiens



Number of annotations by evidence

Species filter: Escherichia coli K-12

GO Enrichment Analysis

Powered by PANTHER

APOH
APP
CND2
COL3A1
COL5A2
CXCL6

biological process

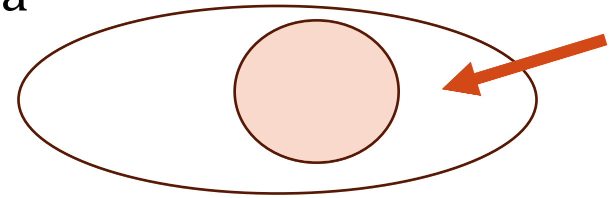
Homo sapiens

Gene set example: genes up-regulated during formation of blood vessels/angiogenesis (source: msigdb)

https://geneontology.org/

АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

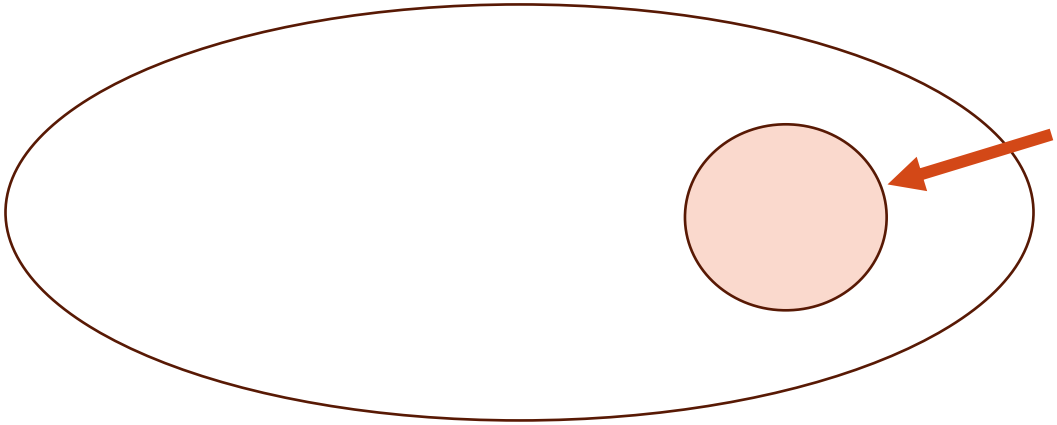
Есть список генов.
Перепредставлены ли в нем митохондриальные гены?



Гены из нашего списка
N=50

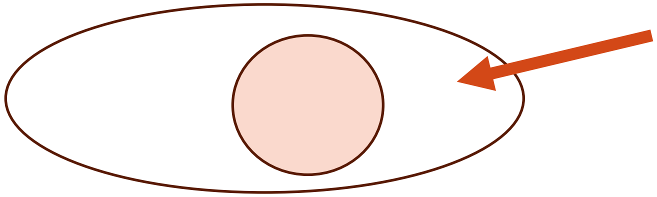
Митохондриальные
N=15

Это много или мало? С чем сравнить?



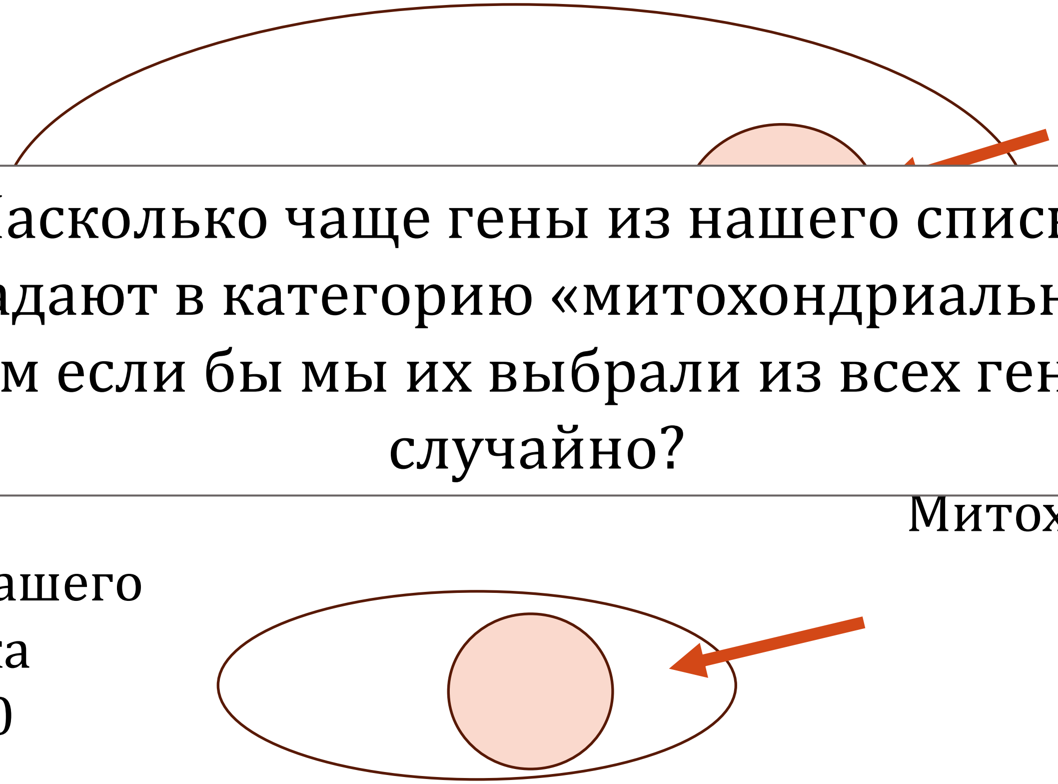
Все гены в геноме
N=20 000

Митохондриальные
N=50



Гены из нашего
списка
N=50

Митохондриальные
N=15



Все гены в геноме
N=20 000

Митохондриальные
N=50

Насколько чаще гены из нашего списка попадают в категорию «митохондриальные», чем если бы мы их выбрали из всех генов случайно?

Митохондриальные
N=15

Гены из нашего списка
N=50

АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

- Для GO термина «митохондриальные» можно все гены разделить на:
  - Гены из нашего списка из GO термина «митохондриальные»
  - Гены из нашего списка вне GO термина «митохондриальные»
  - Все гены из базы (background) из GO термина «митохондриальные»
  - Все гены из базы (background) вне GO термина «митохондриальные»

АНАЛИЗ ОБОГАЩЕНИЯ ТЕРМИНАМИ GO

- В GO тысячи терминов!

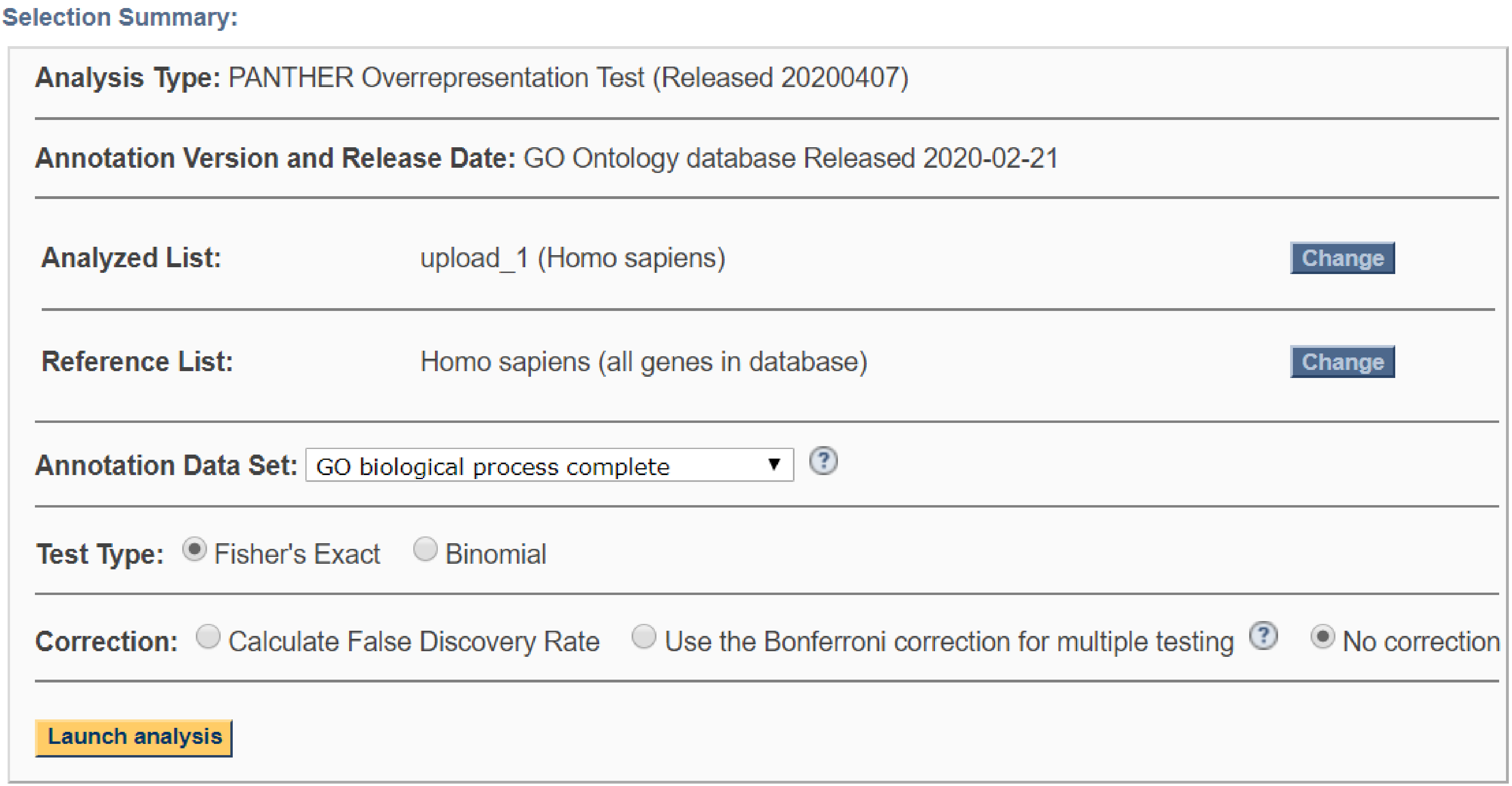
РАЗНЫЕ СЕРВИСЫ ИСПОЛЬЗУЮТ РАЗНЫЕ СТАТИСТИЧЕСКИЕ ТЕСТЫ

(Выбор теста зависит от задачи)

- Точный тест Фишера
- Биномиальный тест
- Критерий хи-квадрат
- и другие

P-value

Всё, кроме теста Фишера, лучше применять, если у вас много генов http://geneontology.org/docs/go-enrichment-analysis/



Selection Summary:

Analysis Type: PANTHER Overrepresentation Test (Released 20200407)

Annotation Version and Release Date: GO Ontology database Released 2020-02-21

Analyzed List: upload_1 (Homo sapiens)

Reference List: Homo sapiens (all genes in database)

Annotation Data Set: GO biological process complete ▼

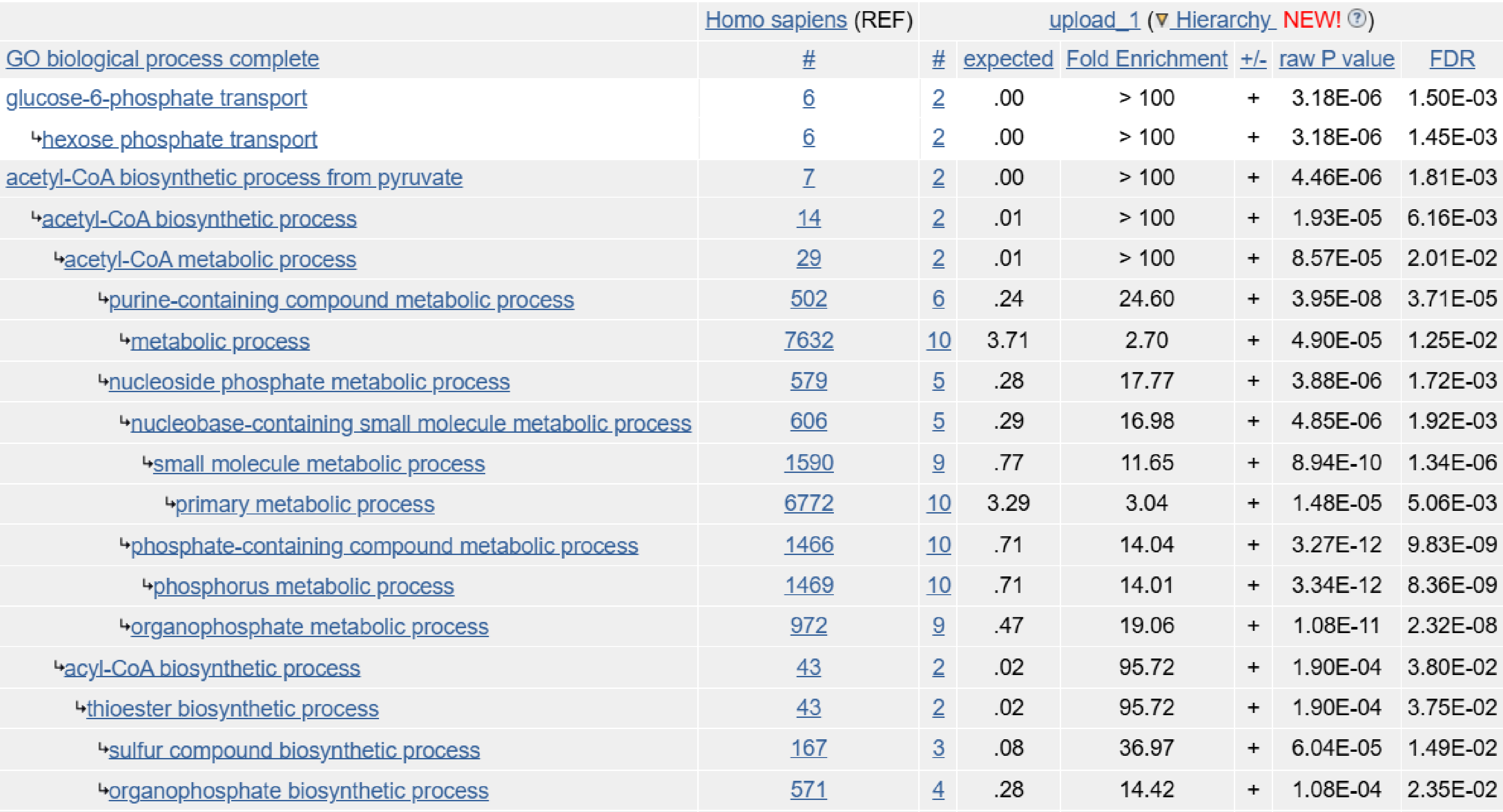
Test Type: ● Fisher's Exact ○ Binomial

Correction: ○ Calculate False Discovery Rate ○ Use the Bonferroni correction for multiple testing ● No correction

Число генов, приписанных к процессу

Число генов из нашего набора

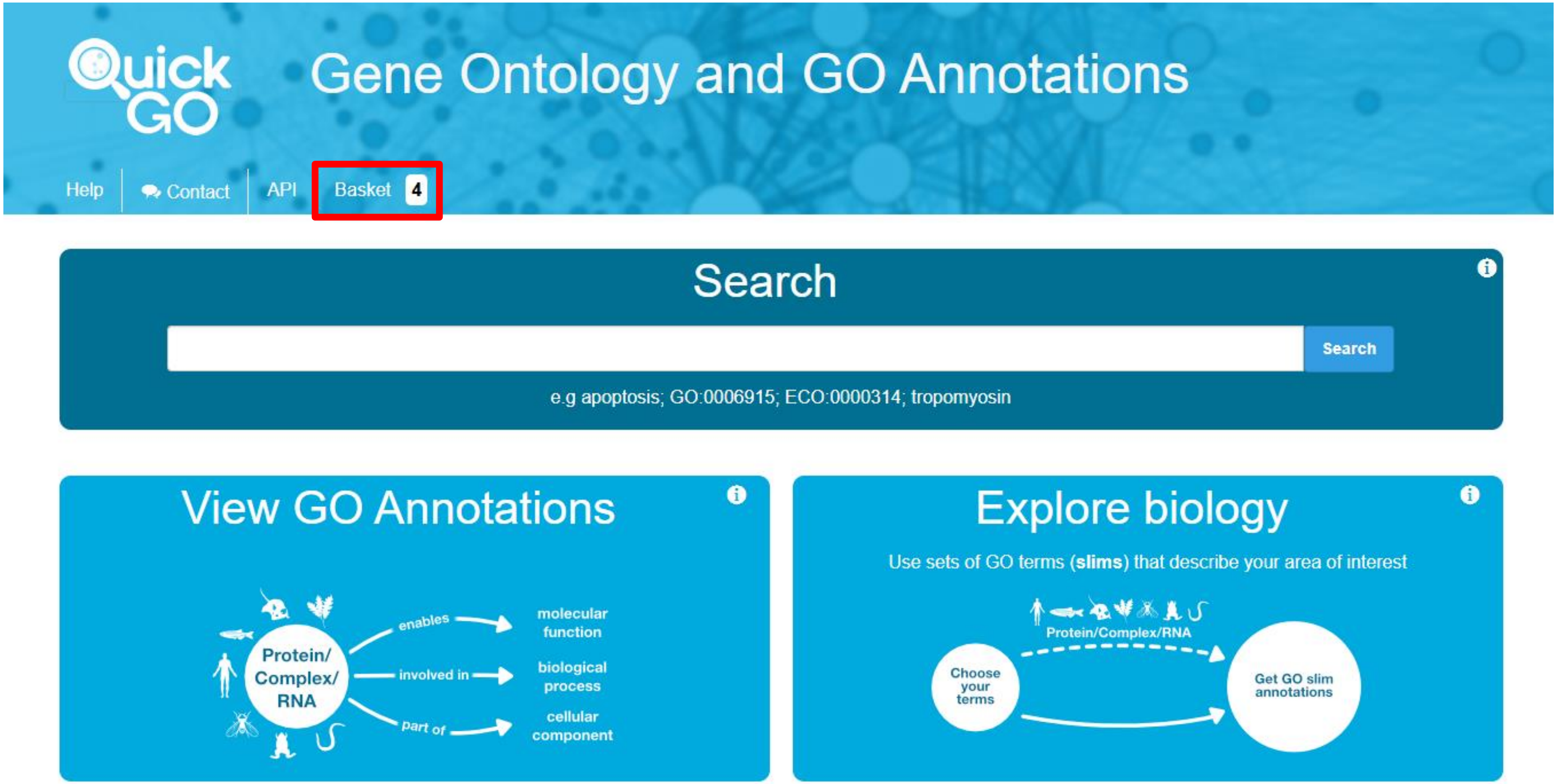
P-value после поправки
Нужно отсортировать!



	Homo sapiens (REF)	upload_1 (▽ Hierarchy NEW! ?)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
glucose-6-phosphate transport	6	2	.00	> 100	+	3.18E-06	1.50E-03
↳hexose phosphate transport	6	2	.00	> 100	+	3.18E-06	1.45E-03
acetyl-CoA biosynthetic process from pyruvate	7	2	.00	> 100	+	4.46E-06	1.81E-03
↳acetyl-CoA biosynthetic process	14	2	.01	> 100	+	1.93E-05	6.16E-03
↳acetyl-CoA metabolic process	29	2	.01	> 100	+	8.57E-05	2.01E-02
↳purine-containing compound metabolic process	502	6	.24	24.60	+	3.95E-08	3.71E-05
↳metabolic process	7632	10	3.71	2.70	+	4.90E-05	1.25E-02
↳nucleoside phosphate metabolic process	579	5	.28	17.77	+	3.88E-06	1.72E-03
↳nucleobase-containing small molecule metabolic process	606	5	.29	16.98	+	4.85E-06	1.92E-03
↳small molecule metabolic process	1590	9	.77	11.65	+	8.94E-10	1.34E-06
↳primary metabolic process	6772	10	3.29	3.04	+	1.48E-05	5.06E-03
↳phosphate-containing compound metabolic process	1466	10	.71	14.04	+	3.27E-12	9.83E-09
↳phosphorus metabolic process	1469	10	.71	14.01	+	3.34E-12	8.36E-09
↳organophosphate metabolic process	972	9	.47	19.06	+	1.08E-11	2.32E-08
↳acyl-CoA biosynthetic process	43	2	.02	95.72	+	1.90E-04	3.80E-02
↳thioester biosynthetic process	43	2	.02	95.72	+	1.90E-04	3.75E-02
↳sulfur compound biosynthetic process	167	3	.08	36.97	+	6.04E-05	1.49E-02
↳organophosphate biosynthetic process	571	4	.28	14.42	+	1.08E-04	2.35E-02

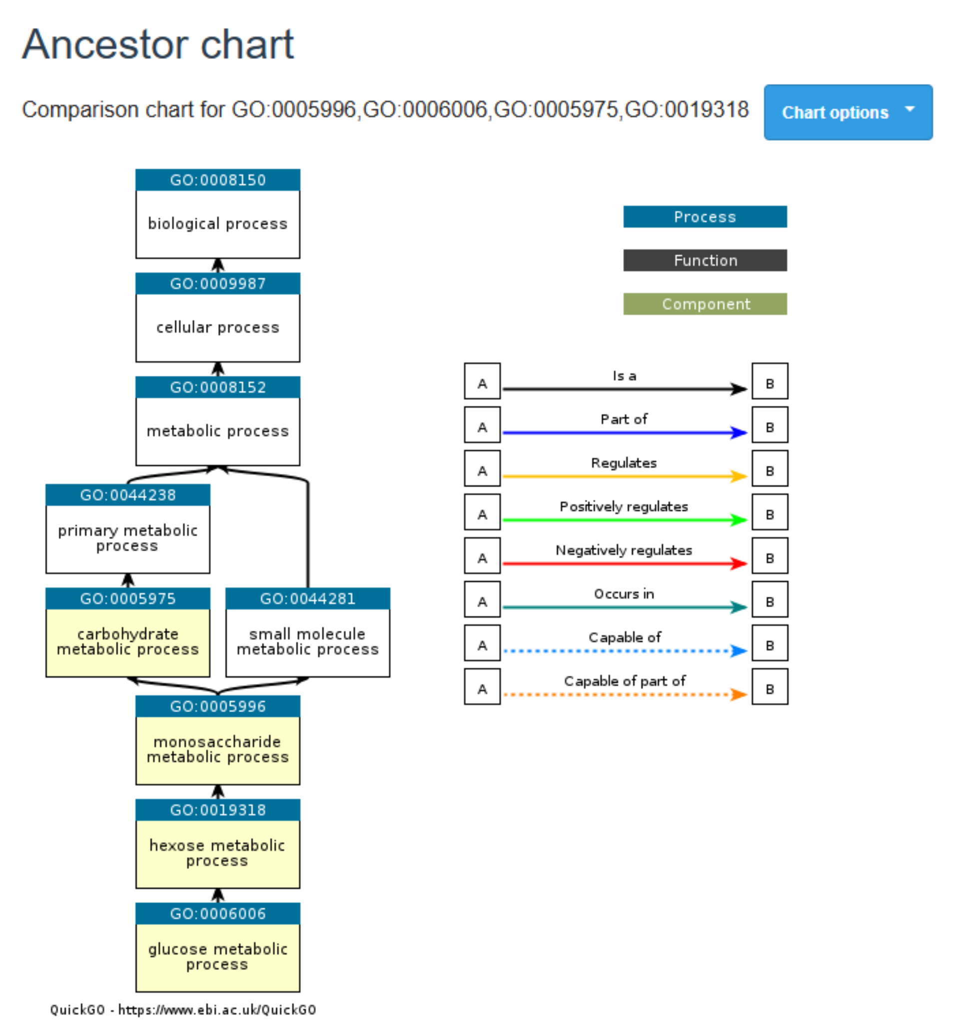
https://www.ebi.ac.uk/QuickGO/

ВИЗУАЛИЗИРОВАТЬ ОТНОШЕНИЯ МЕЖДУ ТЕРМИНАМИ МОЖНО В QUICK GO



https://www.ebi.ac.uk/QuickGO/

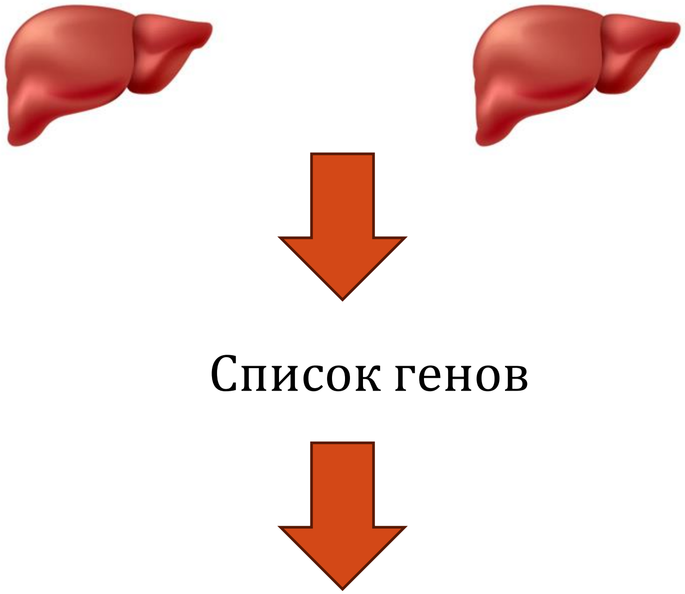
ВИЗУАЛИЗИРОВАТЬ ОТНОШЕНИЯ МЕЖДУ ТЕРМИНАМИ МОЖНО В QUICK GO



Ancestor chart

Comparison chart for GO:0005996,GO:0006006,GO:0005975,GO:0019318

СМОТРИМ ДИФФЕРЕНЦИАЛЬНУЮ ЭКСПРЕССИЮ В ОБРАЗЦАХ СПЕЦИФИЧЕСКОЙ ТКАНИ



Sample 1

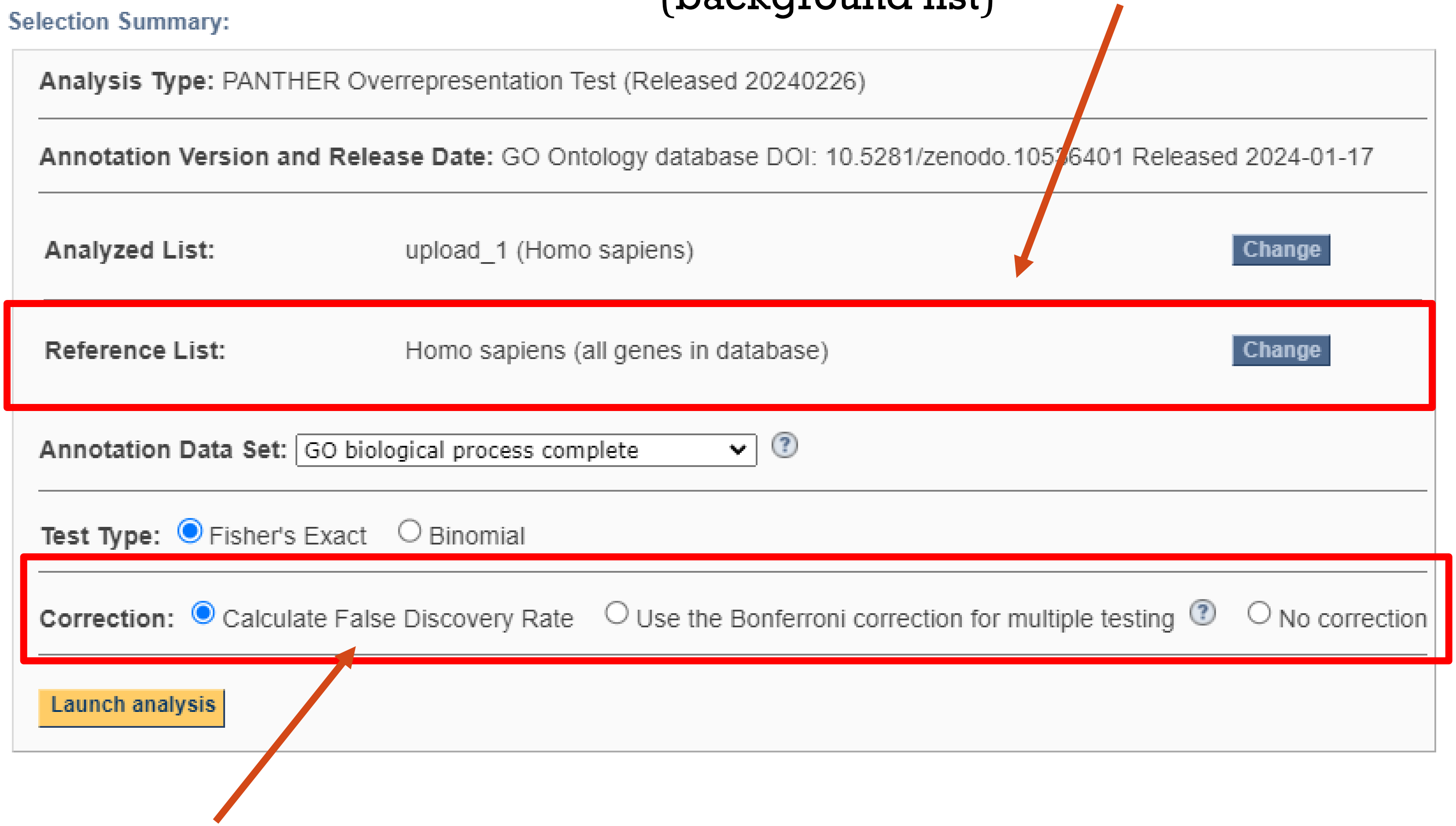
Sample 2

Список генов

GO Enrichment

- Сравниваем частоты терминов GO с частотами по всем генам
- Какие термины будут обогащены?

Можно указать свой список генов, с которыми будут сравниваться частоты терминов
(**background list**)



Selection Summary:

Analysis Type: PANTHER Overrepresentation Test (Released 20240226)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.10536401 Released 2024-01-17

Analyzed List: upload_1 (Homo sapiens)

Reference List: Homo sapiens (all genes in database)

Annotation Data Set: GO biological process complete

Test Type: ● Fisher's Exact ○ Binomial

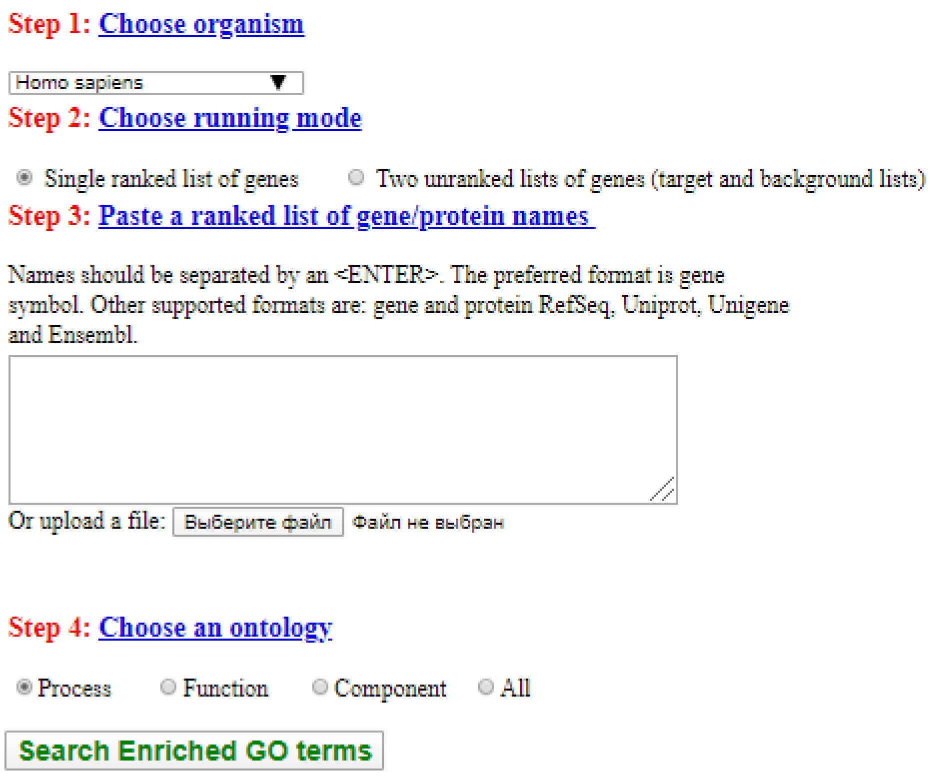
Correction: ● Calculate False Discovery Rate ○ Use the Bonferroni correction for multiple testing ○ No correction

Для чего это нужно?

GORILLA

***G**ene **O**ntology en**RI**chment ana**L**ysis and visua**L**iz**A**tion tool*

GORILLA



Step 1: Choose organism

Homo sapiens ▼

Step 2: Choose running mode

● Single ranked list of genes ○ Two unranked lists of genes (target and background lists)

Step 3: Paste a ranked list of gene/protein names

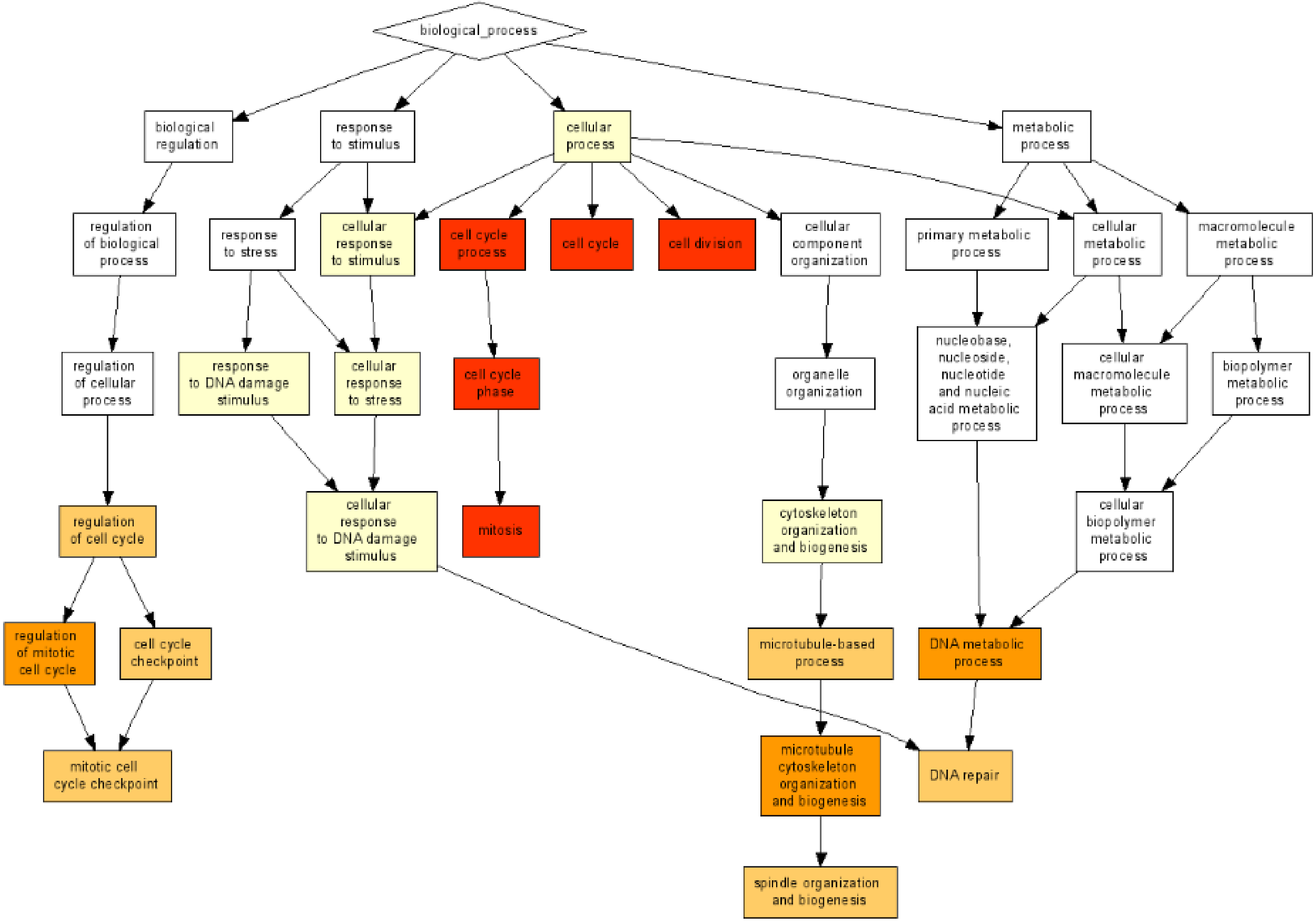
Names should be separated by an <ENTER>. The preferred format is gene symbol. Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and Ensembl.

Or upload a file: Выберите файл Файл не выбран

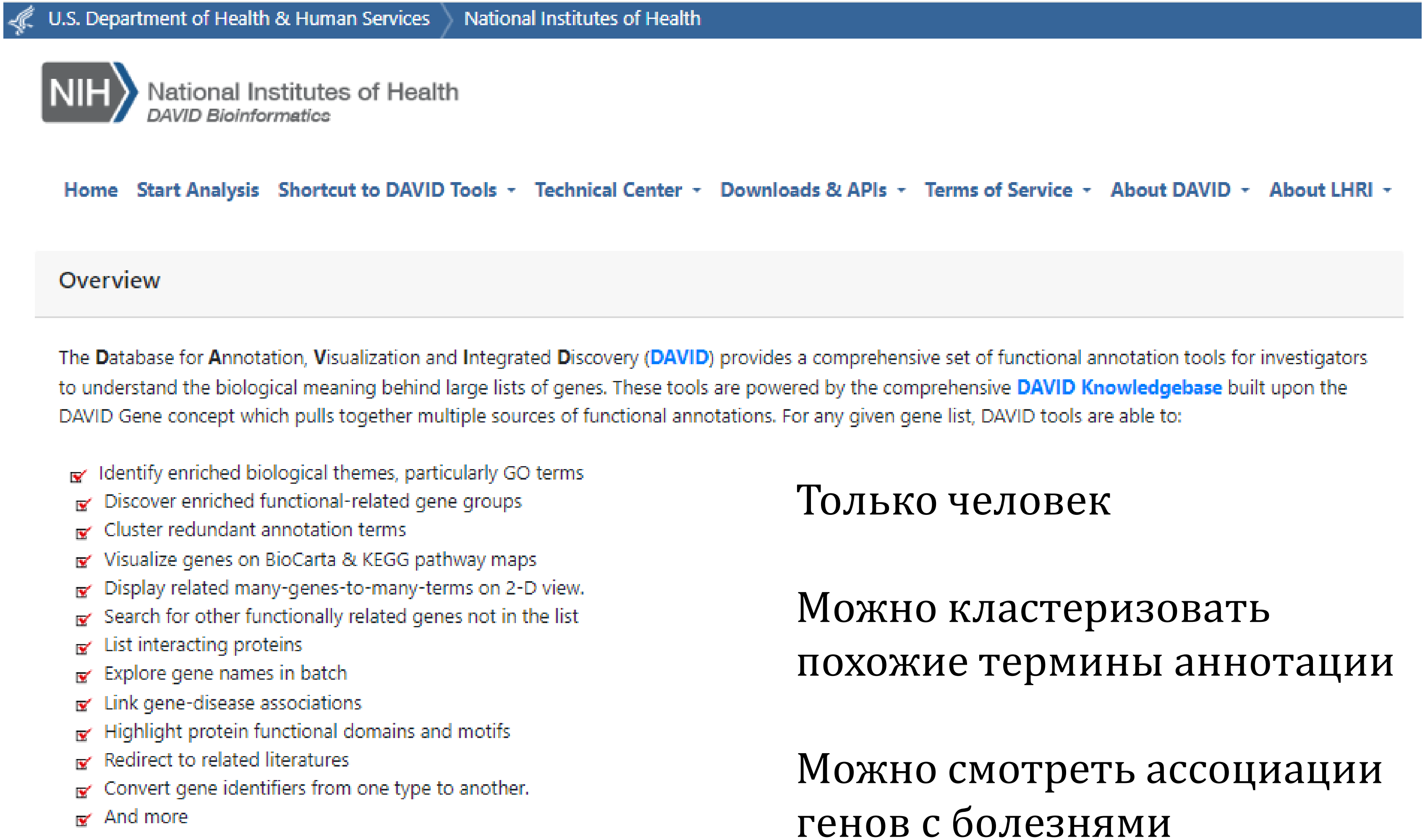
Step 4: Choose an ontology

● Process ○ Function ○ Component ○ All

GORILLA


biological_process
biological regulation
response to stimulus
cellular process
metabolic process
regulation of biological process
response to stress
cellular response to stimulus
cell cycle process
cell cycle
cell division
cellular component organization
primary metabolic process
cellular metabolic process
macromolecule metabolic process
regulation of cellular process
response to DNA damage stimulus
cellular response to stress
cell cycle phase
organelle organization
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
cellular macromolecule metabolic process
biopolymer metabolic process
regulation of cell cycle
cellular response to DNA damage stimulus
mitosis
cytoskeleton organization and biogenesis
cellular biopolymer metabolic process
regulation of mitotic cell cycle
cell cycle checkpoint
microtubule-based process
DNA metabolic process
mitotic cell cycle checkpoint
microtubule cytoskeleton organization and biogenesis
DNA repair
spindle organization and biogenesis

https://davidbioinformatics.nih.gov/



Overview

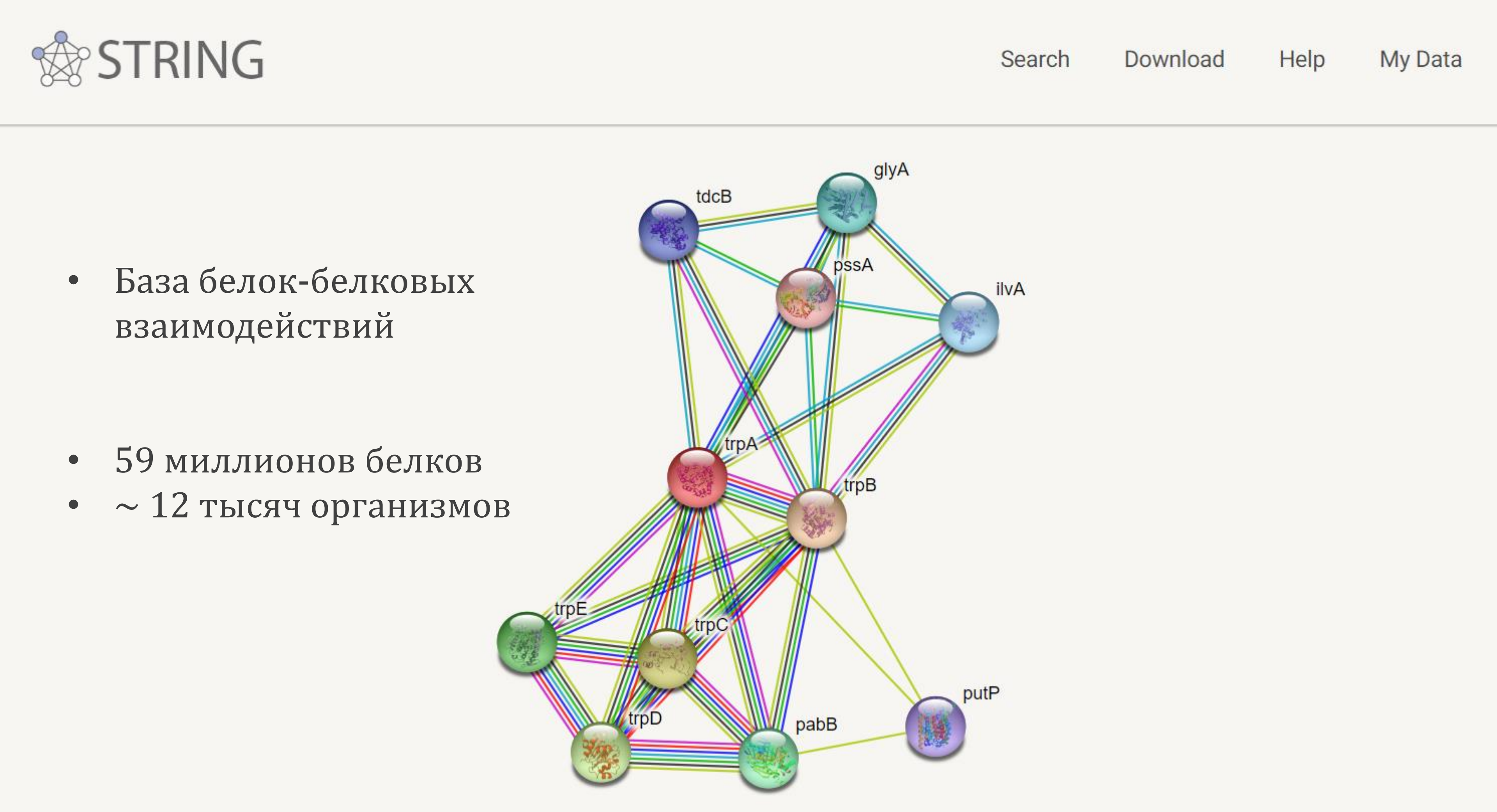
The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind large lists of genes. These tools are powered by the comprehensive **DAVID Knowledgebase** built upon the DAVID Gene concept which pulls together multiple sources of functional annotations. For any given gene list, DAVID tools are able to:

- ☑ Identify enriched biological themes, particularly GO terms
- ☑ Discover enriched functional-related gene groups
- ☑ Cluster redundant annotation terms
- ☑ Visualize genes on BioCarta & KEGG pathway maps
- ☑ Display related many-genes-to-many-terms on 2-D view.
- ☑ Search for other functionally related genes not in the list
- ☑ List interacting proteins
- ☑ Explore gene names in batch
- ☑ Link gene-disease associations
- ☑ Highlight protein functional domains and motifs
- ☑ Redirect to related literatures
- ☑ Convert gene identifiers from one type to another.
- ☑ And more

Только человек

Можно кластеризовать похожие термины аннотации

Можно смотреть ассоциации генов с болезнями

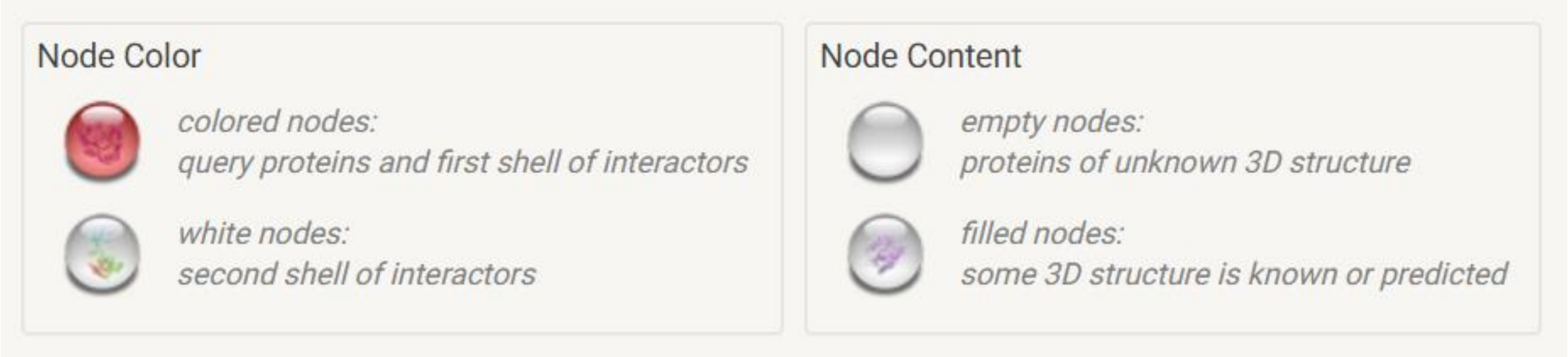


- База белок-белковых взаимодействий

- 59 миллионов белков
- ~ 12 тысяч организмов

https://string-db.org/

STRING



Node Color

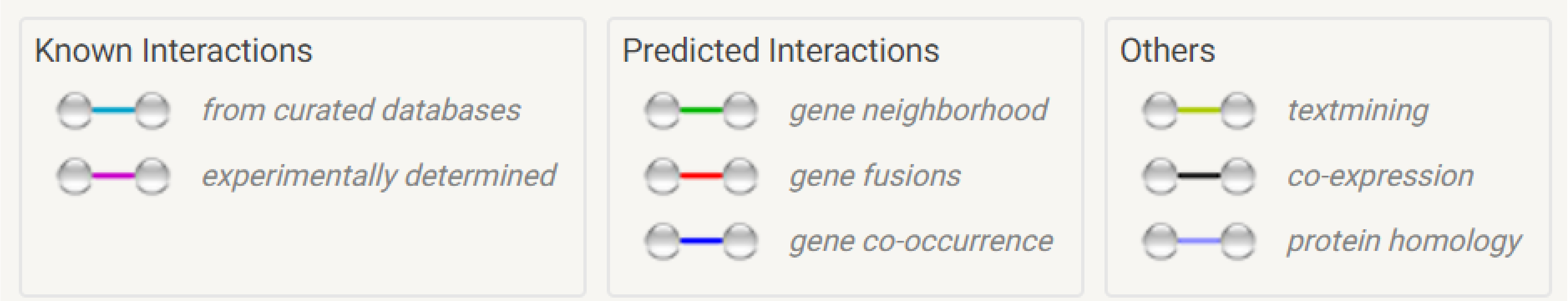
colored nodes:
query proteins and first shell of interactors

white nodes:
second shell of interactors

Node Content

empty nodes:
proteins of unknown 3D structure

filled nodes:
some 3D structure is known or predicted



Known Interactions

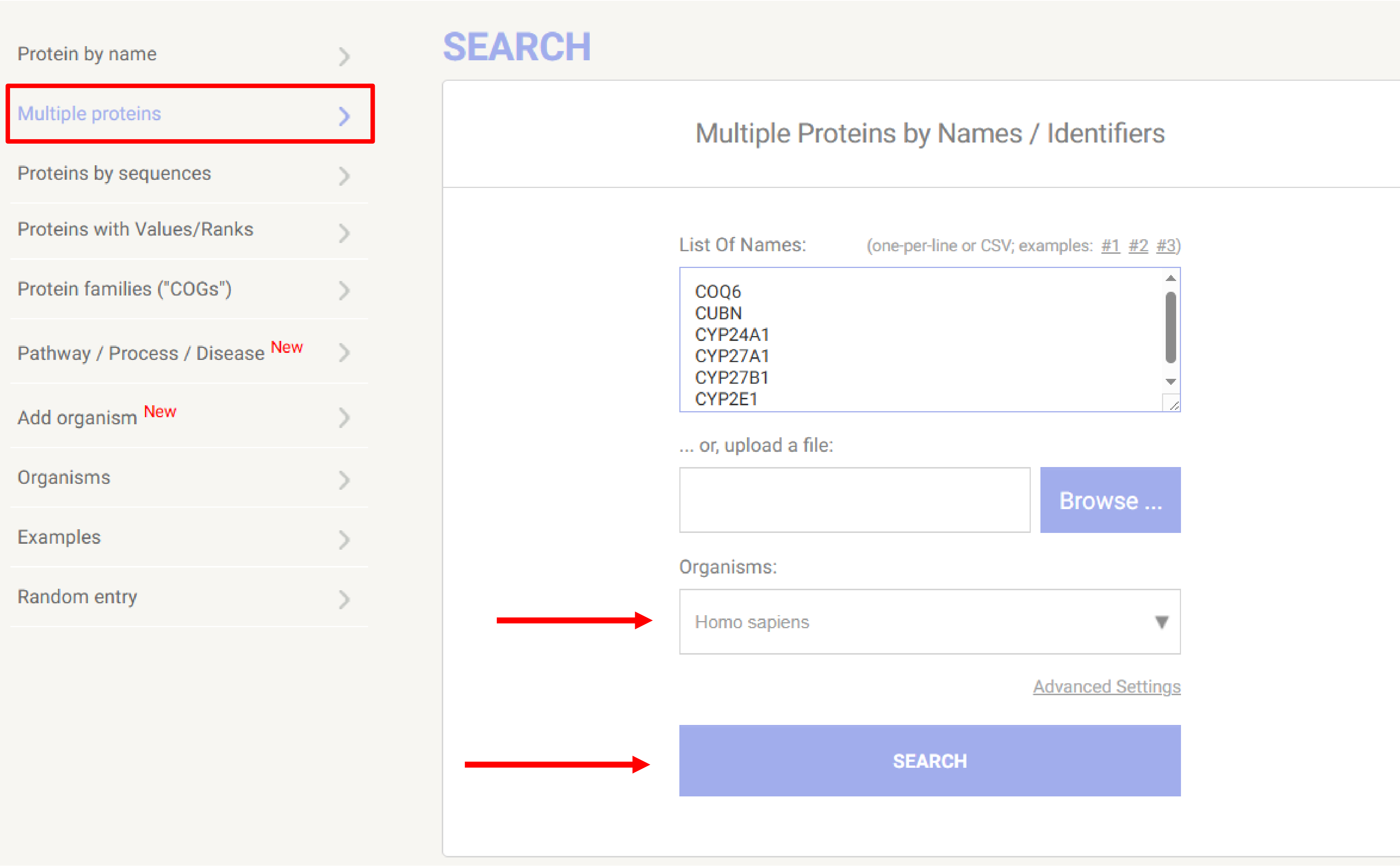
- *from curated databases*
- *experimentally determined*

Predicted Interactions

- *gene neighborhood*
- *gene fusions*
- *gene co-occurrence*

Others

- *textmining*
- *co-expression*
- *protein homology*



SEARCH

Multiple Proteins by Names / Identifiers

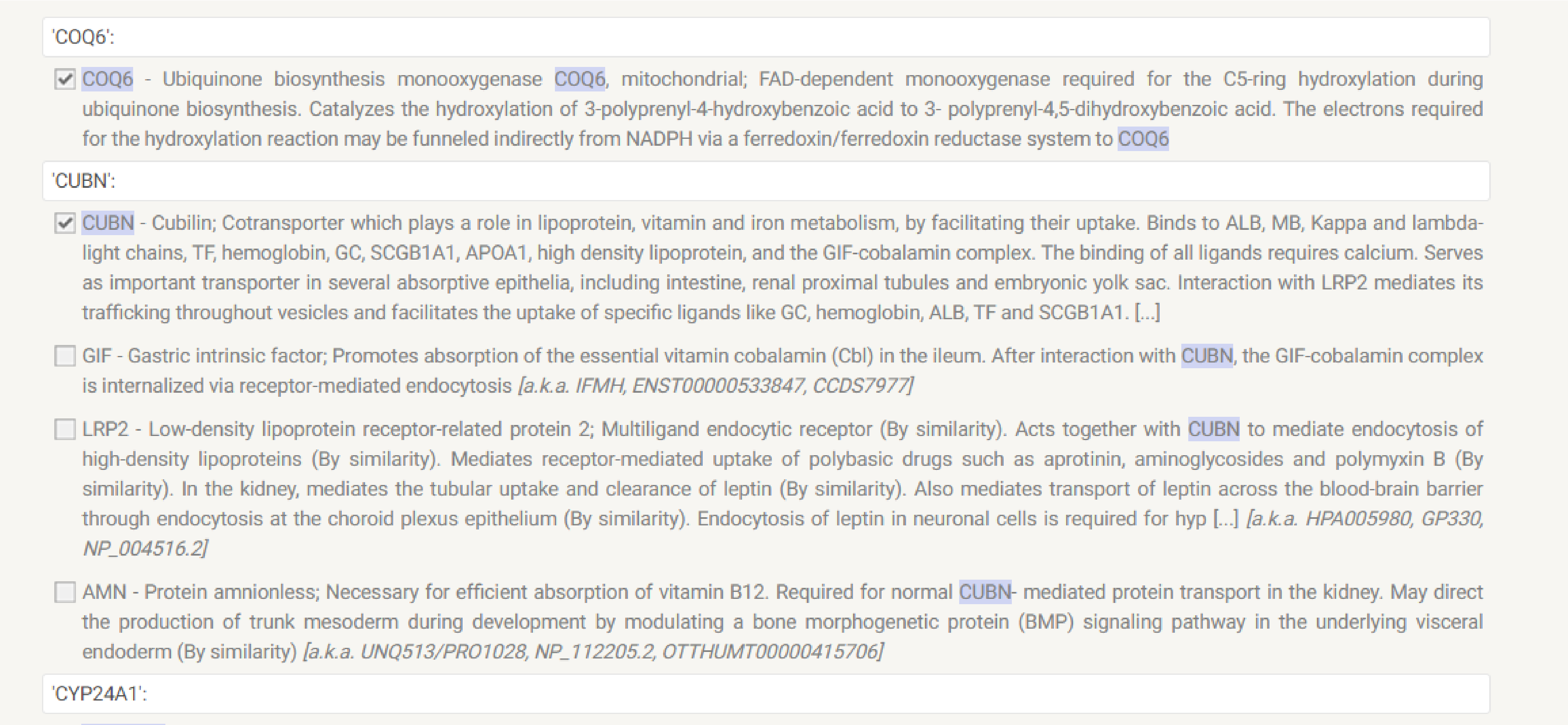
List Of Names: (one-per-line or CSV; examples: #1 #2 #3)

COQ6
CUBN
CYP24A1
CYP27A1
CYP27B1
CYP2E1

... or, upload a file:

Organisms: Homo sapiens

STRING



'COQ6':

☑ COQ6 - Ubiquinone biosynthesis monooxygenase COQ6, mitochondrial; FAD-dependent monooxygenase required for the C5-ring hydroxylation during ubiquinone biosynthesis. Catalyzes the hydroxylation of 3-polyprenyl-4-hydroxybenzoic acid to 3- polyprenyl-4,5-dihydroxybenzoic acid. The electrons required for the hydroxylation reaction may be funneled indirectly from NADPH via a ferredoxin/ferredoxin reductase system to COQ6

'CUBN':

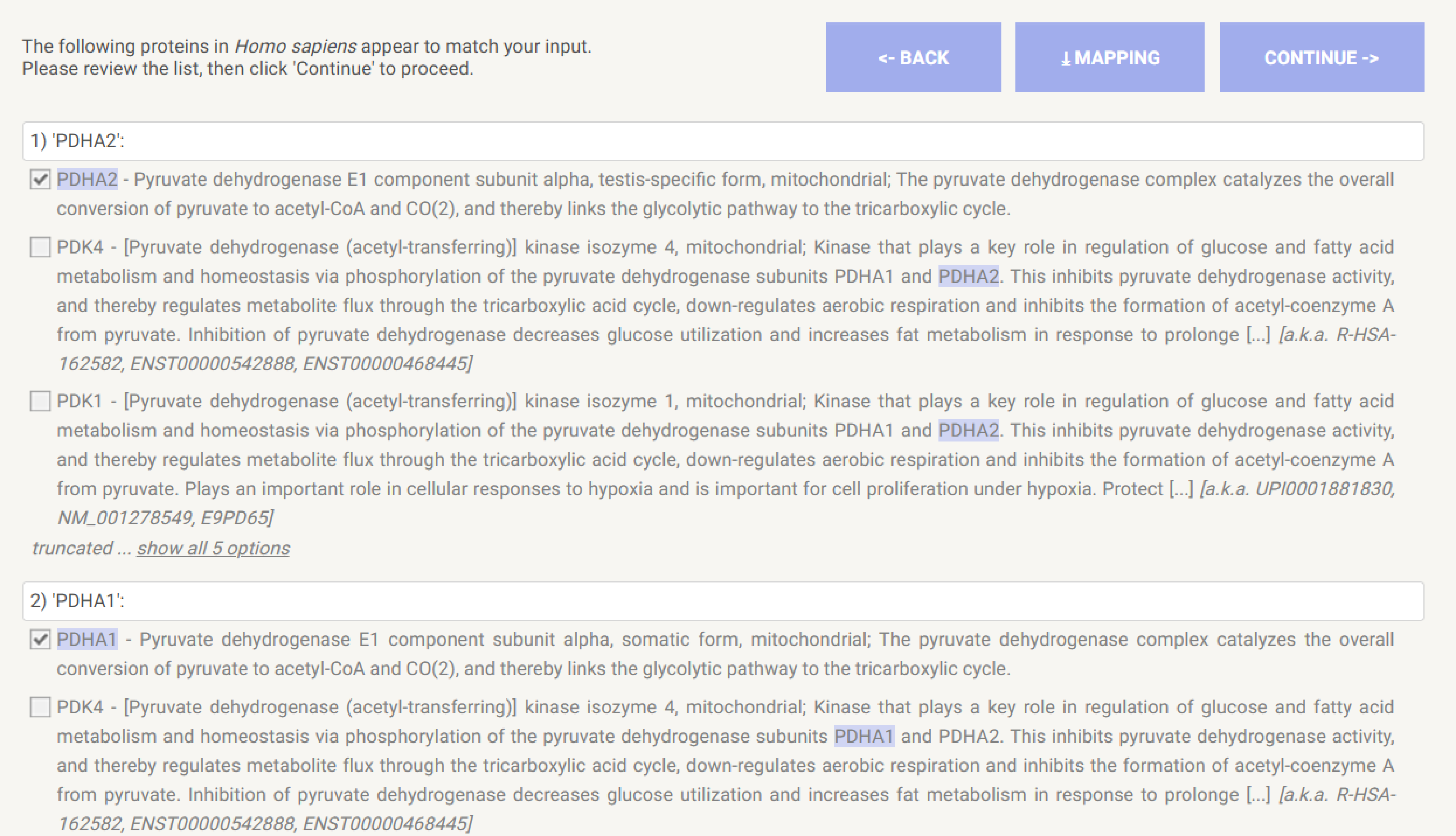
☑ CUBN - Cubilin; Cotransporter which plays a role in lipoprotein, vitamin and iron metabolism, by facilitating their uptake. Binds to ALB, MB, Kappa and lambda-light chains, TF, hemoglobin, GC, SCGB1A1, APOA1, high density lipoprotein, and the GIF-cobalamin complex. The binding of all ligands requires calcium. Serves as important transporter in several absorptive epithelia, including intestine, renal proximal tubules and embryonic yolk sac. Interaction with LRP2 mediates its trafficking throughout vesicles and facilitates the uptake of specific ligands like GC, hemoglobin, ALB, TF and SCGB1A1. [...]

☐ GIF - Gastric intrinsic factor; Promotes absorption of the essential vitamin cobalamin (Cbl) in the ileum. After interaction with CUBN, the GIF-cobalamin complex is internalized via receptor-mediated endocytosis *[a.k.a. IFMH, ENST00000533847, CCDS7977]*

☐ LRP2 - Low-density lipoprotein receptor-related protein 2; Multiligand endocytic receptor (By similarity). Acts together with CUBN to mediate endocytosis of high-density lipoproteins (By similarity). Mediates receptor-mediated uptake of polybasic drugs such as aprotinin, aminoglycosides and polymyxin B (By similarity). In the kidney, mediates the tubular uptake and clearance of leptin (By similarity). Also mediates transport of leptin across the blood-brain barrier through endocytosis at the choroid plexus epithelium (By similarity). Endocytosis of leptin in neuronal cells is required for hyp [...] *[a.k.a. HPA005980, GP330, NP_004516.2]*

☐ AMN - Protein amnionless; Necessary for efficient absorption of vitamin B12. Required for normal CUBN- mediated protein transport in the kidney. May direct the production of trunk mesoderm during development by modulating a bone morphogenetic protein (BMP) signaling pathway in the underlying visceral endoderm (By similarity) *[a.k.a. UNQ513/PRO1028, NP_112205.2, OTTHUMT00000415706]*

'CYP24A1':



The following proteins in *Homo sapiens* appear to match your input.
Please review the list, then click 'Continue' to proceed.

1) 'PDHA2':

☑ PDHA2 - Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial; The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and CO(2), and thereby links the glycolytic pathway to the tricarboxylic cycle.

☐ PDK4 - [Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial; Kinase that plays a key role in regulation of glucose and fatty acid metabolism and homeostasis via phosphorylation of the pyruvate dehydrogenase subunits PDHA1 and PDHA2. This inhibits pyruvate dehydrogenase activity, and thereby regulates metabolite flux through the tricarboxylic acid cycle, down-regulates aerobic respiration and inhibits the formation of acetyl-coenzyme A from pyruvate. Inhibition of pyruvate dehydrogenase decreases glucose utilization and increases fat metabolism in response to prolonge [...] *[a.k.a. R-HSA-162582, ENST00000542888, ENST00000468445]*

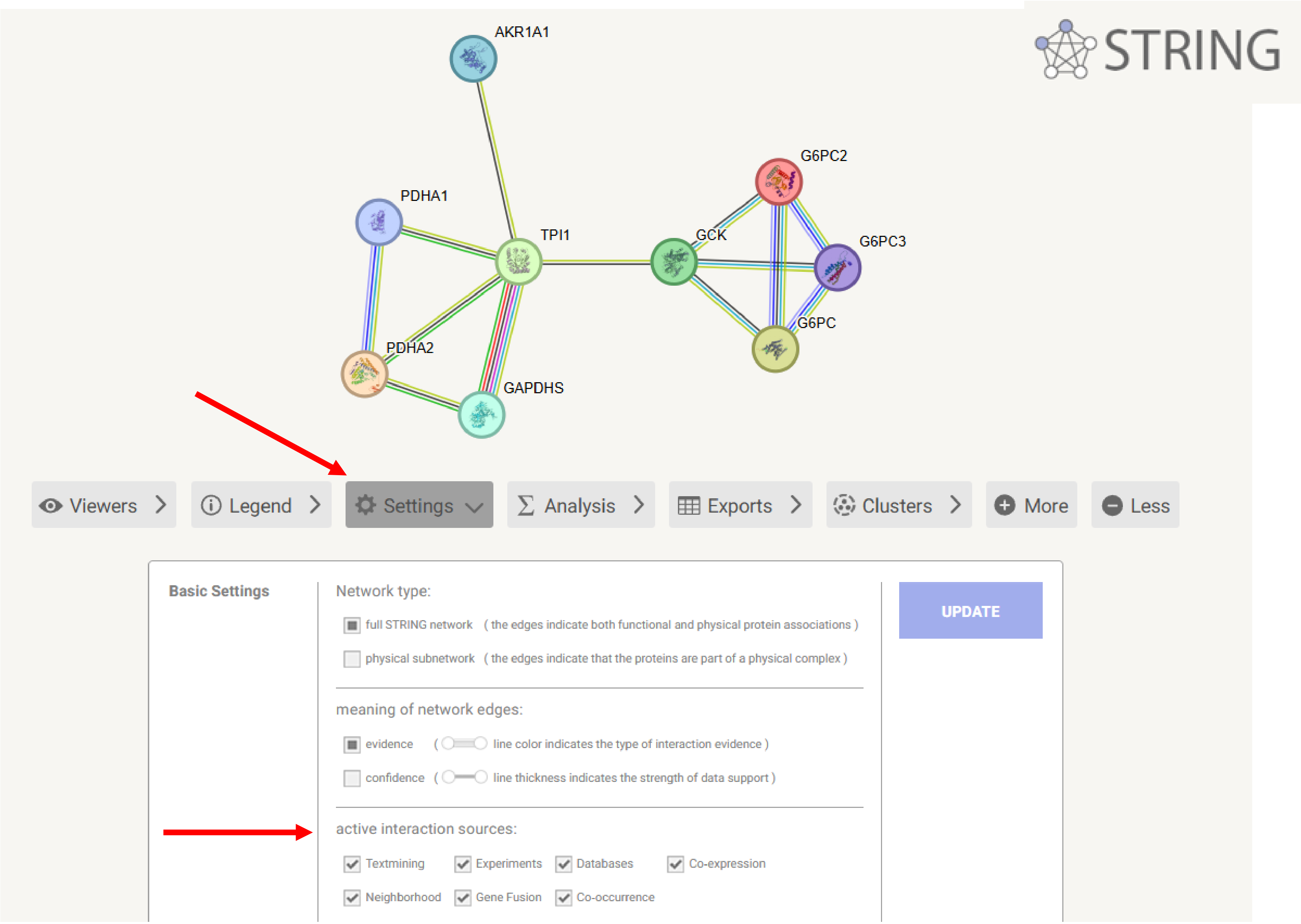
☐ PDK1 - [Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 1, mitochondrial; Kinase that plays a key role in regulation of glucose and fatty acid metabolism and homeostasis via phosphorylation of the pyruvate dehydrogenase subunits PDHA1 and PDHA2. This inhibits pyruvate dehydrogenase activity, and thereby regulates metabolite flux through the tricarboxylic acid cycle, down-regulates aerobic respiration and inhibits the formation of acetyl-coenzyme A from pyruvate. Plays an important role in cellular responses to hypoxia and is important for cell proliferation under hypoxia. Protect [...] *[a.k.a. UPI0001881830, NM_001278549, E9PD65]*

truncated ... show all 5 options

2) 'PDHA1':

☑ PDHA1 - Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial; The pyruvate dehydrogenase complex catalyzes the overall conversion of pyruvate to acetyl-CoA and CO(2), and thereby links the glycolytic pathway to the tricarboxylic cycle.

☐ PDK4 - [Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 4, mitochondrial; Kinase that plays a key role in regulation of glucose and fatty acid metabolism and homeostasis via phosphorylation of the pyruvate dehydrogenase subunits PDHA1 and PDHA2. This inhibits pyruvate dehydrogenase activity, and thereby regulates metabolite flux through the tricarboxylic acid cycle, down-regulates aerobic respiration and inhibits the formation of acetyl-coenzyme A from pyruvate. Inhibition of pyruvate dehydrogenase decreases glucose utilization and increases fat metabolism in response to prolonge [...] *[a.k.a. R-HSA-162582, ENST00000542888, ENST00000468445]*



Basic Settings

Network type:

☒ full STRING network (the edges indicate both functional and physical protein associations)

☐ physical subnetwork (the edges indicate that the proteins are part of a physical complex)

meaning of network edges:

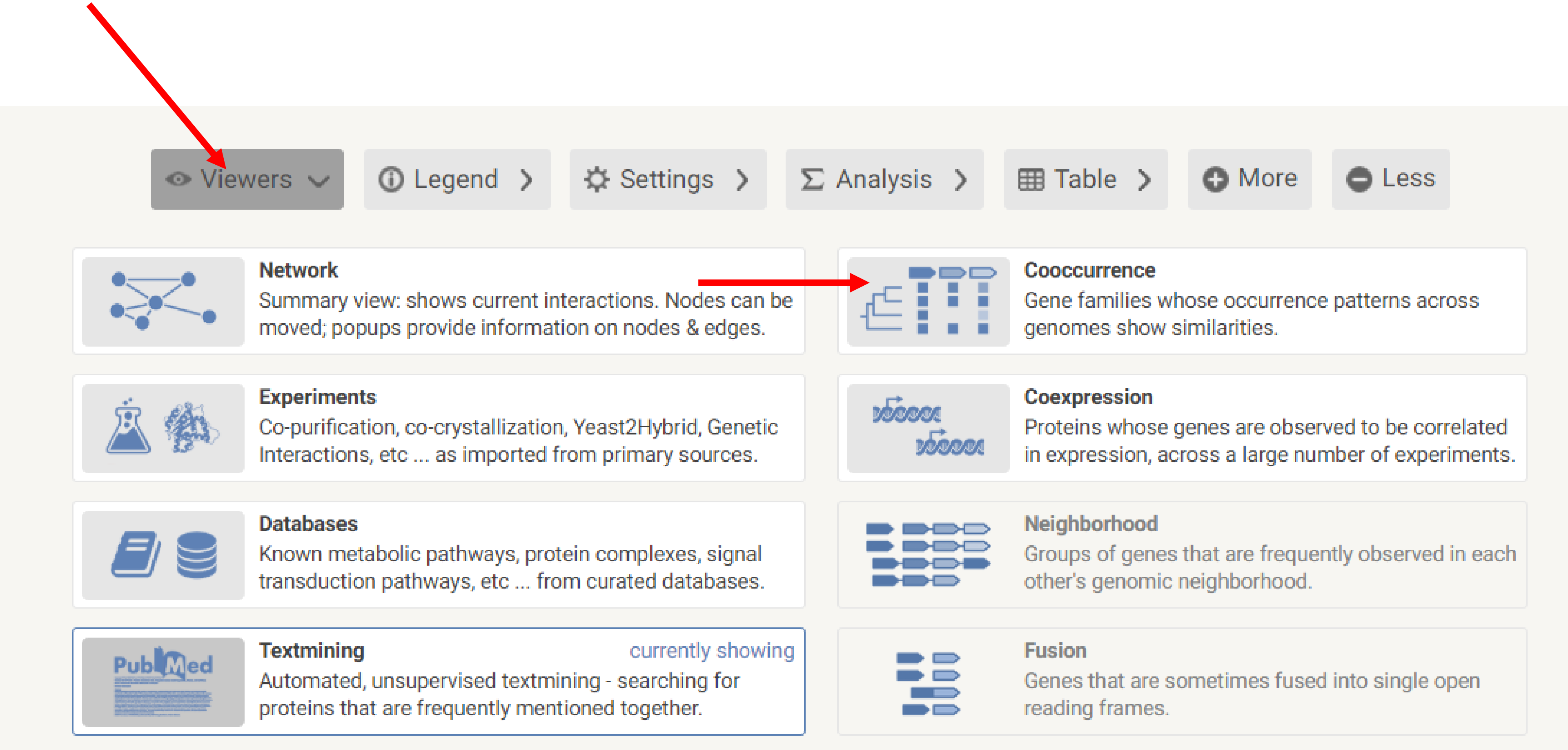
☒ evidence (line color indicates the type of interaction evidence)

☐ confidence (line thickness indicates the strength of data support)

active interaction sources:

☑ Textmining ☑ Experiments ☑ Databases ☑ Co-expression

☑ Neighborhood ☑ Gene Fusion ☑ Co-occurrence



Network
Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.

Cooccurrence
Gene families whose occurrence patterns across genomes show similarities.

Experiments
Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.

Coexpression
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.

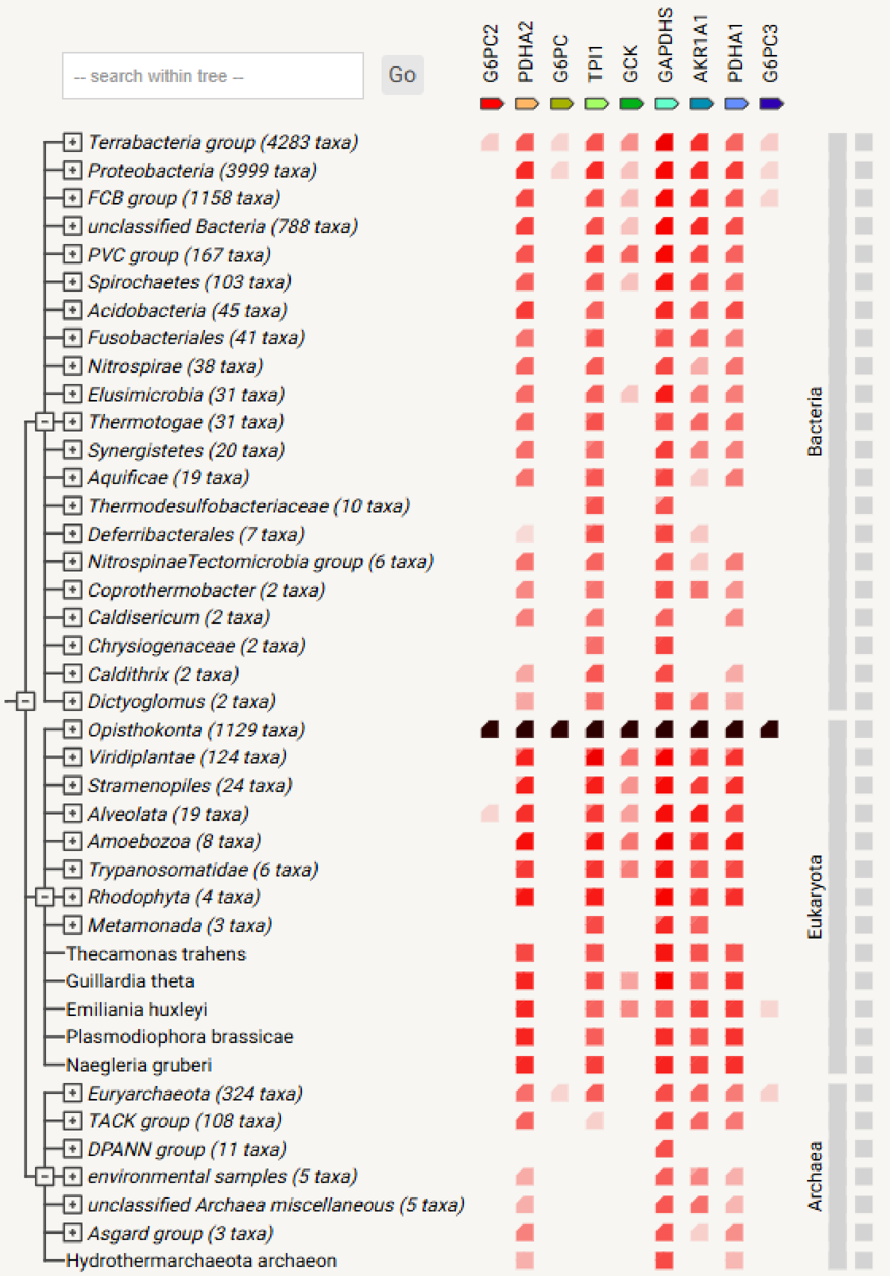
Databases
Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.

Neighborhood
Groups of genes that are frequently observed in each other's genomic neighborhood.

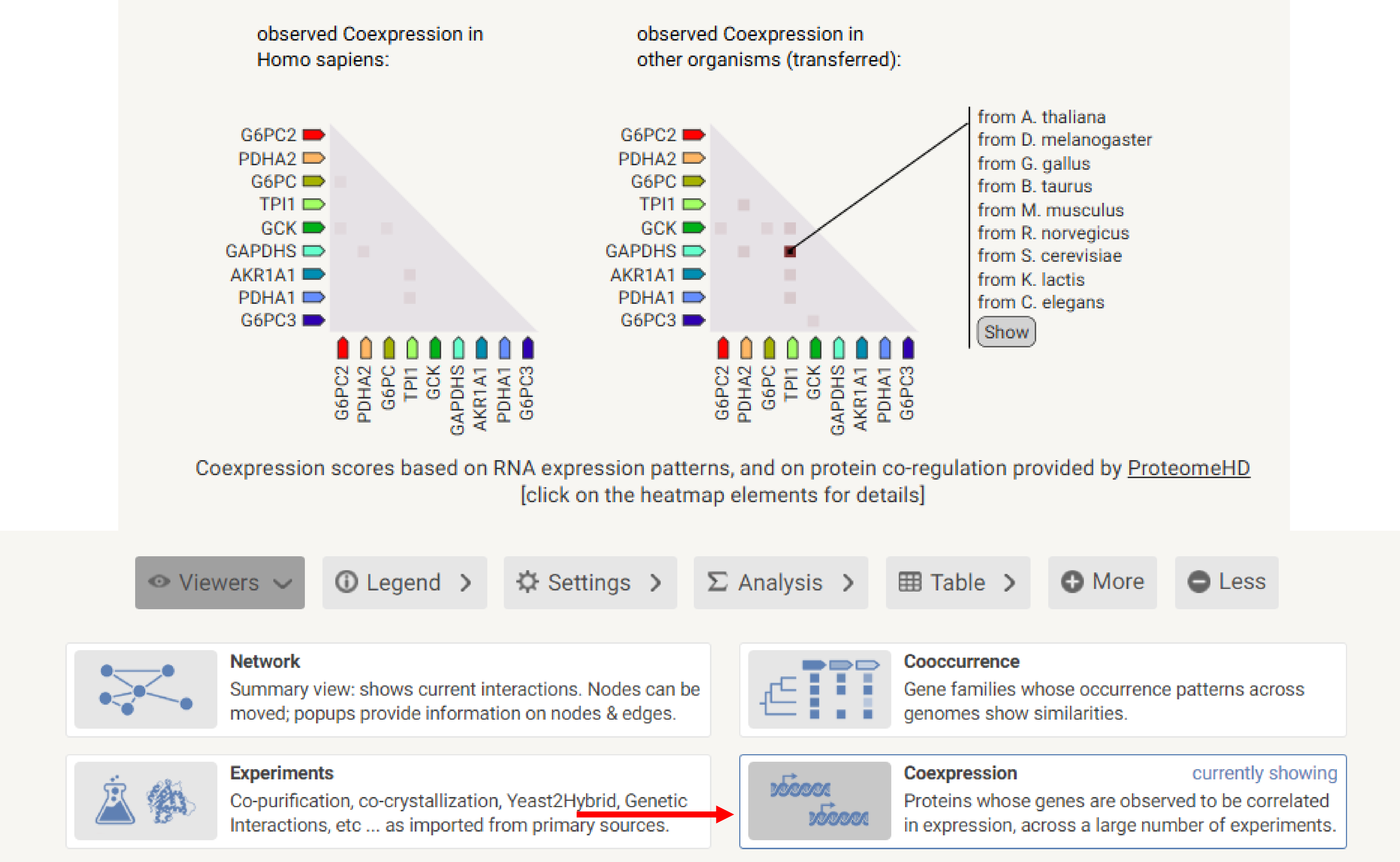
Textmining currently showing
Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.

Fusion
Genes that are sometimes fused into single open reading frames.

GENE COOCCURRENCE



GENE COEXPRESSION



observed Coexpression in Homo sapiens:

observed Coexpression in other organisms (transferred):

Coexpression scores based on RNA expression patterns, and on protein co-regulation provided by ProteomeHD
[click on the heatmap elements for details]

Network
Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.

Cooccurrence
Gene families whose occurrence patterns across genomes show similarities.

Experiments
Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.

Coexpression
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.

STRING

Все белки из нашего списка вошли в эту категорию

Functional enrichments in your network

explain columns

Biological Process (Gene Ontology)

GO-term	description	count in network	strength	signal	false discovery rate
GO:0006006	Glucose metabolic process	8 of 116	2.18	5.84	1.37e-13
GO:0005996	Monosaccharide metabolic process	9 of 179	2.04	5.71	8.46e-15
GO:0046364	Monosaccharide biosynthetic process	5 of 51	2.33	3.9	4.32e-08
GO:0006090	Pyruvate metabolic process	5 of 66	2.22	3.57	1.28e-07
GO:0051156	Glucose 6-phosphate metabolic process	4 of 22	2.6	3.56	4.68e-07

(more ...)

Molecular Function (Gene Ontology)

GO-term	description	count in network	strength	signal	false discovery rate
GO:0004346	glucose-6-phosphatase activity	3 of 3	3.34	3.06	6.49e-06
GO:0034604	Pyruvate dehydrogenase (NAD+) activity	2 of 2	3.34	1.8	0.00091
GO:0004739	Pyruvate dehydrogenase (acetyl-transferring) activity	2 of 3	3.16	1.77	0.0010
GO:0016620	Oxidoreductase activity, acting on the aldehyde or oxo group of don...	3 of 39	2.23	1.75	0.00074
GO:0003824	Catalytic activity	9 of 5522	0.55	0.41	0.0046

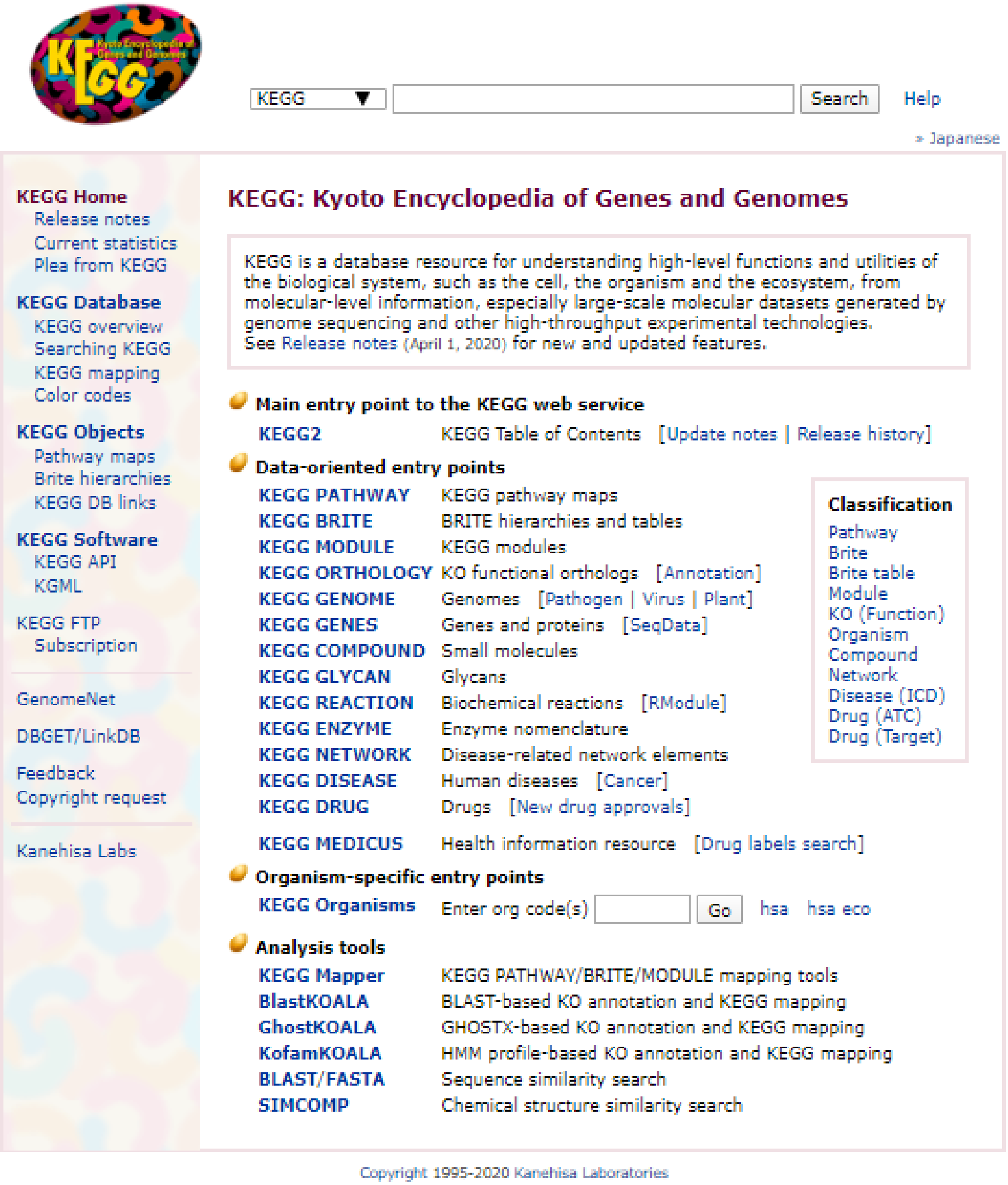
Reference Publications (PubMed)

Встречаемость в публикациях!

publication	(year) title	count in network	strength	signal	false discovery rate
PMID:24052902	(2013) Mutually exclusive acetylation and ubiquitylation among en...	5 of 13	2.93	4.04	1.05e-07
PMID:25859558	(2015) Analysis and interpretation of transcriptomic data obtained ...	7 of 101	2.18	3.85	2.19e-08
PMID:26283222	(2015) Anopheles stephensi p38 MAPK signaling regulates innate i...	5 of 18	2.78	3.82	2.15e-07
PMID:28202506	(2017) Recurrent patterns of DNA copy number alterations in tumo...	6 of 52	2.4	3.78	1.05e-07
PMID:31166604	(2019) Fructose co-ingestion to increase carbohydrate availability i...	5 of 20	2.74	3.75	2.71e-07

(more ...)

Метаболические пути



KEGG Home
- Release notes
- Current statistics
- Plea from KEGG

KEGG Database
- KEGG overview
- Searching KEGG
- KEGG mapping
- Color codes

KEGG Objects
- Pathway maps
- Brite hierarchies
- KEGG DB links

KEGG Software
- KEGG API
- KGML

KEGG FTP
- Subscription

GenomeNet

DBGET/LinkDB

Feedback
Copyright request

Kanehisa Labs

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.
See Release notes (April 1, 2020) for new and updated features.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents [Update notes | Release history]

Data-oriented entry points

KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
KEGG ORTHOLOGY	KO functional orthologs [Annotation]
KEGG GENOME	Genomes [Pathogen | Virus | Plant]
KEGG GENES	Genes and proteins [SeqData]
KEGG COMPOUND	Small molecules
KEGG GLYCAN	Glycans
KEGG REACTION	Biochemical reactions [RModule]
KEGG ENZYME	Enzyme nomenclature
KEGG NETWORK	Disease-related network elements
KEGG DISEASE	Human diseases [Cancer]
KEGG DRUG	Drugs [New drug approvals]
KEGG MEDICUS	Health information resource [Drug labels search]

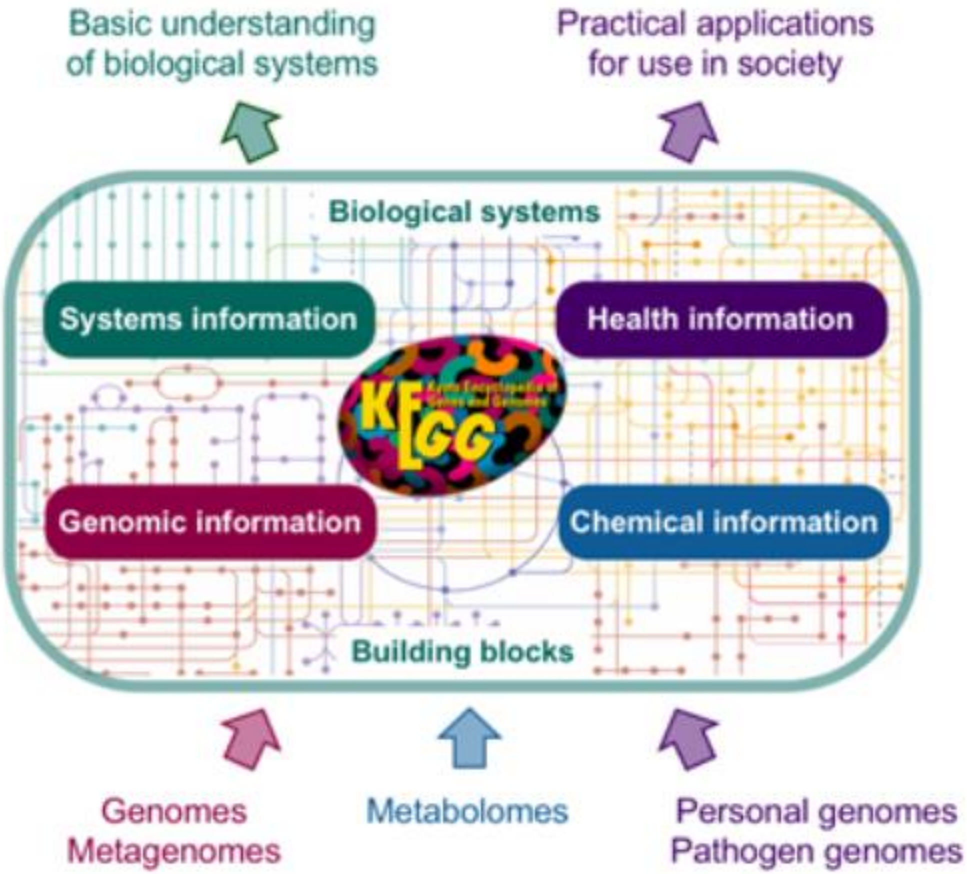
Classification
- Pathway
- Brite
- Brite table
- Module
- KO (Function)
- Organism
- Compound
- Network
- Disease (ICD)
- Drug (ATC)
- Drug (Target)

Organism-specific entry points

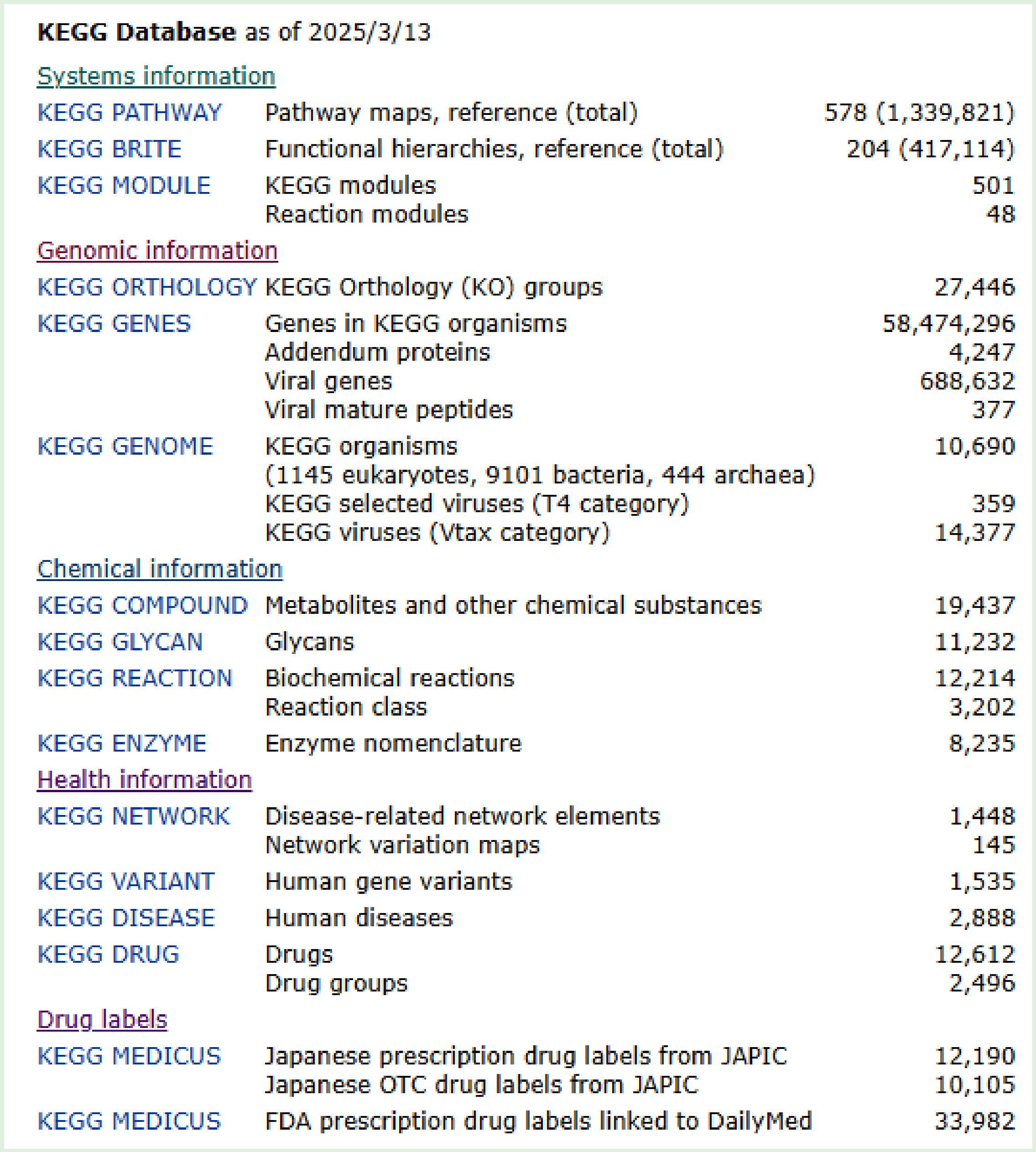
KEGG Organisms Enter org code(s) [ ] Go hsa hsa eco

Analysis tools

KEGG Mapper	KEGG PATHWAY/BRITE/MODULE mapping tools
BlastKOALA	BLAST-based KO annotation and KEGG mapping
GhostKOALA	GHOSTX-based KO annotation and KEGG mapping
KofamKOALA	HMM profile-based KO annotation and KEGG mapping
BLAST/FASTA	Sequence similarity search
SIMCOMP	Chemical structure similarity search

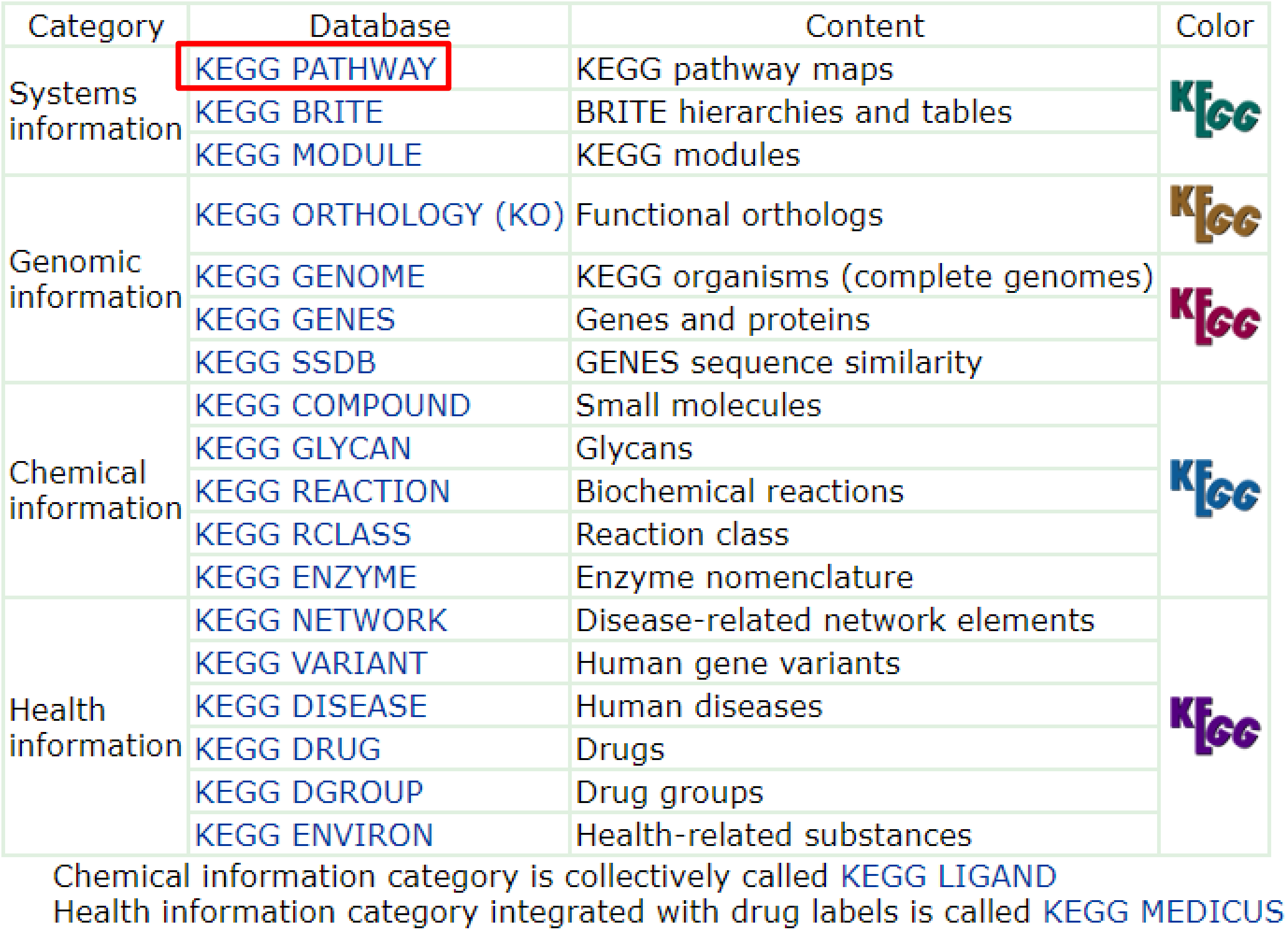

Basic understanding of biological systems
Practical applications for use in society
Biological systems
Systems information
Health information
KEGG
Genomic information
Chemical information
Building blocks
Genomes
Metagenomes
Metabolomes
Personal genomes
Pathogen genomes

Current Statistics



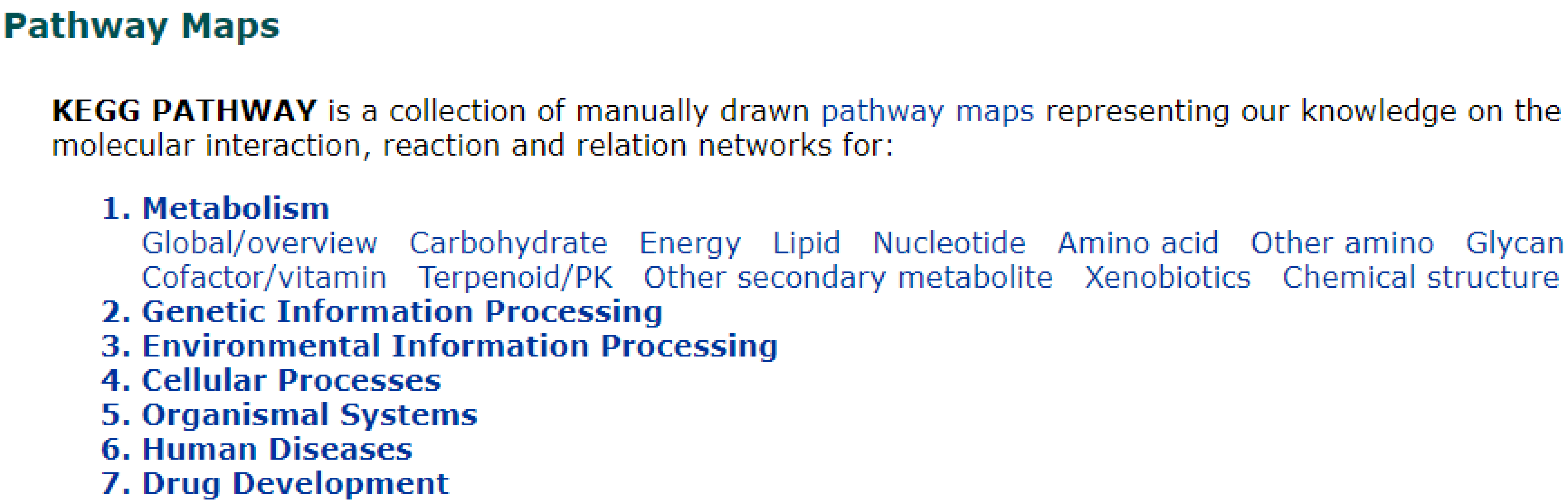
KEGG Database as of 2025/3/13

Systems information		
KEGG PATHWAY	Pathway maps, reference (total)	578 (1,339,821)
KEGG BRITE	Functional hierarchies, reference (total)	204 (417,114)
KEGG MODULE	KEGG modules	501
	Reaction modules	48
Genomic information		
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	27,446
KEGG GENES	Genes in KEGG organisms	58,474,296
	Addendum proteins	4,247
	Viral genes	688,632
	Viral mature peptides	377
KEGG GENOME	KEGG organisms (1145 eukaryotes, 9101 bacteria, 444 archaea)	10,690
	KEGG selected viruses (T4 category)	359
	KEGG viruses (Vtax category)	14,377
Chemical information		
KEGG COMPOUND	Metabolites and other chemical substances	19,437
KEGG GLYCAN	Glycans	11,232
KEGG REACTION	Biochemical reactions	12,214
	Reaction class	3,202
KEGG ENZYME	Enzyme nomenclature	8,235
Health information		
KEGG NETWORK	Disease-related network elements	1,448
	Network variation maps	145
KEGG VARIANT	Human gene variants	1,535
KEGG DISEASE	Human diseases	2,888
KEGG DRUG	Drugs	12,612
	Drug groups	2,496
Drug labels		
KEGG MEDICUS	Japanese prescription drug labels from JAPIC	12,190
	Japanese OTC drug labels from JAPIC	10,105
KEGG MEDICUS	FDA prescription drug labels linked to DailyMed	33,982



Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	KEGG
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	KEGG
	KEGG GENOME	KEGG organisms (complete genomes)	KEGG
	KEGG GENES	Genes and proteins	
	KEGG SSDB	GENES sequence similarity	
Chemical information	KEGG COMPOUND	Small molecules	KEGG
	KEGG GLYCAN	Glycans	
	KEGG REACTION	Biochemical reactions	
	KEGG RCLASS	Reaction class	
	KEGG ENZYME	Enzyme nomenclature	
Health information	KEGG NETWORK	Disease-related network elements	KEGG
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	
	KEGG DRUG	Drugs	
	KEGG DGROUP	Drug groups	
	KEGG ENVIRON	Health-related substances	

Chemical information category is collectively called KEGG LIGAND
Health information category integrated with drug labels is called KEGG MEDICUS

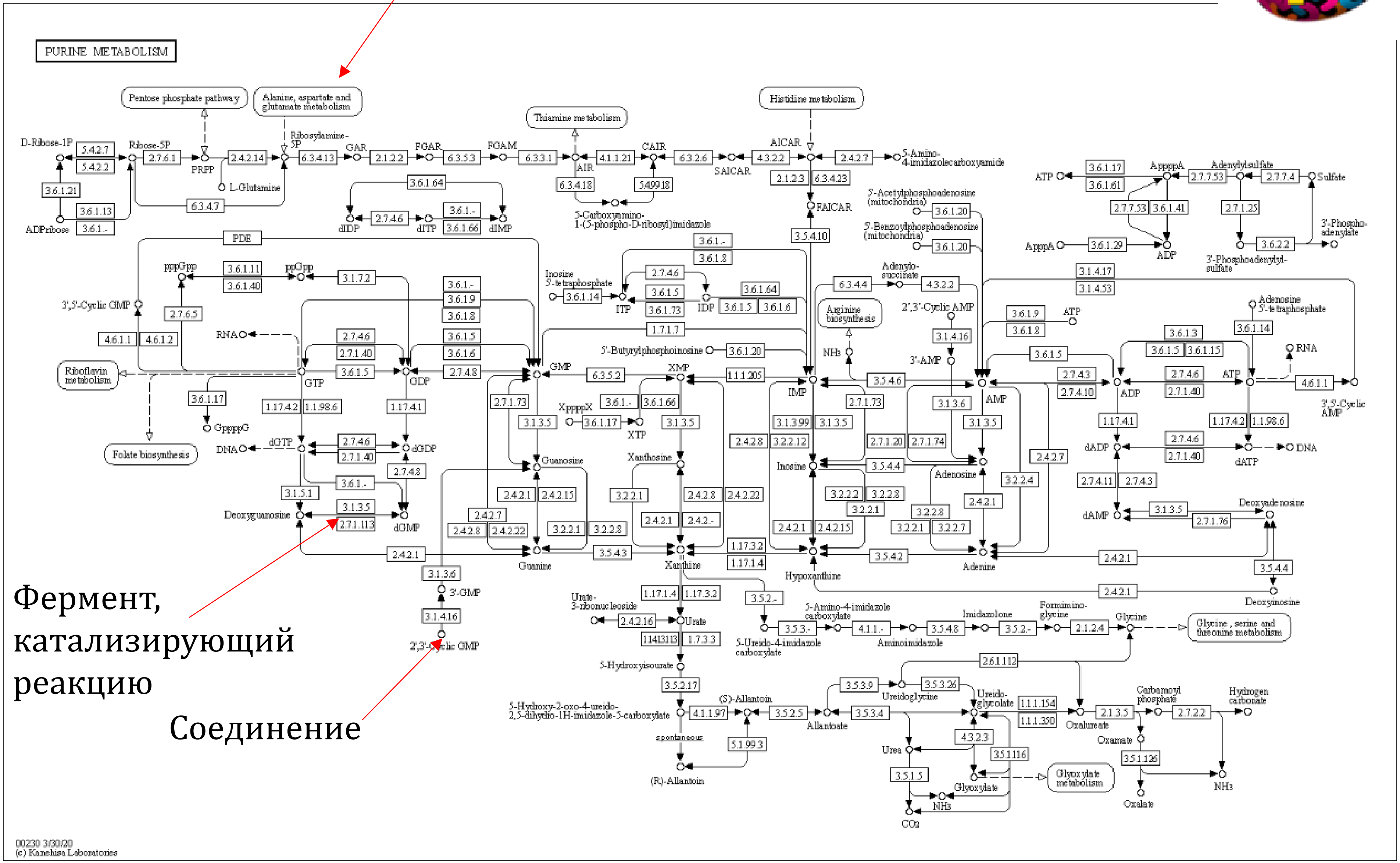


Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction, reaction and relation networks for:

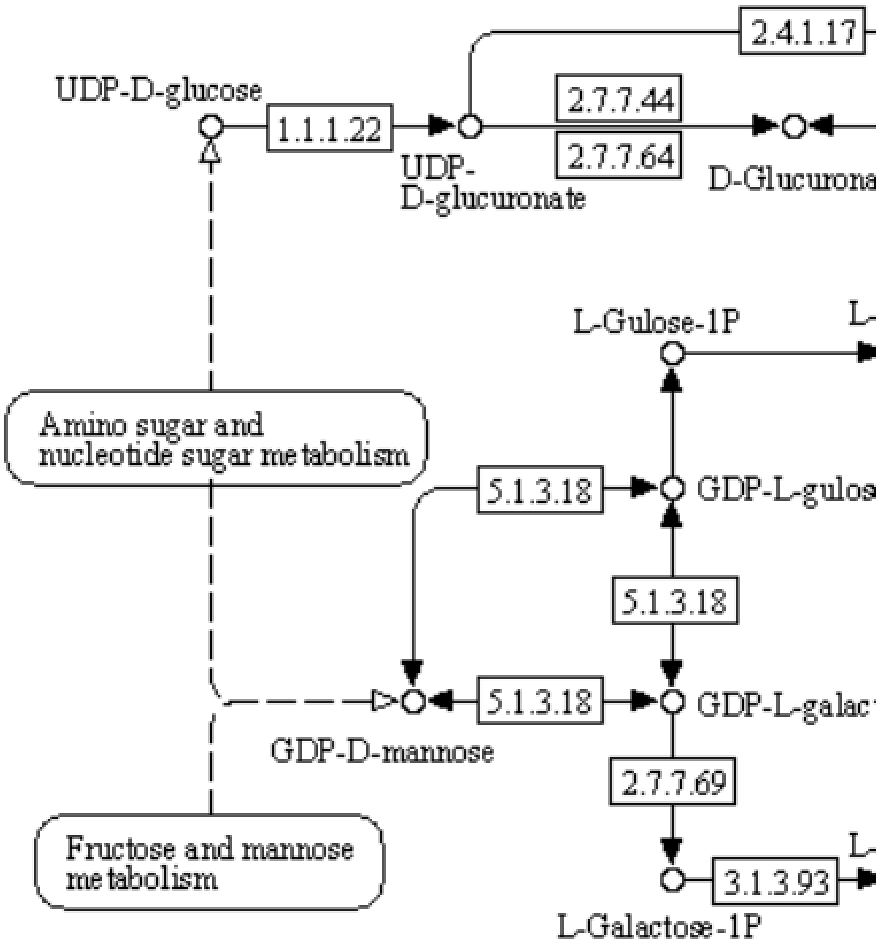
1. **Metabolism**
   Global/overview   Carbohydrate   Energy   Lipid   Nucleotide   Amino acid   Other amino   Glycan
   Cofactor/vitamin   Terpenoid/PK   Other secondary metabolite   Xenobiotics   Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**
7. **Drug Development**

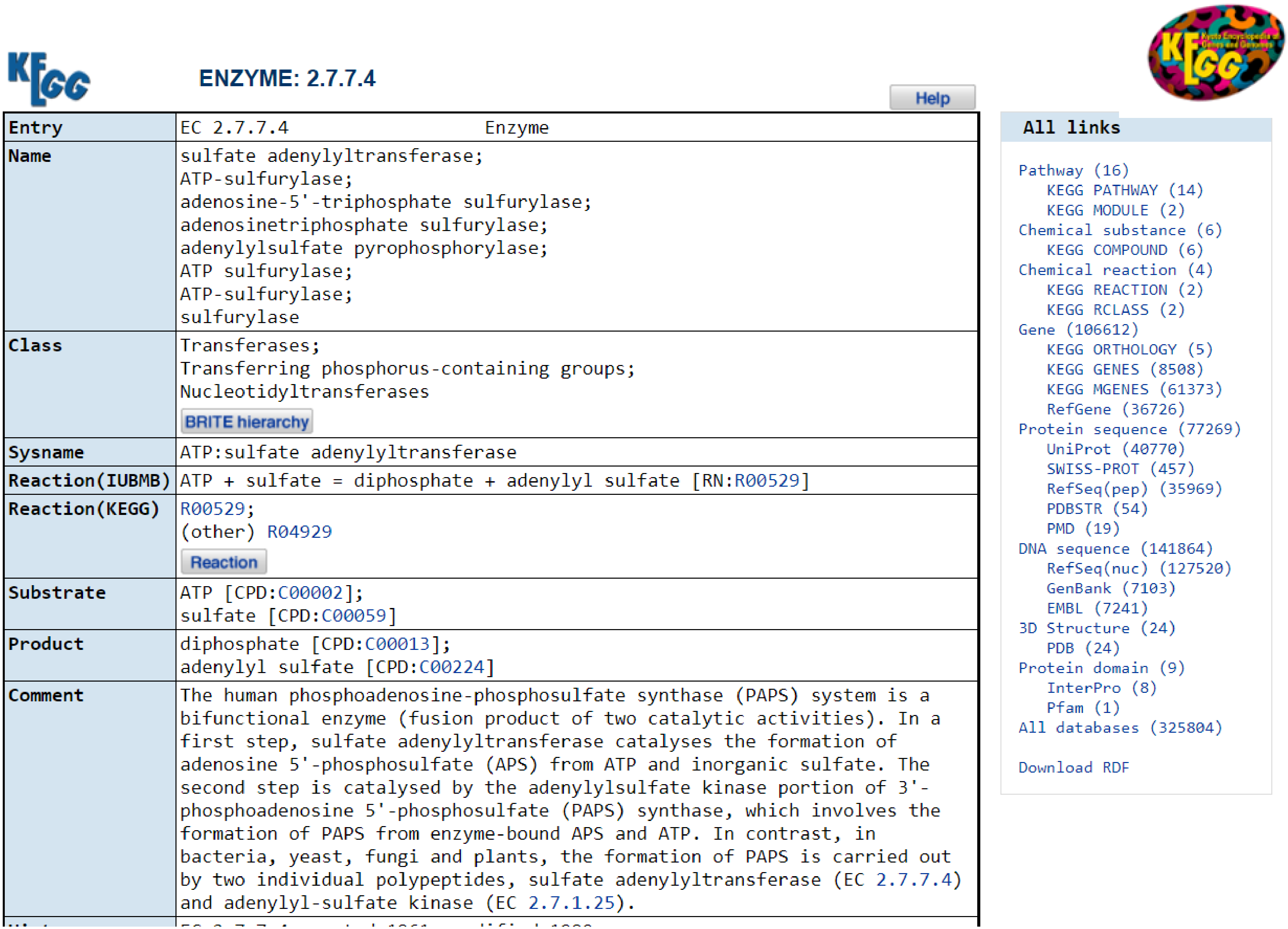
Другой метаболический путь



Фермент, катализирующий реакцию

Соединение


UDP-D-glucose
1.1.1.22
UDP-
D-glucuronate
2.4.1.17
2.7.7.44
2.7.7.64
D-Glucurona
L-Gulose-1P
L-
Amino sugar and
nucleotide sugar metabolism
5.1.3.18
GDP-L-gulos
5.1.3.18
5.1.3.18
GDP-L-galac
GDP-D-mannose
2.7.7.69
L-
3.1.3.93
L-Galactose-1P
Fructose and mannose
metabolism



ENZYME: 2.7.7.4

Entry	EC 2.7.7.4 Enzyme
Name	sulfate adenylyltransferase; ATP-sulfurylase; adenosine-5'-triphosphate sulfurylase; adenosinetriphosphate sulfurylase; adenylylsulfate pyrophosphorylase; ATP sulfurylase; ATP-sulfurylase; sulfurylase
Class	Transferases; Transferring phosphorus-containing groups; Nucleotidyltransferases BRITE hierarchy
Sysname	ATP:sulfate adenylyltransferase
Reaction(IUBMB)	ATP + sulfate = diphosphate + adenylyl sulfate [RN:R00529]
Reaction(KEGG)	R00529; (other) R04929 Reaction
Substrate	ATP [CPD:C00002]; sulfate [CPD:C00059]
Product	diphosphate [CPD:C00013]; adenylyl sulfate [CPD:C00224]
Comment	The human phosphoadenosine-phosphosulfate synthase (PAPS) system is a bifunctional enzyme (fusion product of two catalytic activities). In a first step, sulfate adenylyltransferase catalyses the formation of adenosine 5'-phosphosulfate (APS) from ATP and inorganic sulfate. The second step is catalysed by the adenylylsulfate kinase portion of 3'-phosphoadenosine 5'-phosphosulfate (PAPS) synthase, which involves the formation of PAPS from enzyme-bound APS and ATP. In contrast, in bacteria, yeast, fungi and plants, the formation of PAPS is carried out by two individual polypeptides, sulfate adenylyltransferase (EC 2.7.7.4) and adenylyl-sulfate kinase (EC 2.7.1.25).

All links

- Pathway (16)
  - KEGG PATHWAY (14)
  - KEGG MODULE (2)
- Chemical substance (6)
  - KEGG COMPOUND (6)
- Chemical reaction (4)
  - KEGG REACTION (2)
  - KEGG RCLASS (2)
- Gene (106612)
  - KEGG ORTHOLOGY (5)
  - KEGG GENES (8508)
  - KEGG MGENES (61373)
  - RefGene (36726)
- Protein sequence (77269)
  - UniProt (40770)
  - SWISS-PROT (457)
  - RefSeq(pep) (35969)
  - PDBSTR (54)
  - PMD (19)
- DNA sequence (141864)
  - RefSeq(nuc) (127520)
  - GenBank (7103)
  - EMBL (7241)
- 3D Structure (24)
  - PDB (24)
- Protein domain (9)
  - InterPro (8)
  - Pfam (1)
- All databases (325804)

Download RDF

КОД ФЕРМЕНТА

- **Класс** – код первого уровня: 1-7. Номер одного из семи главных классов ферментов
- **Подкласс** – код второго уровня. Номер характеризует основные виды субстратов, участвующих в данном виде химических реакций.
- **Под-подкласс** – код третьего уровня. Номер определяет более частные подгруппы, отличающиеся природой химических соединений доноров или акцепторов, участвующих в данной подгруппе реакций.
- **Код четвертого уровня** – все ферменты, относящиеся к данному под-подклассу получают свои порядковые номера.

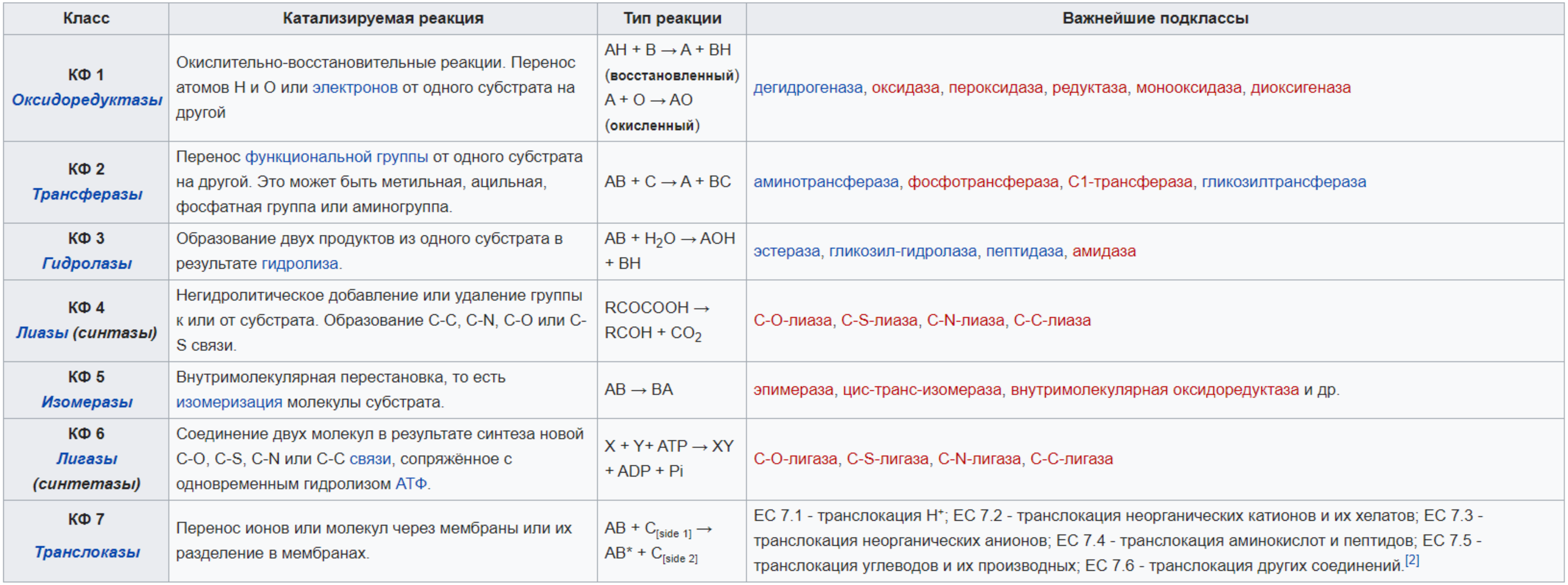
- **Пример**: глюкозоксидаза – 1.1.3.4

- **КФ 1** — Оксидоредуктазы
- **КФ 1.1** — Алкогольоксидоредуктазы
- **КФ 1.1.3** — Оксидоредуктазы, окисляющие группу CH-OH и восстанавливающие кислород.
- **КФ 1.1.3.4** — Оксидоредуктазы, окисляющие глюкозу в присутствии кислорода. Всего известно 3 глюкозоксидазы из разных организмов.

ref

ОСНОВНЫЕ КЛАССЫ ФЕРМЕНТОВ

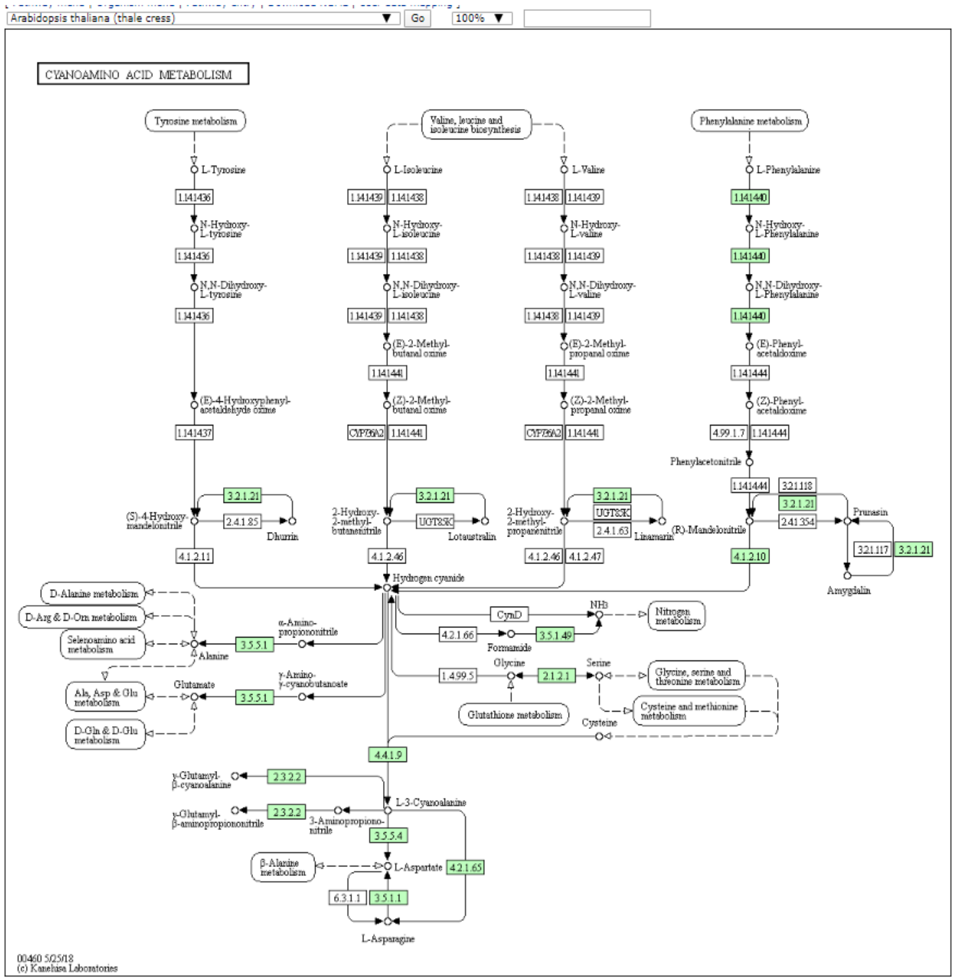
- Коды первого уровня



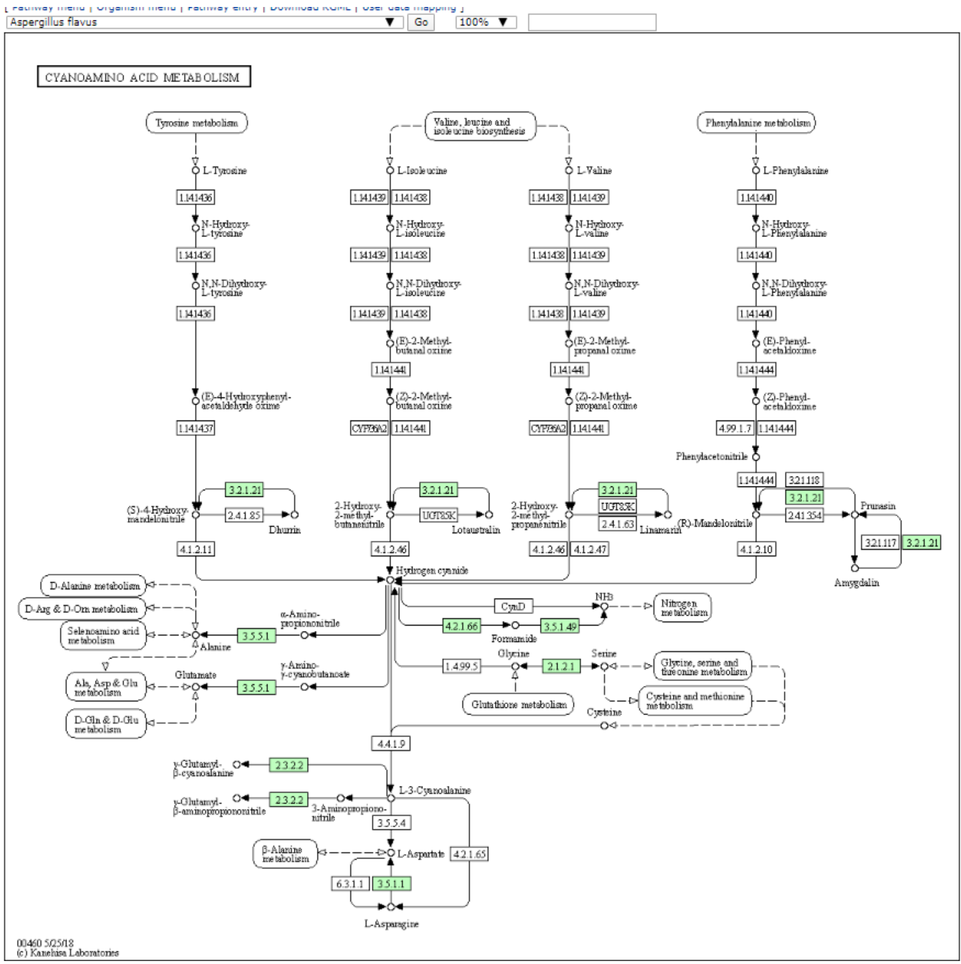
Класс	Катализируемая реакция	Тип реакции	Важнейшие подклассы
КФ 1 ***Оксидоредуктазы***	Окислительно-восстановительные реакции. Перенос атомов H и O или электронов от одного субстрата на другой	AH + B → A + BH **(восстановленный)** A + O → AO **(окисленный)**	дегидрогеназа, оксидаза, пероксидаза, редуктаза, монооксидаза, диоксигеназа
КФ 2 ***Трансферазы***	Перенос функциональной группы от одного субстрата на другой. Это может быть метильная, ацильная, фосфатная группа или аминогруппа.	AB + C → A + BC	аминотрансфераза, фосфотрансфераза, C1-трансфераза, гликозилтрансфераза
КФ 3 ***Гидролазы***	Образование двух продуктов из одного субстрата в результате гидролиза.	AB + H_2O → AOH + BH	эстераза, гликозил-гидролаза, пептидаза, амидаза
КФ 4 ***Лиазы (синтазы)***	Негидролитическое добавление или удаление группы к или от субстрата. Образование C-C, C-N, C-O или C-S связи.	RCOCOOH → RCOH + CO_2	C-O-лиаза, C-S-лиаза, C-N-лиаза, C-C-лиаза
КФ 5 ***Изомеразы***	Внутримолекулярная перестановка, то есть изомеризация молекулы субстрата.	AB → BA	эпимераза, цис-транс-изомераза, внутримолекулярная оксидоредуктаза и др.
КФ 6 ***Лигазы (синтетазы)***	Соединение двух молекул в результате синтеза новой C-O, C-S, C-N или C-C связи, сопряжённое с одновременным гидролизом АТФ.	X + Y+ ATP → XY + ADP + Pi	C-O-лигаза, C-S-лигаза, C-N-лигаза, C-C-лигаза
КФ 7 ***Транслоказы***	Перенос ионов или молекул через мембраны или их разделение в мембранах.	AB + $C_{[side\ 1]}$ → AB* + $C_{[side\ 2]}$	EC 7.1 - транслокация H^+; EC 7.2 - транслокация неорганических катионов и их хелатов; EC 7.3 - транслокация неорганических анионов; EC 7.4 - транслокация аминокислот и пептидов; EC 7.5 - транслокация углеводов и их производных; EC 7.6 - транслокация других соединений.[2]

ref

Сравнительная геномика

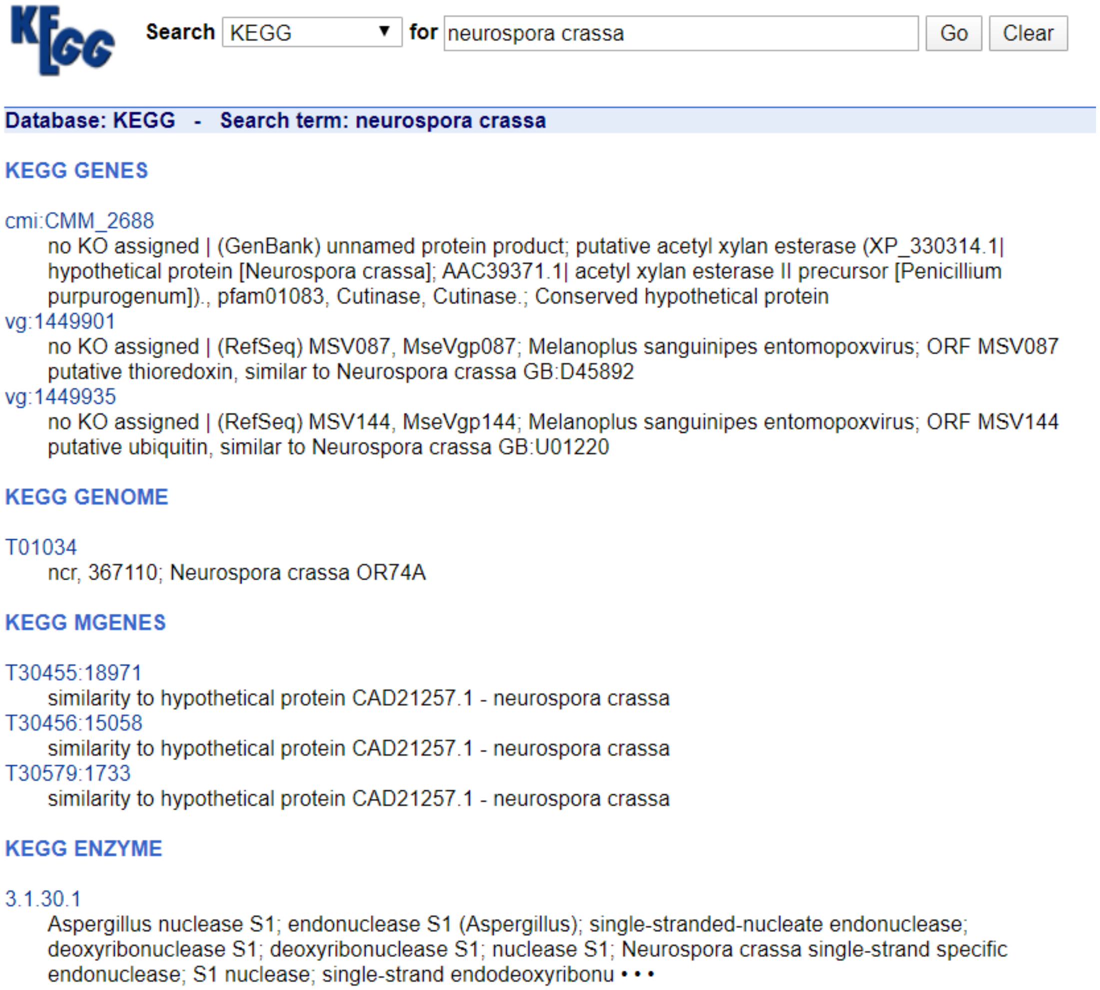


Arabidopsis thaliana (thale cress) ▼ Go 100% ▼



Aspergillus flavus ▼ Go 100% ▼

ПОИСК ОТ ОРГАНИЗМА



Database: KEGG - Search term: neurospora crassa

KEGG GENES

cmi:CMM_2688
no KO assigned | (GenBank) unnamed protein product; putative acetyl xylan esterase (XP_330314.1| hypothetical protein [Neurospora crassa]; AAC39371.1| acetyl xylan esterase II precursor [Penicillium purpurogenum])., pfam01083, Cutinase, Cutinase.; Conserved hypothetical protein
vg:1449901
no KO assigned | (RefSeq) MSV087, MseVgp087; Melanoplus sanguinipes entomopoxvirus; ORF MSV087 putative thioredoxin, similar to Neurospora crassa GB:D45892
vg:1449935
no KO assigned | (RefSeq) MSV144, MseVgp144; Melanoplus sanguinipes entomopoxvirus; ORF MSV144 putative ubiquitin, similar to Neurospora crassa GB:U01220

KEGG GENOME

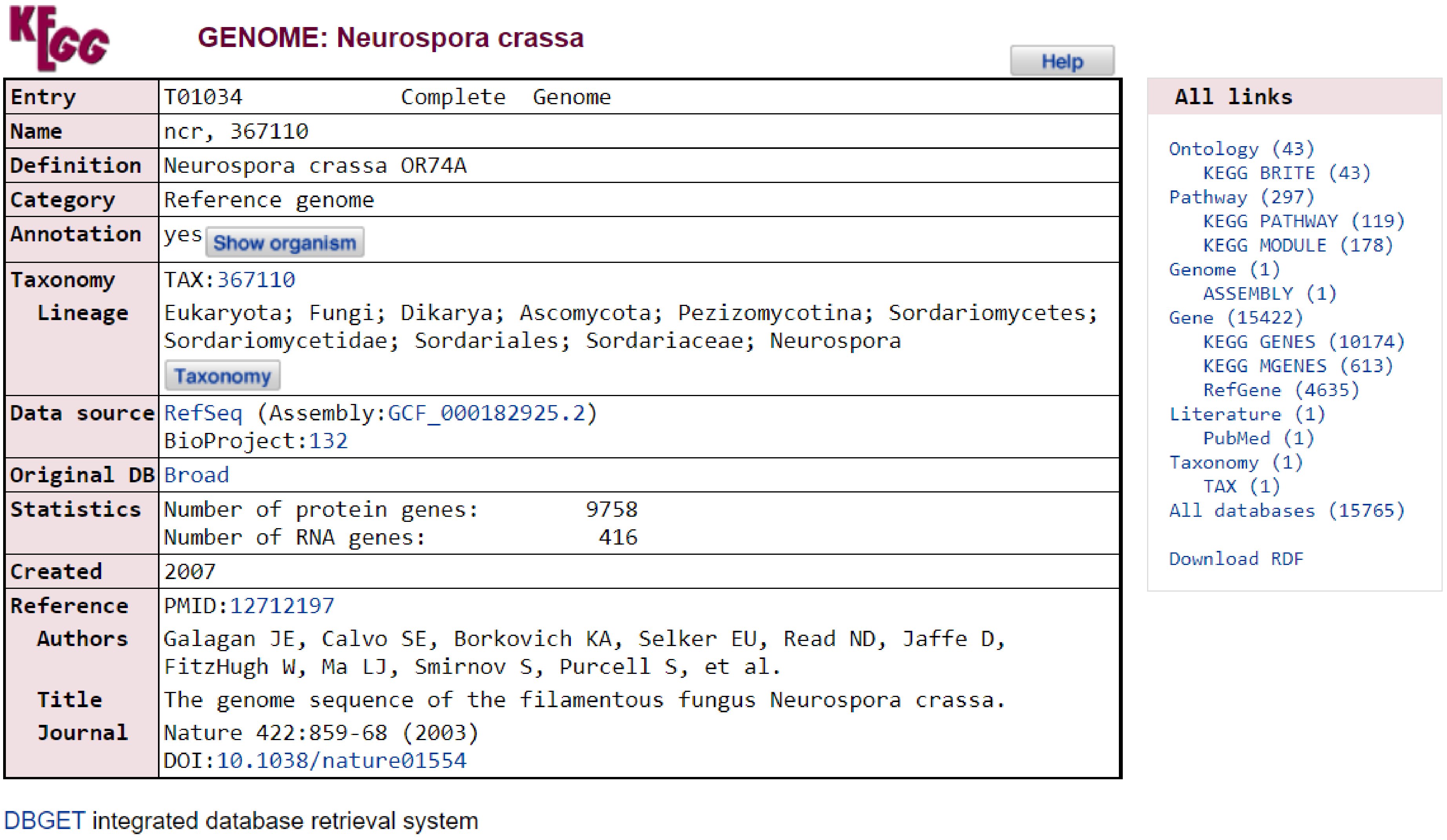
T01034
ncr, 367110; Neurospora crassa OR74A

KEGG MGENES

T30455:18971
similarity to hypothetical protein CAD21257.1 - neurospora crassa
T30456:15058
similarity to hypothetical protein CAD21257.1 - neurospora crassa
T30579:1733
similarity to hypothetical protein CAD21257.1 - neurospora crassa

KEGG ENZYME

3.1.30.1
Aspergillus nuclease S1; endonuclease S1 (Aspergillus); single-stranded-nucleate endonuclease; deoxyribonuclease S1; deoxyribonuclease S1; nuclease S1; Neurospora crassa single-strand specific endonuclease; S1 nuclease; single-strand endodeoxyribonu • • •



GENOME: Neurospora crassa

Entry	T01034 Complete Genome
Name	ncr, 367110
Definition	Neurospora crassa OR74A
Category	Reference genome
Annotation	yes
Taxonomy	TAX:367110
Lineage	Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora
Data source	RefSeq (Assembly:GCF_000182925.2) BioProject:132
Original DB	Broad
Statistics	Number of protein genes: 9758 Number of RNA genes: 416
Created	2007
Reference	PMID:12712197
Authors	Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, et al.
Title	The genome sequence of the filamentous fungus Neurospora crassa.
Journal	Nature 422:859-68 (2003) DOI:10.1038/nature01554

All links

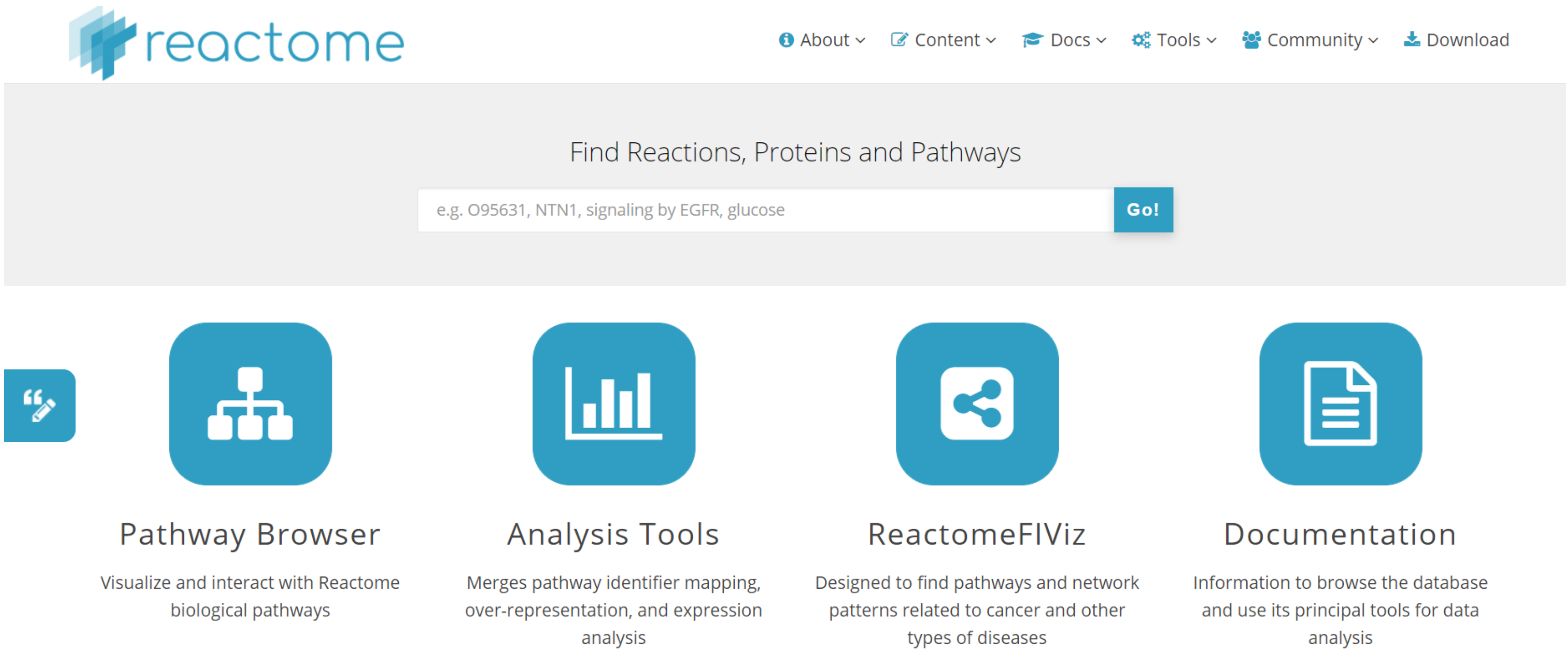
- Ontology (43)
  - KEGG BRITE (43)
- Pathway (297)
  - KEGG PATHWAY (119)
  - KEGG MODULE (178)
- Genome (1)
  - ASSEMBLY (1)
- Gene (15422)
  - KEGG GENES (10174)
  - KEGG MGENES (613)
  - RefGene (4635)
- Literature (1)
  - PubMed (1)
- Taxonomy (1)
  - TAX (1)
- All databases (15765)

Download RDF

DBGET integrated database retrieval system

REACTOME

- https://reactome.org/



Find Reactions, Proteins and Pathways

Pathway Browser

Visualize and interact with Reactome biological pathways

Analysis Tools

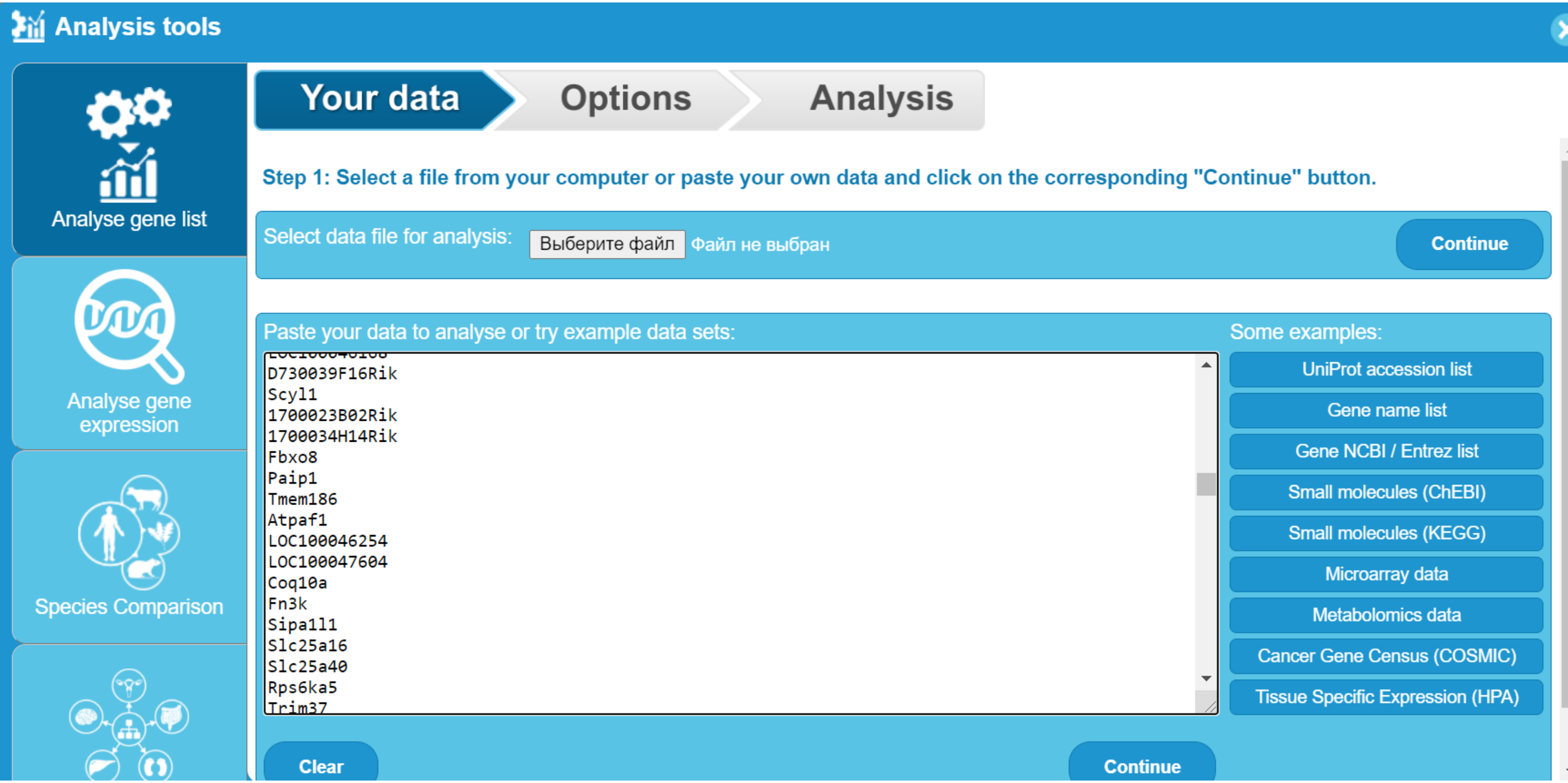
Merges pathway identifier mapping, over-representation, and expression analysis

ReactomeFIViz

Designed to find pathways and network patterns related to cancer and other types of diseases

Documentation

Information to browse the database and use its principal tools for data analysis

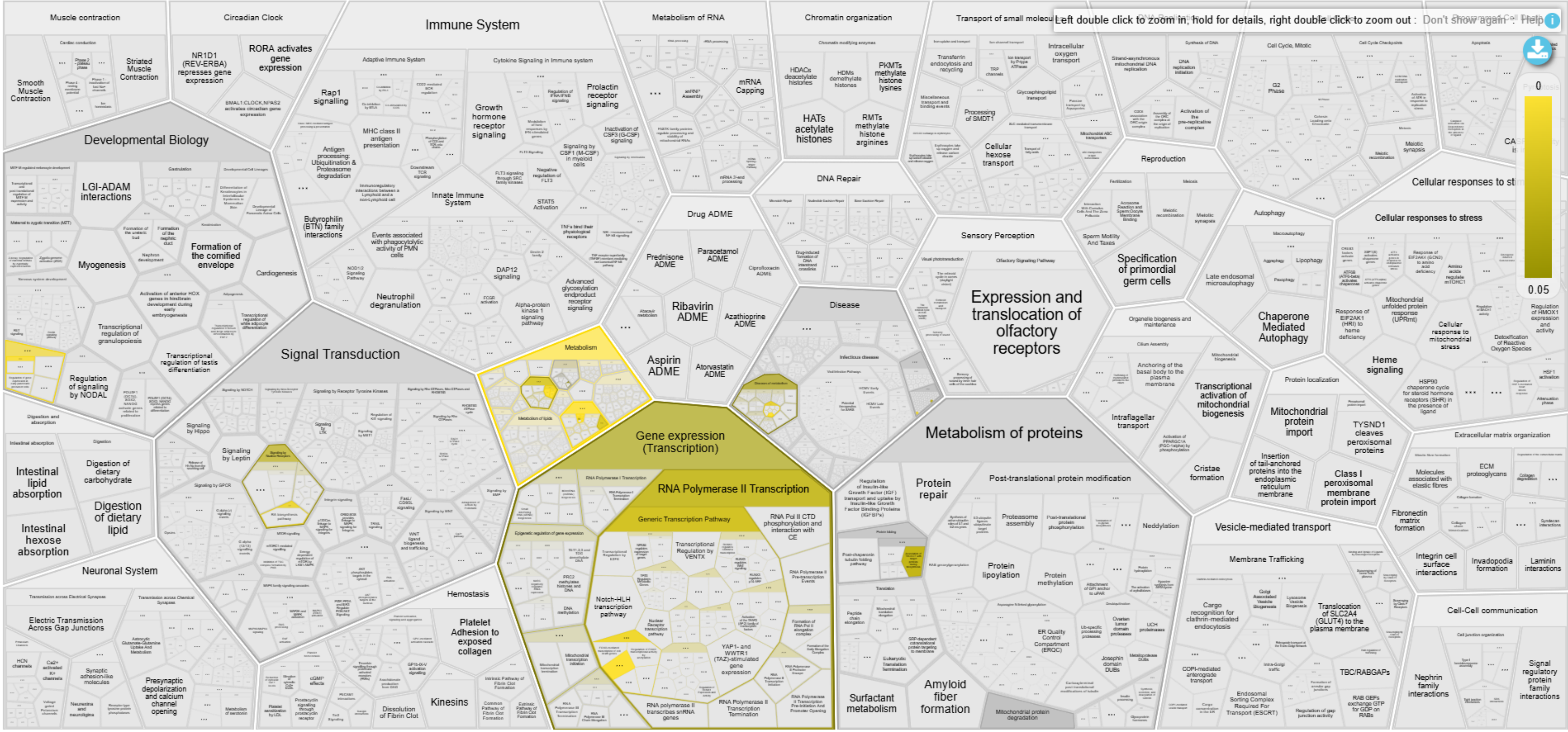


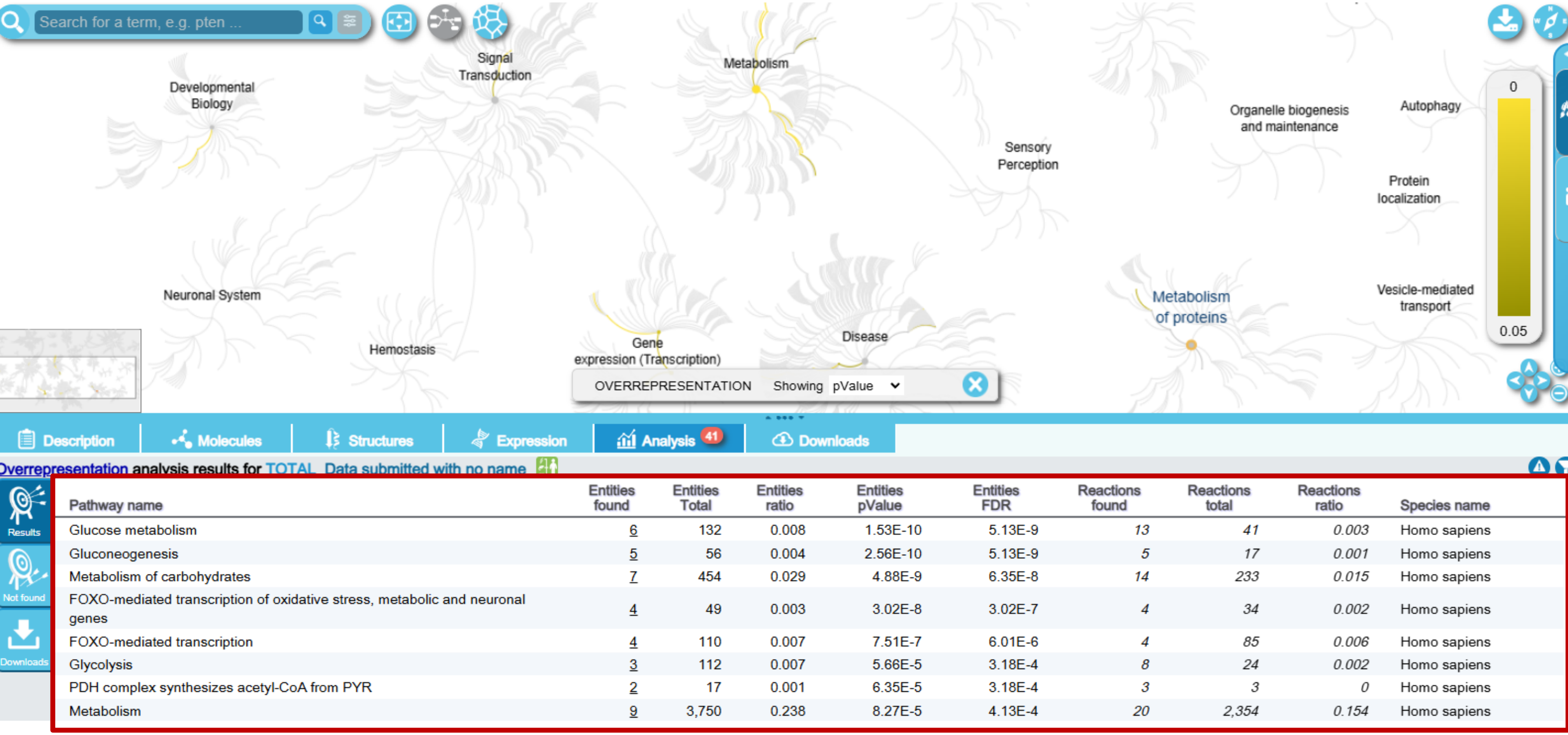
Step 1: Select a file from your computer or paste your own data and click on the corresponding "Continue" button.

Select data file for analysis: Выберите файл Файл не выбран

Paste your data to analyse or try example data sets:

```
D730039F16Rik
Scyl1
1700023B02Rik
1700034H14Rik
Fbxo8
Paip1
Tmem186
Atpaf1
LOC100046254
LOC100047604
Coq10a
Fn3k
Sipa1l1
Slc25a16
Slc25a40
Rps6ka5
Trim37
```

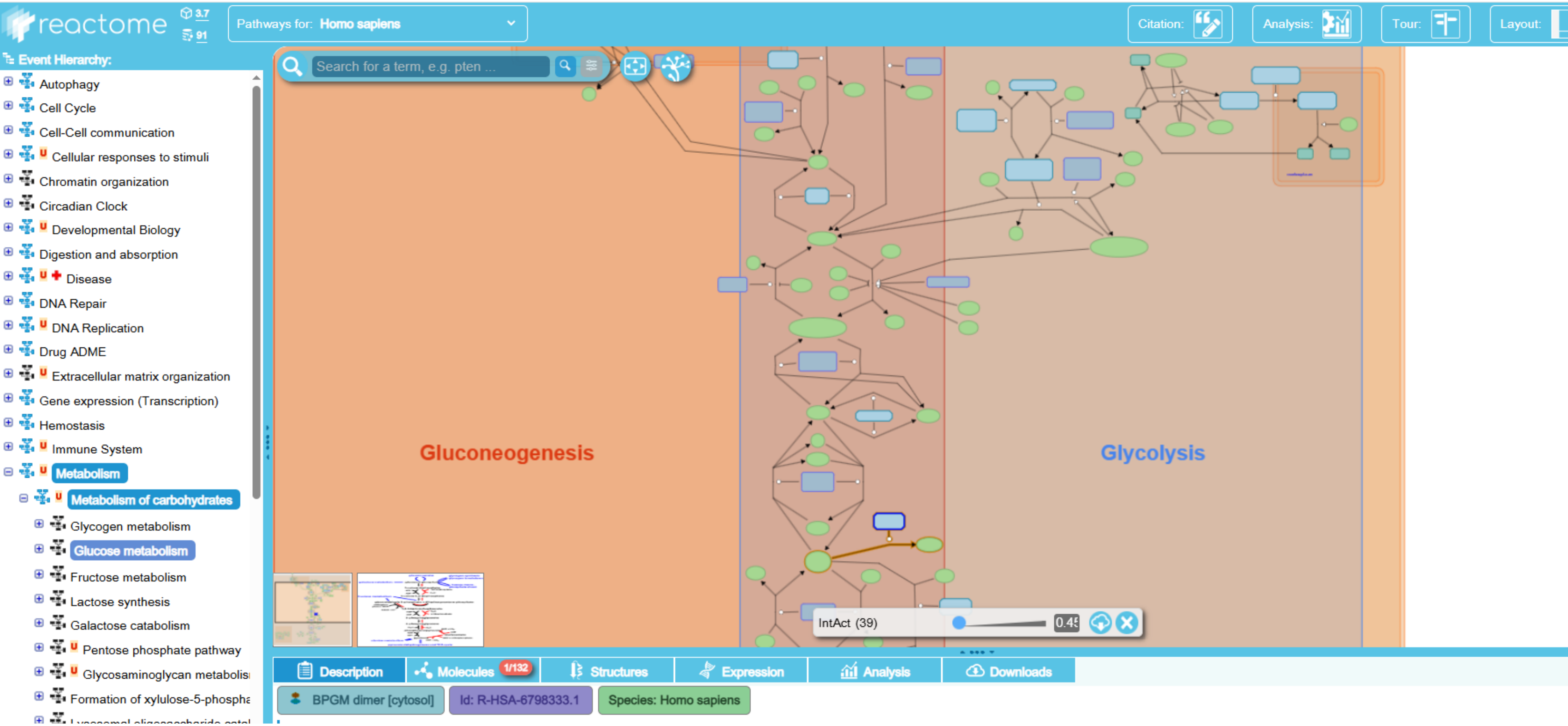

Left double click to zoom in, hold for details, right double click to zoom out : Don't show again : Help
0
0.05
Muscle contraction
Smooth Muscle Contraction
Striated Muscle Contraction
Circadian Clock
NR1D1 (REV-ERBA) represses gene expression
RORA activates gene expression
Immune System
Adaptive Immune System
Cytokine Signaling in Immune system
Rap1 signalling
MHC class II antigen presentation
Growth hormone receptor signaling
Prolactin receptor signaling
Antigen processing: Ubiquitination & Proteasome degradation
Signaling by CSF1 (M-CSF) in myeloid cells
Innate Immune System
Butyrophilin (BTN) family interactions
Neutrophil degranulation
DAP12 signaling
Metabolism of RNA
mRNA Capping
Chromatin organization
HDACs deacetylate histones
HDMs demethylate histones
PKMTs methylate histone lysines
HATs acetylate histones
RMTs methylate histone arginines
DNA Repair
Transport of small molecu
Intracellular oxygen transport
Cellular hexose transport
Cell Cycle, Mitotic
Cell Cycle Checkpoints
G2 Phase
Apoptosis
Reproduction
Cellular responses to st
Cellular responses to stress
Developmental Biology
LGI-ADAM interactions
Myogenesis
Formation of the cornified envelope
Cardiogenesis
Transcriptional regulation of granulopoiesis
Transcriptional regulation of testis differentiation
Regulation of signaling by NODAL
Signal Transduction
Signaling by Hippo
Signaling by Leptin
Drug ADME
Prednisone ADME
Paracetamol ADME
Ribavirin ADME
Azathioprine ADME
Aspirin ADME
Atorvastatin ADME
Metabolism
Metabolism of lipids
Gene expression (Transcription)
RNA Polymerase II Transcription
Generic Transcription Pathway
RNA Pol II CTD phosphorylation and interaction with CE
Transcriptional Regulation by VENTX
Notch-HLH transcription pathway
YAP1- and WWTR1 (TAZ)-stimulated gene expression
RNA polymerase II transcribes snRNA genes
RNA Polymerase II Transcription Termination
Disease
Sensory Perception
Expression and translocation of olfactory receptors
Specification of primordial germ cells
Autophagy
Late endosomal microautophagy
Chaperone Mediated Autophagy
Heme signaling
Cellular response to mitochondrial stress
Transcriptional activation of mitochondrial biogenesis
Intraflagellar transport
Protein localization
Mitochondrial protein import
TYSND1 cleaves peroxisomal proteins
Class I peroxisomal membrane protein import
Cristae formation
Metabolism of proteins
Protein repair
Post-translational protein modification
Neddylation
Protein lipoylation
Protein methylation
Surfactant metabolism
Amyloid fiber formation
Vesicle-mediated transport
Membrane Trafficking
Translocation of SLC2A4 (GLUT4) to the plasma membrane
TBC/RABGAPs
Cargo recognition for clathrin-mediated endocytosis
Extracellular matrix organization
ECM proteoglycans
Fibronectin matrix formation
Integrin cell surface interactions
Invadopodia formation
Laminin interactions
Cell-Cell communication
Nephrin family interactions
Signal regulatory protein family interactions
Digestion and absorption
Intestinal lipid absorption
Digestion of dietary carbohydrate
Digestion of dietary lipid
Intestinal hexose absorption
Neuronal System
Electric Transmission Across Gap Junctions
Presynaptic depolarization and calcium channel opening
Hemostasis
Platelet Adhesion to exposed collagen
Kinesins
Dissolution of Fibrin Clot



OVERREPRESENTATION Showing pValue

Overrepresentation analysis results for TOTAL Data submitted with no name

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Glucose metabolism	6	132	0.008	1.53E-10	5.13E-9	13	41	0.003	Homo sapiens
Gluconeogenesis	5	56	0.004	2.56E-10	5.13E-9	5	17	0.001	Homo sapiens
Metabolism of carbohydrates	7	454	0.029	4.88E-9	6.35E-8	14	233	0.015	Homo sapiens
FOXO-mediated transcription of oxidative stress, metabolic and neuronal genes	4	49	0.003	3.02E-8	3.02E-7	4	34	0.002	Homo sapiens
FOXO-mediated transcription	4	110	0.007	7.51E-7	6.01E-6	4	85	0.006	Homo sapiens
Glycolysis	3	112	0.007	5.66E-5	3.18E-4	8	24	0.002	Homo sapiens
PDH complex synthesizes acetyl-CoA from PYR	2	17	0.001	6.35E-5	3.18E-4	3	3	0	Homo sapiens
Metabolism	9	3,750	0.238	8.27E-5	4.13E-4	20	2,354	0.154	Homo sapiens



Pathways for: Homo sapiens

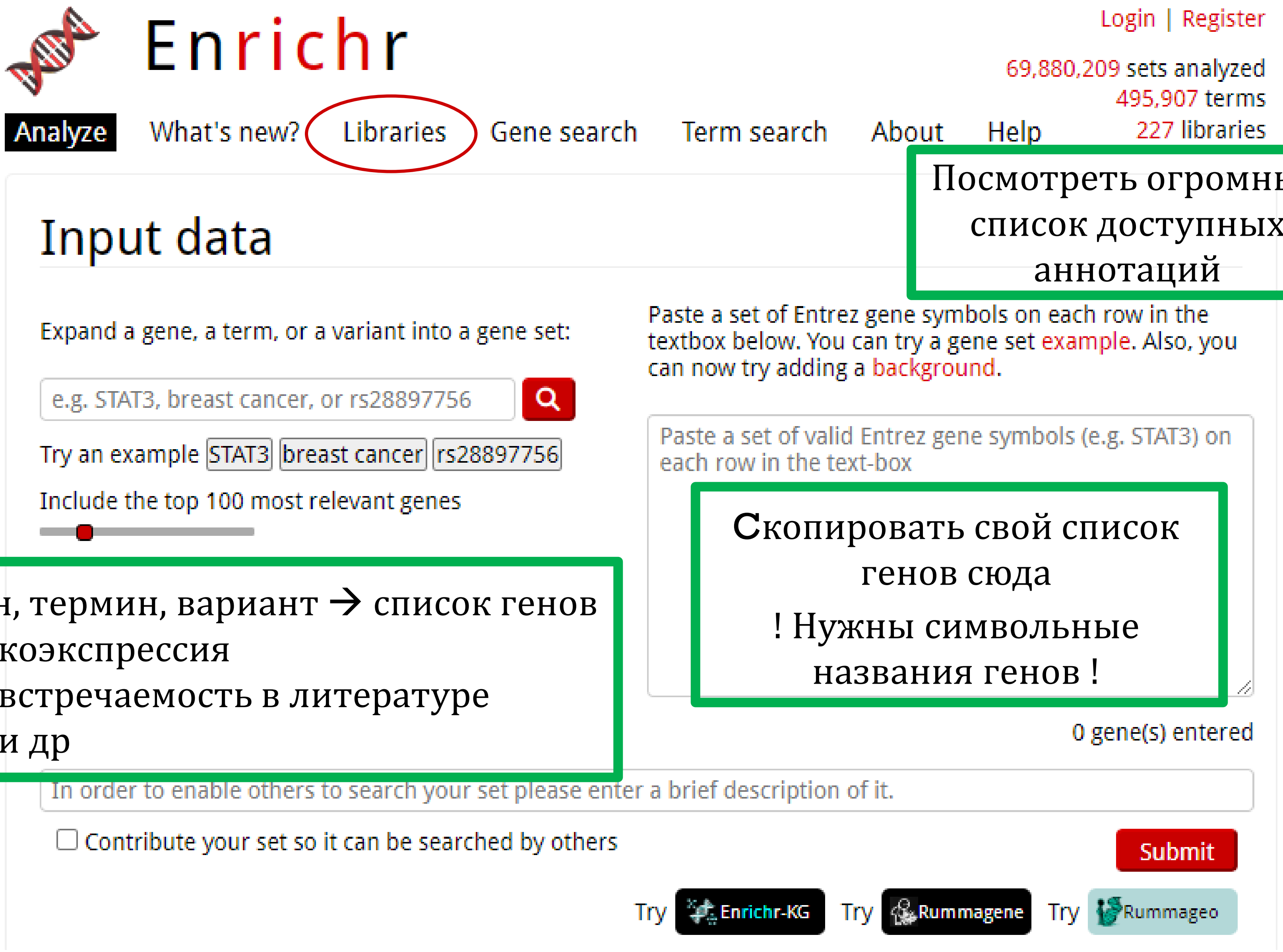
IntAct (39) 0.45

BPGM dimer [cytosol]

Id: R-HSA-6798333.1

Species: Homo sapiens

https://maayanlab.cloud/Enrichr/



Input data

Expand a gene, a term, or a variant into a gene set:

e.g. STAT3, breast cancer, or rs28897756

Include the top 100 most relevant genes

Paste a set of Entrez gene symbols on each row in the textbox below. You can try a gene set example. Also, you can now try adding a background.

Paste a set of valid Entrez gene symbols (e.g. STAT3) on each row in the text-box

0 gene(s) entered

In order to enable others to search your set please enter a brief description of it.

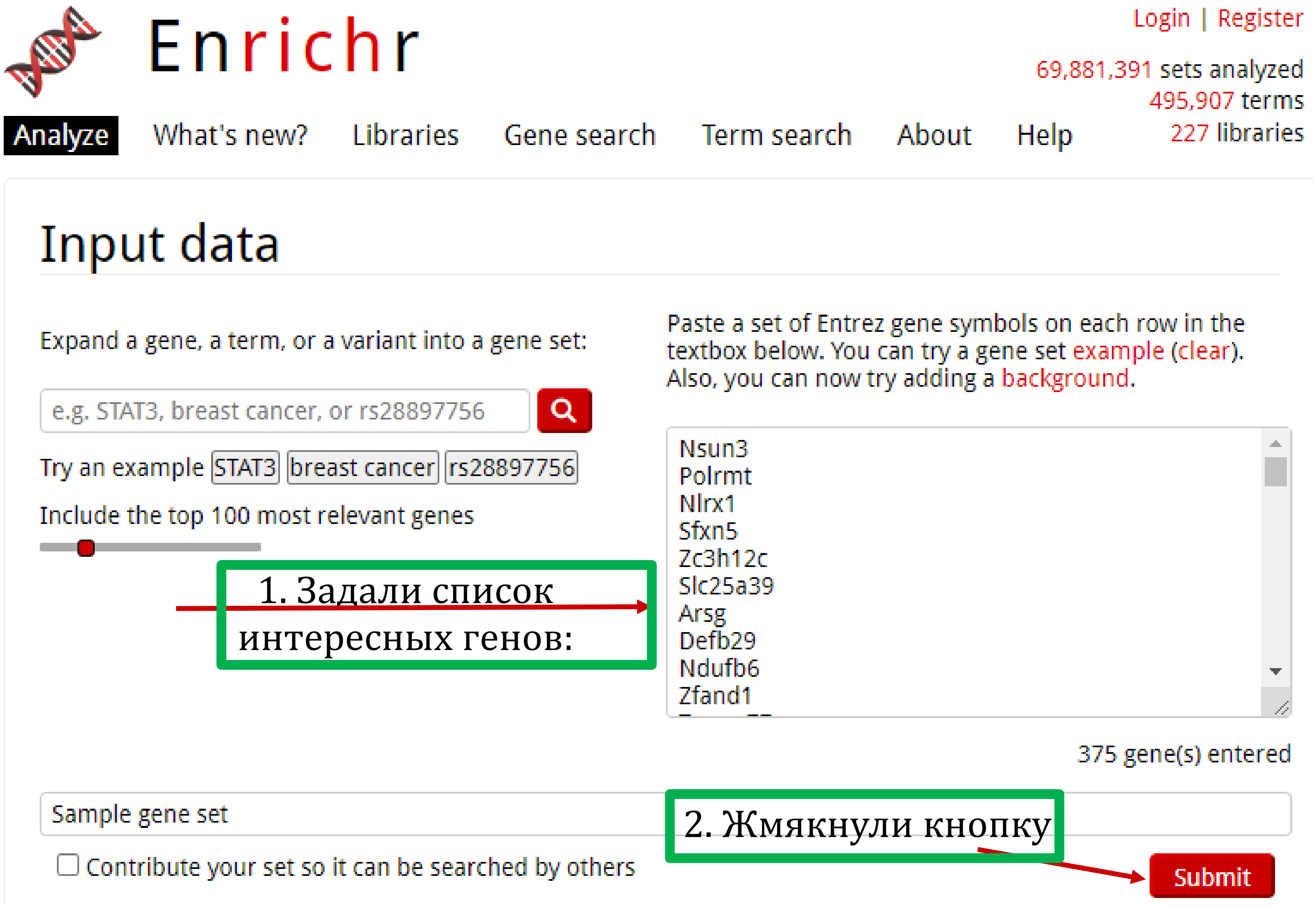
☐ Contribute your set so it can be searched by others

Посмотреть огромный список доступных аннотаций

Скопировать свой список генов сюда

! Нужны символьные названия генов !

Ген, термин, вариант → список генов
- коэкспрессия
- встречаемость в литературе
- и др



69,881,391 sets analyzed
495,907 terms
227 libraries

Input data

Expand a gene, a term, or a variant into a gene set:

e.g. STAT3, breast cancer, or rs28897756

Try an example STAT3 breast cancer rs28897756

Include the top 100 most relevant genes

Paste a set of Entrez gene symbols on each row in the textbox below. You can try a gene set example (clear). Also, you can now try adding a background.

```
Nsun3
Polrmt
Nlrx1
Sfxn5
Zc3h12c
Slc25a39
Arsg
Defb29
Ndufb6
Zfand1
```

375 gene(s) entered

Sample gene set

☐ Contribute your set so it can be searched by others

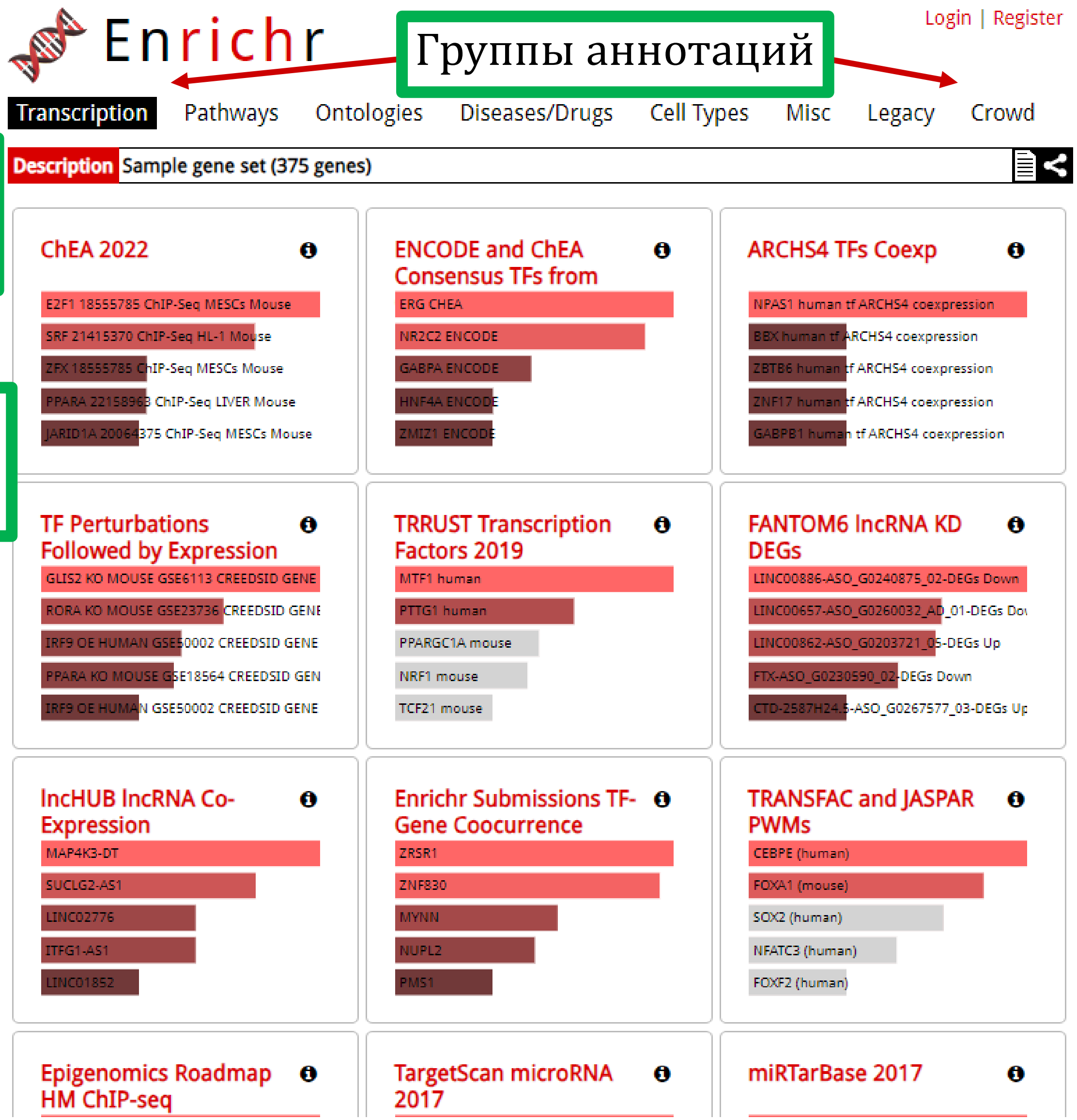
1. Задали список интересных генов:

2. Жмякнули кнопку

Получили результат - радуемся!

Еще больше результатов по разным аннотациям

Группы аннотаций



Description Sample gene set (375 genes)

ChEA 2022
- E2F1 18555785 ChIP-Seq MESCs Mouse
- SRF 21415370 ChIP-Seq HL-1 Mouse
- ZFX 18555785 ChIP-Seq MESCs Mouse
- PPARA 22158963 ChIP-Seq LIVER Mouse
- JARID1A 20064375 ChIP-Seq MESCs Mouse

ENCODE and ChEA Consensus TFs from
- ERG CHEA
- NR2C2 ENCODE
- GABPA ENCODE
- HNF4A ENCODE
- ZMIZ1 ENCODE

ARCHS4 TFs Coexp
- NPAS1 human tf ARCHS4 coexpression
- BBX human tf ARCHS4 coexpression
- ZBTB6 human tf ARCHS4 coexpression
- ZNF17 human tf ARCHS4 coexpression
- GABPB1 human tf ARCHS4 coexpression

TF Perturbations Followed by Expression
- GLIS2 KO MOUSE GSE6113 CREEDSID GENE
- RORA KO MOUSE GSE23736 CREEDSID GENI
- IRF9 OE HUMAN GSE50002 CREEDSID GENE
- PPARA KO MOUSE GSE18564 CREEDSID GEN
- IRF9 OE HUMAN GSE50002 CREEDSID GENE

TRRUST Transcription Factors 2019
- MTF1 human
- PTTG1 human
- PPARGC1A mouse
- NRF1 mouse
- TCF21 mouse

FANTOM6 lncRNA KD DEGs
- LINC00886-ASO_G0240875_02-DEGs Down
- LINC00657-ASO_G0260032_AD_01-DEGs Do
- LINC00862-ASO_G0203721_05-DEGs Up
- FTX-ASO_G0230590_02-DEGs Down
- CTD-2587H24.5-ASO_G0267577_03-DEGs Up

lncHUB lncRNA Co-Expression
- MAP4K3-DT
- SUCLG2-AS1
- LINC02776
- ITFG1-AS1
- LINC01852

Enrichr Submissions TF-Gene Coocurrence
- ZRSR1
- ZNF830
- MYNN
- NUPL2
- PMS1

TRANSFAC and JASPAR PWMs
- CEBPE (human)
- FOXA1 (mouse)
- SOX2 (human)
- NFATC3 (human)
- FOXF2 (human)

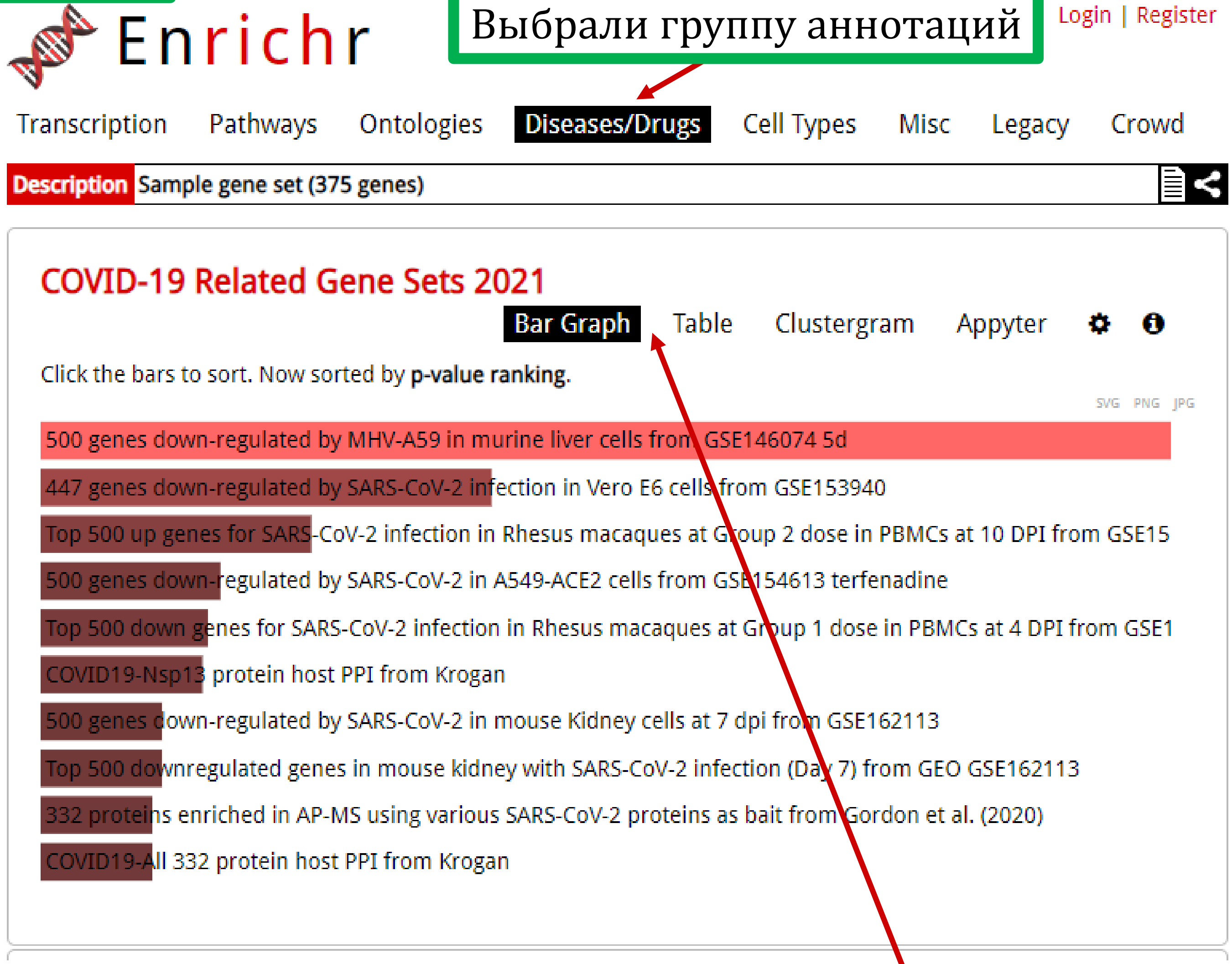
Epigenomics Roadmap HM ChIP-seq

TargetScan microRNA 2017

miRTarBase 2017

Изучили результаты по нужным аннотациям!

Выбрали группу аннотаций



Description Sample gene set (375 genes)

COVID-19 Related Gene Sets 2021

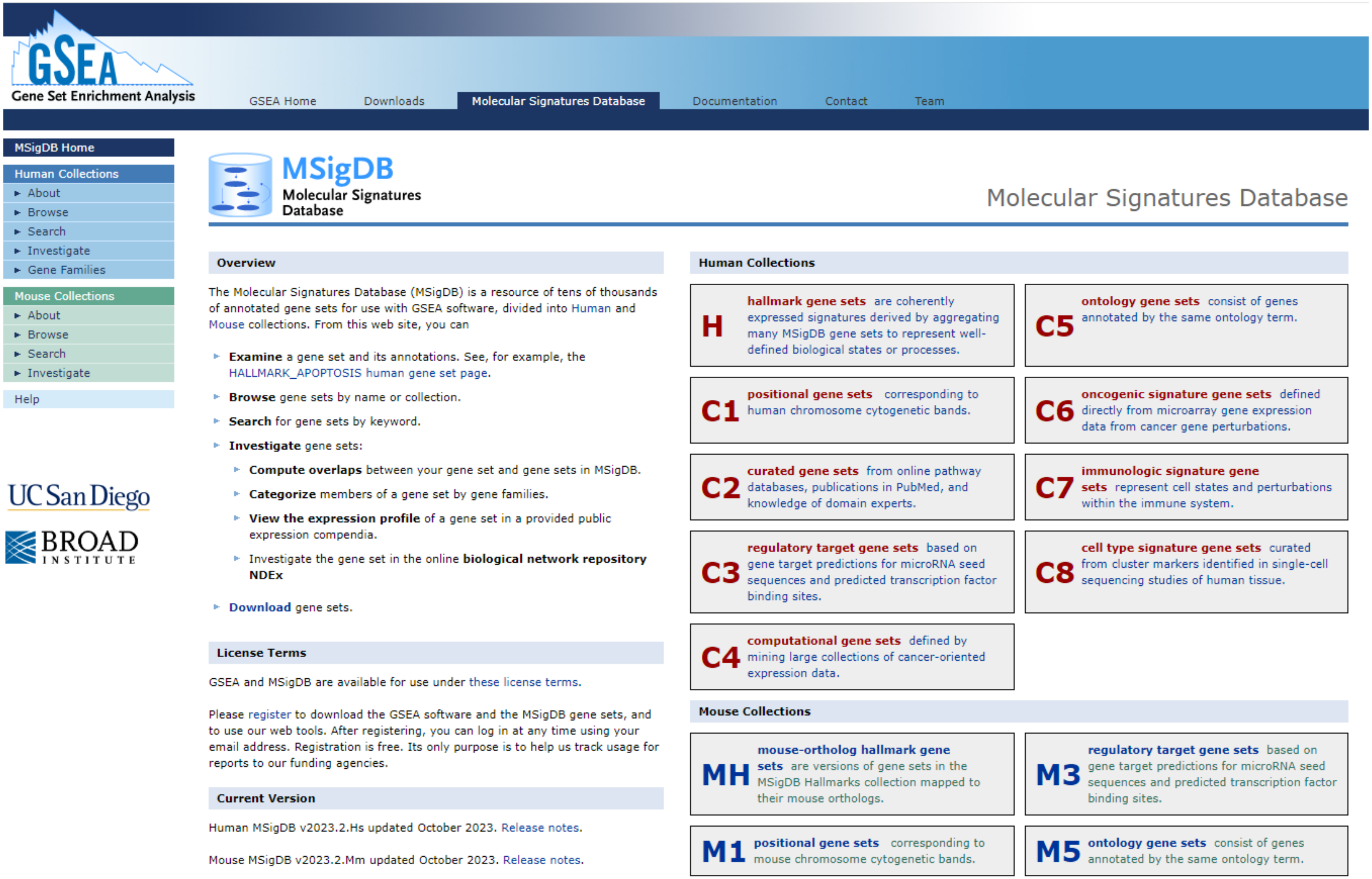
Click the bars to sort. Now sorted by **p-value ranking**.

500 genes down-regulated by MHV-A59 in murine liver cells from GSE146074 5d
447 genes down-regulated by SARS-CoV-2 infection in Vero E6 cells from GSE153940
Top 500 up genes for SARS-CoV-2 infection in Rhesus macaques at Group 2 dose in PBMCs at 10 DPI from GSE15
500 genes down-regulated by SARS-CoV-2 in A549-ACE2 cells from GSE154613 terfenadine
Top 500 down genes for SARS-CoV-2 infection in Rhesus macaques at Group 1 dose in PBMCs at 4 DPI from GSE1
COVID19-Nsp13 protein host PPI from Krogan
500 genes down-regulated by SARS-CoV-2 in mouse Kidney cells at 7 dpi from GSE162113
Top 500 downregulated genes in mouse kidney with SARS-CoV-2 infection (Day 7) from GEO GSE162113
332 proteins enriched in AP-MS using various SARS-CoV-2 proteins as bait from Gordon et al. (2020)
COVID19-All 332 protein host PPI from Krogan

Обогащенные категории

Выбрали вид представления результатов

http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp



Molecular Signatures Database

Overview

The Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use with GSEA software, divided into Human and Mouse collections. From this web site, you can

- **Examine** a gene set and its annotations. See, for example, the HALLMARK_APOPTOSIS human gene set page.
- **Browse** gene sets by name or collection.
- **Search** for gene sets by keyword.
- **Investigate** gene sets:
  - **Compute overlaps** between your gene set and gene sets in MSigDB.
  - **Categorize** members of a gene set by gene families.
  - **View the expression profile** of a gene set in a provided public expression compendia.
  - Investigate the gene set in the online **biological network repository NDEx**
- **Download** gene sets.

License Terms

GSEA and MSigDB are available for use under these license terms.

Please register to download the GSEA software and the MSigDB gene sets, and to use our web tools. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Current Version

Human MSigDB v2023.2.Hs updated October 2023. Release notes.

Mouse MSigDB v2023.2.Mm updated October 2023. Release notes.

Human Collections

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** corresponding to human chromosome cytogenetic bands.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3 **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

C4 **computational gene sets** defined by mining large collections of cancer-oriented expression data.

C5 **ontology gene sets** consist of genes annotated by the same ontology term.

C6 **oncogenic signature gene sets** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic signature gene sets** represent cell states and perturbations within the immune system.

C8 **cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of human tissue.

Mouse Collections

MH **mouse-ortholog hallmark gene sets** are versions of gene sets in the MSigDB Hallmarks collection mapped to their mouse orthologs.

M1 **positional gene sets** corresponding to mouse chromosome cytogenetic bands.

M3 **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

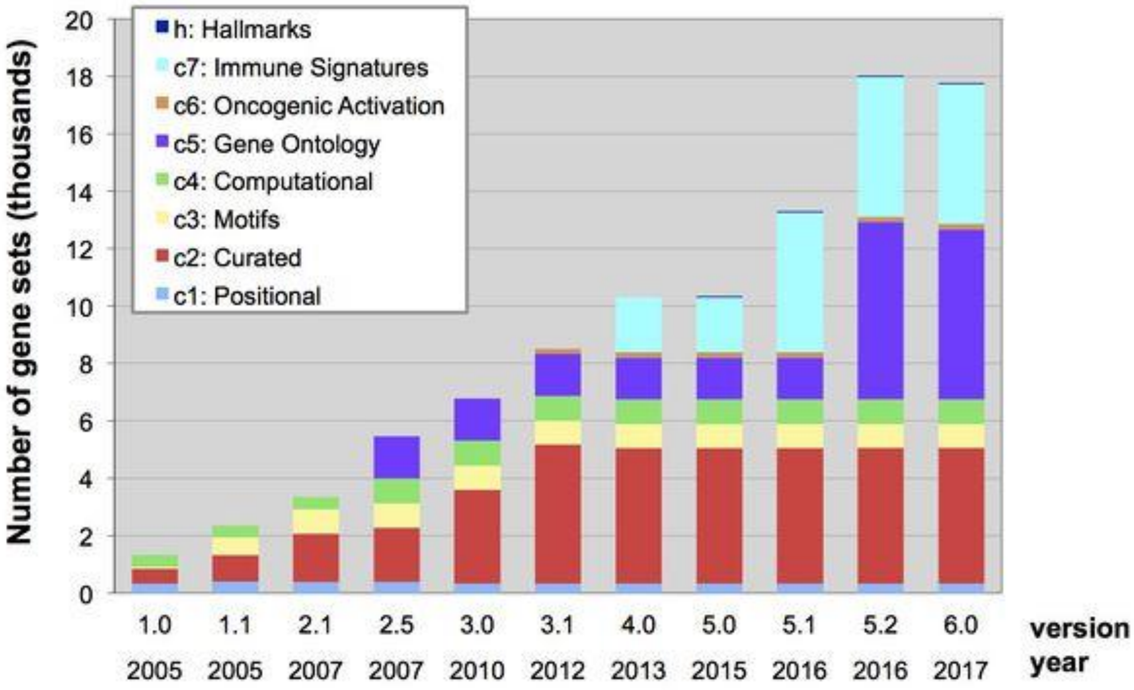
M5 **ontology gene sets** consist of genes annotated by the same ontology term.

Аннотированные гены человека и и мыши сгруппированы по 9 и 6 коллекциям

Можно скачать все сразу, либо отдельные коллекции или аннотации внутри коллекций.

Коллекции включают в себя как широко известные GO, KEGG, Reactome…

Так и более специализированные коллекции:
- связанные с иммунной системой
- раком
- вакцинами…

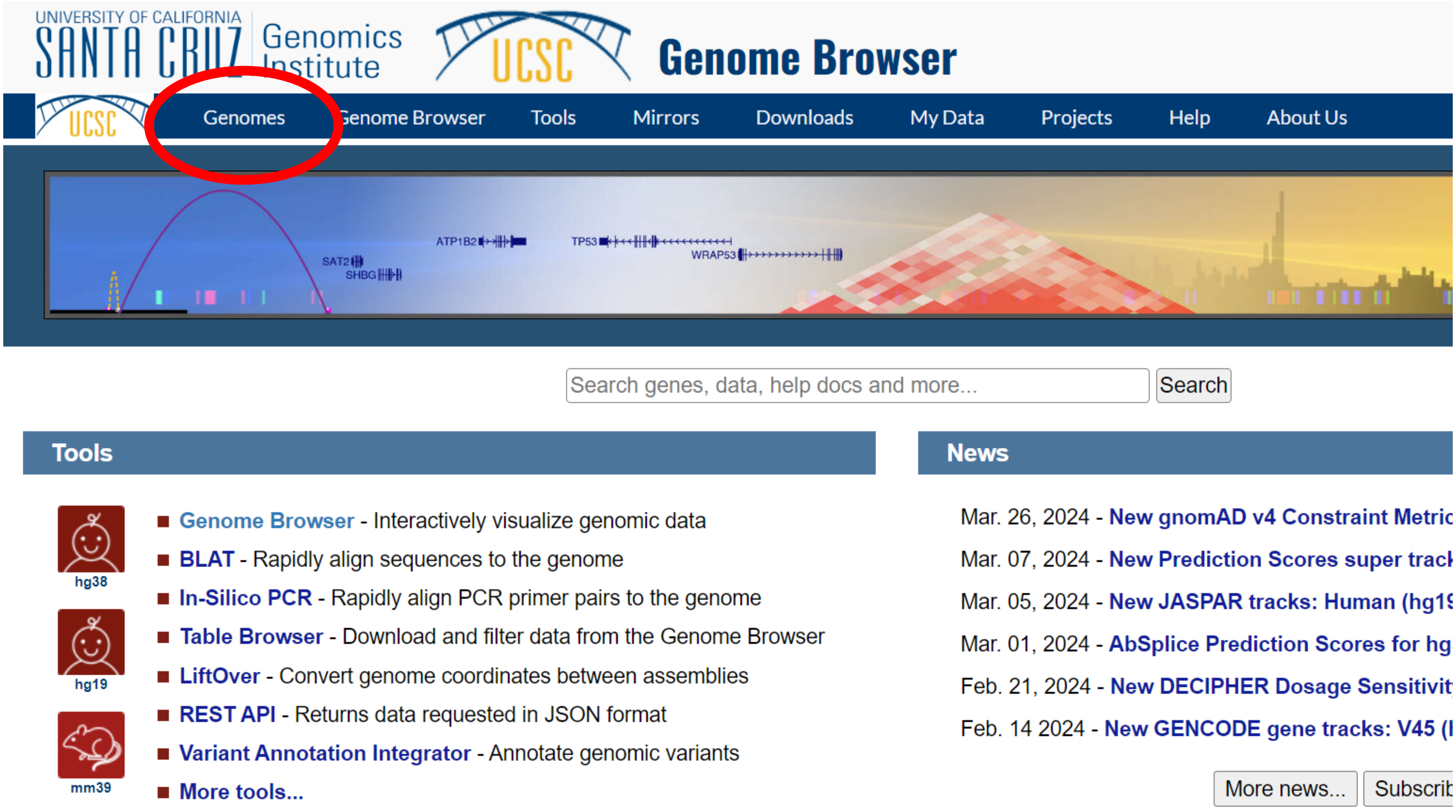


Есть и маркеры клеточных типов, собранные по нескольким scRNA-seq публикациям.

На этом же сайте можно скачать ПО для GSEA анализа.

Так же можно использовать веб-сервис https://www.genepattern.org/

https://genome.ucsc.edu/



Tools

- **Genome Browser** - Interactively visualize genomic data
- **BLAT** - Rapidly align sequences to the genome
- **In-Silico PCR** - Rapidly align PCR primer pairs to the genome
- **Table Browser** - Download and filter data from the Genome Browser
- **LiftOver** - Convert genome coordinates between assemblies
- **REST API** - Returns data requested in JSON format
- **Variant Annotation Integrator** - Annotate genomic variants
- **More tools...**

News

Mar. 26, 2024 - **New gnomAD v4 Constraint Metric**

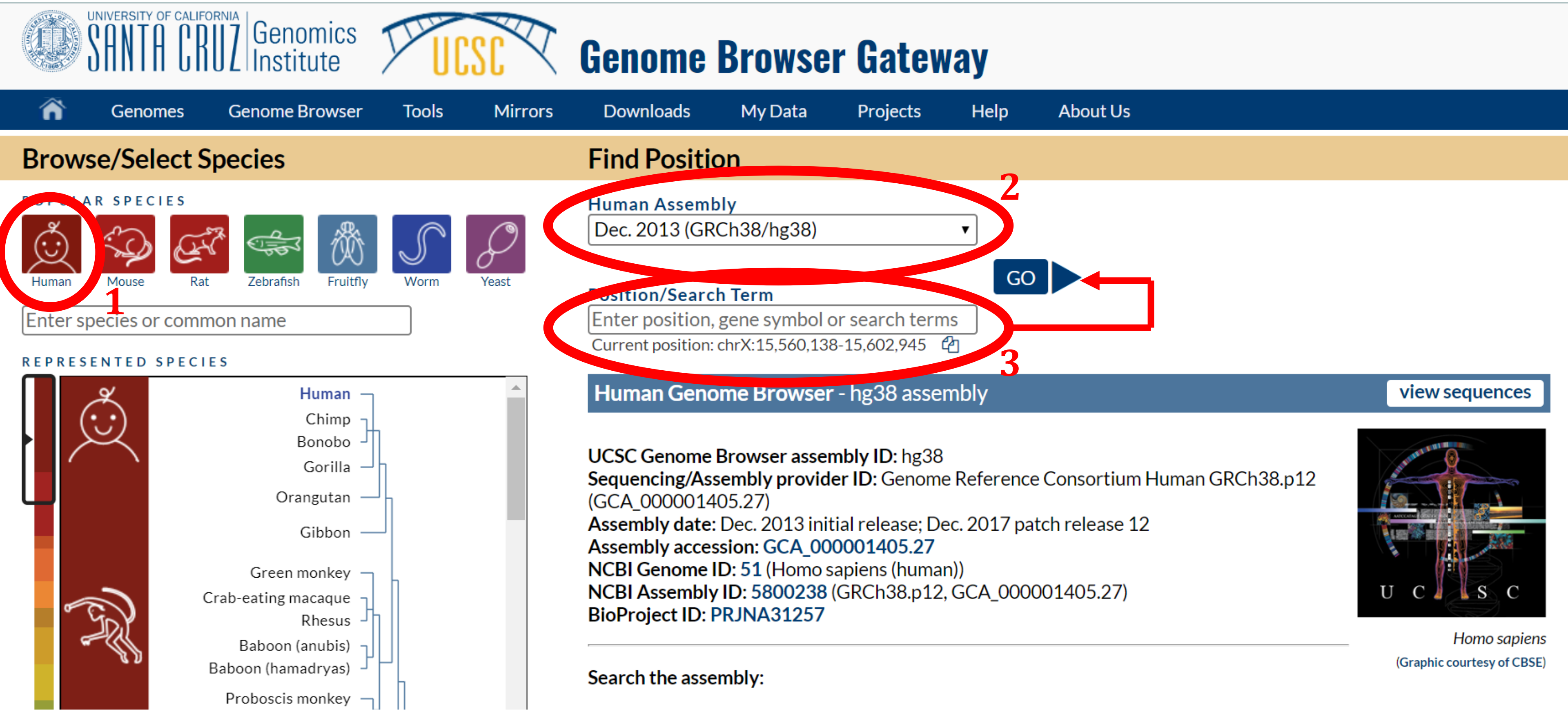
Mar. 07, 2024 - **New Prediction Scores super track**

Mar. 05, 2024 - **New JASPAR tracks: Human (hg1**

Mar. 01, 2024 - **AbSplice Prediction Scores for hg**

Feb. 21, 2024 - **New DECIPHER Dosage Sensitivit**

Feb. 14 2024 - **New GENCODE gene tracks: V45 (**



Enter species or common name

Human Assembly: Dec. 2013 (GRCh38/hg38)

Position/Search Term: Enter position, gene symbol or search terms

Current position: chrX:15,560,138-15,602,945

Human Genome Browser - hg38 assembly

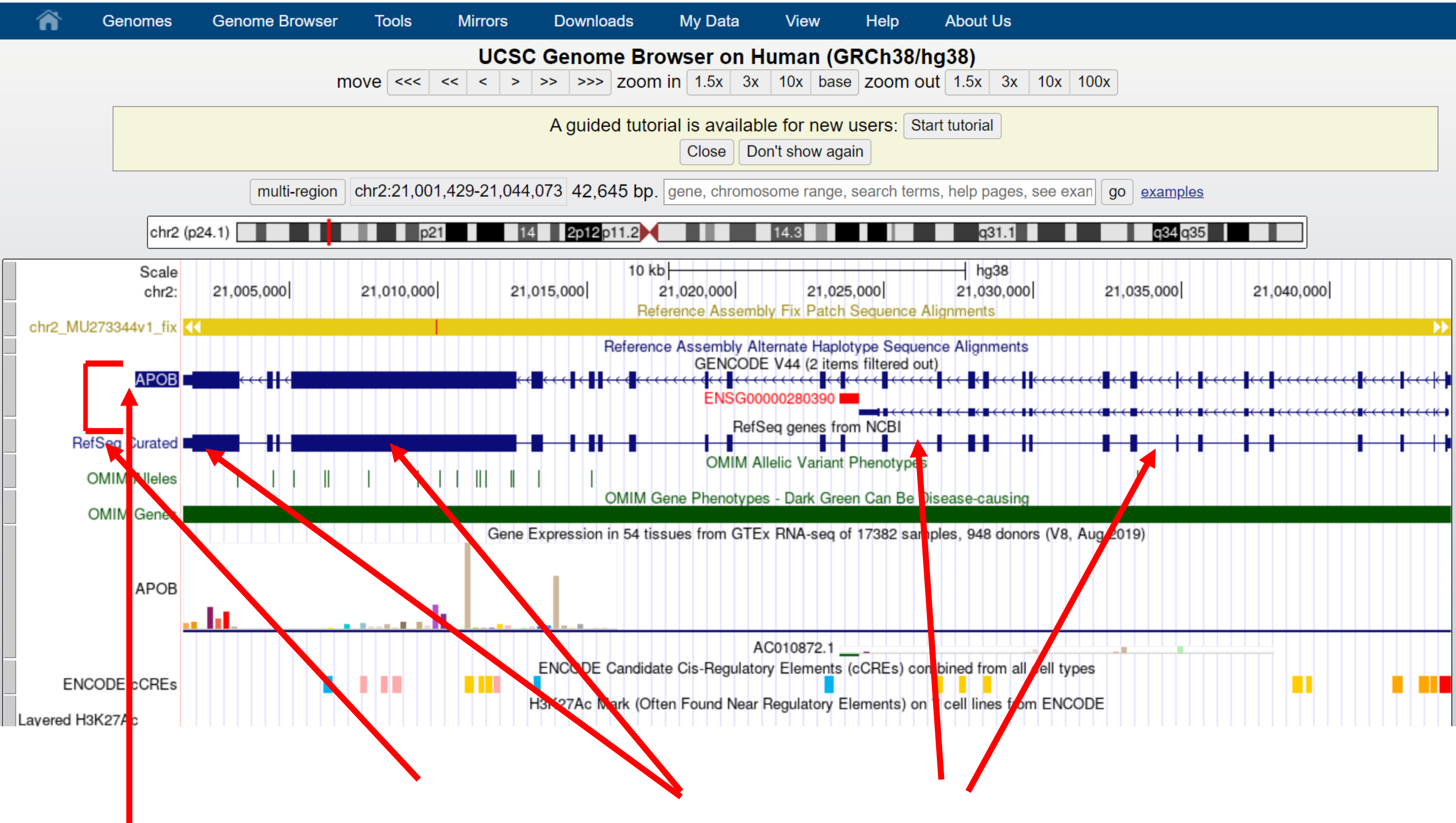
UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p12 (GCA_000001405.27)
Assembly date: Dec. 2013 initial release; Dec. 2017 patch release 12
Assembly accession: GCA_000001405.27
NCBI Genome ID: 51 (Homo sapiens (human))
NCBI Assembly ID: 5800238 (GRCh38.p12, GCA_000001405.27)
BioProject ID: PRJNA31257

Search the assembly:

1) Выбираем человека – Human

2) Версия генома человека – GRCh38/hg38

3) В раздел Position/Search Term вводим название своего гена или координаты локуса



UCSC Genome Browser on Human (GRCh38/hg38)

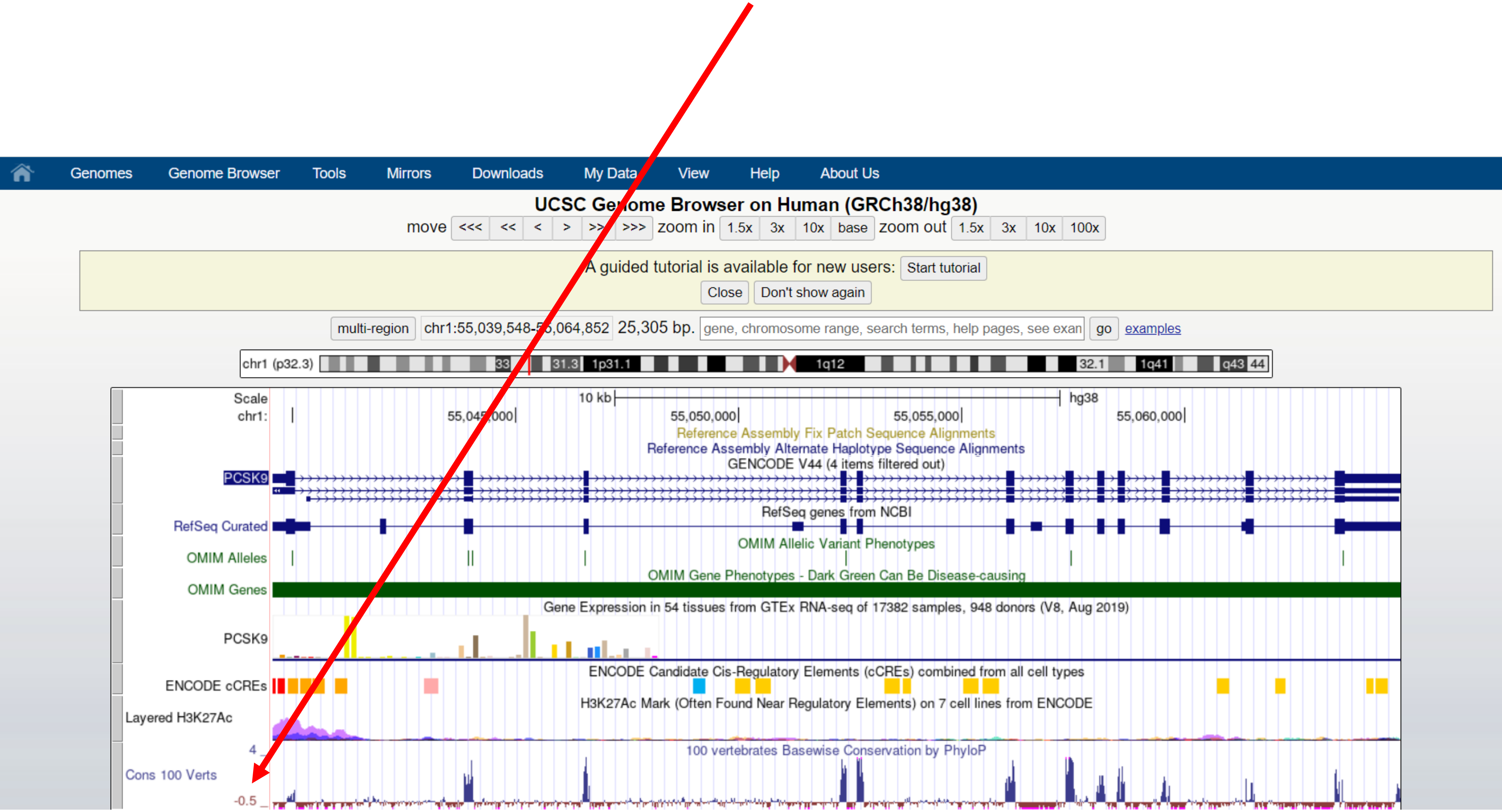
A guided tutorial is available for new users: Start tutorial

chr2:21,001,429-21,044,073 42,645 bp.

Ген Экзоны Интроны

Нажимаем на название гена и попадаем на страницу его подробного описания

Обратите внимание на трек консервативности



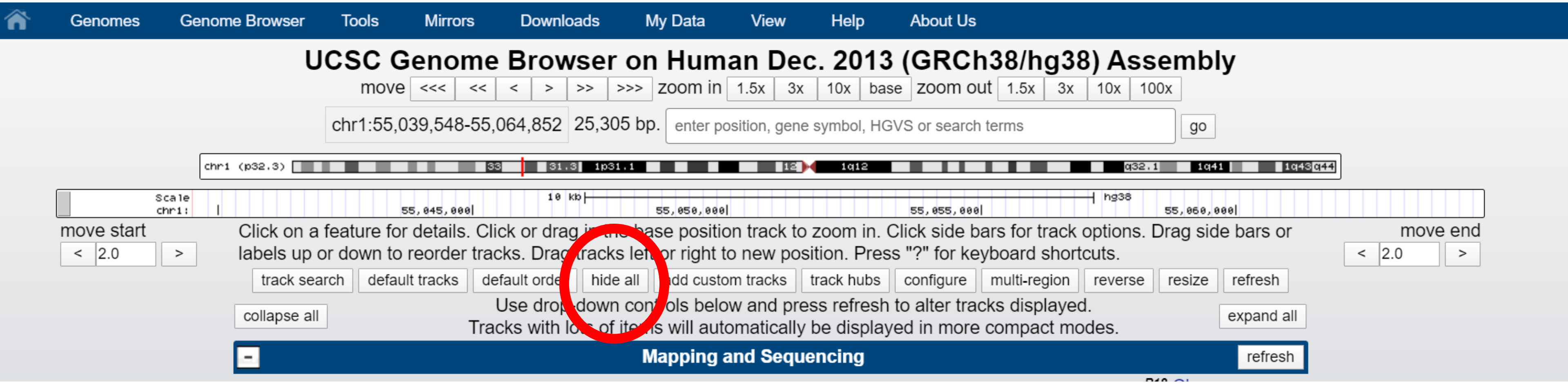
UCSC Genome Browser on Human (GRCh38/hg38)

A guided tutorial is available for new users:

chr1:55,039,548-55,064,852 25,305 bp.

gene, chromosome range, search terms, help pages, see exan

Можно спрятать все треки



UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr1:55,039,548-55,064,852 25,305 bp.

enter position, gene symbol, HGVS or search terms

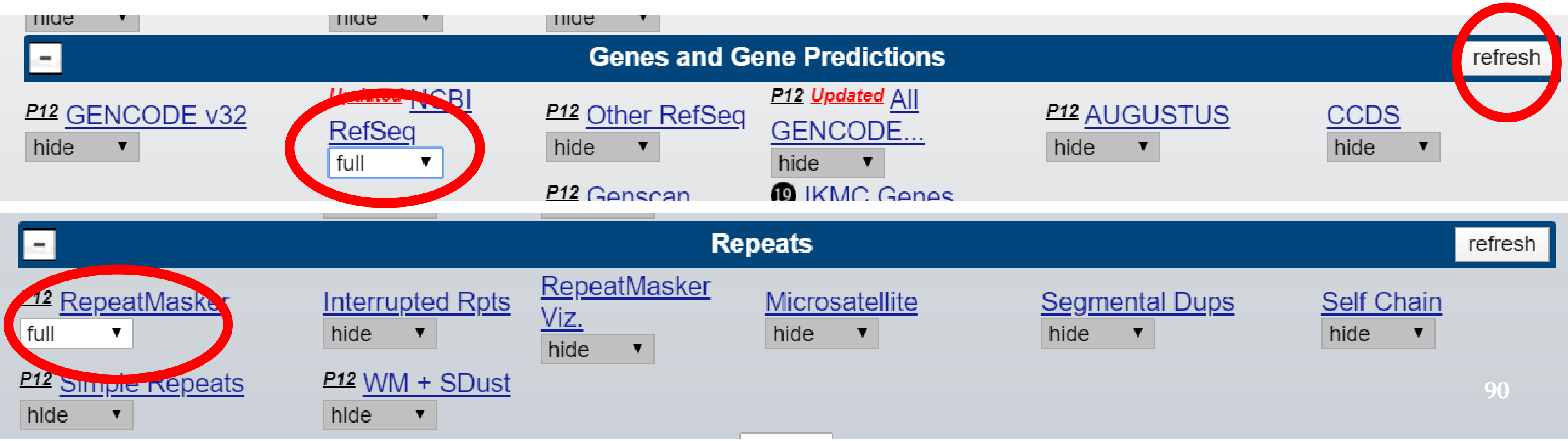
move start: 2.0

move end: 2.0

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Открыть только некоторые



P12 GENCODE v32: hide

Updated NCBI RefSeq: full

P12 Other RefSeq: hide

P12 Updated All GENCODE...: hide

P12 AUGUSTUS: hide

CCDS: hide

P12 Genscan

IKMC Genes

P12 RepeatMasker: full

Interrupted Rpts: hide

RepeatMasker Viz.: hide

Microsatellite: hide

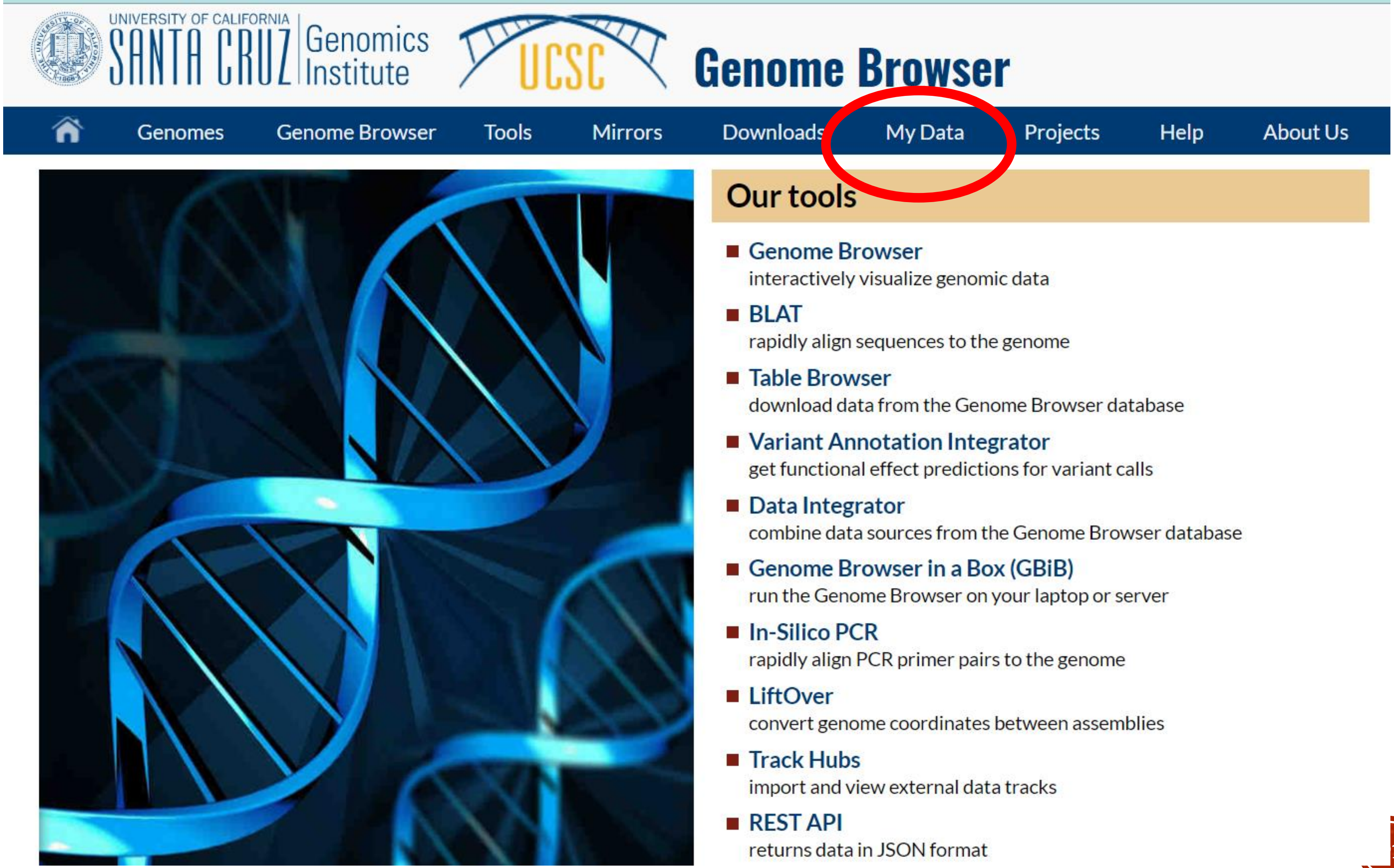
Segmental Dups: hide

Self Chain: hide

P12 Simple Repeats: hide

P12 WM + SDust: hide

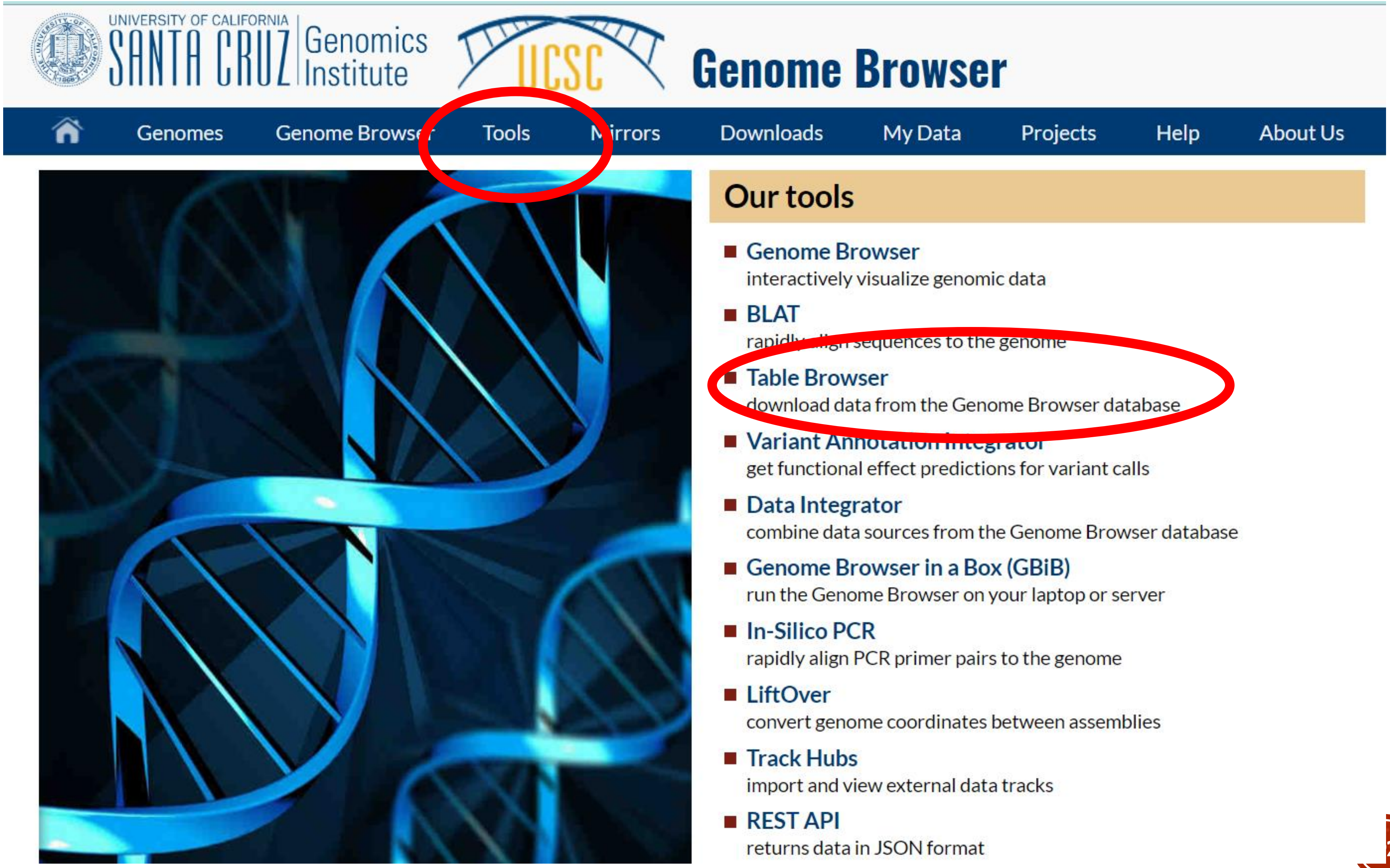
https://genome.ucsc.edu/



Our tools

- **Genome Browser**
  interactively visualize genomic data
- **BLAT**
  rapidly align sequences to the genome
- **Table Browser**
  download data from the Genome Browser database
- **Variant Annotation Integrator**
  get functional effect predictions for variant calls
- **Data Integrator**
  combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**
  run the Genome Browser on your laptop or server
- **In-Silico PCR**
  rapidly align PCR primer pairs to the genome
- **LiftOver**
  convert genome coordinates between assemblies
- **Track Hubs**
  import and view external data tracks
- **REST API**
  returns data in JSON format

https://genome.ucsc.edu/



Our tools

- **Genome Browser**
  interactively visualize genomic data
- **BLAT**
  rapidly align sequences to the genome
- **Table Browser**
  download data from the Genome Browser database
- **Variant Annotation Integrator**
  get functional effect predictions for variant calls
- **Data Integrator**
  combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**
  run the Genome Browser on your laptop or server
- **In-Silico PCR**
  rapidly align PCR primer pairs to the genome
- **LiftOver**
  convert genome coordinates between assemblies
- **Track Hubs**
  import and view external data tracks
- **REST API**
  returns data in JSON format

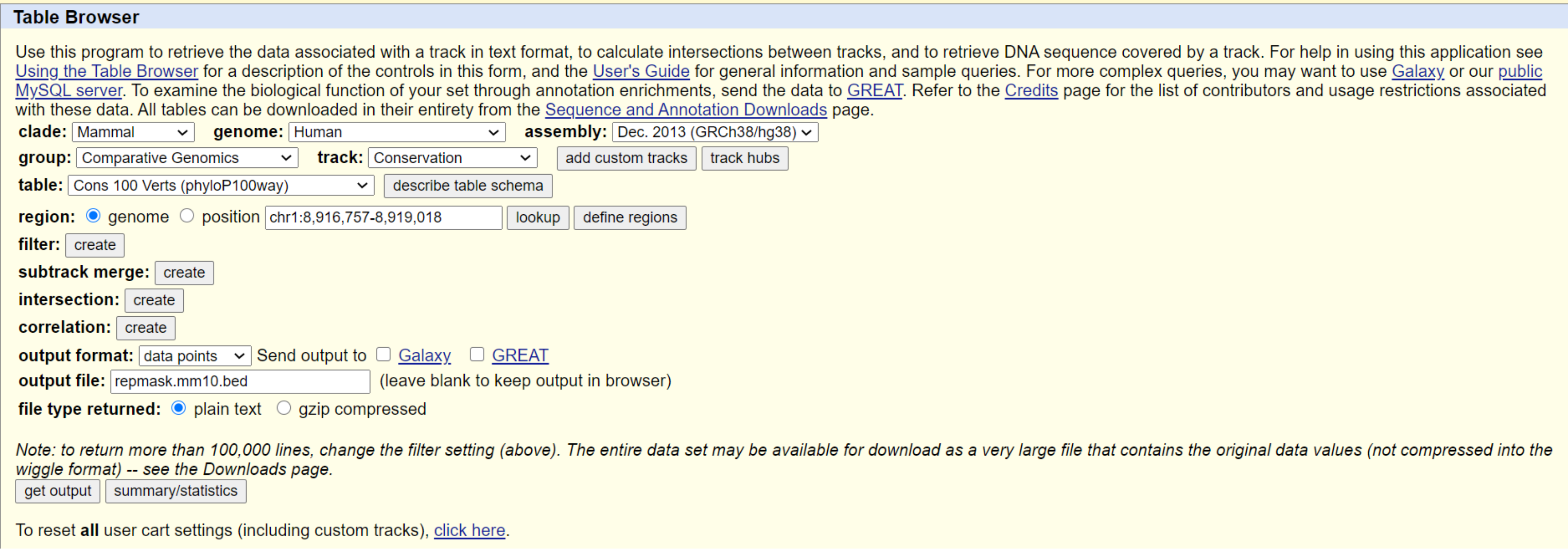


Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, and the User's Guide for general information and sample queries. For more complex queries, you may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page.

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)

group: Comparative Genomics **track:** Conservation add custom tracks track hubs

table: Cons 100 Verts (phyloP100way) describe table schema

region: ● genome ○ position chr1:8,916,757-8,919,018 lookup define regions

filter: create

subtrack merge: create

intersection: create

correlation: create

output format: data points Send output to ☐ Galaxy ☐ GREAT

output file: repmask.mm10.bed (leave blank to keep output in browser)

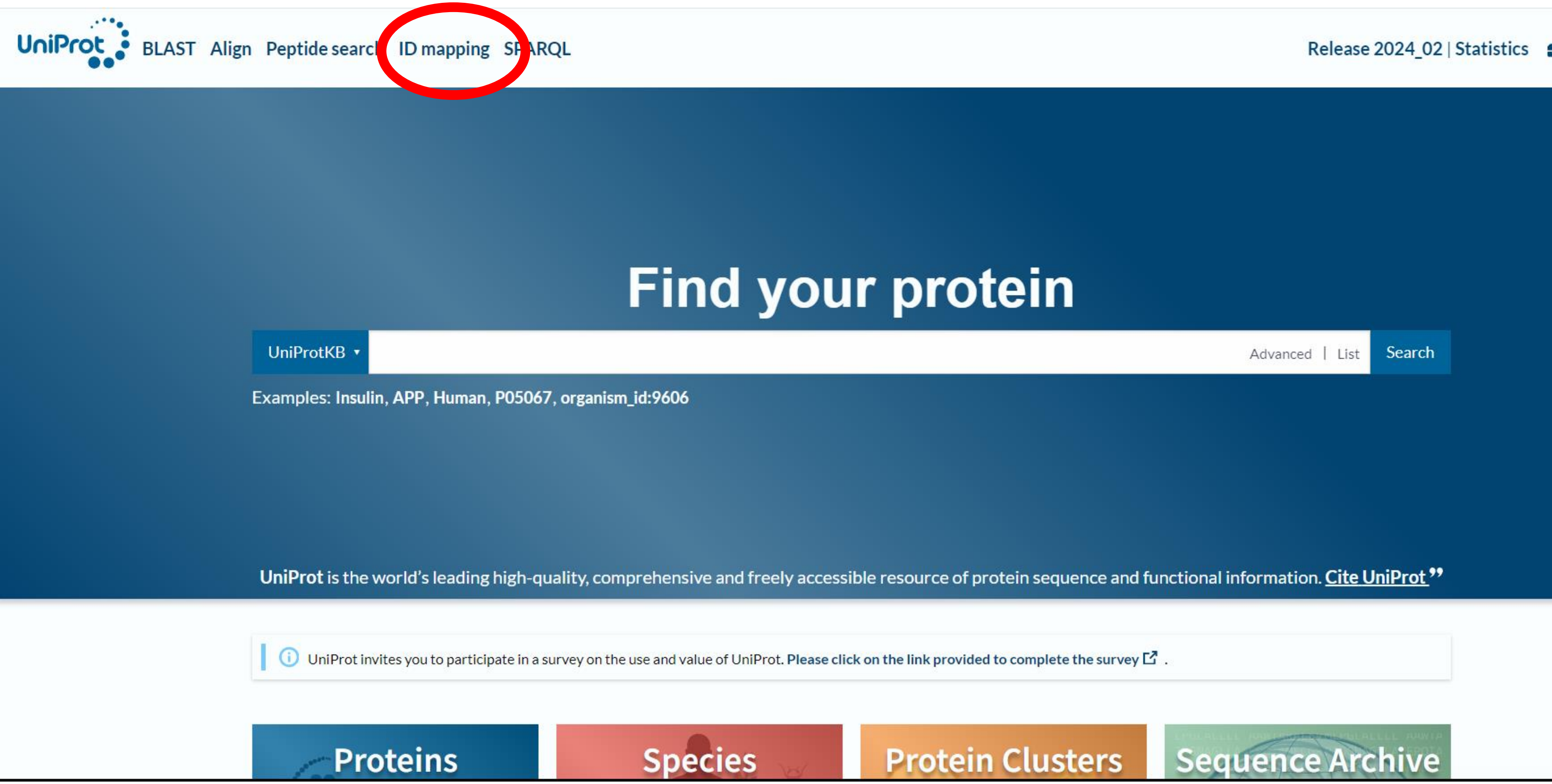
file type returned: ● plain text ○ gzip compressed

Note: to return more than 100,000 lines, change the filter setting (above). The entire data set may be available for download as a very large file that contains the original data values (not compressed into the wiggle format) -- see the Downloads page.

get output summary/statistics

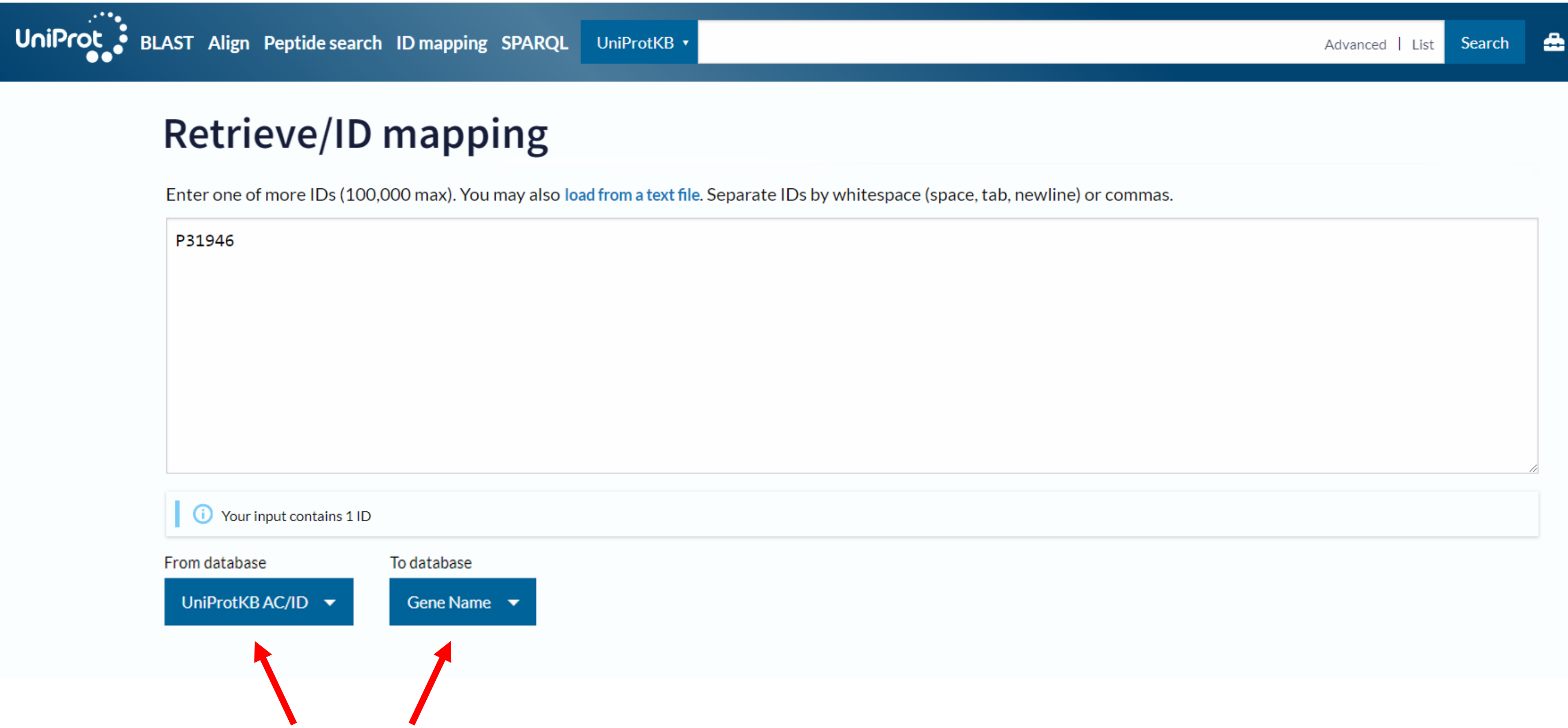
To reset **all** user cart settings (including custom tracks), click here.

https://www.uniprot.org/



UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. Cite UniProt

UniProt invites you to participate in a survey on the use and value of UniProt. Please click on the link provided to complete the survey .



Retrieve/ID mapping

Enter one of more IDs (100,000 max). You may also load from a text file. Separate IDs by whitespace (space, tab, newline) or commas.

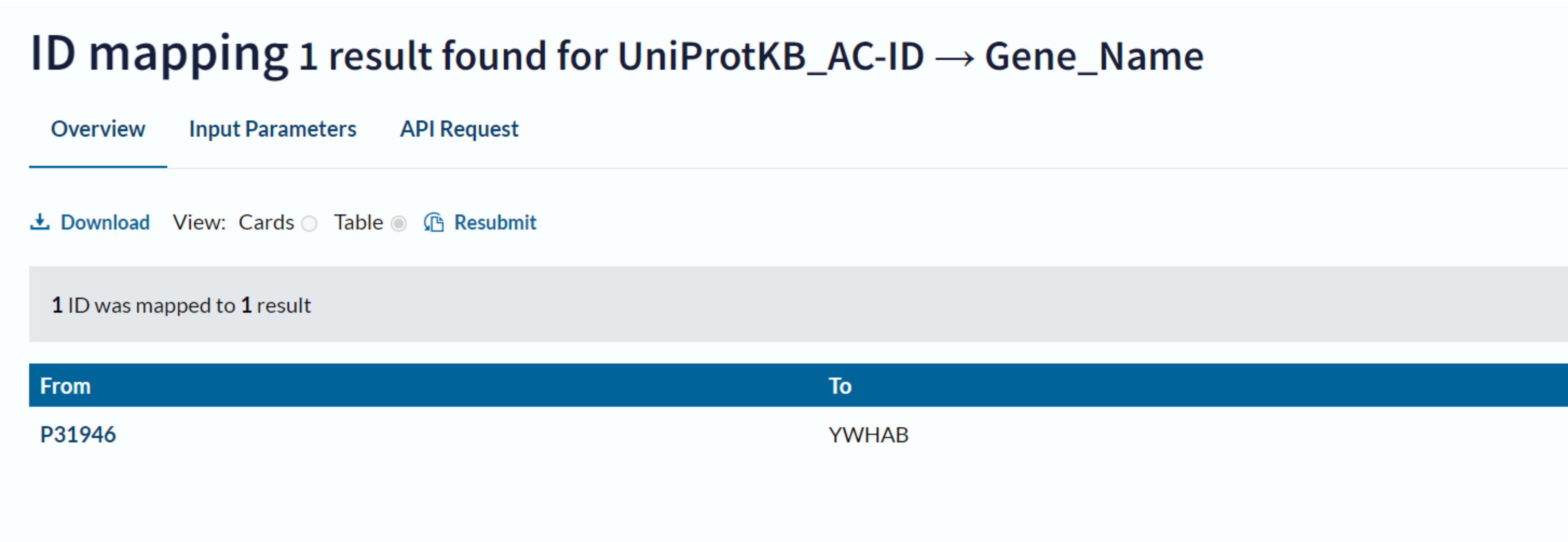
```
P31946
```

Your input contains 1 ID

From database: UniProtKB AC/ID

To database: Gene Name

Указываем базы данных



ID mapping 1 result found for UniProtKB_AC-ID → Gene_Name

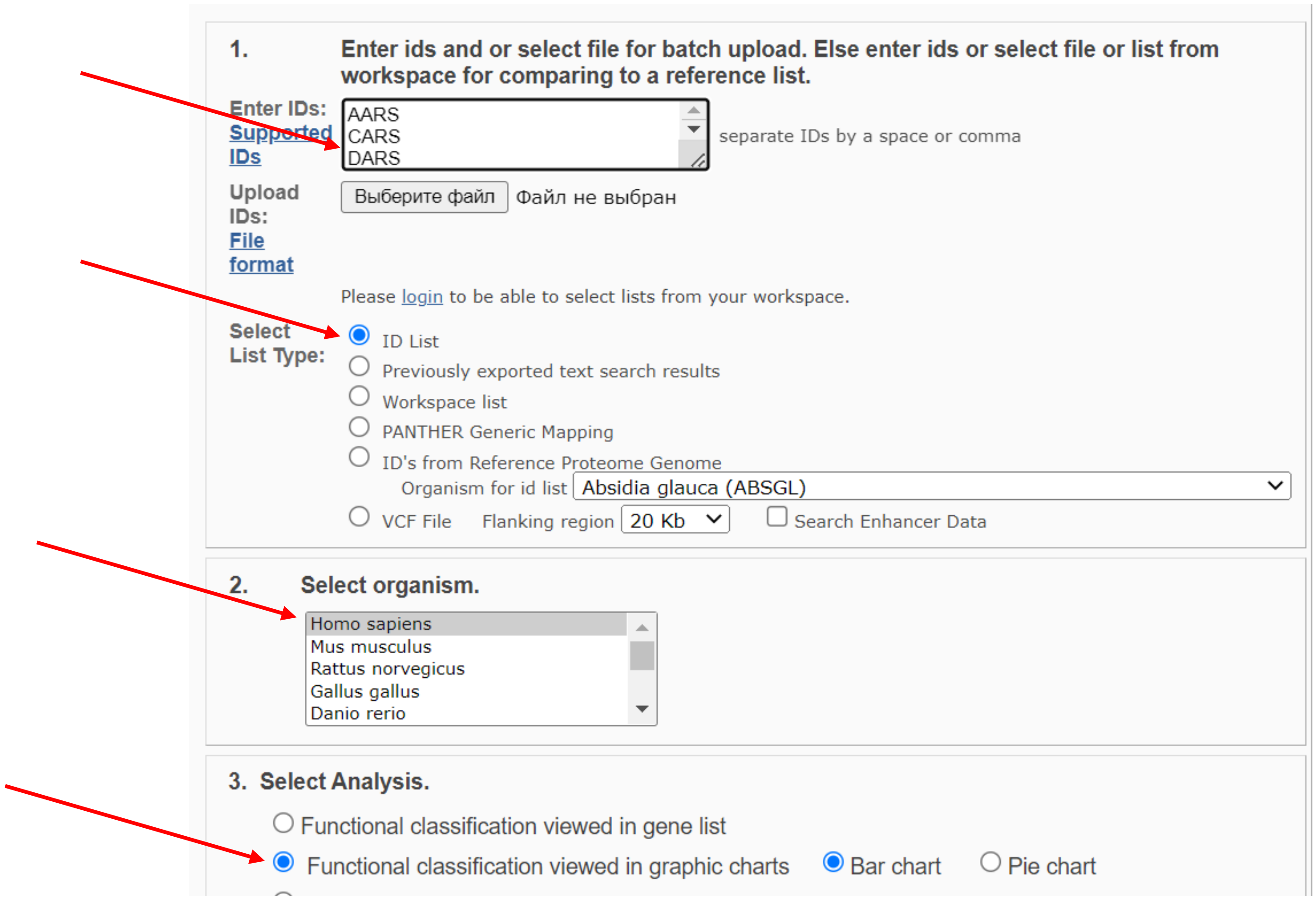
View: Cards ○ Table ●

1 ID was mapped to 1 result

From	To
P31946	YWHAB

PANTHER
Classification System

http://www.pantherdb.org/



1. **Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.**

Enter IDs: Supported IDs

AARS
CARS
DARS

separate IDs by a space or comma

Upload IDs: File format

Выберите файл Файл не выбран

Please login to be able to select lists from your workspace.

Select List Type:

- ● ID List
- ○ Previously exported text search results
- ○ Workspace list
- ○ PANTHER Generic Mapping
- ○ ID's from Reference Proteome Genome
  - Organism for id list: Absidia glauca (ABSGL)
- ○ VCF File Flanking region 20 Kb ☐ Search Enhancer Data

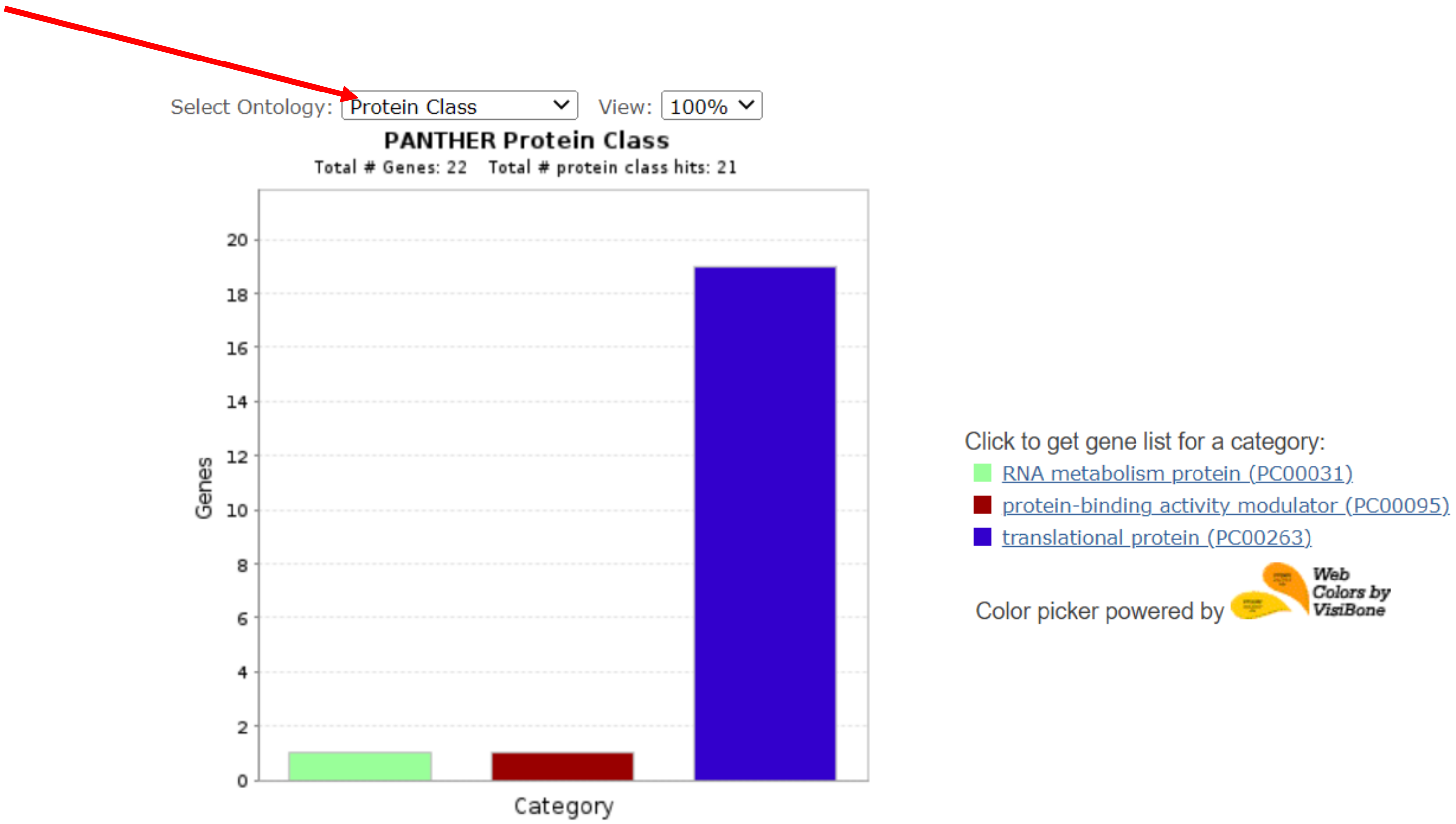
2. **Select organism.**

Homo sapiens
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio

3. **Select Analysis.**

- ○ Functional classification viewed in gene list
- ● Functional classification viewed in graphic charts ● Bar chart ○ Pie chart

http://www.pantherdb.org/



Select Ontology: Protein Class View: 100%

PANTHER Protein Class

Total # Genes: 22 Total # protein class hits: 21

Click to get gene list for a category:

- 🟩 RNA metabolism protein (PC00031)
- 🟥 protein-binding activity modulator (PC00095)
- 🟦 translational protein (PC00263)

Color picker powered by Web Colors by VisiBone