Gene Ontology

Дано: набор белков или генов

Дано: набор белков или генов

Откуда их взять?

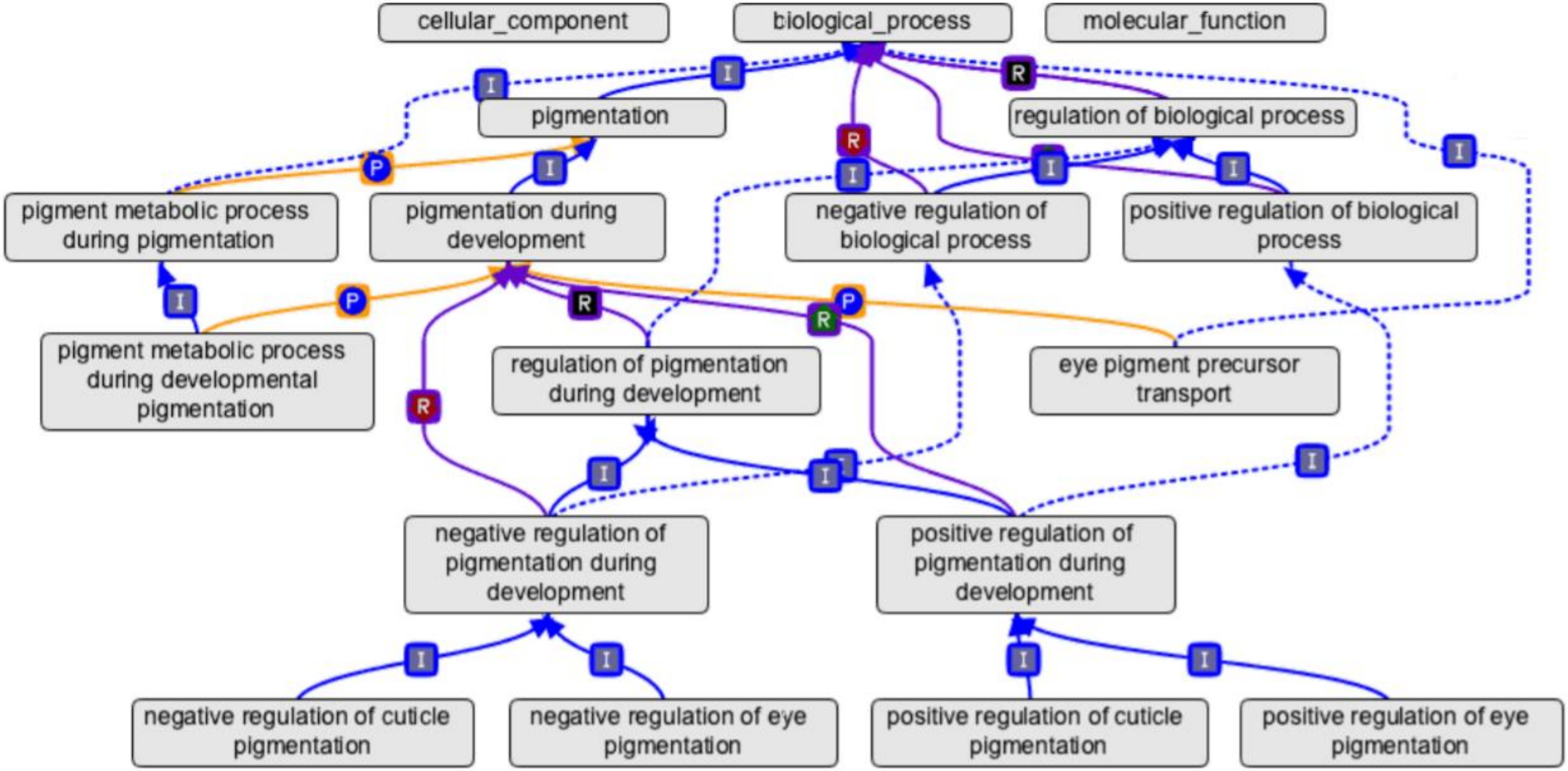
Экспрессия
Дифференциальная экспрессия
Масспектрометрия
Любимый набор
…

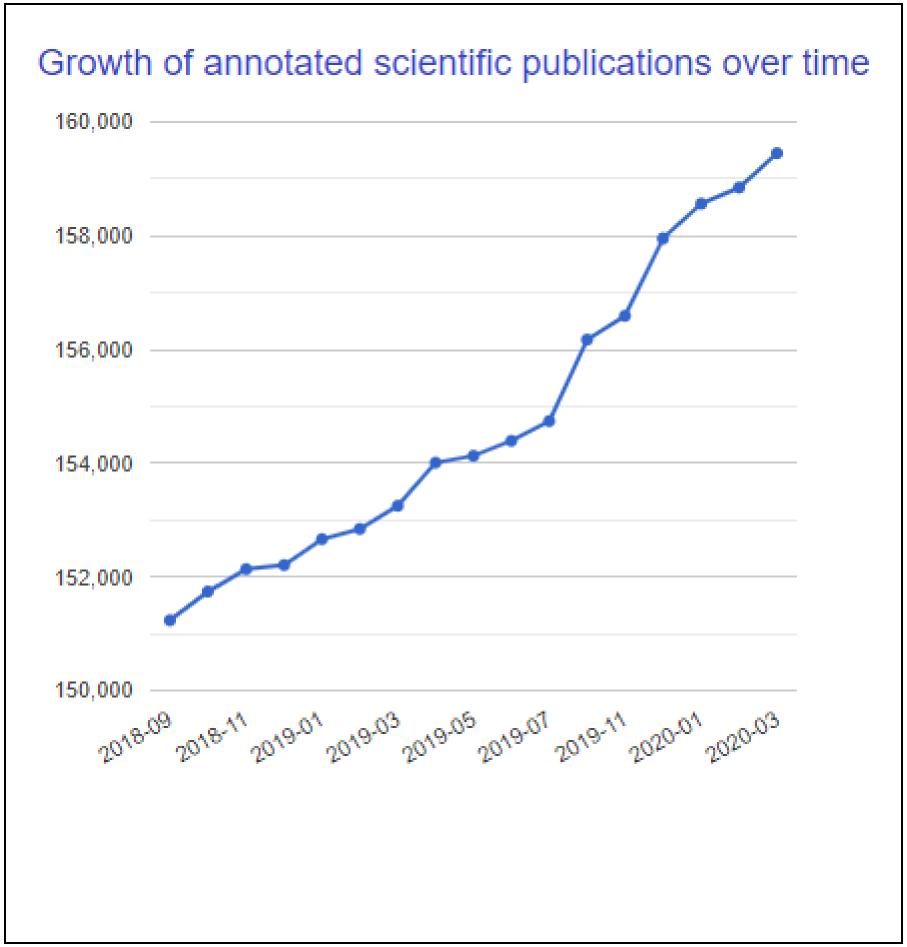
Что дальше делать?

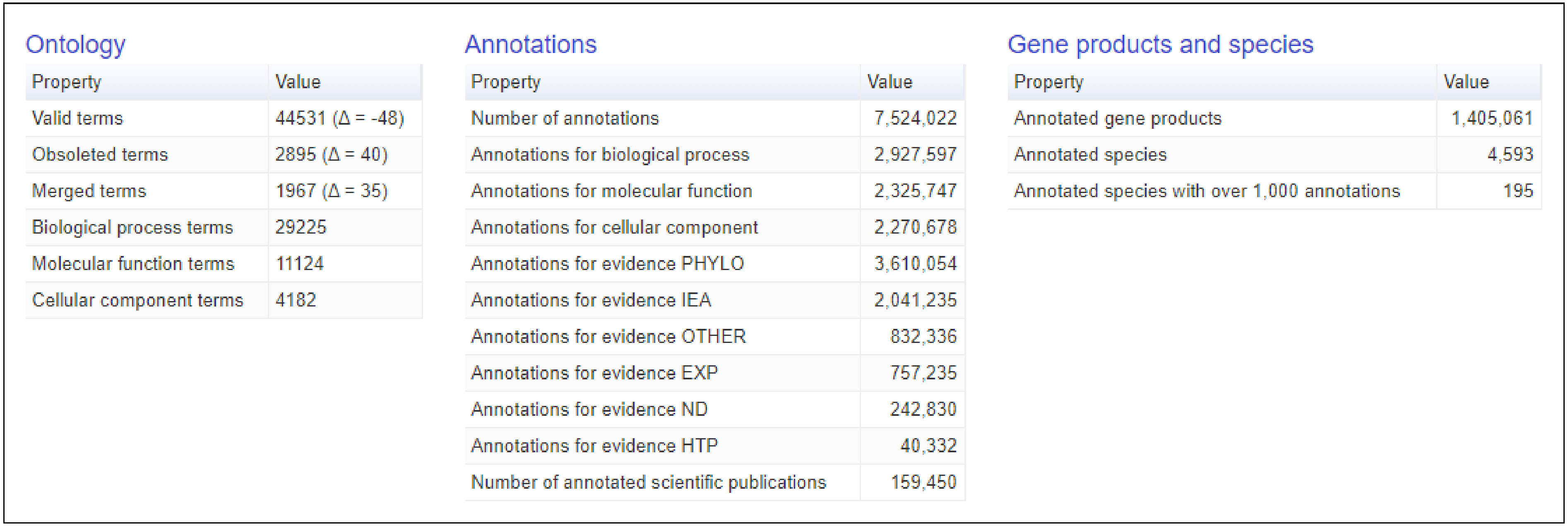
Gene Ontology – GO

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
I
P
R


Growth of annotated scientific publications over time
160,000
158,000
156,000
154,000
152,000
150,000
2018-09
2018-11
2019-01
2019-03
2019-05
2019-07
2019-11
2020-01
2020-03



Ontology

Property	Value
Valid terms	44531 (Δ = -48)
Obsoleted terms	2895 (Δ = 40)
Merged terms	1967 (Δ = 35)
Biological process terms	29225
Molecular function terms	11124
Cellular component terms	4182

Annotations

Property	Value
Number of annotations	7,524,022
Annotations for biological process	2,927,597
Annotations for molecular function	2,325,747
Annotations for cellular component	2,270,678
Annotations for evidence PHYLO	3,610,054
Annotations for evidence IEA	2,041,235
Annotations for evidence OTHER	832,336
Annotations for evidence EXP	757,235
Annotations for evidence ND	242,830
Annotations for evidence HTP	40,332
Number of annotated scientific publications	159,450

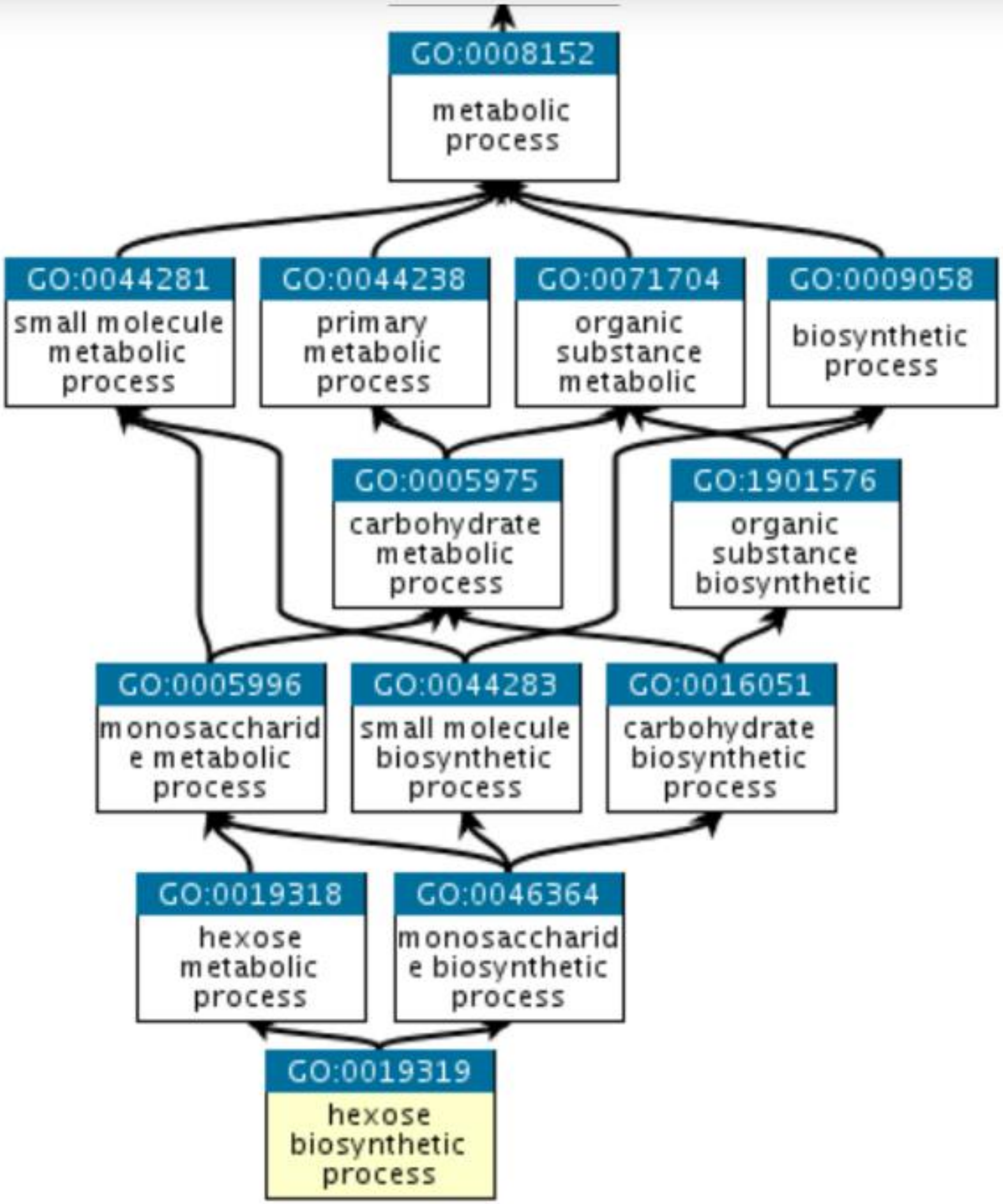
Gene products and species

Property	Value
Annotated gene products	1,405,061
Annotated species	4,593
Annotated species with over 1,000 annotations	195

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



GO:0008152
metabolic process

GO:0044281
small molecule metabolic process

GO:0044238
primary metabolic process

GO:0071704
organic substance metabolic

GO:0009058
biosynthetic process

GO:0005975
carbohydrate metabolic process

GO:1901576
organic substance biosynthetic

GO:0005996
monosaccharide metabolic process

GO:0044283
small molecule biosynthetic process

GO:0016051
carbohydrate biosynthetic process

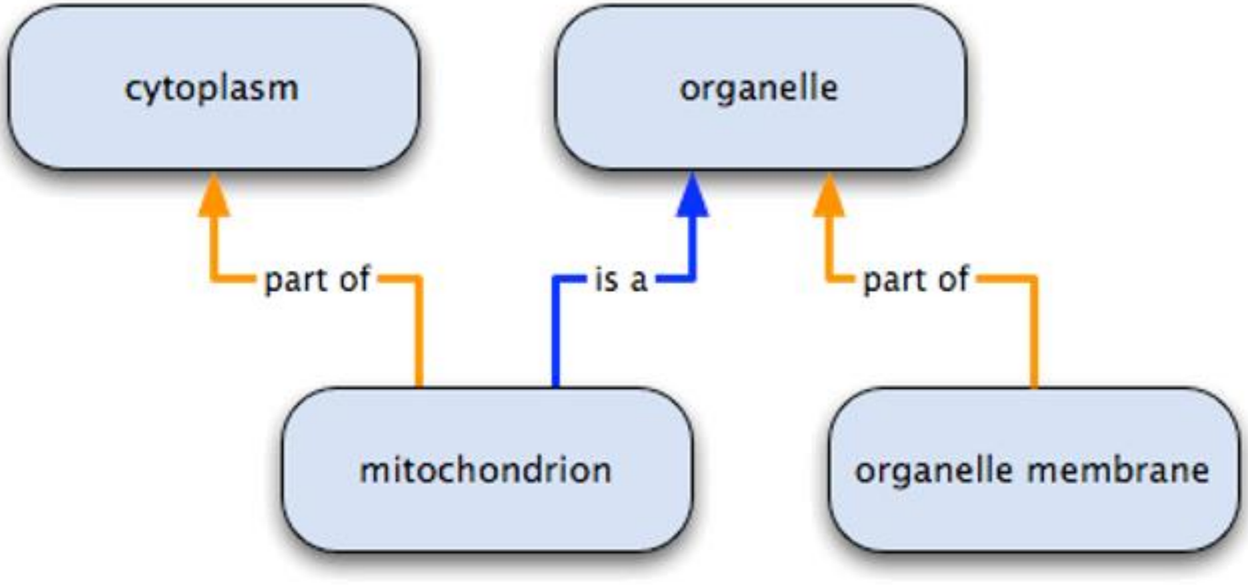
GO:0019318
hexose metabolic process

GO:0046364
monosaccharide biosynthetic process

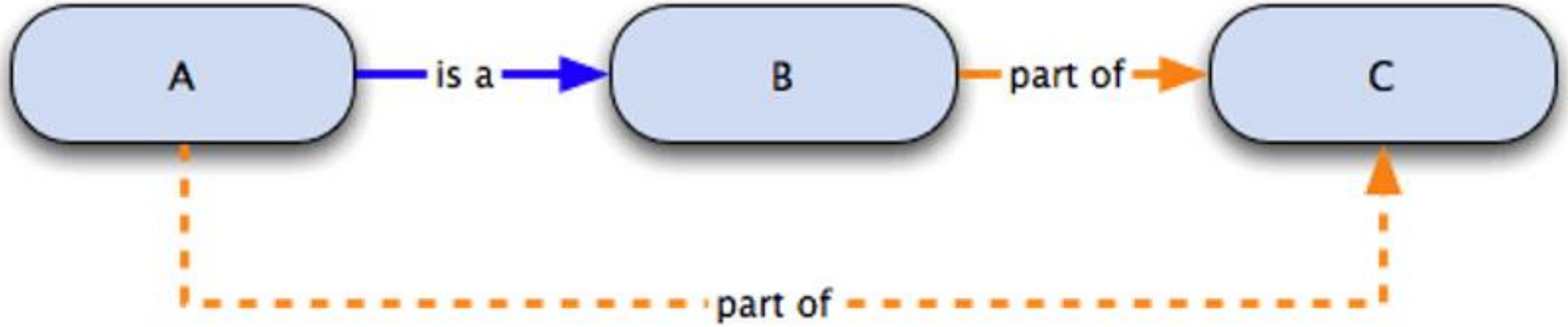
GO:0019319
hexose biosynthetic process

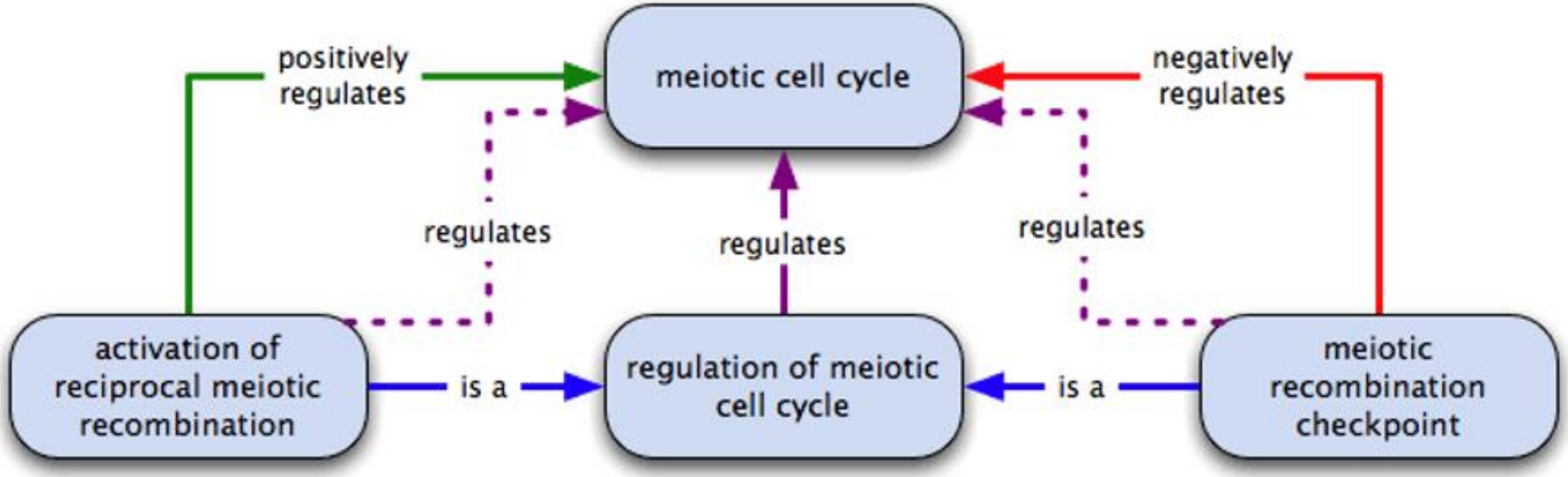
Термин GO (GO term) – узел графа – может иметь любое количество связей с любыми другими узлами

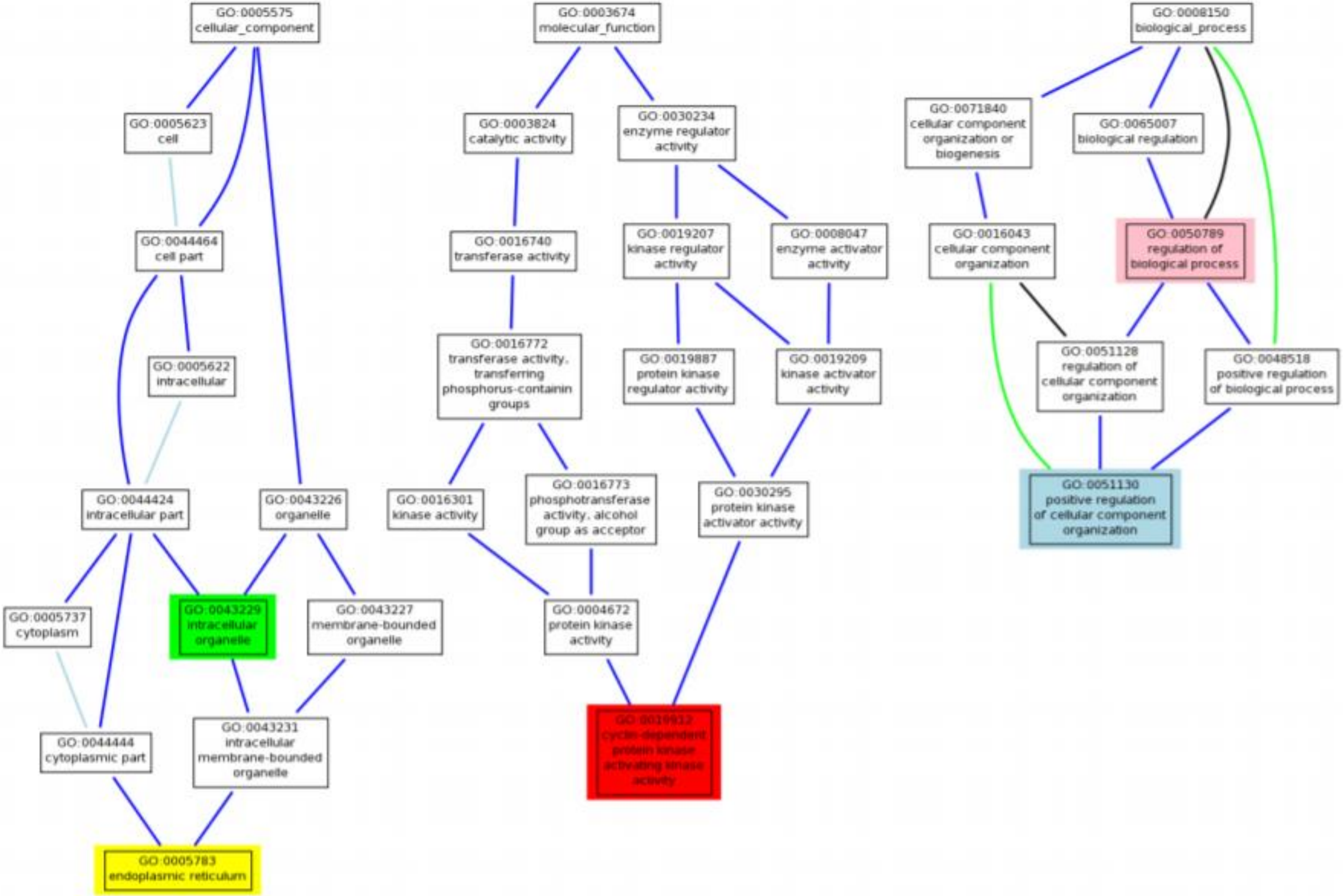
Отношения между узлами – ребра графа бывают разного типа

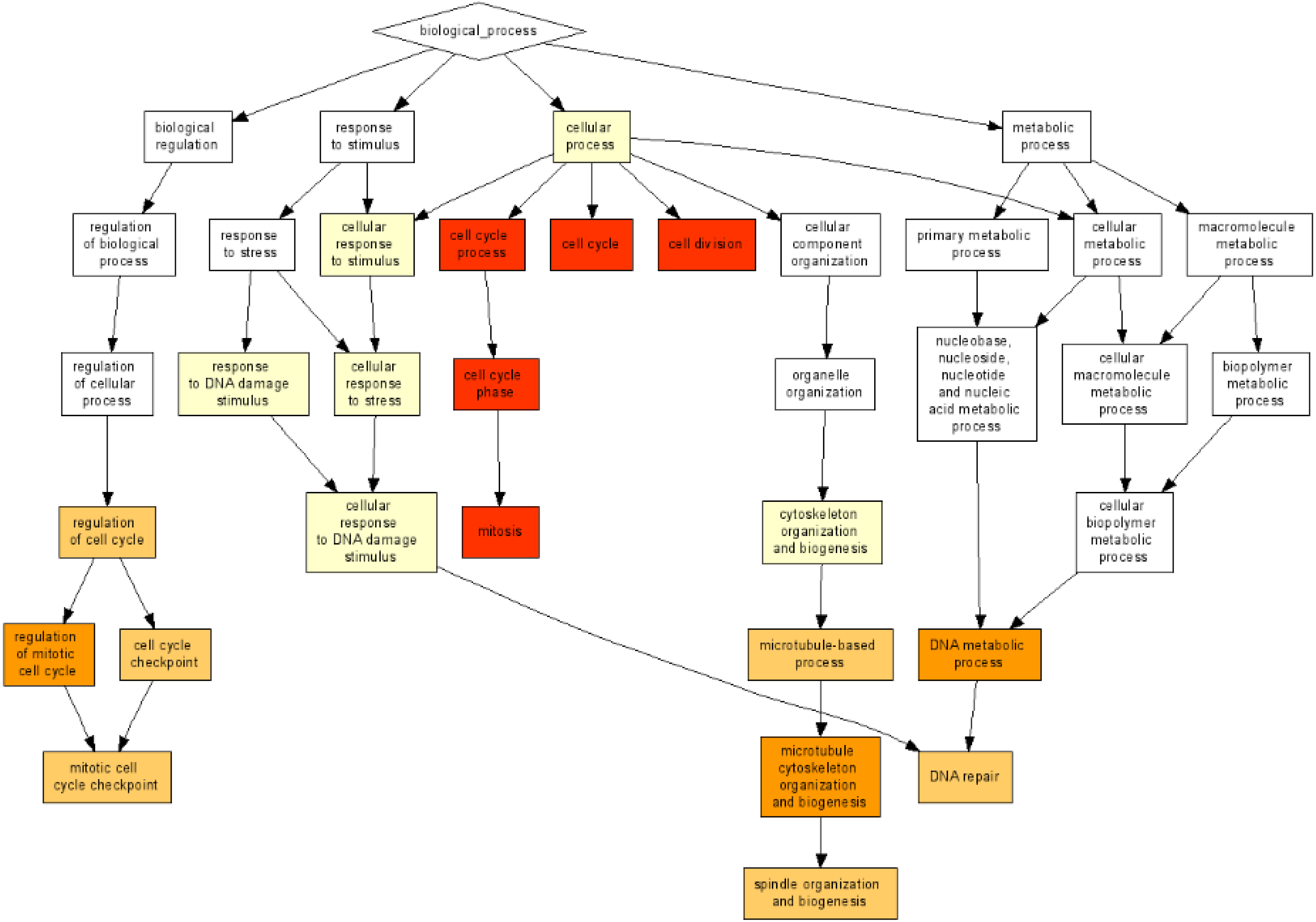


A *is a* B → Указанные
B *part of* C → непосредственно
Значит, A *part of* C ⇢ Логически вытекающее



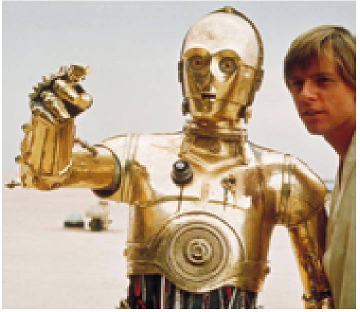

meiotic cell cycle
positively regulates
negatively regulates
regulates
regulates
regulates
activation of reciprocal meiotic recombination
is a
regulation of meiotic cell cycle
is a
meiotic recombination checkpoint


GO:0005575 cellular_component
GO:0003674 molecular_function
GO:0008150 biological_process
GO:0005623 cell
GO:0003824 catalytic activity
GO:0030234 enzyme regulator activity
GO:0071840 cellular component organization or biogenesis
GO:0065007 biological regulation
GO:0044464 cell part
GO:0016740 transferase activity
GO:0019207 kinase regulator activity
GO:0008047 enzyme activator activity
GO:0016043 cellular component organization
GO:0050789 regulation of biological process
GO:0005622 intracellular
GO:0016772 transferase activity, transferring phosphorus-containin groups
GO:0019887 protein kinase regulator activity
GO:0019209 kinase activator activity
GO:0051128 regulation of cellular component organization
GO:0048518 positive regulation of biological process
GO:0044424 intracellular part
GO:0043226 organelle
GO:0016301 kinase activity
GO:0016773 phosphotransferase activity, alcohol group as acceptor
GO:0030295 protein kinase activator activity
GO:0051130 positive regulation of cellular component organization
GO:0005737 cytoplasm
GO:0043229 intracellular organelle
GO:0043227 membrane-bounded organelle
GO:0004672 protein kinase activity
GO:0044444 cytoplasmic part
GO:0043231 intracellular membrane-bounded organelle
GO:0019912 cyclin-dependent protein kinase activating kinase activity
GO:0005783 endoplasmic reticulum


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

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**)

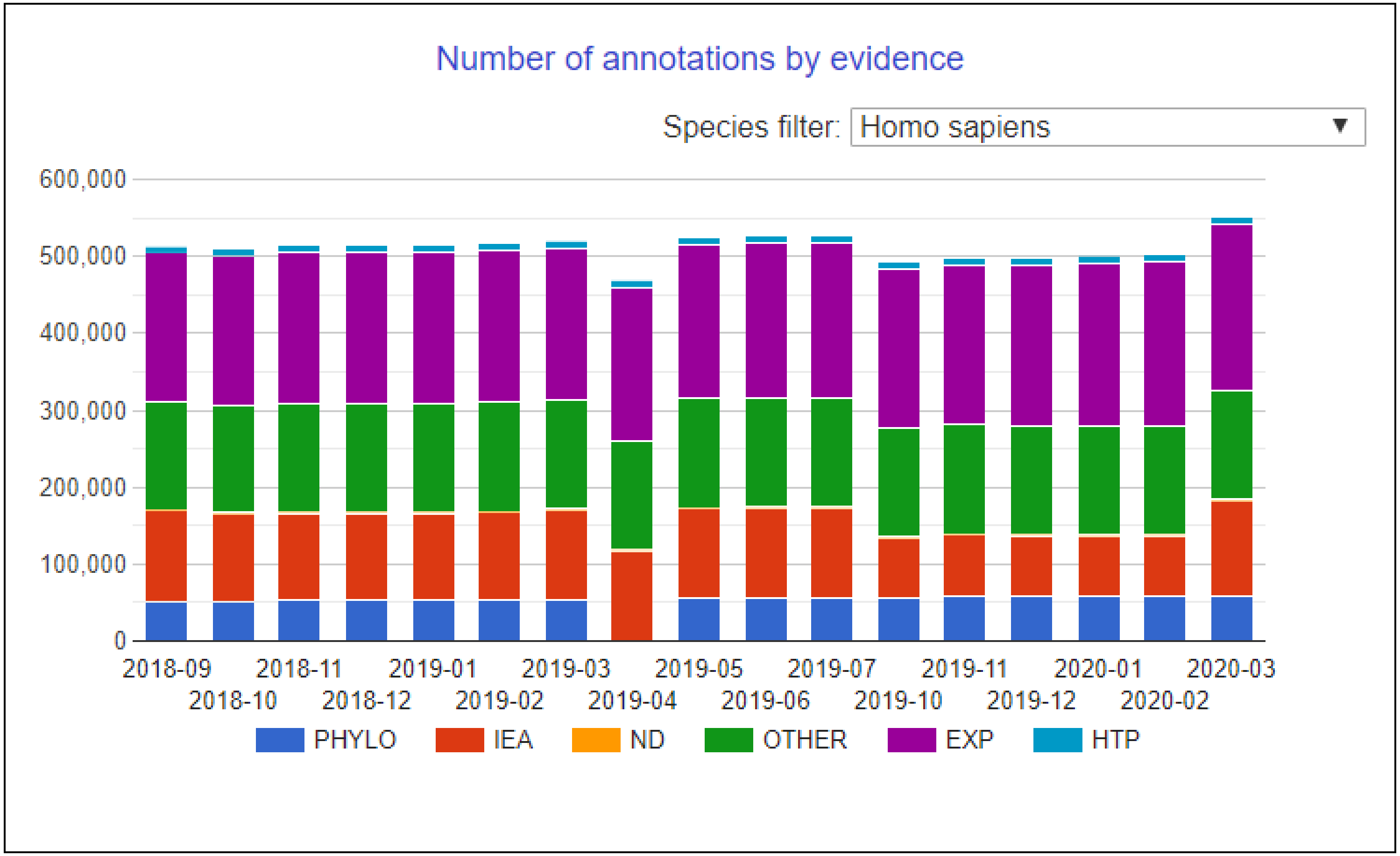
- Inferred from Direct Assay (**IDA**)
- Inferred from Physical Interaction (**IPI**)
- Inferred from Mutant Phenotype (**IMP**)
- Inferred from Genetic Interaction (**IGI**)
- Inferred from Expression Pattern (**IEP**)

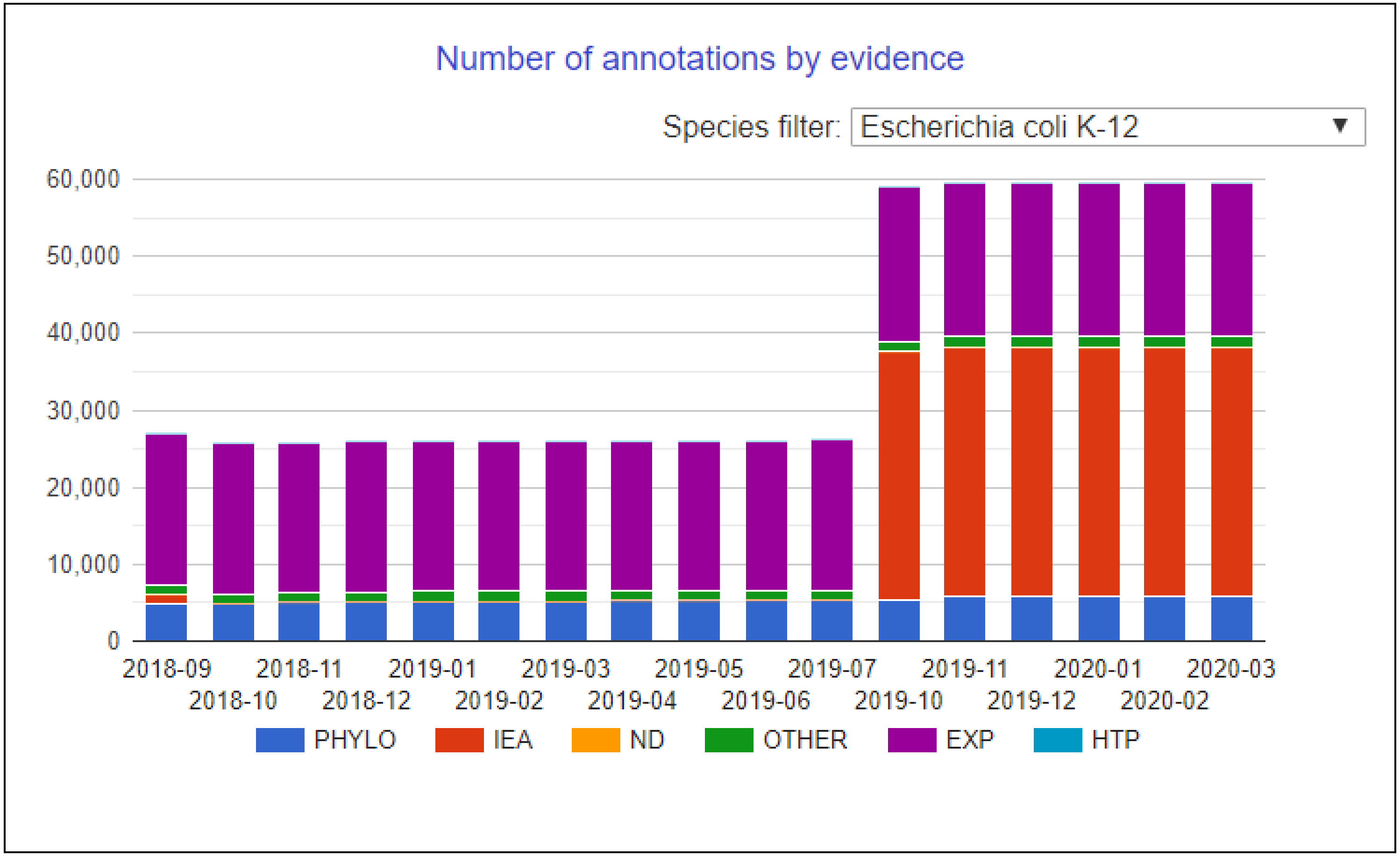
Traceable Author Statement (**TAS**)
Non-traceable Author Statement (**NAS**)

🟥 Плохо
🟨
🟩 Хорошо

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

☐	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 ▼
600,000
500,000
400,000
300,000
200,000
100,000
0
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07
2019-10
2019-11
2019-12
2020-01
2020-02
2020-03
PHYLO
IEA
ND
OTHER
EXP
HTP



Species filter: Escherichia coli K-12 ▼

GO Enrichment Analysis

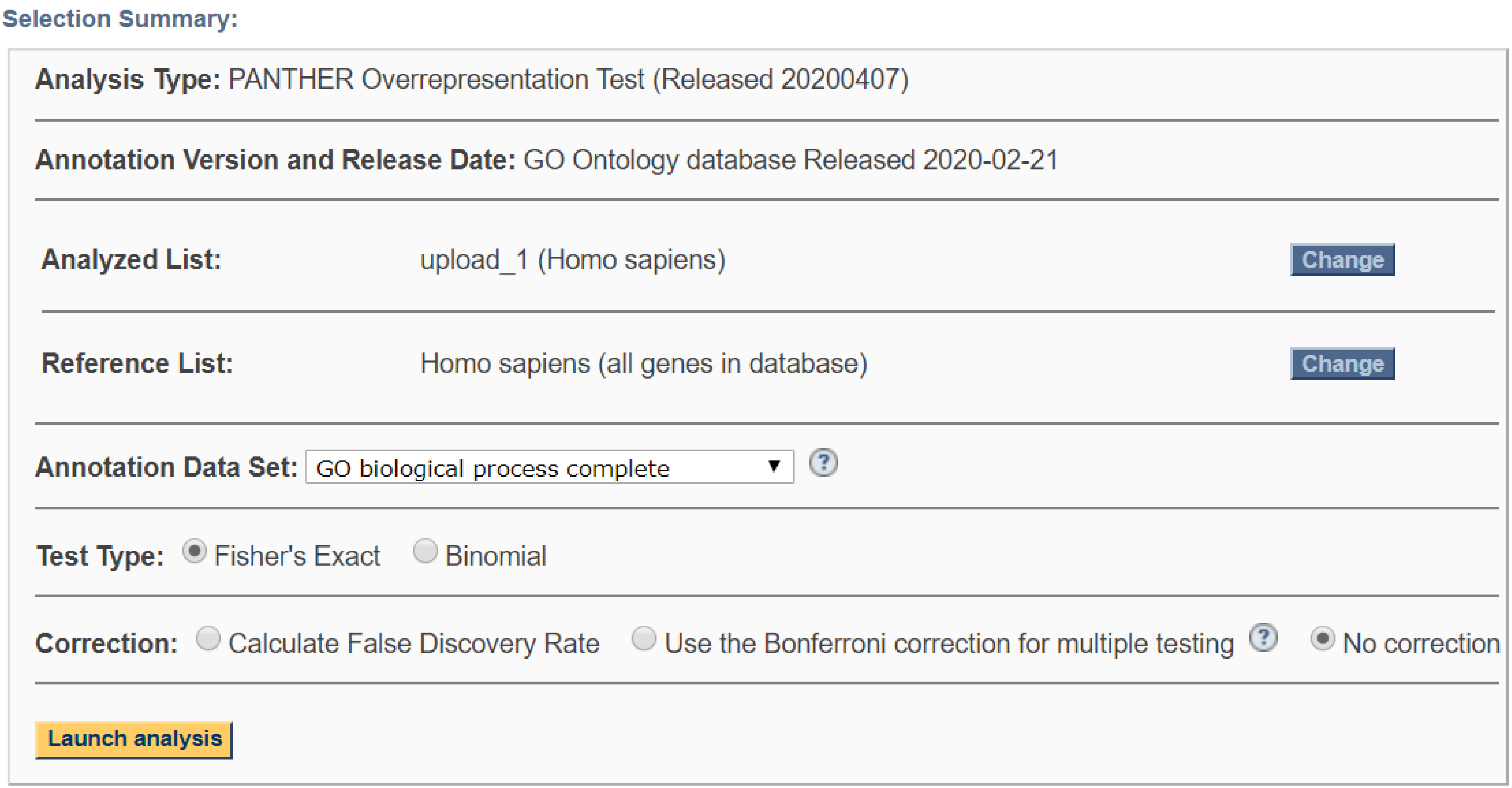
Powered by PANTHER

APOH
APP
CND2
COL3A1
COL5A2
CXCL6

biological process

Homo sapiens

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

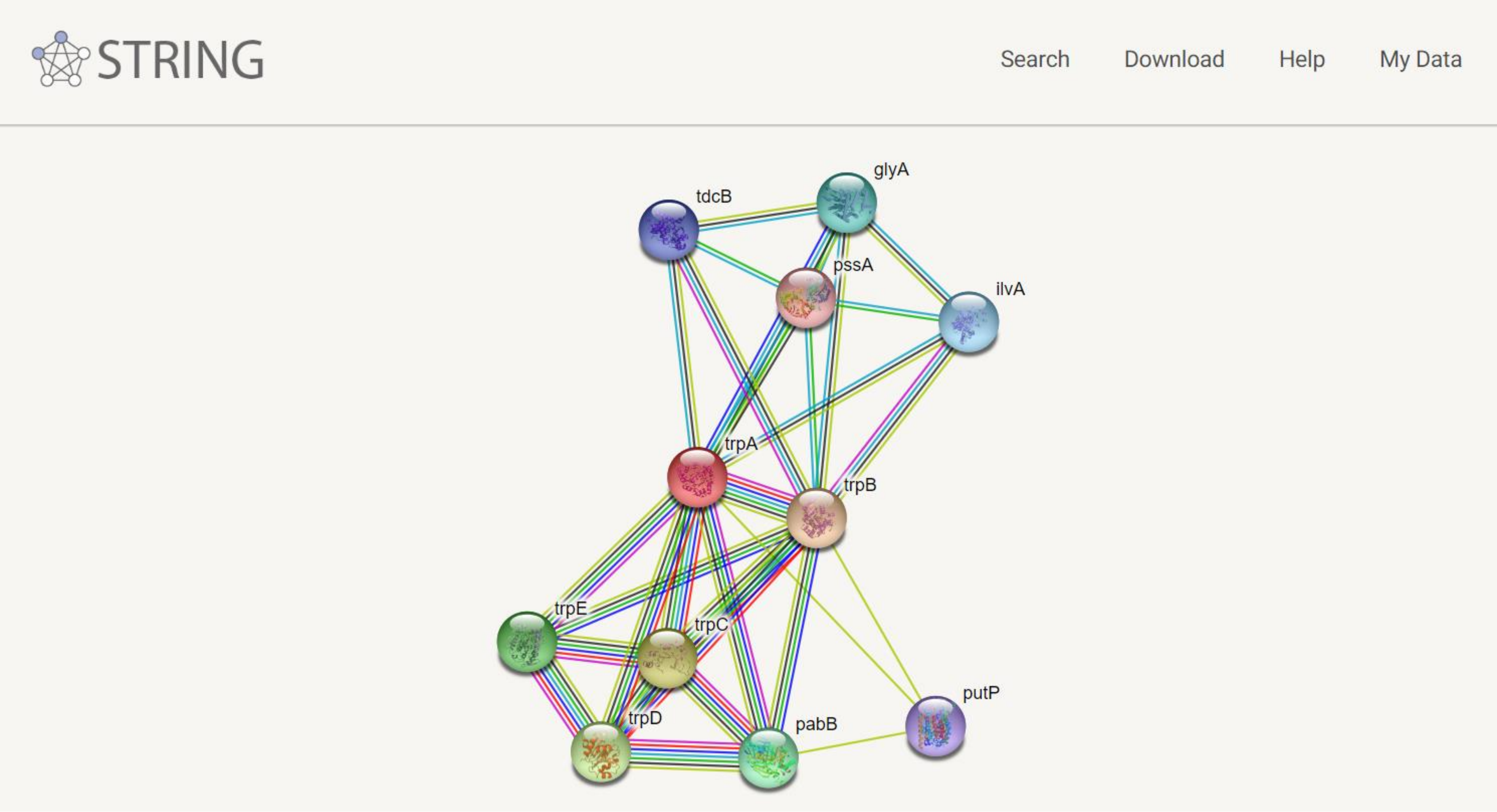
Analyzed List: upload_1 (Homo sapiens)

Reference List: Homo sapiens (all genes in database)

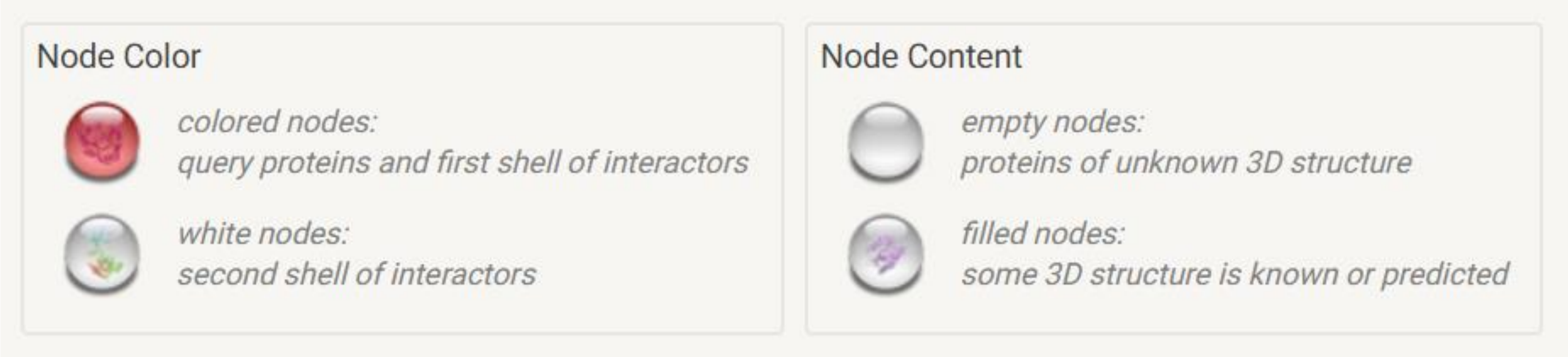
Annotation Data Set: GO biological process complete

Test Type: ● Fisher's Exact ○ Binomial

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



https://string-db.org/



Node Color

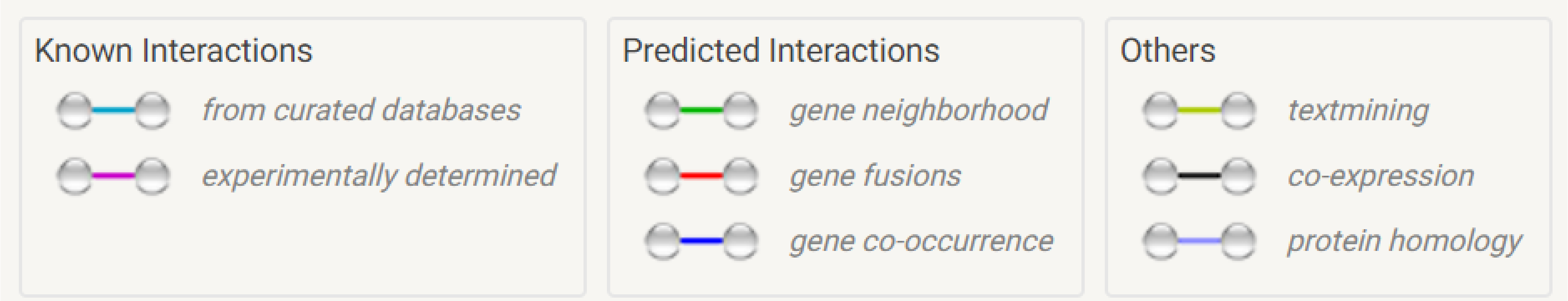
colored nodes:
query proteins and first shell of interactors

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

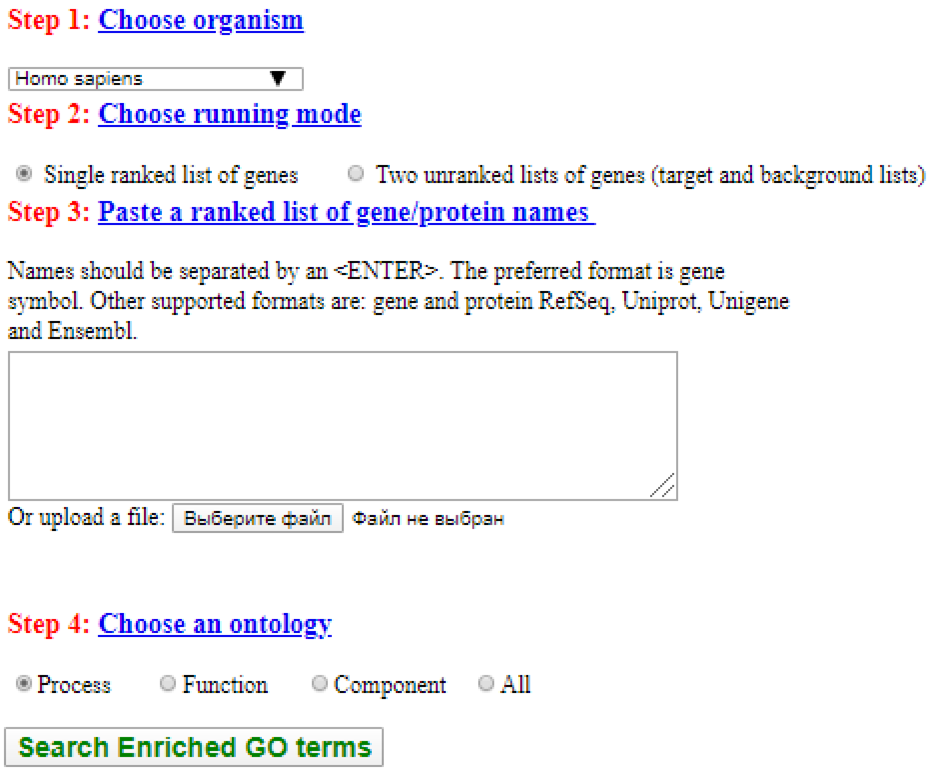
textmining

co-expression

protein homology

GORILLA

Gene Ontology enRIchment anaLysis and visuaLizAtion tool



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

https://david.ncifcrf.gov/

DAVID Bioinformatics Resources 6.8
Laboratory of Human Retrovirology and Immunoinformatics (LHRI)

*** Welcome to DAVID 6.8 ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Shortcut to DAVID Tools

Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering , BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. More

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More

Gene Name Batch Viewer

Recommending: A paper published in ***Nature Protocols*** describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.8

2003 - 2020

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) v6.8 comprises a full Knowledgebase update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

What's Important in DAVID?

- Cite DAVID
- IDs of Affy Exon and Gene arrays supported
- Novel Classification Algorithms
- Pre-built Affymetrix and Illumina backgrounds
- User's customized gene background
- Enhanced calculating speed

Statistics of DAVID