Задание к лекции по Gene Ontology

Каждому из вас дан набор ID в файле **listN.txt – см. в ведомости**

Мы будем рассматривать эти ID как имена генов или РНК или белков в зависимости от изучаемой базы данных.

В жизни необходимо понимать природу ваших данных и не путать гены с белками, белки с РНК, РНК с ДНК!

=)

Часть 1 - GO Enrichment Analysis

Проведите GO Enrichment Analysis ID из вашего listN.txt с помощью сервиса http://geneontology.org/
Используйте следующие настройки:

GO Enrichment Analysis

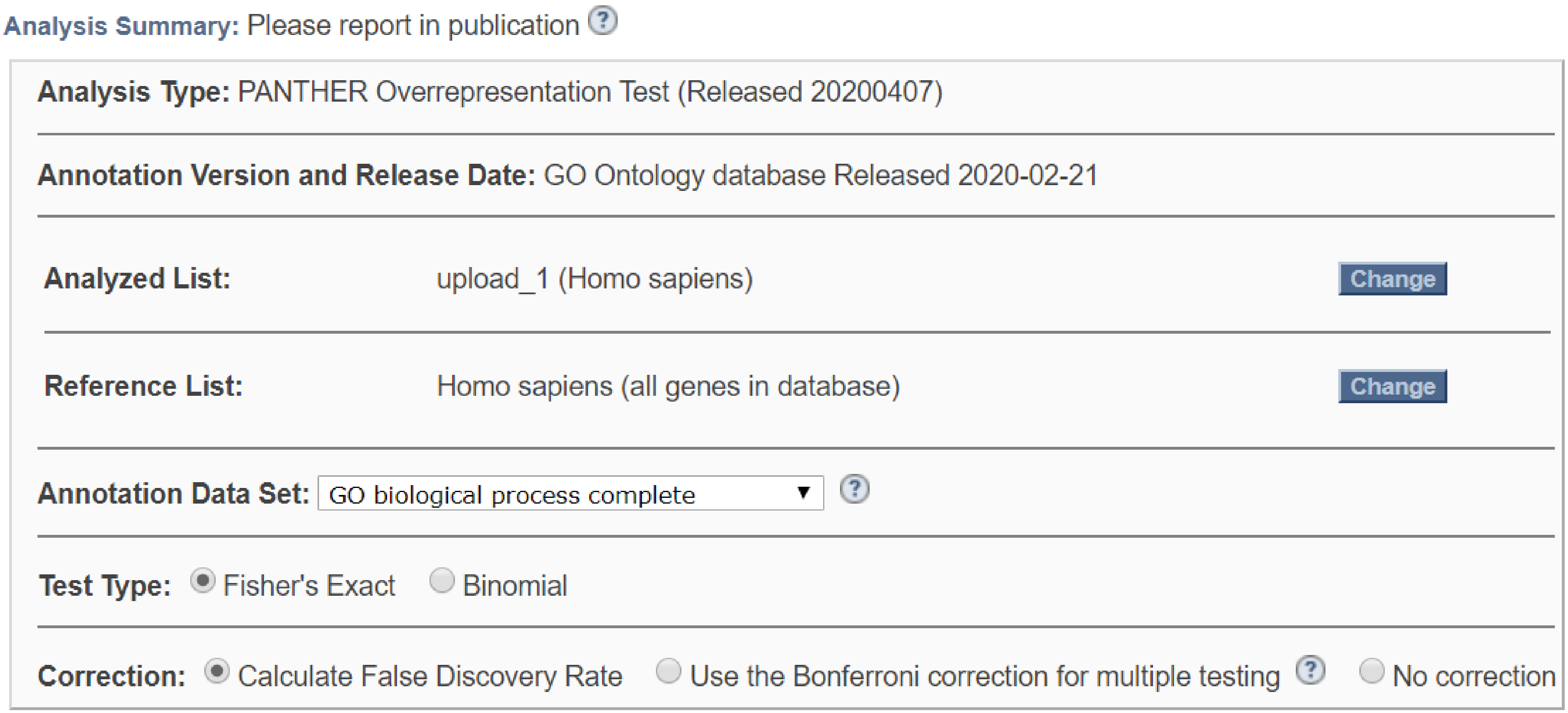
Powered by PANTHER

biological process

Homo sapiens

Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs

Вы будете переведены на сайт http://pantherdb.org/
Используйте следующие настройки:



Analysis Summary: Please report in publication

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)

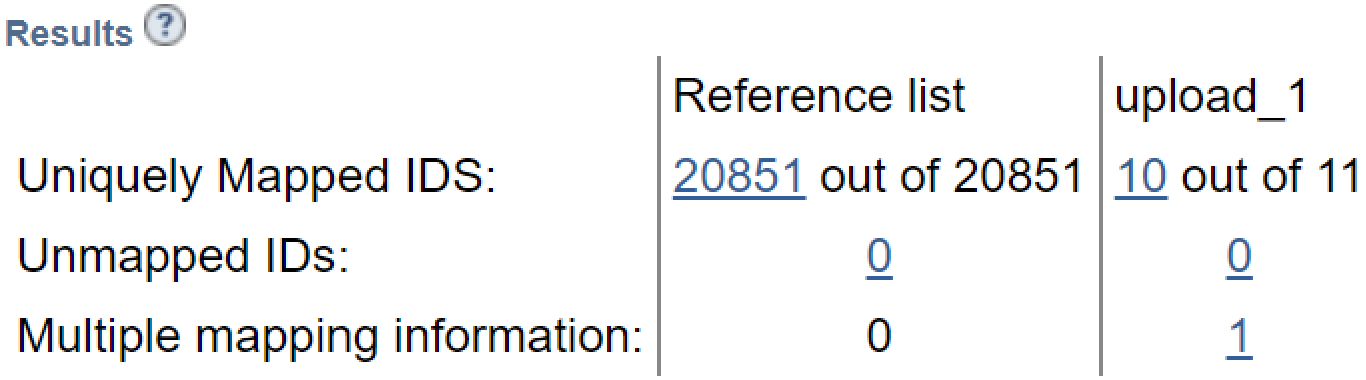
Reference List: Homo sapiens (all genes in database)

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

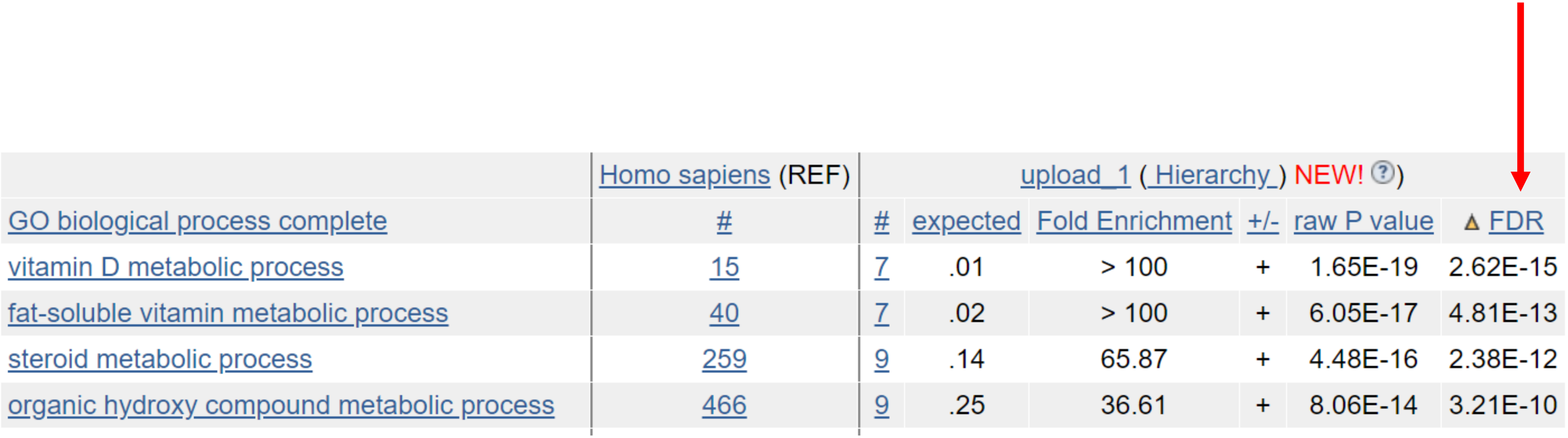
Обратите внимание, что не все ID из вашего списка будут участвовать в анализе обогащения. Это происходит из-за возможного несоответствия ID между различными базами данных. Сервис указывает количество прошедших в итоге ID и выдает отчет по непрошедшим.



Results

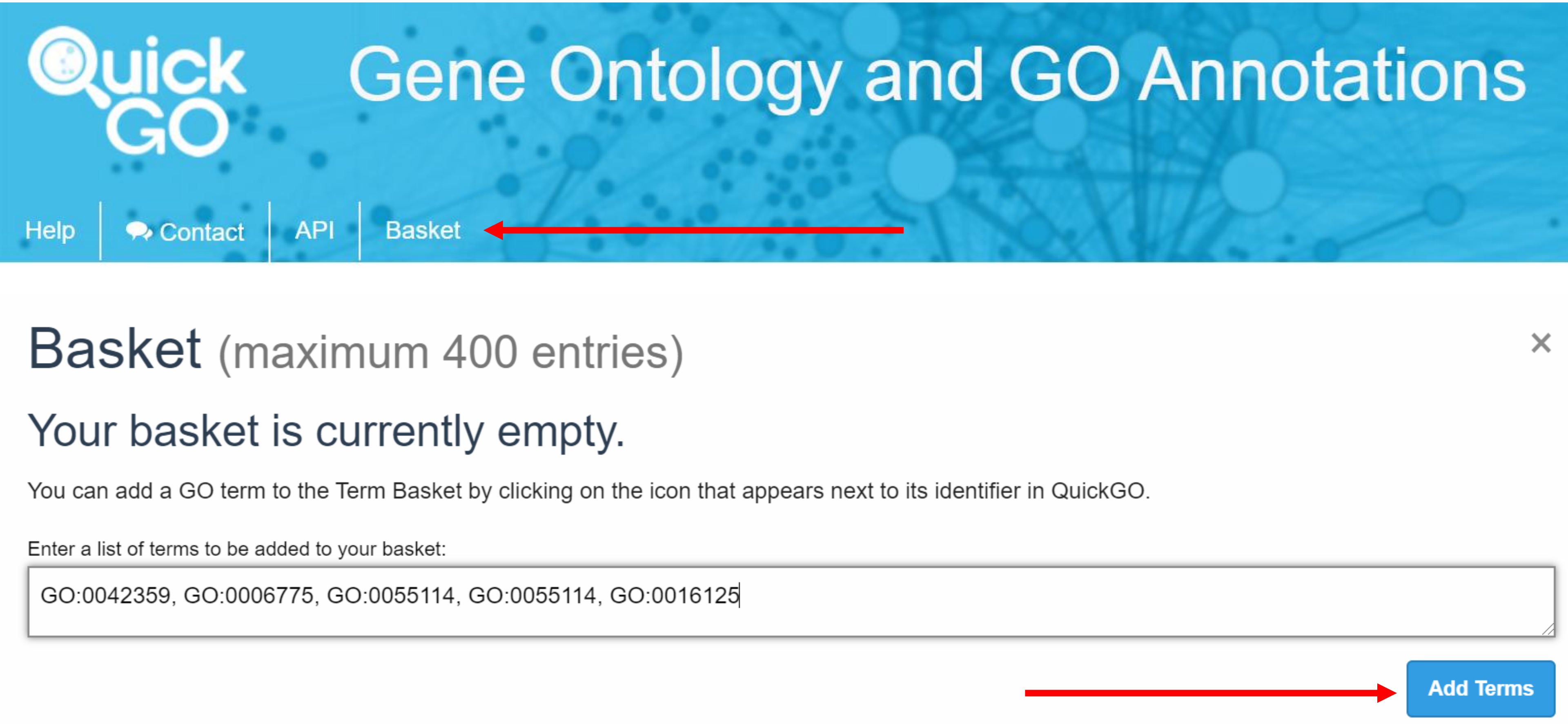
	Reference list	upload_1
Uniquely Mapped IDS:	20851 out of 20851	10 out of 11
Unmapped IDs:	0	0
Multiple mapping information:	0	1

Отсортируйте выдачу по поправленному p-value так, чтобы наверху оказались самые значимые находки



	Homo sapiens (REF)	upload_1 (Hierarchy) NEW! ?)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	▲ FDR
vitamin D metabolic process	15	7	.01	> 100	+	1.65E-19	2.62E-15
fat-soluble vitamin metabolic process	40	7	.02	> 100	+	6.05E-17	4.81E-13
steroid metabolic process	259	9	.14	65.87	+	4.48E-16	2.38E-12
organic hydroxy compound metabolic process	466	9	.25	36.61	+	8.06E-14	3.21E-10

Затем возьмите 5 самых значимых GO terms и визуализируйте их на графе с помощью сервиса https://www.ebi.ac.uk/QuickGO/



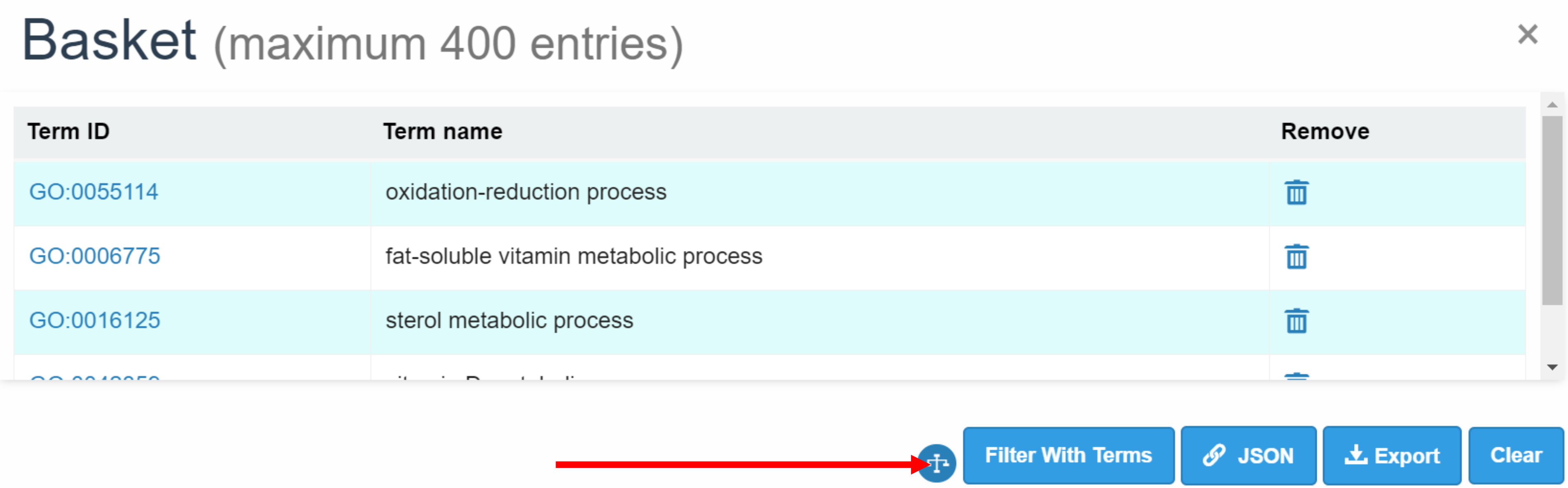
Basket (maximum 400 entries)

Your basket is currently empty.

You can add a GO term to the Term Basket by clicking on the icon that appears next to its identifier in QuickGO.

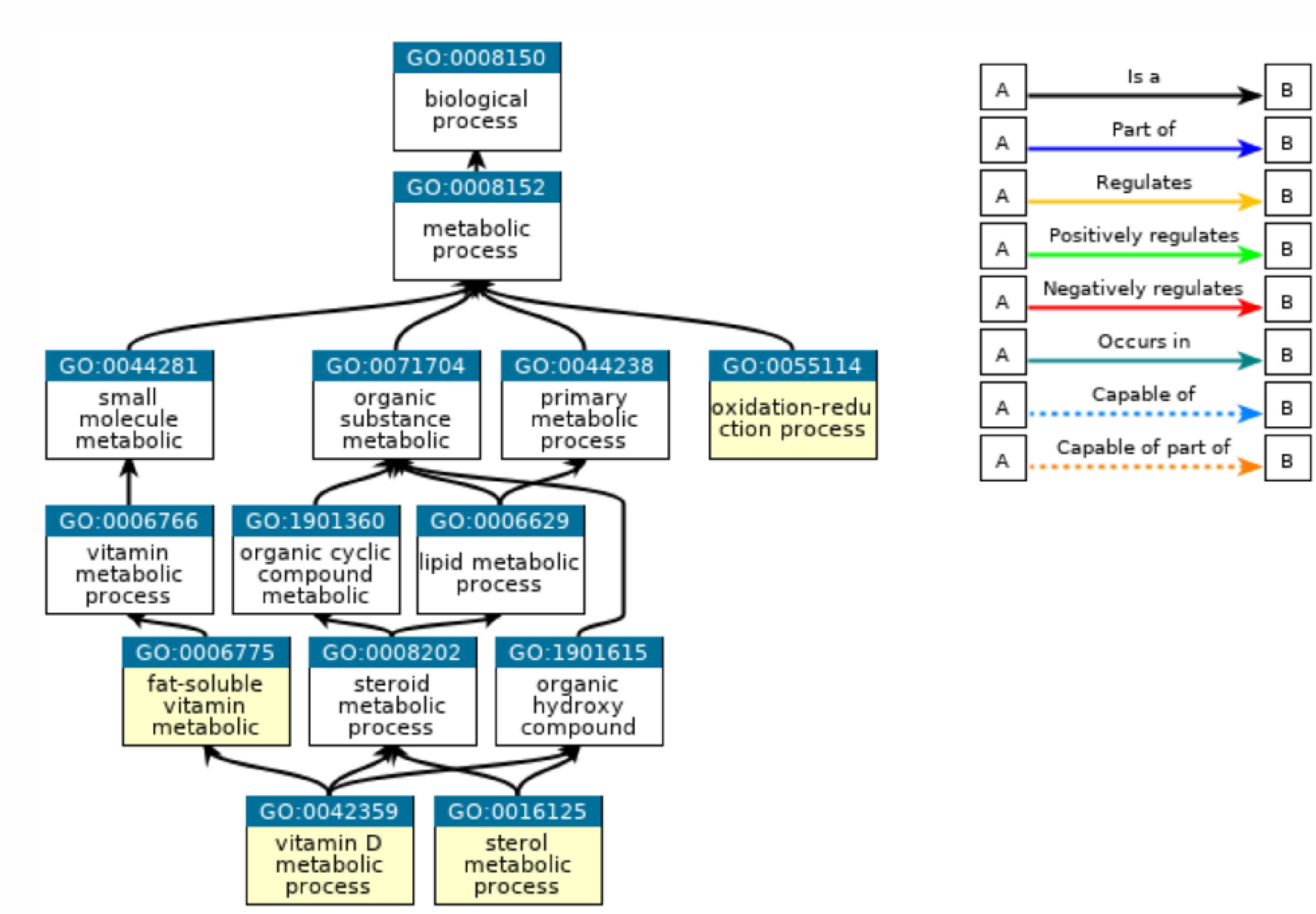
Enter a list of terms to be added to your basket:

GO:0042359, GO:0006775, GO:0055114, GO:0055114, GO:0016125



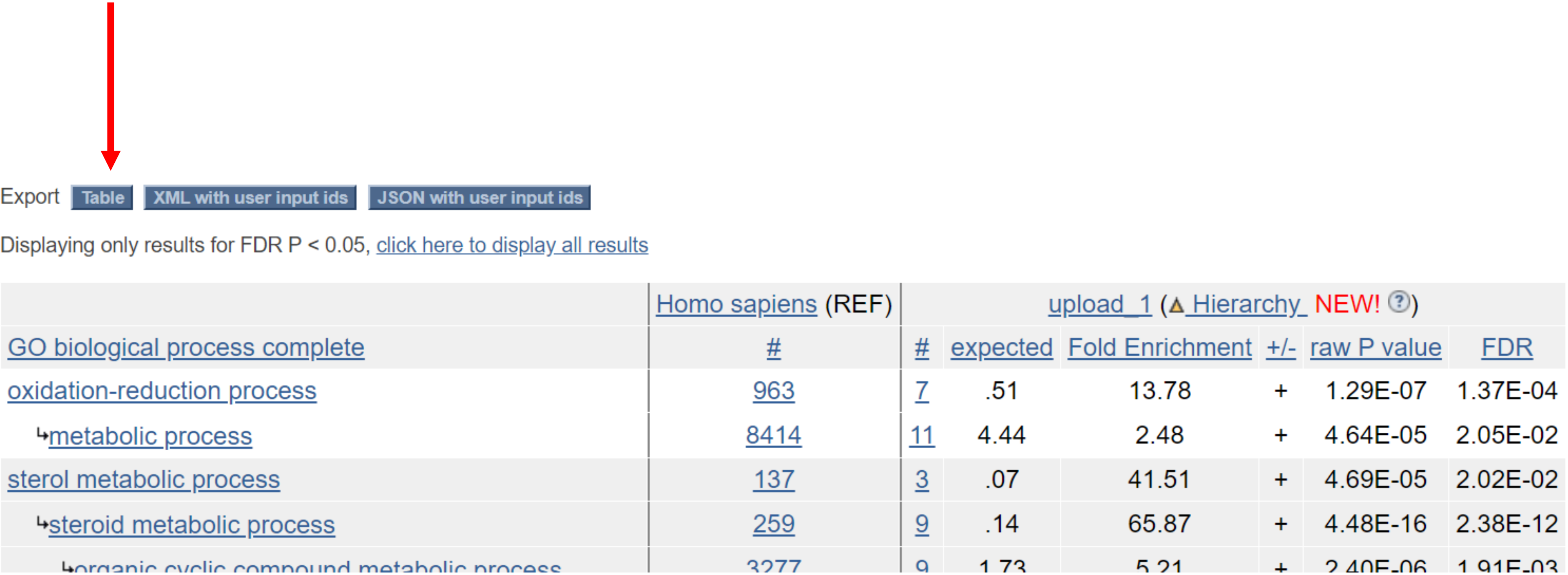
Basket (maximum 400 entries)

Term ID	Term name	Remove
GO:0055114	oxidation-reduction process	🗑
GO:0006775	fat-soluble vitamin metabolic process	🗑
GO:0016125	sterol metabolic process	🗑



Если вы затрудняетесь в дальнейшей интерпретации, можно попробовать отрисовать большее количество категорий

Чтобы посчитать число категорий, можно скачать таблицу



Displaying only results for FDR P < 0.05, click here to display all results

	Homo sapiens (REF)	upload_1 (▲ Hierarchy NEW! ?)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
oxidation-reduction process	963	7	.51	13.78	+	1.29E-07	1.37E-04
↳metabolic process	8414	11	4.44	2.48	+	4.64E-05	2.05E-02
sterol metabolic process	137	3	.07	41.51	+	4.69E-05	2.02E-02
↳steroid metabolic process	259	9	.14	65.87	+	4.48E-16	2.38E-12
↳organic cyclic compound metabolic process	3277	9	1.73	5.21	+	2.40E-06	1.91E-03

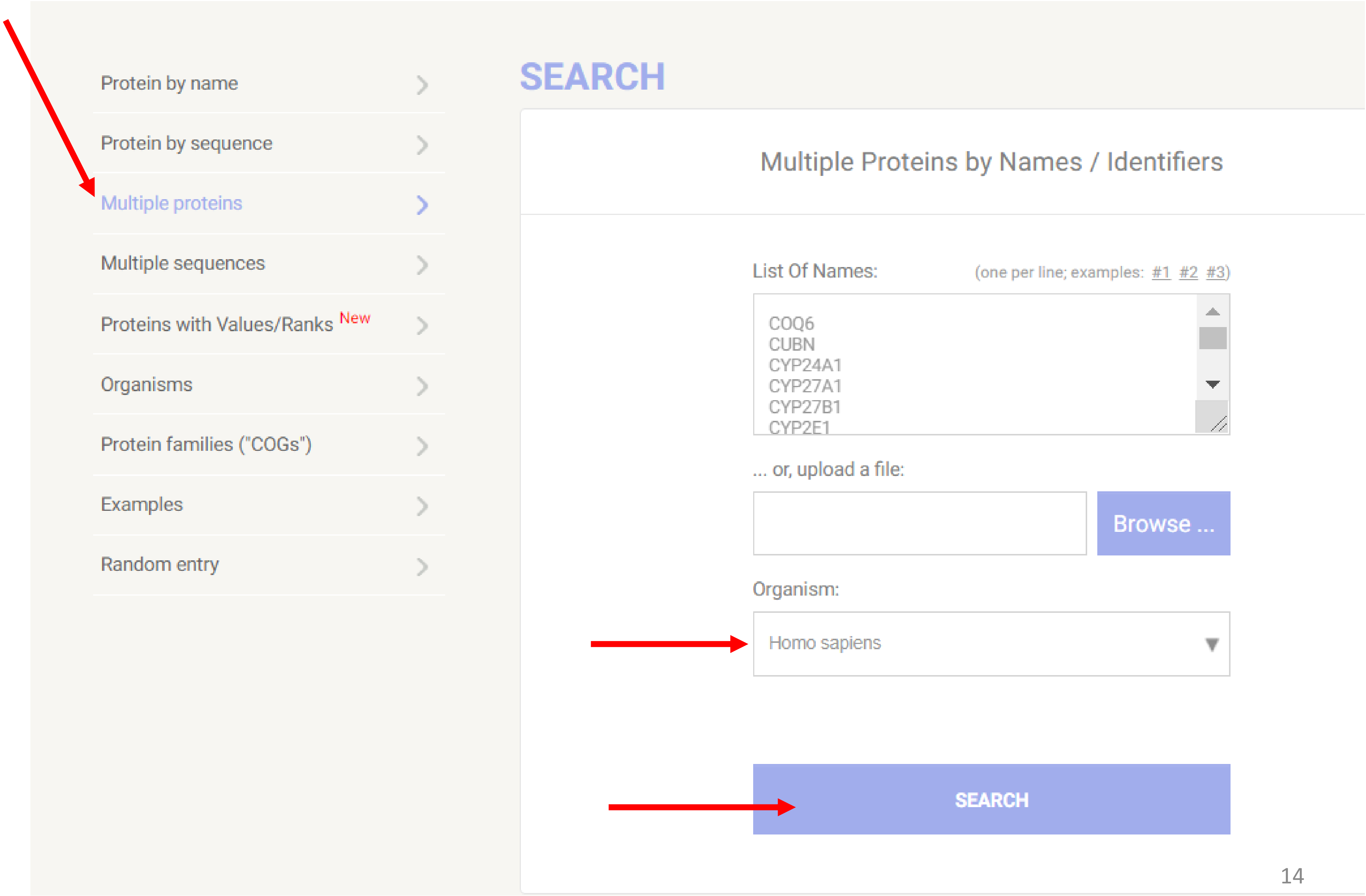
Часть 1 – GO Enrichment Analysis

В отчете укажите:

1. Название вашего файла с ID и количество ID в списке
2. Сколько ID участвовало в результате в анализе обогащения? Если меньше, чем было в списке изначально – укажите причину
3. Сколько GO terms оказалось в выдаче? Почему именно столько?
4. Приведите список десяти самых значимых GO terms
5. Приведите картинку графа, где отмечены 5 самых значимых GO terms
6. Проинтерпретируйте полученные результаты. Какими отношениями соединены узлы полученного графа?
7. Что объединяет ID вашего списка? Охарактеризуйте ваш список ID, воспользовавшись любым количеством GO terms

Часть 2 - String

Проанализируйте ваш список ID с помощью сервиса https://string-db.org/

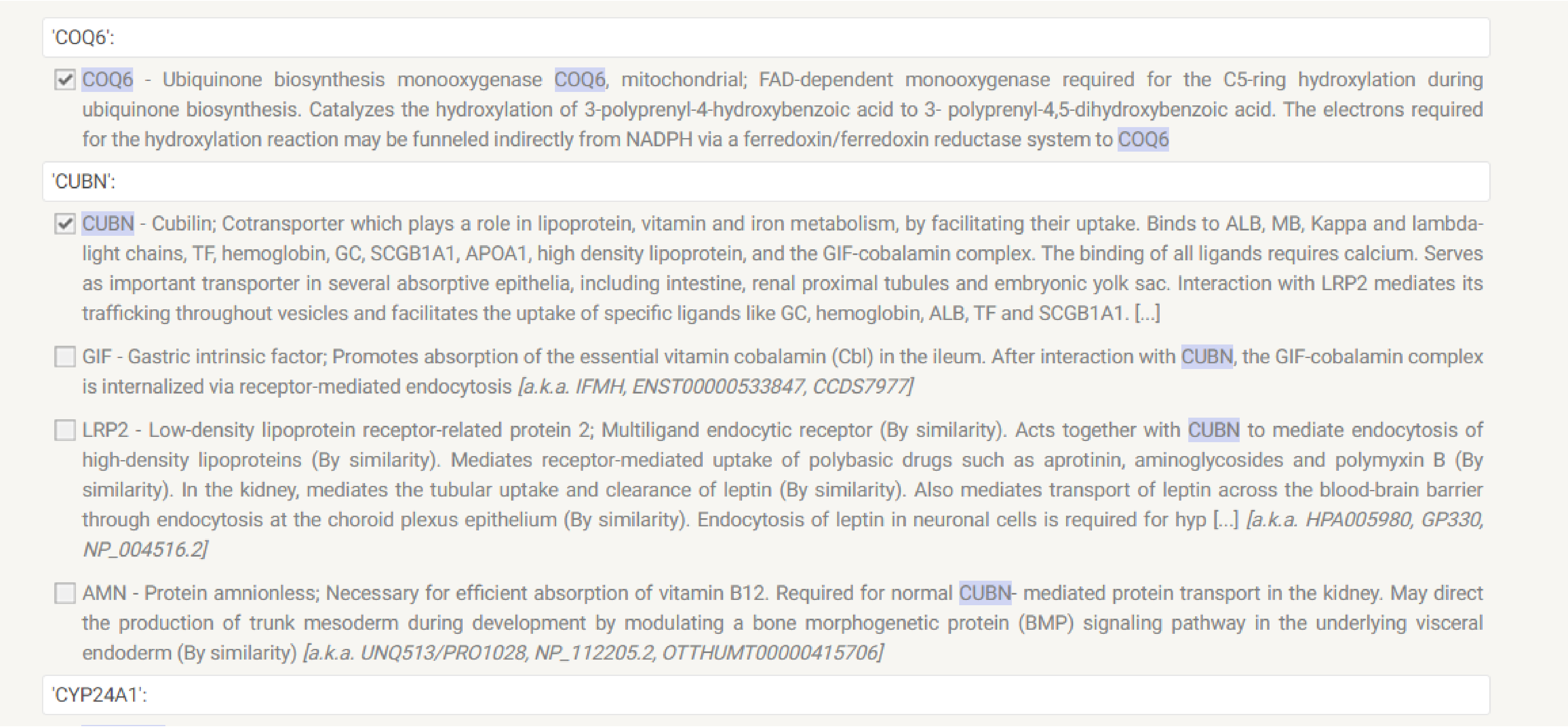


List Of Names:

COQ6
CUBN
CYP24A1
CYP27A1
CYP27B1
CYP2E1

Organism: Homo sapiens

Для некоторых ID String будет предлагать альтернативные или похожие названия. Не пугайтесь, убедитесь, что выбраны именно те ID, что указаны в вашем списке



'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. [...]

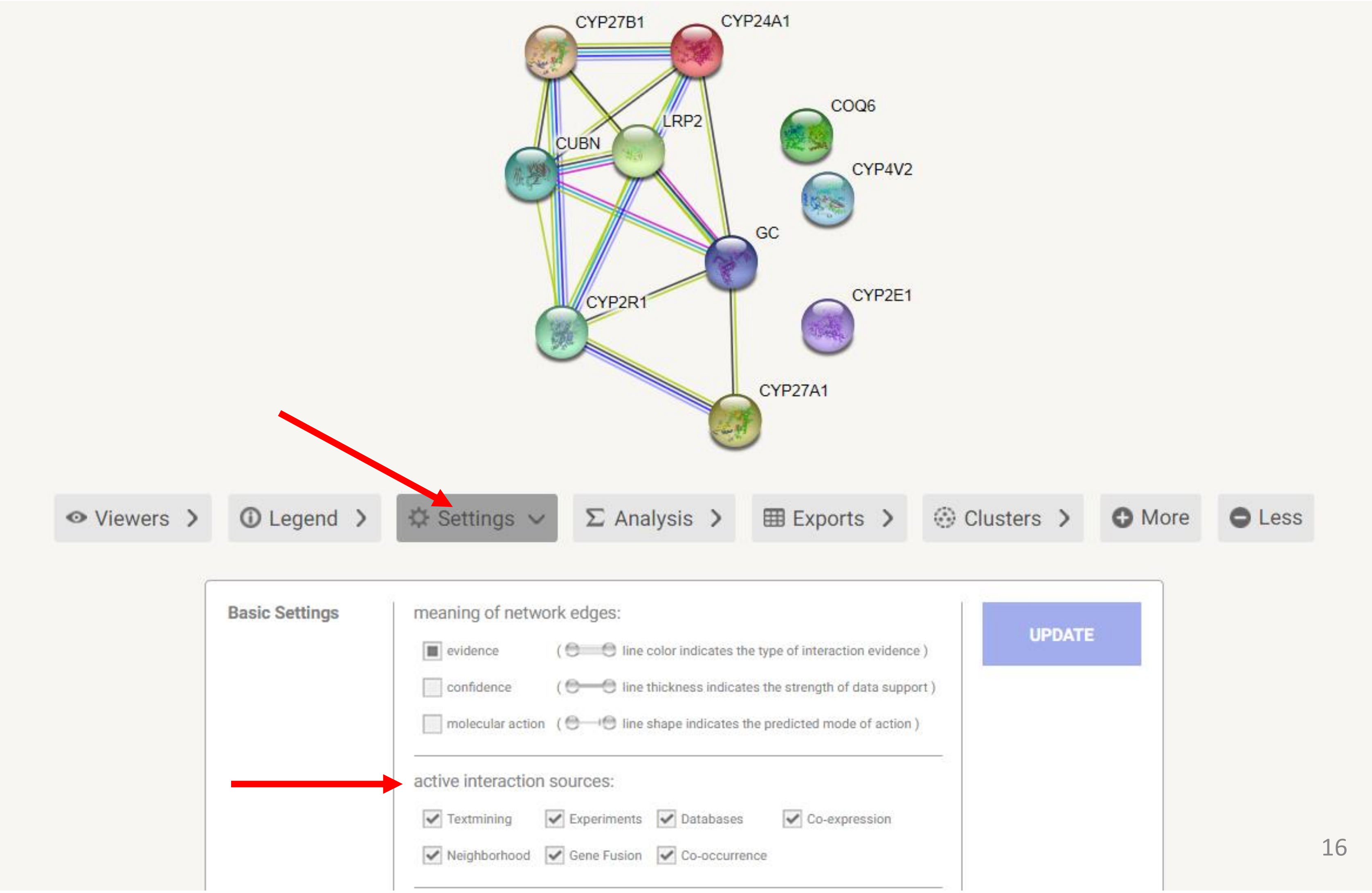
☐ GIF - Gastric intrinsic factor; Promotes absorption of the essential vitamin cobalamin (Cbl) in the ileum. After interaction with CUBN, the GIF-cobalamin complex is internalized via receptor-mediated endocytosis *[a.k.a. IFMH, ENST00000533847, CCDS7977]*

☐ 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':

В результате вы получите граф
Убедитесь, что в настройках включены все типы связей



Basic Settings

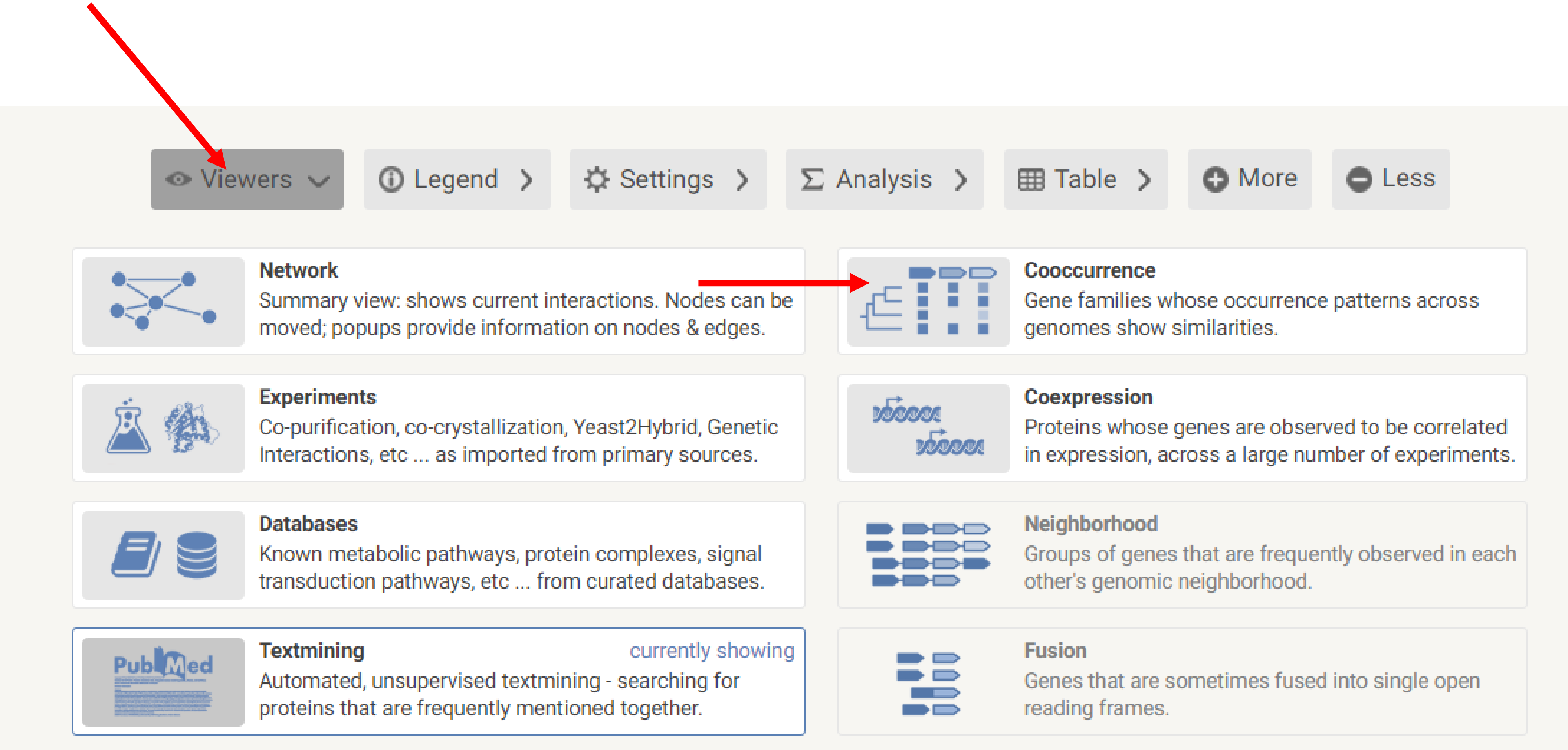
meaning of network edges:

■ evidence (line color indicates the type of interaction evidence)
☐ confidence (line thickness indicates the strength of data support)
☐ molecular action (line shape indicates the predicted mode of action)

active interaction sources:

☑ Textmining ☑ Experiments ☑ Databases ☑ Co-expression
☑ Neighborhood ☑ Gene Fusion ☑ Co-occurrence

Изучите представленность ID вашего списка в различных организмах



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.

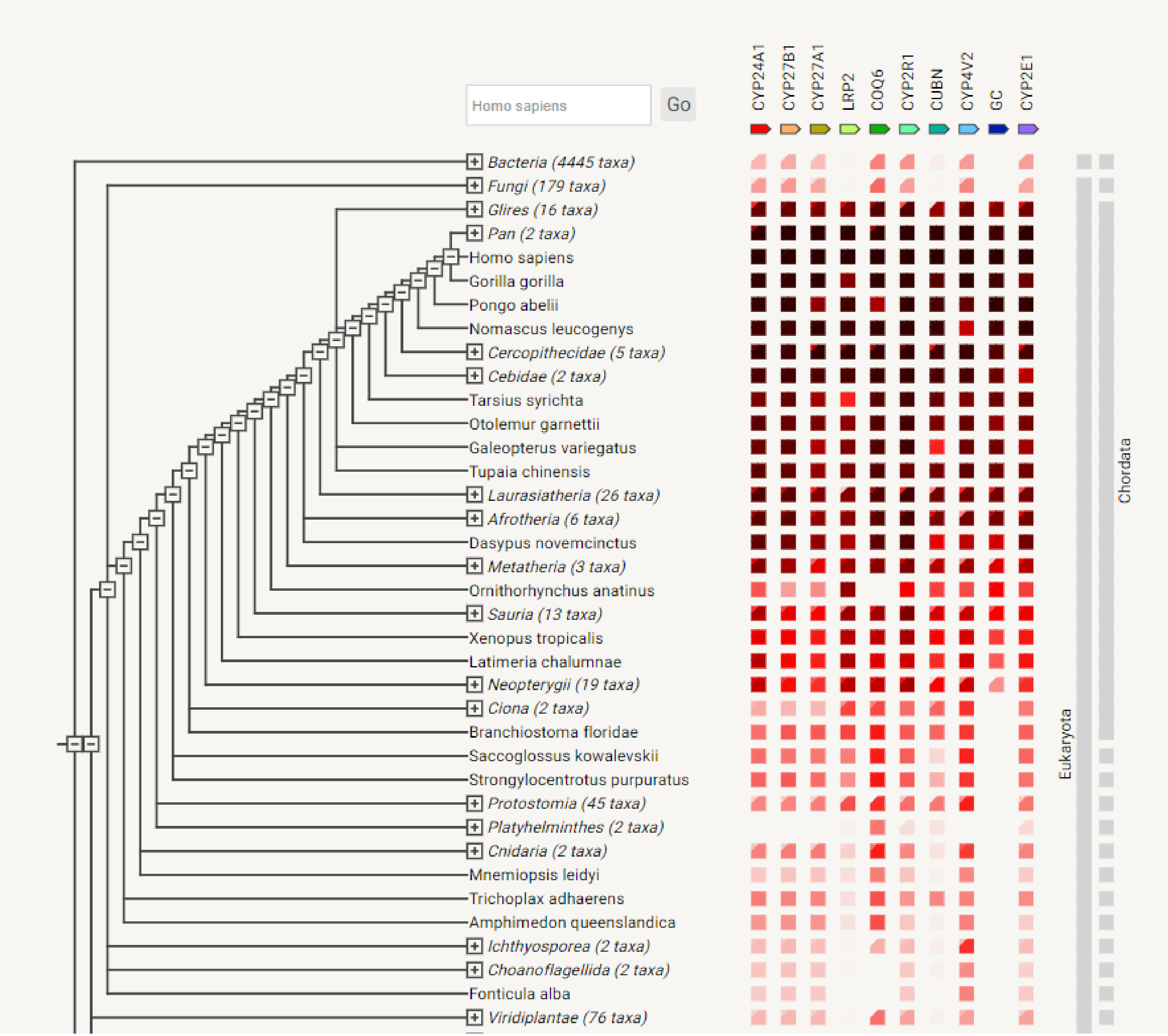
Coexpression
Proteins whose genes are observed to be correlated in expression, across a large number of experiments.

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

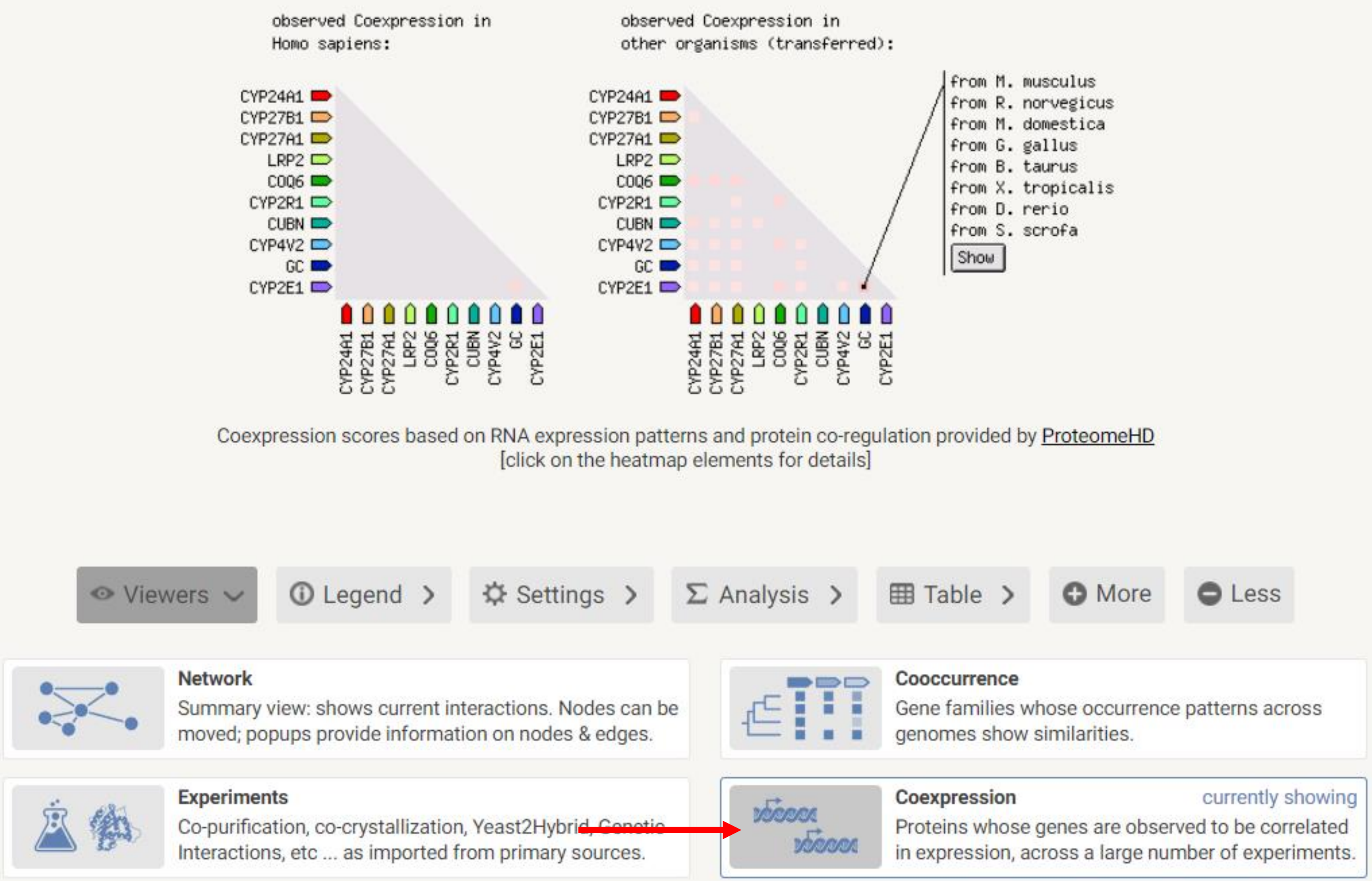
Neighborhood
Groups of genes that are frequently observed in each other's genomic neighborhood.

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.


Homo sapiens
Go
CYP24A1
CYP27B1
CYP27A1
LRP2
COQ6
CYP2R1
CUBN
CYP4V2
GC
CYP2E1
Bacteria (4445 taxa)
Fungi (179 taxa)
Glires (16 taxa)
Pan (2 taxa)
Homo sapiens
Gorilla gorilla
Pongo abelii
Nomascus leucogenys
Cercopithecidae (5 taxa)
Cebidae (2 taxa)
Tarsius syrichta
Otolemur garnettii
Galeopterus variegatus
Tupaia chinensis
Laurasiatheria (26 taxa)
Afrotheria (6 taxa)
Dasypus novemcinctus
Metatheria (3 taxa)
Ornithorhynchus anatinus
Sauria (13 taxa)
Xenopus tropicalis
Latimeria chalumnae
Neopterygii (19 taxa)
Ciona (2 taxa)
Branchiostoma floridae
Saccoglossus kowalevskii
Strongylocentrotus purpuratus
Protostomia (45 taxa)
Platyhelminthes (2 taxa)
Cnidaria (2 taxa)
Mnemiopsis leidyi
Trichoplax adhaerens
Amphimedon queenslandica
Ichthyosporea (2 taxa)
Choanoflagellida (2 taxa)
Fonticula alba
Viridiplantae (76 taxa)
Chordata
Eukaryota

GENE COEXPRESSION



Coexpression scores based on RNA expression patterns and 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.

Часть 2 – String

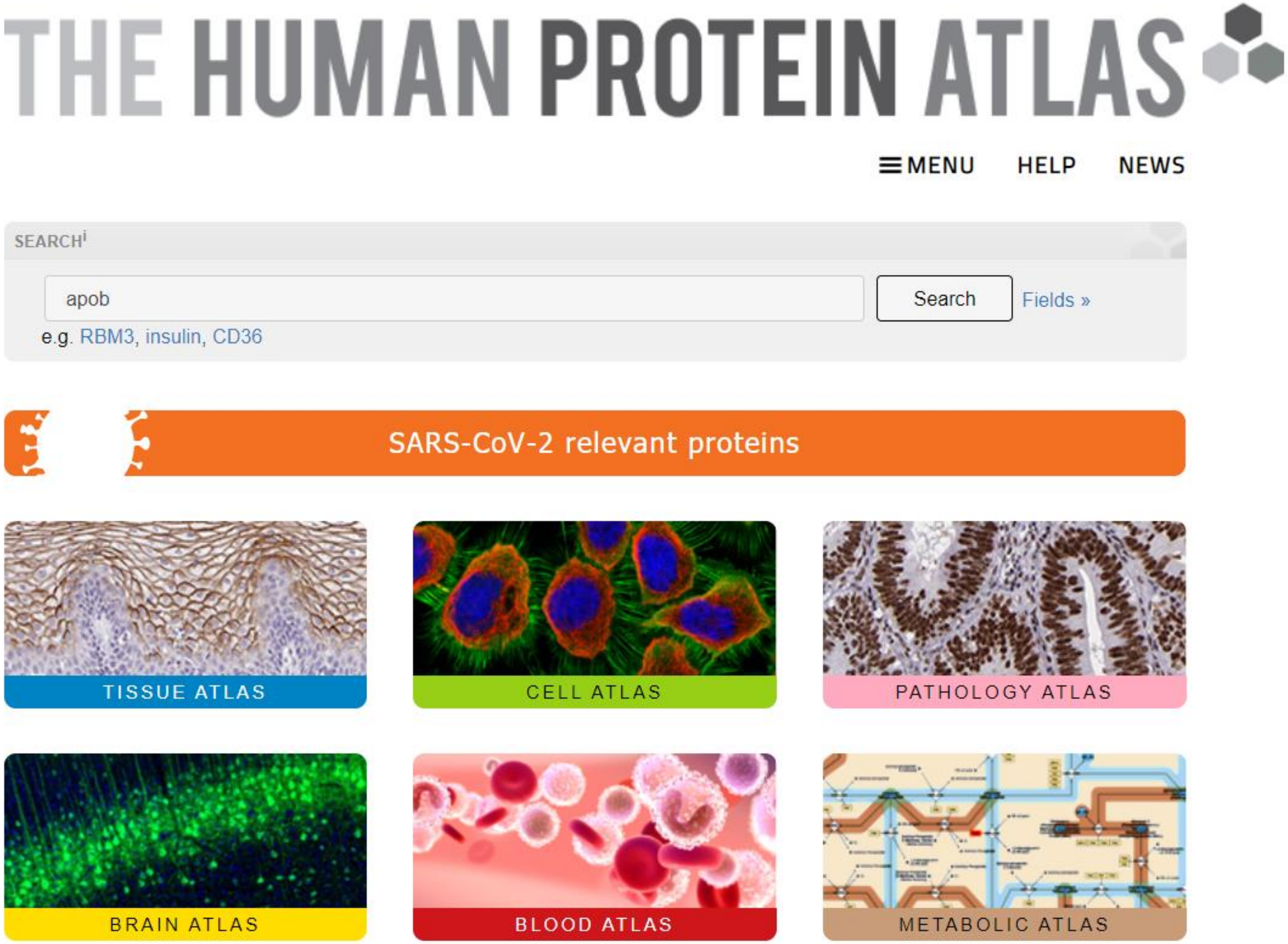
В отчете укажите:

1. Получившийся граф, вставьте картинку
2. Для скольких узлов указано наличие 3D-структур (смотрите подсказки в Legend)?
3. Какими типами взаимодействий связаны узлы вашего графа?
4. Обсудите консервативность ID вашего списка. Приведите картинку (cooccrurrence)
5. Обсудите совместную экспрессию ID вашего списка, а также отличие паттернов коэкспрессии у человека и других видов. Приведите картинку (coexpression)

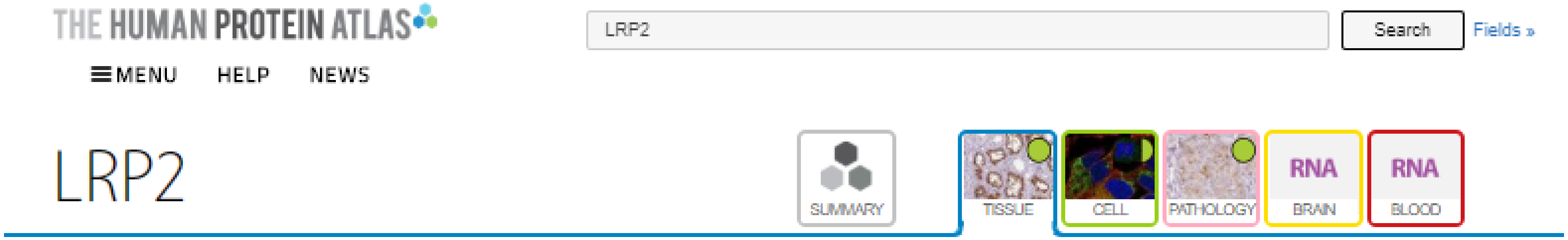
Если у вас нет какого-то типа взаимодействий – нажмите more (можно несколько раз) – появятся новые связи между узлами или новые узлы и новые связи – опишите ваши действия

Часть 3 – Human Protein Atlas

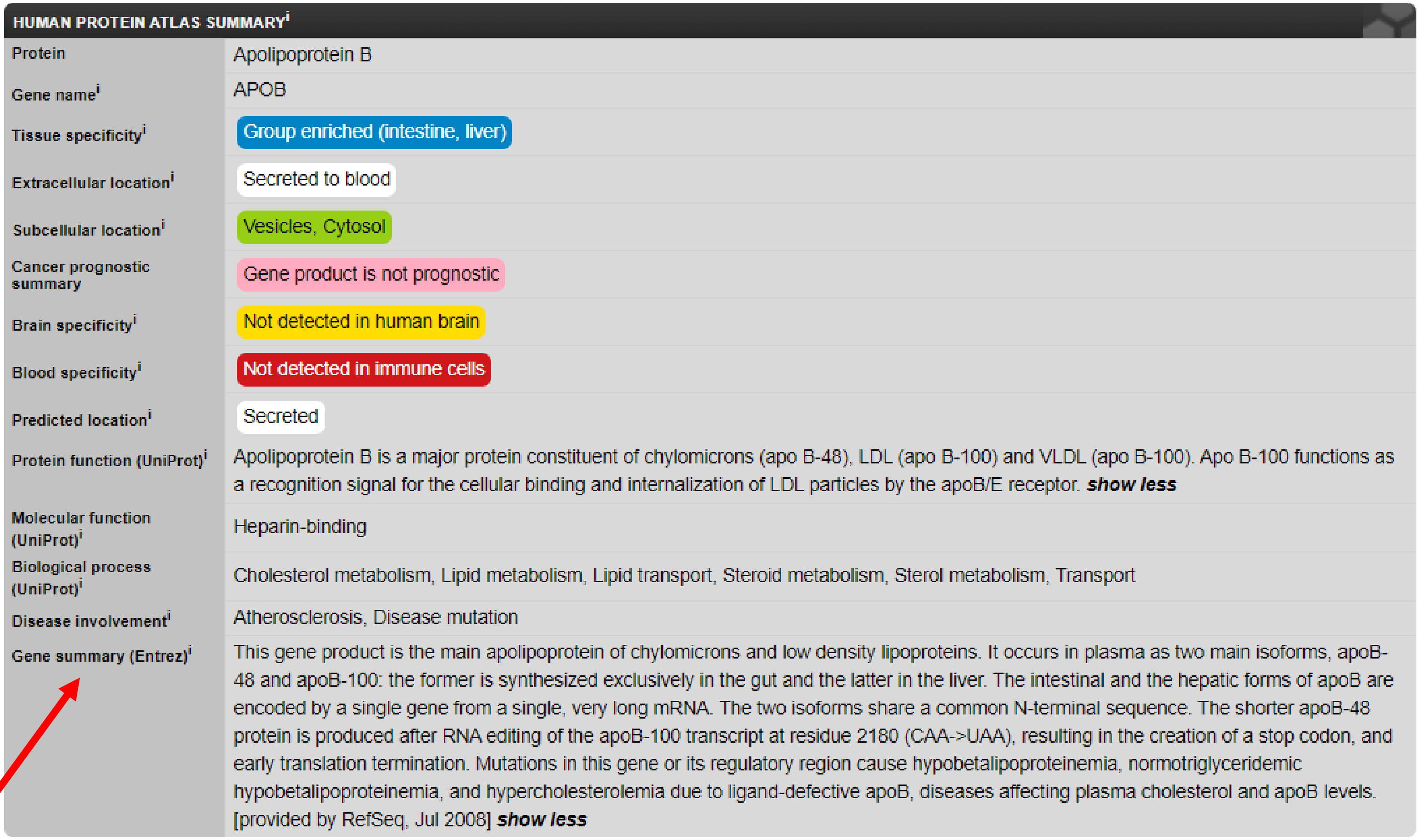
Возьмите один любой ID из вашего списка и изучите его с помощью https://www.proteinatlas.org/



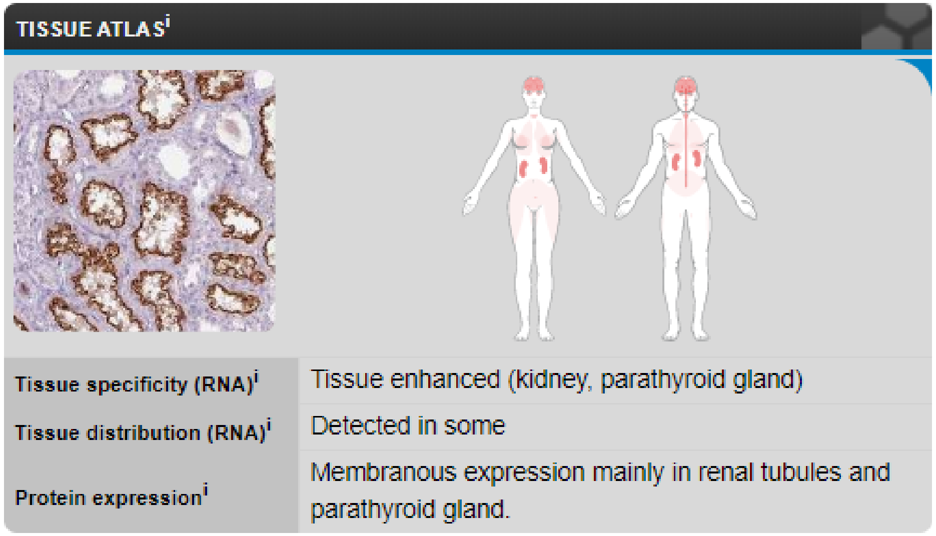
При поиске вы можете обнаружить несколько вкладок Нам понадобится Summary, Tissue и Cell



На странице каждого белка в разделе Summary можно увидеть следующую информацию:

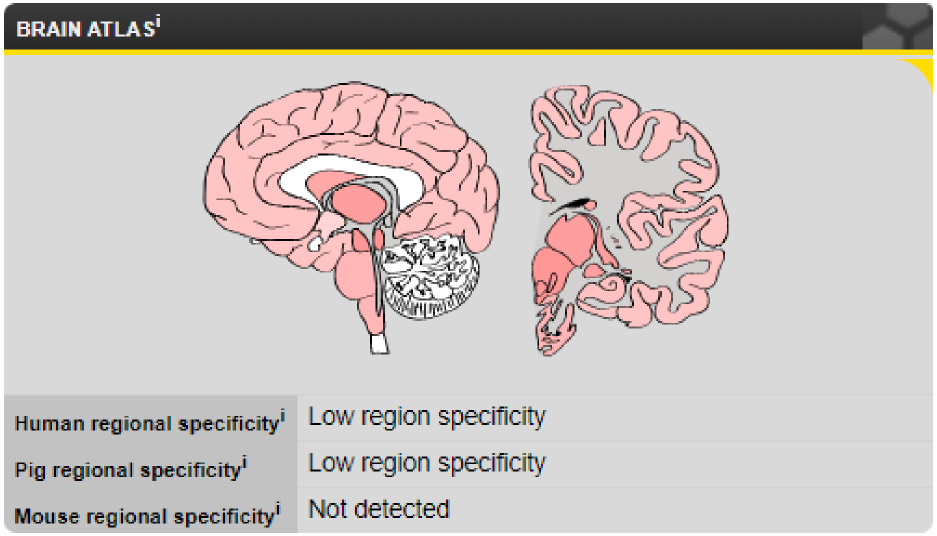


HUMAN PROTEIN ATLAS SUMMARY	
Protein	Apolipoprotein B
Gene name	APOB
Tissue specificity	Group enriched (intestine, liver)
Extracellular location	Secreted to blood
Subcellular location	Vesicles, Cytosol
Cancer prognostic summary	Gene product is not prognostic
Brain specificity	Not detected in human brain
Blood specificity	Not detected in immune cells
Predicted location	Secreted
Protein function (UniProt)	Apolipoprotein B is a major protein constituent of chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo B-100 functions as a recognition signal for the cellular binding and internalization of LDL particles by the apoB/E receptor. ***show less***
Molecular function (UniProt)	Heparin-binding
Biological process (UniProt)	Cholesterol metabolism, Lipid metabolism, Lipid transport, Steroid metabolism, Sterol metabolism, Transport
Disease involvement	Atherosclerosis, Disease mutation
Gene summary (Entrez)	This gene product is the main apolipoprotein of chylomicrons and low density lipoproteins. It occurs in plasma as two main isoforms, apoB-48 and apoB-100: the former is synthesized exclusively in the gut and the latter in the liver. The intestinal and the hepatic forms of apoB are encoded by a single gene from a single, very long mRNA. The two isoforms share a common N-terminal sequence. The shorter apoB-48 protein is produced after RNA editing of the apoB-100 transcript at residue 2180 (CAA->UAA), resulting in the creation of a stop codon, and early translation termination. Mutations in this gene or its regulatory region cause hypobetalipoproteinemia, normotriglyceridemic hypobetalipoproteinemia, and hypercholesterolemia due to ligand-defective apoB, diseases affecting plasma cholesterol and apoB levels. [provided by RefSeq, Jul 2008] ***show less***



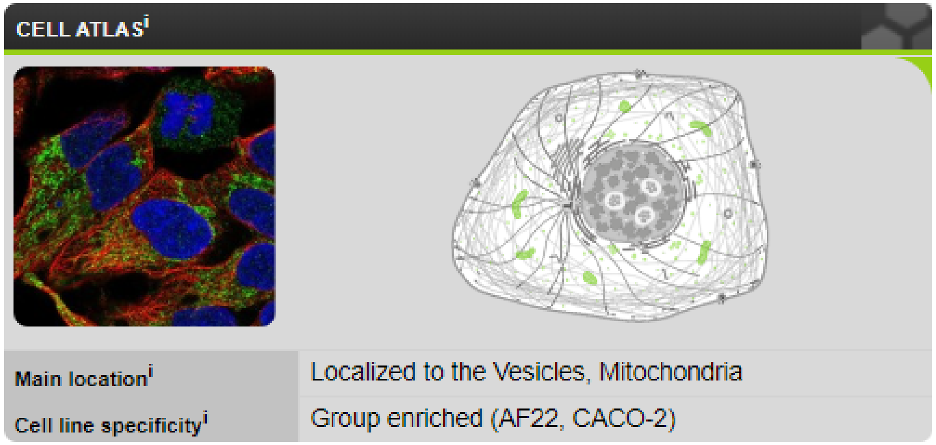
TISSUE ATLASⓘ

Tissue specificity (RNA)ⓘ	Tissue enhanced (kidney, parathyroid gland)
Tissue distribution (RNA)ⓘ	Detected in some
Protein expressionⓘ	Membranous expression mainly in renal tubules and parathyroid gland.



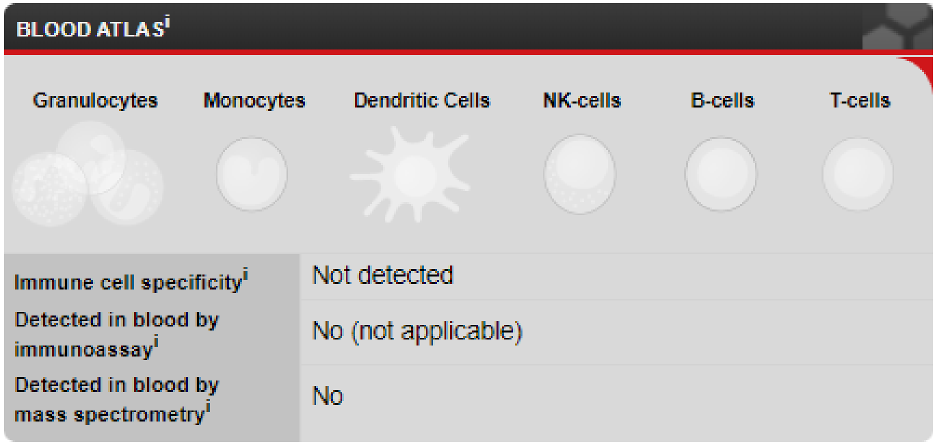
BRAIN ATLASⓘ

Human regional specificityⓘ	Low region specificity
Pig regional specificityⓘ	Low region specificity
Mouse regional specificityⓘ	Not detected



CELL ATLASⓘ

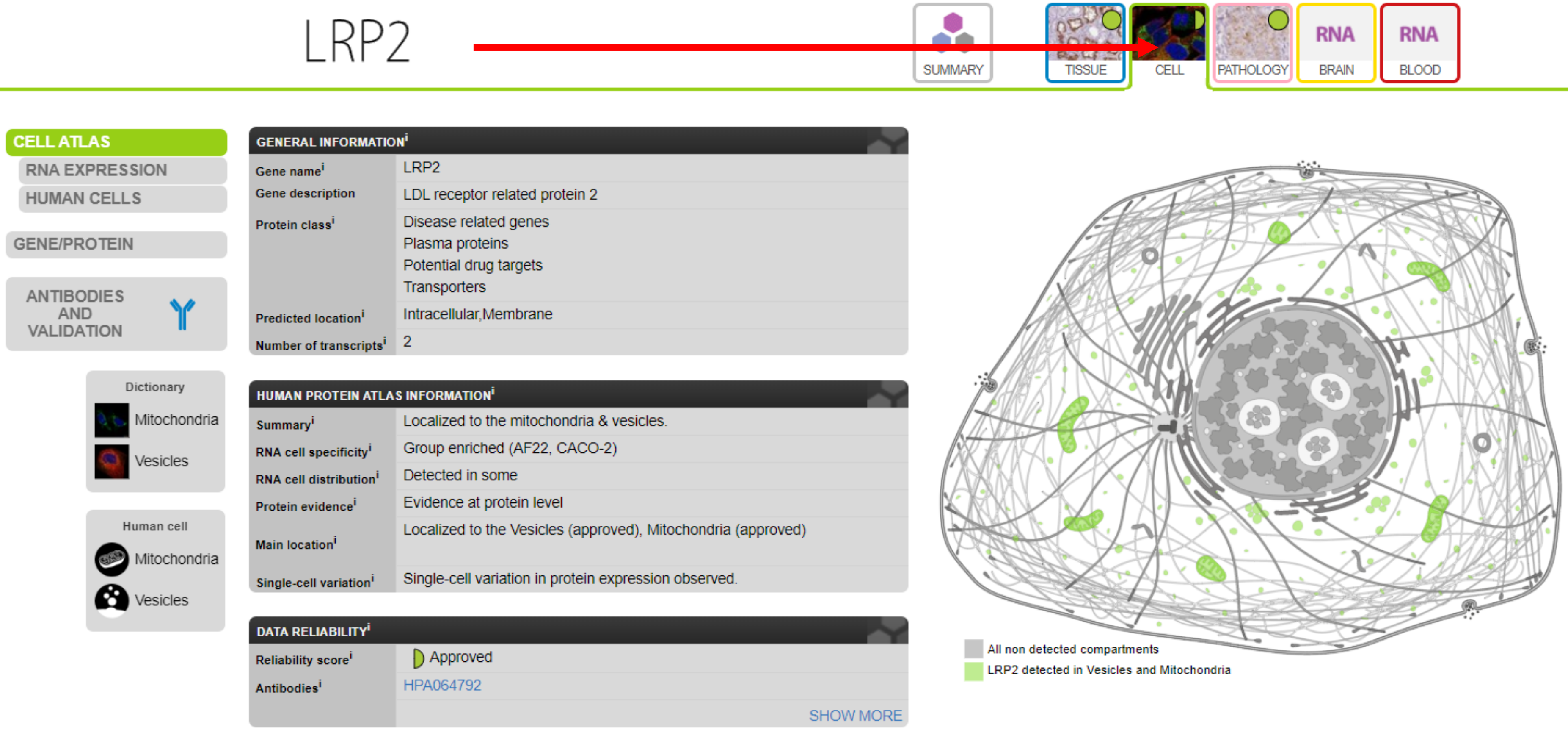
Main locationⓘ	Localized to the Vesicles, Mitochondria
Cell line specificityⓘ	Group enriched (AF22, CACO-2)



BLOOD ATLASⓘ

Immune cell specificityⓘ	Not detected
Detected in blood by immunoassayⓘ	No (not applicable)
Detected in blood by mass spectrometryⓘ	No

Ваш ID может быть специфичен для какой-либо ткани, может присутствовать везде или нигде не быть детектирован. Постарайтесь найти из списка что-то специфичное для какой-либо ткани или клеточного компартмента



GENERAL INFORMATION

Gene name	LRP2
Gene description	LDL receptor related protein 2
Protein class	Disease related genes Plasma proteins Potential drug targets Transporters
Predicted location	Intracellular,Membrane
Number of transcripts	2

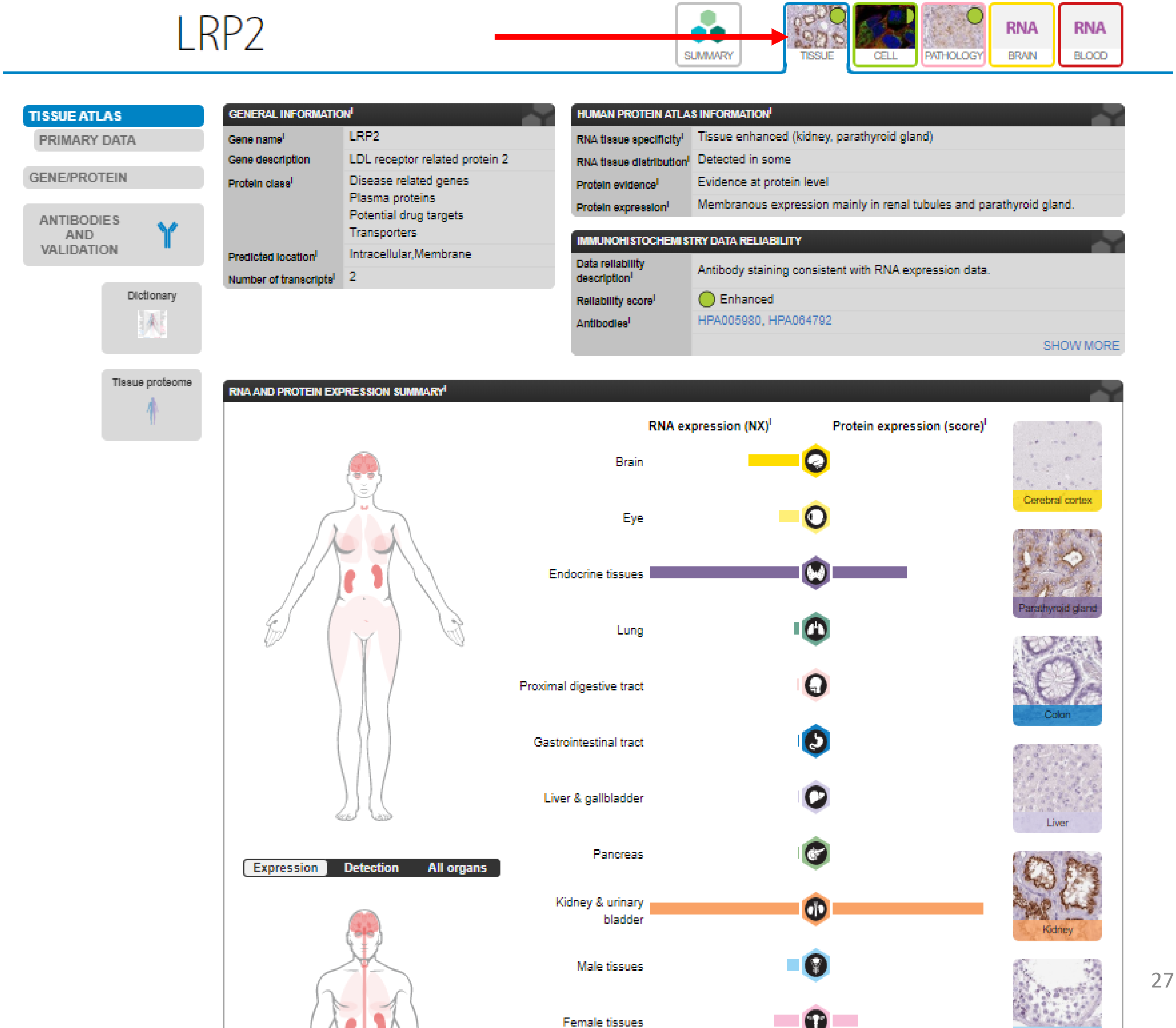
HUMAN PROTEIN ATLAS INFORMATION

Summary	Localized to the mitochondria & vesicles.
RNA cell specificity	Group enriched (AF22, CACO-2)
RNA cell distribution	Detected in some
Protein evidence	Evidence at protein level
Main location	Localized to the Vesicles (approved), Mitochondria (approved)
Single-cell variation	Single-cell variation in protein expression observed.

DATA RELIABILITY

Reliability score	Approved
Antibodies	HPA064792

- ⬜ All non detected compartments
- 🟩 LRP2 detected in Vesicles and Mitochondria



LRP2

GENERAL INFORMATION

Gene name	LRP2
Gene description	LDL receptor related protein 2
Protein class	Disease related genes Plasma proteins Potential drug targets Transporters
Predicted location	Intracellular,Membrane
Number of transcripts	2

HUMAN PROTEIN ATLAS INFORMATION

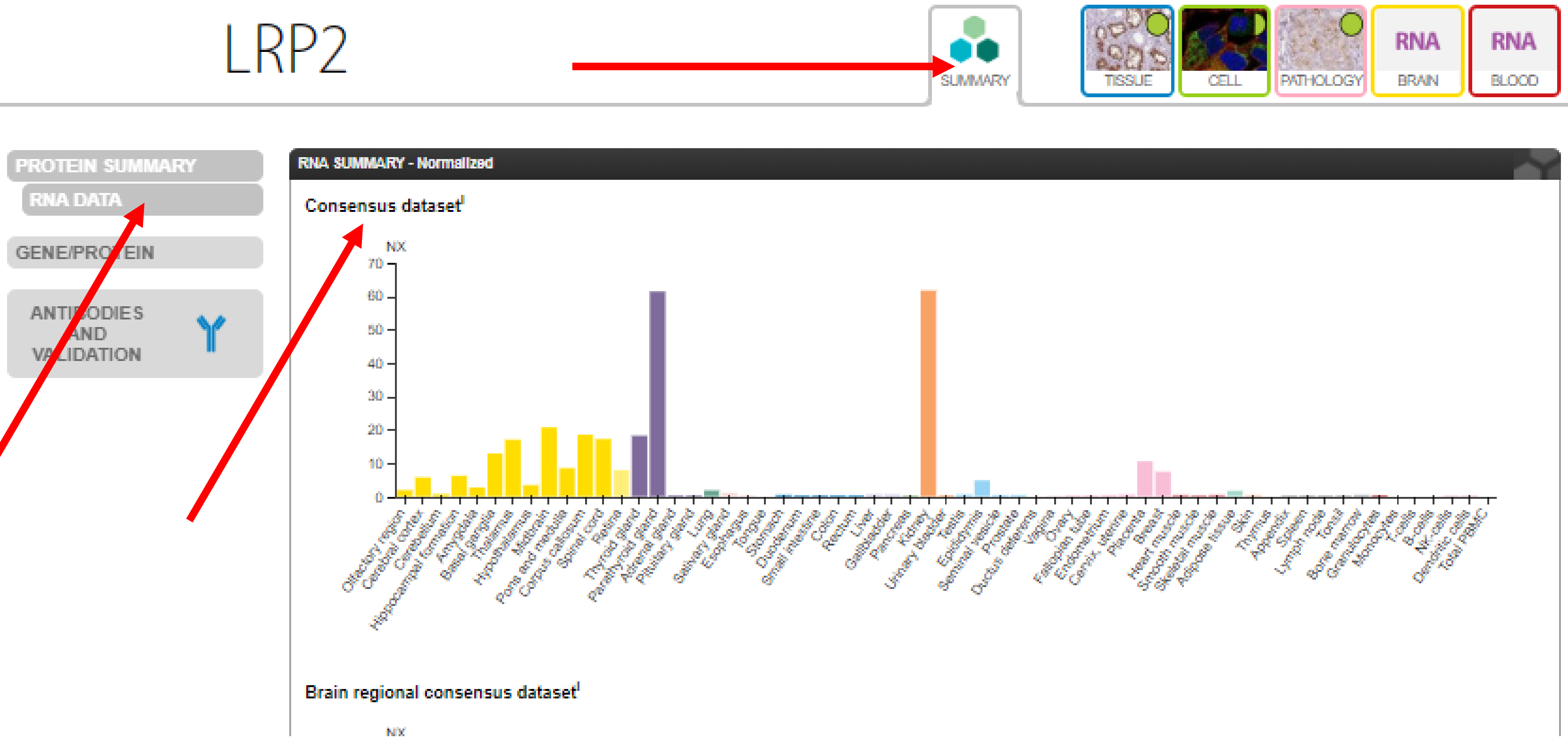
RNA tissue specificity	Tissue enhanced (kidney, parathyroid gland)
RNA tissue distribution	Detected in some
Protein evidence	Evidence at protein level
Protein expression	Membranous expression mainly in renal tubules and parathyroid gland.

IMMUNOHISTOCHEMISTRY DATA RELIABILITY

Data reliability description	Antibody staining consistent with RNA expression data.
Reliability score	● Enhanced
Antibodies	HPA005980, HPA064792

SHOW MORE

RNA AND PROTEIN EXPRESSION SUMMARY



LRP2

RNA SUMMARY - Normalized

Consensus dataset

Brain regional consensus dataset

Часть 3 – Human protein atlas

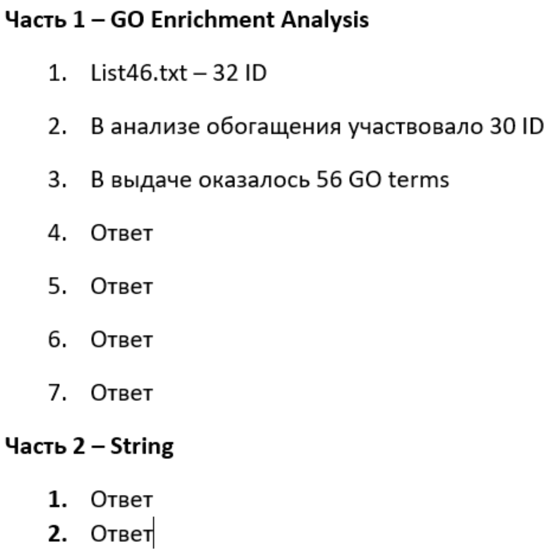
В отчете укажите:

1. Какой ID вы выбрали?
2. Опишите ваш ID (Gene summary, слайд 24)
3. Специфичен ли ваш ID для какого-либо участка мозга (вкладка Summary – Brain atlas, приведите картинку, слайд 25)?
4. Обладает ли ваш ID специфической субклеточной локализацией (вкладка Cell, слайд 26, приведите картинку)?
5. Различается ли ваш ID по экспрессии РНК и белка (вкладка TISSUE, приведите картинку, слайд 27)
6. В каких тканях экспрессируется ваш ID, если рассматривать его как РНК (consensus dataset, слайд 28, приведите картинку)?

Оформление отчета

Убедительная просьба сохранить структуру отчета для облегчения проверки,
т.е. разделение отчета на части, а внутри каждой части на вопросы
Обязательным является выполнения любых ДВУХ частей из трех
При выполнении трех частей будут начислены дополнительные баллы
Для получения зачета по этому заданию вам необходимо на каждый выделенный пункт отчета выбранных частей написать хоть что-то!!!

Если где-то запутаетесь – спрашивайте
Я постаралась сделать максимально подробные подсказки
Надеюсь, вам будет интересно



Часть 1 – GO Enrichment Analysis

1. List46.txt – 32 ID
2. В анализе обогащения участвовало 30 ID
3. В выдаче оказалось 56 GO terms
4. Ответ
5. Ответ
6. Ответ
7. Ответ

Часть 2 – String

1. Ответ
2. Ответ