

# ДОМАШНЕЕ ЗАДАНИЕ К ЛЕКЦИИ №5

Анастасия Жарикова - 2022

# Набор генов

- Возьмите один из набора генов из списка

[https://makarich.fbb.msu.ru/suvorova/students/lists\\_go/](https://makarich.fbb.msu.ru/suvorova/students/lists_go/)

Чтобы файлы не повторялись у разных студентов, предлагается отсчитать N-ый файл из списка, взяв за N день вашего рождения

**ДЕДЛАЙН – 10 апреля в 23:59**

# Варианты задания

Выбрав набор генов, вы можете пойти двумя путями

Вариант 1: анализировать набор генов

Вариант 2: выбрать из набора один ген и изучать только его

Можно сделать оба варианта

Обратите внимание, что разные базы данных ориентированы на гены или на белки. В нашей учебной задаче мы будем считать, что наш список состоит из ID генов или белков, в зависимости от того, какой базой будем пользоваться. В жизни нужно четко понимать, с чем мы работаем.

В качестве ответа напишите отчет на 3-5 страниц, включающий описание ваших действий, полученные результаты в виде списков и\или картинок, а также обсуждение полученных результатов.

Отчет отправьте на почту [azharikova89@gmail.com](mailto:azharikova89@gmail.com), указав ваши имя, фамилию и название дисциплины в теме письма и названии файла.

# Вариант 1

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

The screenshot shows the STRING database search interface. On the left, a sidebar contains navigation options: 'Protein by name', 'Protein by sequence', 'Multiple proteins' (highlighted with a red arrow), 'Multiple sequences', 'Proteins with Values/Ranks <sup>New</sup>', 'Organisms', 'Protein families ("COGs")', 'Examples', and 'Random entry'. The main area is titled 'SEARCH' and 'Multiple Proteins by Names / Identifiers'. It features a 'List Of Names:' input field with a list of protein identifiers: COQ6, CUBN, CYP24A1, CYP27A1, CYP27B1, and CYP2E1. Below this is an option to '... or, upload a file:' with a 'Browse ...' button. The 'Organism:' dropdown menu is set to 'Homo sapiens' (indicated by a red arrow). At the bottom, a large blue 'SEARCH' button is visible (indicated by a red arrow).

# STRING

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**

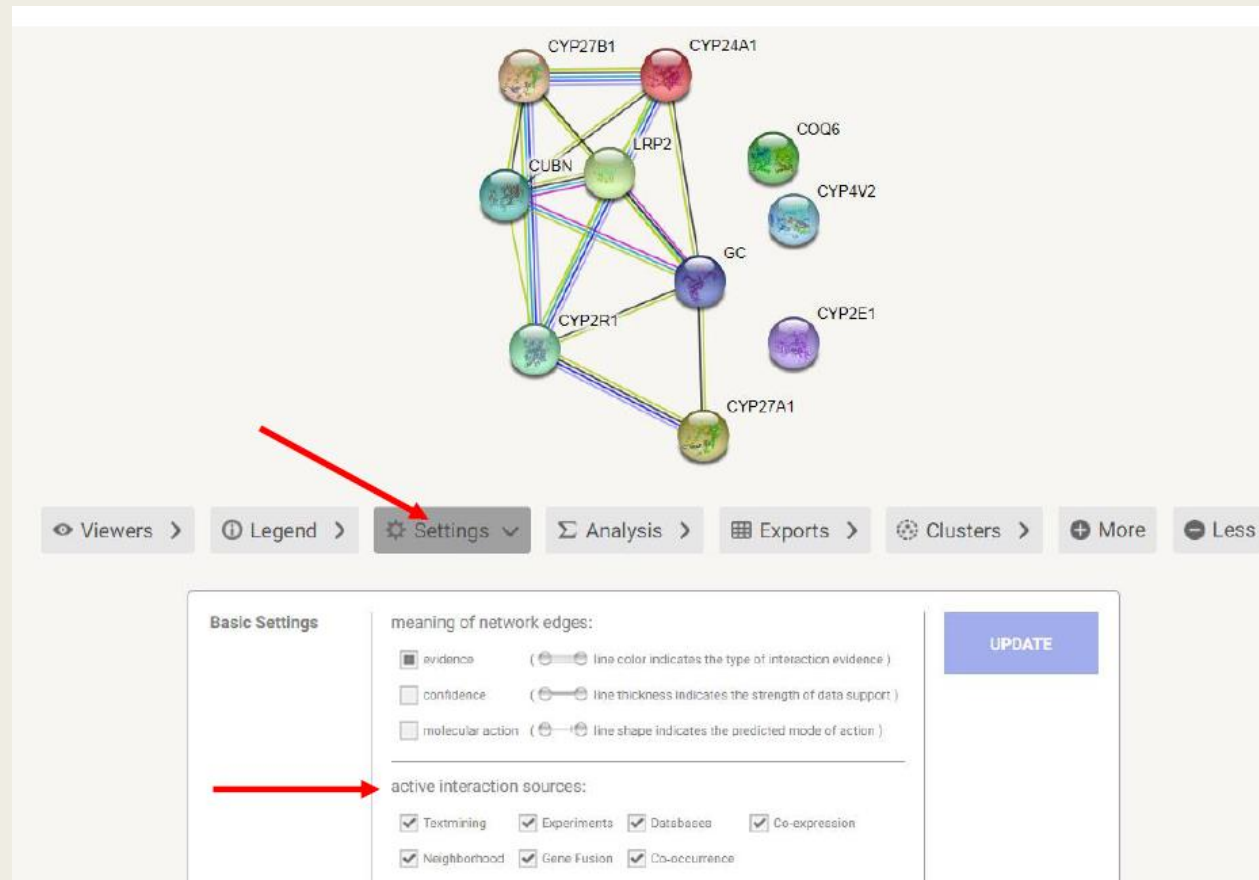
'CUBN':

- 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

# STRING

В результате вы получите граф

Убедитесь, что в настройках включены все типы связей



The screenshot displays the STRING database interface. At the top, a network graph shows interactions between several proteins: CYP27B1, CYP24A1, COQ6, CYP4V2, CYP2E1, CYP27A1, GC, LRP2, CUBN, and CYP2R1. The nodes are represented by colored spheres, and the edges are lines of varying colors and thicknesses. Below the graph is a navigation bar with buttons for Viewers, Legend, Settings, Analysis, Exports, Clusters, More, and Less. A red arrow points to the Settings button. Below the navigation bar is the Settings panel, which is divided into two sections: Basic Settings and meaning of network edges. The Basic Settings section has a red arrow pointing to the active interaction sources section. The meaning of network edges section explains the meaning of network edges: evidence (line color indicates the type of interaction evidence), confidence (line thickness indicates the strength of data support), and molecular action (line shape indicates the predicted mode of action). The active interaction sources section lists several sources that are checked: Textmining, Experiments, Databases, Co-expression, Neighborhood, Gene Fusion, and Co-occurrence. An UPDATE button is located to the right of the settings panel.

Viewers > Legend > **Settings** > Analysis > Exports > Clusters > More > Less

**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)

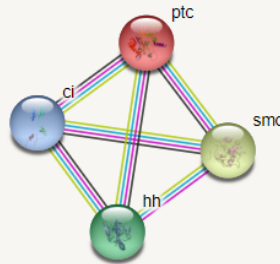
active interaction sources:

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

UPDATE

# STRING

Обратите внимание на раздел **Functional enrichments in your network**



3 из 4 генов нашей выборки принадлежат категории Anterior/posterior lineage restriction, imaginal disk

Viewers > Legend > Settings > **Analysis** > Exports > Clusters > More > Less >

## Network Stats

number of nodes: 4  
number of edges: 6  
average node degree: 3  
avg. local clustering coefficient: 1

expected number of edges: 0  
PPI enrichment p value: 7.22e-07  
*your network has significantly more interactions than expected (what does that mean?)*

## Functional enrichments in your network

[explain columns](#)

Biological Process (Gene Ontology)				
GO-term	description	count in network	strength	false discovery rate
GO:0048099	Anterior/posterior lineage restriction, imaginal disc	3 of 4	3.42	4.81e-07
GO:0035217	Labial disc development	2 of 4	3.24	0.00020
GO:0007224	Smoothed signaling pathway	4 of 12	3.06	7.18e-09
GO:0001746	Bolwigs organ morphogenesis	3 of 9	3.06	1.73e-06
GO:0060914	Heart formation	2 of 6	3.06	0.00032

(more ...)

Molecular Function (Gene Ontology)				
GO-term	description	count in network	strength	false discovery rate
GO:0005113	Patched binding	2 of 5	3.14	0.0024
GO:0070273	phosphatidylinositol-4-phosphate binding	2 of 6	3.06	0.0024
GO:0005102	Signaling receptor binding	3 of 268	1.59	0.0178

# STRING

Обратите внимание на одну из онтологий, отмеченных красным  
Можно ли что-то сказать о наборе генов, опираясь на полученные обогащения?

Есть ли категории, куда попали все гены из списка?

В какую категорию попало наибольшее количество генов из списка?

Biological Process (Gene Ontology)	<a href="#">download</a>	<i>42 GO-terms significantly enriched; file-format: tab-delimited</i>
Molecular Function (Gene Ontology)	<a href="#">download</a>	<i>3 GO-terms significantly enriched; file-format: tab-delimited</i>
Cellular Component (Gene Ontology)	<a href="#">download</a>	<i>one single GO-term is enriched; file-format: tab-delimited</i>
Reference publications (PubMed)	<a href="#">download</a>	<i>1420 publications significantly enriched; file-format: tab-delimited</i>
Local network cluster (STRING)	<a href="#">download</a>	<i>3 clusters significantly enriched; file-format: tab-delimited</i>
KEGG Pathways	<a href="#">download</a>	<i>one single pathway is enriched; file-format: tab-delimited</i>
Reactome Pathways	<a href="#">download</a>	<i>6 pathways significantly enriched; file-format: tab-delimited</i>
WikiPathways	<a href="#">download</a>	<i>2 pathways significantly enriched; file-format: tab-delimited</i>
Annotated Keywords (UniProt)	<a href="#">download</a>	<i>2 keywords significantly enriched; file-format: tab-delimited</i>
<b>All enriched terms</b> (without PubMed)	<a href="#">download</a>	<i>60 enriched terms in 8 categories; file-format: tab-delimited</i>
Selected terms only	<a href="#">download</a>	<i>no enriched terms selected (click on any term above to select)</i>



# PANTHER

- Проанализируйте ваш набор генов с помощью сервиса

<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:  separate IDs by a space or comma  
[Supported IDs](#)

Upload IDs:  Файл не выбран

[File format](#)

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:   
 VCF File Flanking region:   Search Enhancer Data

2. **Select organism.**

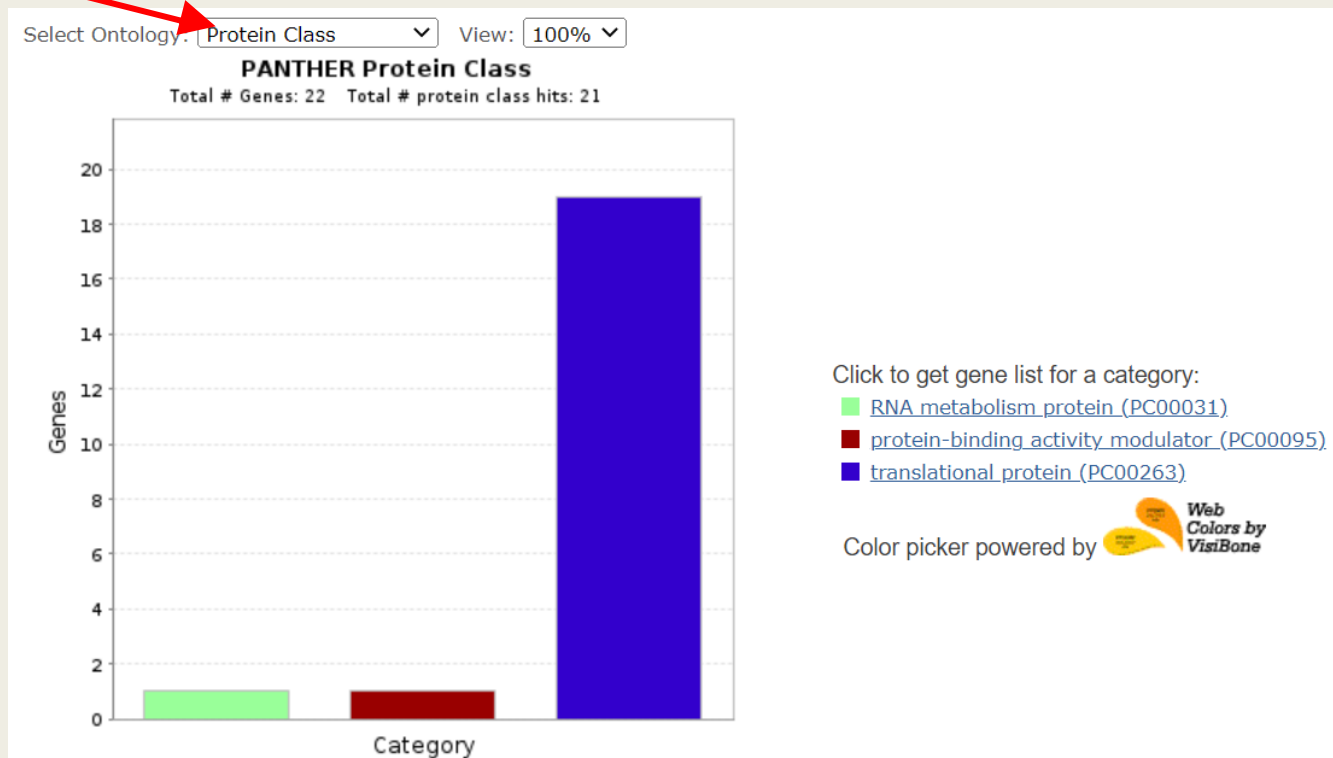
Mus musculus  
Rattus norvegicus  
Gallus gallus  
Danio rerio

3. **Select Analysis.**

Functional classification viewed in gene list  
 Functional classification viewed in graphic charts  Bar chart  Pie chart

# PANTHER

- Опишите, каким классам принадлежат белки из вашего списка

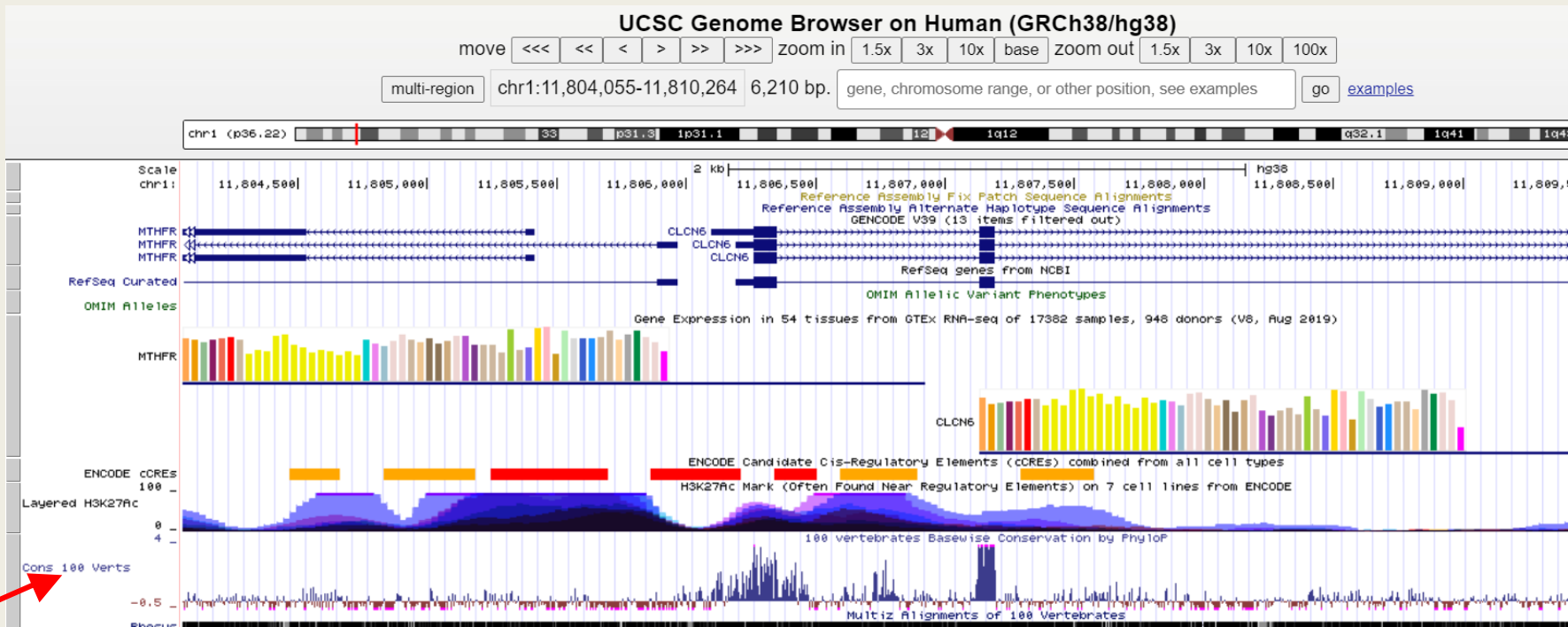


# Вариант 2

- Из выбранного вами списка генов выберите ОДИН ID

# Геномный браузер

- Найдите выбранный ген в геномной браузере:  
<https://genome.ucsc.edu/>



Сколько транскриптов аннотировано для данного гена?  
Обратите внимание на трек консервативности, на какие элементы гена  
приходятся пики трека консервативности?

# GeneCards

- Найдите выбранный ген в GeneCards
- Опишите локализацию и функции гена
- Выберите любой дополнительный раздел и опишите представленную для выбранного гена информацию
- Найдите на странице GeneCards ссылку на Human Protein Atlas.
- Перейдите на вкладку SUBCELL. В каком клеточном компартменте детектирована секреция белка?
- Перейдите на вкладку TISSUE. В каких органах детектированы экспрессия гена, а в каких секреция белка?
- Не забывайте иллюстрировать ответ соответствующими картинками
- Если про выбранный вами ID мало информации, то выберите другой гениз списка

# ВАЖНО

- Если у вас возникли вопросы, не стесняйтесь писать мне на почту!