

Практикум 6. Базы данных KEGG, GO и другие

Краткое описание входных данных

Входные данные представляют собой список из 5 идентификаторов (ID) генов человека: ALDH1A1, TKFC, GLYCK, ALDOB, KHK. Список генов доступен по ссылке: [list51.txt](#). Все перечисленные гены кодируют белки, участвующие в метаболизме углеводов. Анализ функций отдельных генов позволяет предположить, что весь список представляет собой функционально связанную группу, участвующую в едином метаболическом пути — метаболизме фруктозы:

- KHK (кетогексокиназа) катализирует первый этап метаболизма фруктозы — превращение фруктозы в фруктозо-1-фосфат.
- ALDOB (альдолаза B) расщепляет фруктозо-1-фосфат на глицеральдегид и дигидроксиацетонфосфат.
- TKFC (триокиназа) фосфорилирует глицеральдегид, образующийся при расщеплении фруктозо-1-фосфата.
- GLYCK (глицераткиназа) участвует в метаболизме глицерата, образуя 3-фосфоглицерат.
- ALDH1A1 (альдегиддегидрогеназа 1A1) участвует в дальнейших превращениях альдегидов, образующихся в ходе этих реакций.

Таким образом, все гены так или иначе связаны с метаболизмом фруктозы и родственных соединений.

Групповой анализ: GO Enrichment Analysis (geneontology.org)

Какие задачи можно решать?

Задача 1. С помощью GO Enrichment Analysis можно установить, что гены из списка статистически значимо связаны с биологическим процессом *fructose catabolic process* (GO:0006000). Анализ показал, что гены из списка попали в этот термин с экстремально низким значением FDR.

Задача 2. Сервис позволяет выявить более узкие процессы, например, *fructose catabolic process to hydroxyacetone phosphate and glyceraldehyde-3-phosphate* (GO:0006533), который является самым специфичным термином в выдаче.

Задача 3. GO-анализ позволяет оценить силу обогащения через Fold Enrichment. Для большинства терминов, связанных с фруктозой, этот показатель превышает 100, что означает, что гены списка встречаются в этих процессах более чем в 100 раз чаще, чем ожидалось бы случайно.

Общая «тема» списка ID по результатам обогащения

Параметры запуска анализа:

- Сервис: geneontology.org.
- Метод: Fisher's Exact Test (точный тест Фишера).
- Тип аннотации: GO biological process complete.
- Поправка на множественное тестирование: FDR (Benjamini-Hochberg).
- Организм: Homo sapiens.
- Ссылка на таблицу: [analysis](#)

Описание лучших находок (Рис. 1):

	Homo sapiens (REF)		upload_1 (▼ Hierarchy NEW! ⓘ)				
GO biological process complete	#	# expected	Fold Enrichment	+/-	raw P value	FDR	
fructose catabolic process to hydroxyacetone phosphate and glyceraldehyde-3-phosphate	3	3	.00	> 100	+	6.88E-12	3.16E-08
↳ fructose catabolic process	3	3	.00	> 100	+	6.88E-12	4.74E-08
↳ hexose catabolic process	50	3	.01	> 100	+	1.34E-07	2.31E-04
↳ hexose metabolic process	122	4	.03	> 100	+	5.85E-09	1.34E-05
↳ monosaccharide metabolic process	139	4	.03	> 100	+	9.91E-09	1.95E-05
↳ carbohydrate metabolic process	356	4	.09	46.25	+	4.34E-07	5.98E-04
↳ small molecule metabolic process	1386	5	.34	14.85	+	1.38E-06	1.72E-03
↳ monosaccharide catabolic process	54	3	.01	> 100	+	1.70E-07	2.60E-04
↳ small molecule catabolic process	346	3	.08	35.69	+	4.59E-05	4.52E-02
↳ carbohydrate catabolic process	110	3	.03	> 100	+	1.47E-06	1.69E-03
↳ fructose metabolic process	14	4	.00	> 100	+	6.70E-13	9.21E-09
↳ glyceraldehyde-3-phosphate metabolic process	12	3	.00	> 100	+	1.51E-09	4.17E-06
↳ carbohydrate derivative metabolic process	858	4	.21	19.19	+	1.45E-05	1.54E-02
↳ aldehyde metabolic process	68	4	.02	> 100	+	5.44E-10	1.87E-06

Рис.1 Лучшие находки

В выдаче оказалось 14 значимых GO-терминов (FDR < 0.05). Все они связаны с метаболизмом углеводов, моносахаридов и альдегидов. Наиболее значимые находки:

1. GO:0006533 - fructose catabolic process to hydroxyacetone phosphate and glyceraldehyde-3-phosphate

- Включает 3 гена из 5.
- Fold Enrichment > 100, FDR = 3.16×10^{-8} .
- Это самый специфичный термин, указывающий на точный путь расщепления фруктозы.

2. GO:0006000 - fructose catabolic process

- Включает 3 гена из 5.
- Fold Enrichment > 100, FDR = 4.74×10^{-8} .
- Подтверждает, что список генов обогащен ферментами катаболизма фруктозы.

3. GO:0006007 - hexose catabolic process

- Включает 3 гена из 5.
- Fold Enrichment > 100, FDR = 2.31×10^{-4} .
- Более общий термин, охватывающий катаболизм гексоз.

Также в выдаче присутствуют термины fructose metabolic process (4 гена, FDR = 9.21×10^{-9}) и aldehyde metabolic process (4 гена, FDR = 1.87×10^{-6}), что указывает на участие `ALDH1A1` в метаболизме альдегидов.

Обсуждение:

Анализ обогащения продемонстрировал, что исследуемый список генов характеризуется выраженным обогащением терминами, ассоциированными с метаболизмом фруктозы. Наличие генов в термине fructose catabolic process с экстремально высоким уровнем значимости и Fold Enrichment > 100 свидетельствует о неслучайном характере выборки. Список охватывает ключевые этапы метаболизма фруктозы. Такое сочетание генов может помочь в изучении наследственной непереносимости фруктозы и других нарушений углеводного обмена.

Индивидуальный анализ: Human Protein Atlas на примере гена ALDH1A1

Какие задачи можно решать?

Задача 1. Определение тканевой специфичности экспрессии белка ALDH1A1. С помощью HPA можно установить, в каких тканях человека наблюдается повышенный уровень экспрессии гена ALDH1A1 и соответствующего белка.

Задача 2. Установление субклеточной локализации белка ALDH1A1. HPA позволяет определить, в каком компартменте клетки локализуется белок ALDH1A1, что необходимо для понимания его функциональной роли.

Задача 3. Оценка ассоциации гена ALDH1A1 с заболеваниями. HPA содержит информацию о связи изменений экспрессии ALDH1A1 с различными патологическими состояниями.

Иллюстрация решения задачи на примере ALDH1A1

Для анализа был выбран ген ALDH1A1 (альдегиддегидрогеназа 1A1, ENSG00000165092), входящий в исходный список. На странице гена в HPA были получены следующие данные.

- Тканевая экспрессия (Tissue). Согласно графику «Protein expression overview» (Рис. 2), белок ALDH1A1 характеризуется наиболее высоким уровнем экспрессии в печени (Liver). В таких тканях, как эпидидимис, щитовидная железа и желудок, уровень экспрессии оценивается как средний. Сводная категория RNA (consensus) подтверждает: Tissue enhanced (Liver) (Рис. 3). Это согласуется с ролью фермента в метаболизме альдегидов, который наиболее активен в печени.

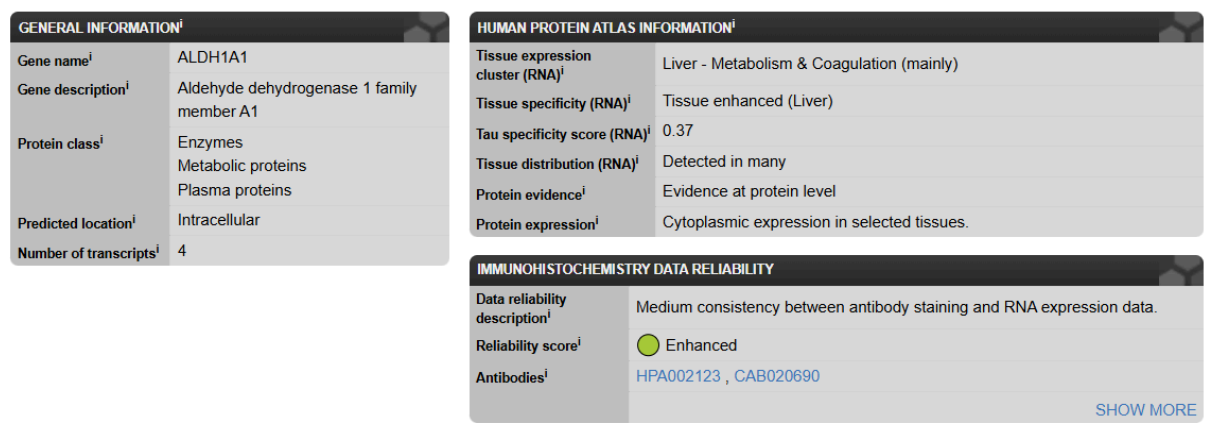


Рис 2.Тканевая экспрессия белка ALDH1A1 по данным Human Protein Atlas.

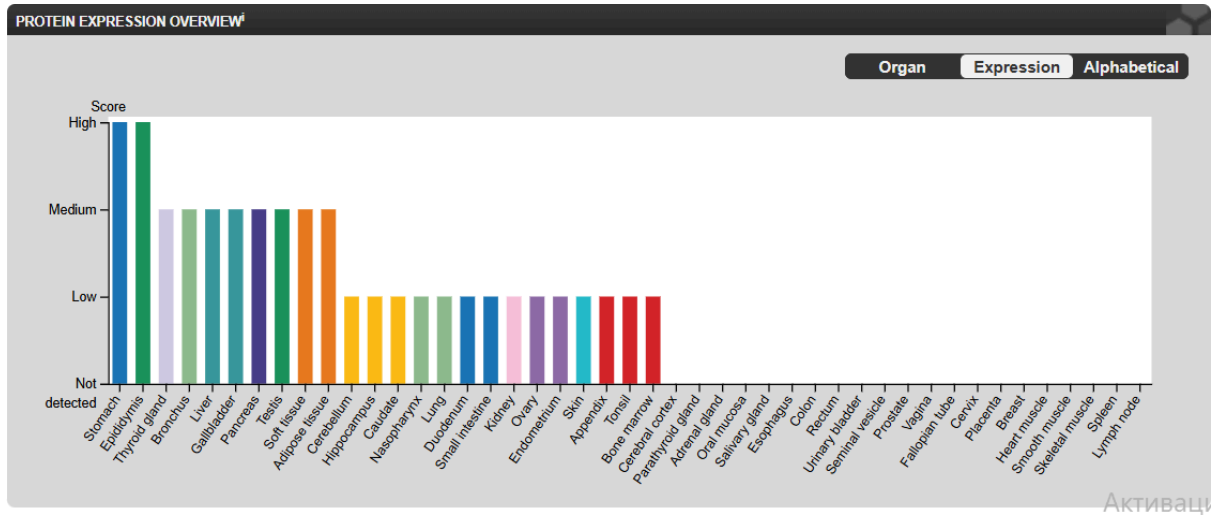
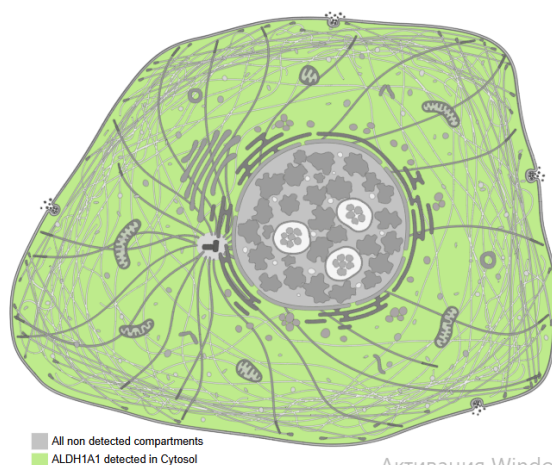


Рис 3.Сводная информация о гене ALDH1A1 в базе Human Protein Atlas.

Субклеточная локализация (Subcell). В разделе «Subcell» указано, что белок ALDH1A1 локализуется преимущественно в цитозоле (Cytosol) (Рис. 4). Это ожидаемый результат, так как фермент участвует в метаболических процессах, протекающих в цитоплазме.

GENERAL INFORMATION ¹	
Gene name ¹	ALDH1A1
Gene description ¹	Aldehyde dehydrogenase 1 family member A1
Protein class ¹	Enzymes Metabolic proteins Plasma proteins
Predicted location ¹	Intracellular
Number of transcripts ¹	4

HUMAN PROTEIN ATLAS INFORMATION ¹	
Main location ¹	Localized to the Cytosol (enhanced)
Single-cell variation ¹	Single-cell variation in protein expression observed.
Reliability score ¹	Enhanced
Antibodies ¹	HPA002123



Активация Windows

Рис 4. Субклеточная локализация белка ALDH1A1.

Связь с заболеваниями (Cancer). В таблице представлены данные по онкологическим заболеваниям (Рис. 5). ALDH1A1 ассоциирован с такими типами рака, как Liver cancer (рак печени), Rhabdoid (рабдоидная опухоль) и Sarcoma (саркома) на уровне клеточных линий, а также с Kidney Renal Papillary Cell Carcinoma и Liver Hepatocellular Carcinoma по данным TCGA. Эти данные подтверждают участие фермента в процессах пролиферации клеток.

GENERAL INFORMATION ¹	
Gene name ¹	ALDH1A1
Gene description ¹	Aldehyde dehydrogenase 1 family member A1
Protein class ¹	Enzymes Metabolic proteins Plasma proteins
Predicted location ¹	Intracellular
Number of transcripts ¹	4

HUMAN PROTEIN ATLAS INFORMATION ¹	
RNA category ¹	Consensus (human tissue): Tissue enhanced (Liver) Detected in many Cancer enhanced (Liver cancer, Rhabdoid, Sarcoma) Cell line (cancer): Detected in many TCGA (cancer tissue): Cancer enhanced (Kidney Renal Papillary Cell Carcinoma, Liver Hepatocellular Carcinoma) Detected in all
Protein evidence ¹	Evidence at protein level
Protein expression normal tissue ¹	Cytoplasmic expression in selected tissues.

IMMUNOHISTOCHEMISTRY DATA RELIABILITY	
Data reliability description ¹	Medium consistency between antibody staining and RNA expression data.
Reliability score - normal tissues ¹	Enhanced
Antibodies ¹	HPA002123 , CAB020690

Рис 5. Ассоциация гена ALDH1A1 с онкологическими заболеваниями по данным Human Protein Atlas.

Соотнесение с результатами группового анализа

В групповом анализе ген ALDH1A1 был идентифицирован как компонент пути aldehyde metabolic process (GO:0006081, FDR = 1.87×10^{-6}), который связан с метаболизмом фруктозы. Индивидуальный анализ в HPA показывает, что этот фермент наиболее активен в печени — органе, который является центральным местом метаболизма фруктозы. Локализация ALDH1A1

в цитозоле соответствует его роли в метаболических реакциях. Ассоциация с различными типами рака, особенно с гепатоцеллюлярной карциномой, указывает на возможную роль ALDH1A1 в опухолевой прогрессии, что может быть связано с изменениями в метаболизме углеводов в раковых клетках.

Обсуждение

Индивидуальный анализ ALDH1A1 в базе Human Protein Atlas продемонстрировал ценность сопоставления данных об экспрессии и субклеточной локализации белка с результатами группового обогащения. Если GO-анализ позволяет ответить на вопрос, какие биологические процессы вовлечены, то HPA уточняет, где и в каком тканевом контексте реализуется функция конкретного гена. Тканевая специфичность (печень) согласуется с физиологической ролью ALDH1A1 в метаболизме альдегидов, образующихся при распаде фруктозы. Локализация в цитозоле подтверждает участие фермента в цитозольных метаболических путях. Ассоциация с онкологическими заболеваниями расширяет представление о системных эффектах ALDH1A1 и открывает перспективы для дальнейших исследований.

Итоговый вывод

В ходе работы был проанализирован список из 5 генов человека. Групповой анализ обогащения (GO) показал, что все они так или иначе связаны с метаболизмом фруктозы: его расщеплением и дальнейшими превращениями продуктов. Индивидуальный анализ гена ALDH1A1 подтвердил его ключевую роль в метаболизме альдегидов и выявил дополнительные особенности: высокую экспрессию в печени, локализацию в цитозоле и связь с некоторыми типами рака. Таким образом, сочетание группового и индивидуального подходов позволило охарактеризовать функцию генов на разных уровнях — от общего пути до конкретного белка в конкретной ткани. Полученные результаты показывают, что исследуемый список не случаен, а представляет собой функционально связанную группу, изучение которой может быть полезно для понимания роли метаболизма фруктозы в норме и при патологии.

Простите пожалуйста, что я сделала не то, не отчисляйте меня пожалуйста, я буду делать все вовремя 😞 😞 😞