

Практикум 6. Базы данных KEGG, GO и другие

1. Описание входных данных

Входные данные представляет список из 5 генов человека (*Homo sapiens*), заданных в символьном формате. Каждый ген кодирует белок, участвующий в метаболических процессах.

Список генов: ALDH1A1, TKFC, GLYCTK, ALDOB, KHK.

2а. Решение задач с помощью сервиса DAVID

1. Определить, какие биологические пути (KEGG pathways) статистически перепредставлены в вашем списке генов.
2. Выявить обогащённые термины Gene Ontology (GO), связанные с биологическими процессами, молекулярными функциями или клеточными компонентами.
3. Кластеризовать функционально связанные аннотации.

2б. Общая «тема» списка по результатам обогащения

- **Метод:** Modified Fisher Exact Test (EASE Score) — модифицированный точный тест Фишера, в котором для консервативности из каждого термина исключается один ген перед расчётом p-значения.

Поправка на множественное тестирование: Бенджамини-Хохберга для контроля ожидаемой доли ложных отклонений.

- В таблице 1 представлены результаты группового анализа, выполненного с помощью сервиса DAVID. Не пустая выдача таблицы предполагает функциональное сходство белков данных генов.

Category	Term	Count	List Total	Pop Hits	Pop Total	P-Value	Benjamini	User Ids
GOTERM_BP_DIRECT	fructose catabolic process to hydroxyacetone phosphate and	3	5	3	19172	9,79E-08	0,00000304	TKFC, GLYCTK, ALDOB
KEGG_PATHWAY	Fructose and mannose metabolism	3	5	35	9375	8,09E-05	0,00097	TKFC, ALDOB, KHK
KEGG_PATHWAY	Metabolic pathways	5	5	1579	9375	0,000802	0,00358	TKFC, GLYCTK, ALDH1A1, ALDOB, KHK
KEGG_PATHWAY	Carbon metabolism	3	5	116	9375	0,000896	0,00358	TKFC, GLYCTK, ALDOB
GOTERM_BP_DIRECT	fructose metabolic process	2	5	10	19172	0,00208	0,0323	ALDOB, KHK
GOTERM_CC_DIRECT	extracellular exosome	4	5	2236	20299	0,0049	0,0344	TKFC, ALDH1A1, ALDOB, KHK
GOTERM_CC_DIRECT	cytosol	5	5	5849	20299	0,00689	0,0344	TKFC, GLYCTK, ALDH1A1, ALDOB, KHK
KEGG_PATHWAY	Pentose phosphate pathway	2	5	31	9375	0,0132	0,0395	GLYCTK, ALDOB
GOTERM_MF_DIRECT	fructose binding	2	5	8	19319	0,00166	0,0447	ALDOB, KHK
KEGG_PATHWAY	Glycerolipid metabolism	2	5	65	9375	0,0275	0,0659	TKFC, GLYCTK
UP_KW_MOLECULAR_FUNCTION	Kinase	3	5	752	12067	0,0214	0,107	TKFC, GLYCTK, KHK
UP_KW_MOLECULAR_FUNCTION	Lyase	2	5	181	12067	0,0587	0,147	TKFC, ALDOB
GOTERM_MF_DIRECT	ATP binding	3	5	1546	19319	0,0344	0,465	TKFC, GLYCTK, KHK
UP_KW_LIGAND	Nucleotide-binding	4	5	1891	7019	0,0624	0,624	TKFC, GLYCTK, ALDH1A1, KHK

Таблица 1. Результаты группового анализа, отсортированные по p-value с поправкой.

- В таблице 1 оказалось 11 уникальных терминов, прошедших порог значимости: 4 KEGG, 3 GO BP, 3 GO CC, 3 GO MF, 2 UP_KW Molecular Function, 1 UP_KW Ligand.

Несмотря на малый размер выборки, совпадение результатов по нескольким независимым категориям аннотаций делает вывод о наличии общей «темы».

- **Лучшие находки:**

Наиболее значимым термином оказался GOTERM_BP_DIRECT: fructose catabolic process to hydroxyacetone phosphate and glyceraldehyde-3-phosphate (Count = 3; Benjamini = 3.04e-06), объединяющий TKFC, GLYCTK и ALDOB.

Второй по значимости — путь KEGG Fructose and mannose metabolism (Count = 3; Benjamini = 8.09e-05), включающий ALDOB, КНК и TKFC.

Все пять генов списка попали в путь Metabolic pathways (Benjamini = 0.000802) и в термин cytosol (Benjamini = 0.00689), что подтверждает их принадлежность к цитозольному метаболизму углеводов.

Дополнительно выявлены пути Carbon metabolism (TKFC, GLYCTK, ALDOB), Pentose phosphate pathway (GLYCTK, ALDOB) и термины, связанные с киназной активностью (Kinase, ATP binding), что отражает участие КНК, TKFC и GLYCTK в фосфорилировании субстратов.

- **Иллюстрации результатов:**

После проведения анализа были просмотрены карты KEGG с помощью Pathway Viewer. Они содержат гены, участвующие в определенных процессах, и являются наглядной иллюстрацией их биологической функции.

На Рис. 1 приведен пример одного из KEGG пути: Pentose phosphate pathway.

Для него было обнаружено участие 2 из 5 генов: ALDOB и GLYCTK.

Пентозофосфатный цикл - это альтернативный путь окисления глюкозы, главная задача которого - производить NADPH и рибозо-5-фосфат, а не АТФ. В окислительной фазе глюкозо-6-фосфат превращается в рибулозо-5-фосфат с выделением 2 NADPH и CO₂, а в неокислительной фазе образуется рибозо-5-фосфат для синтеза нуклеотидов, а избыток пентоз возвращается в гликолиз. NADPH используется для восстановительных синтезов (жирные кислоты,

холестерин) и защиты клеток от окислительного стресса, особенно активно цикл идёт в печени, жировой ткани, эритроцитах и коре надпочечников.

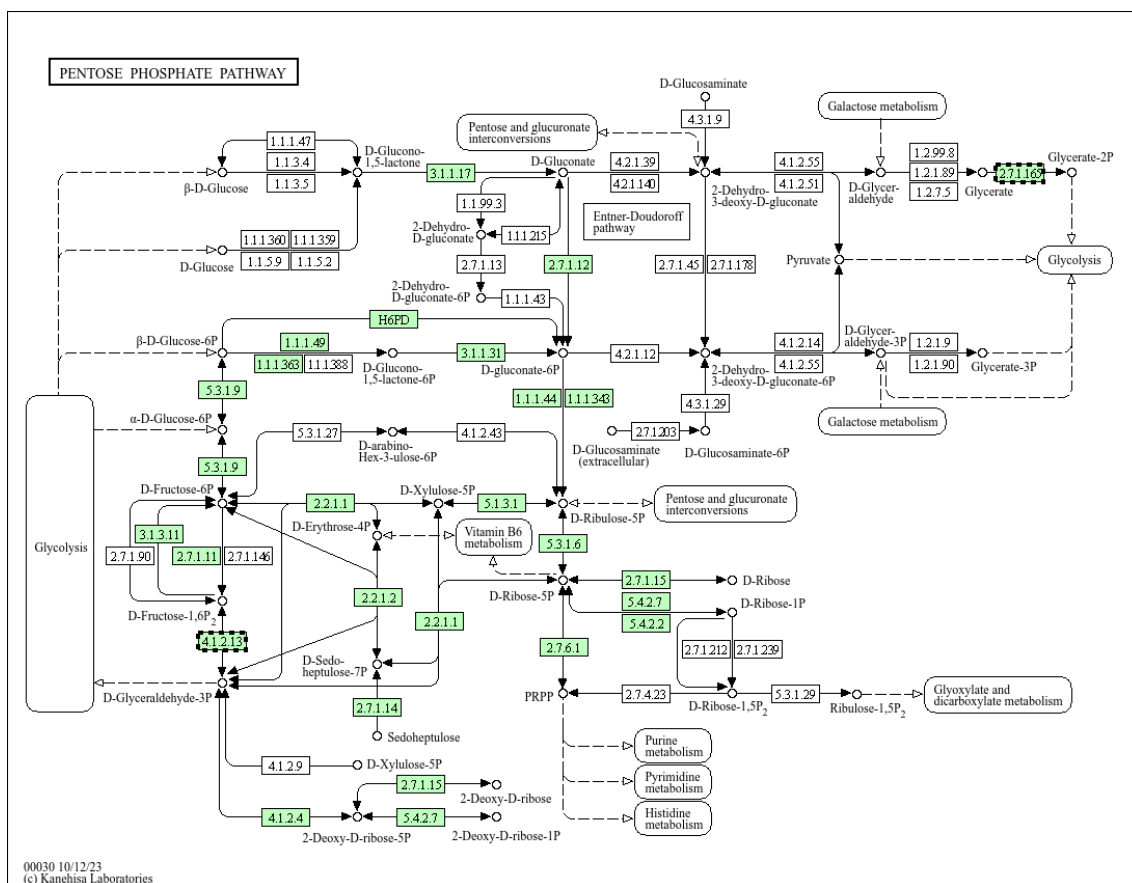


Рис. 1 Карта KEGG путей пентозофосфатного цикла с участием анализируемых генов

На Рис. 1 участие продуктов генов в реакциях пентозофосфатного цикла показано пунктирной рамочкой.

Ген GLYCTK кодирует glycerate kinase, которая катализирует реакцию фосфорилирования глицерата.

Ген ALDOB кодирует aldolase, fructose-bisphosphate В, которая расщепляет фруктозо-1,6-бисфосфат на две трёхуглеродные молекулы: глицеральдегид-3-фосфат и дигидроксиацетонфосфат.

Таким образом, просмотр карты KEGG позволяет удостовериться в правильно найденных путях, в которых принимают участие продукты генов, а также дает понимание их функциональной роли.

3а. Решение задач с помощью сервиса Human Protein Atlas

1. Определение, в каких тканях человека преимущественно экспрессируется ген, и подтверждение его тканеспецифичной локализации.
2. Верификация внутриклеточной локализации белка и сопоставление её с функциональной ролью.
3. Выявление связи гена с наследственными заболеваниями.

3б. Тканевая экспрессия ALDOB

По данным Human Protein Atlas белок преимущественно экспрессируется в четырёх тканях: двенадцатиперстная кишка, почка, печень и тонкий кишечник. RNA-данные на Рис. 2 показывают наибольшую экспрессию в печени около 10 500 nTPM (нормализованные транскрипты на миллион).

3с. Соотношение характеристик с результатами пункта 2

Тканевая специфичность ALDOB, установленная по данным Human Protein Atlas, согласуется с результатами функционального обогащения в DAVID, где ALDOB аннотирован терминами «Fructose and mannose metabolism» (Benjamini = 8.09×10^{-5}) и «Fructose metabolic process» (Benjamini = 0.00208). Поскольку печень, почка и кишечник являются основными сайтами метаболизма фруктозы у человека, наблюдаемая тканевая локализация фермента подтверждает, что выявленная на групповом уровне тема отражает реальный метаболический процесс.

Распределение РНК-экспрессии ALDOB в клетках печени на Рис. 3 демонстрирует гепатоцит-специфичную экспрессию, которая более чем на порядок выше в гепатоцитах (≈ 3800 nCPM) по сравнению со всеми остальными типами клеток (холангиоциты, клетки Купфера, звездчатые клетки, иммунные клетки), где она практически отсутствует. Это подтверждает, что метаболизм фруктозы, опосредованный ALDOB, локализован именно в гепатоцитах.

3д. Иллюстрация результатов

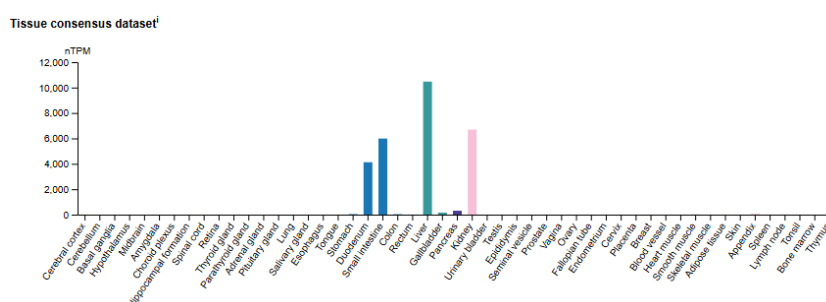


Рис. 2 Тканевой консенсусный набор по данным о РНК экспрессии гена ALDOB по данным Human Protein Atlas

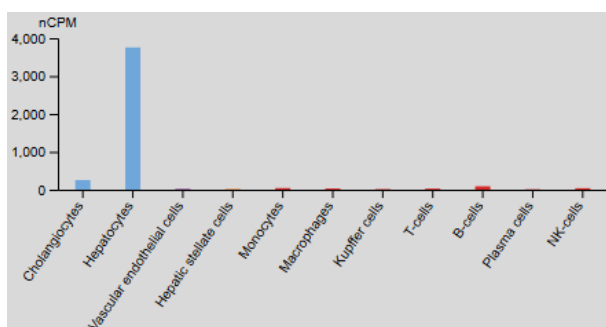


Рис. 3 Распределение РНК экспрессии гена ALDOB среди групп клеток печени по данным Human Protein Atlas

3е. Обсуждение результатов

Индивидуальный анализ ALDOB в Human Protein Atlas подтверждает результаты группового обогащения в DAVID: тканевая специфичность фермента (печень, почка, кишечник) точно соответствует органам, где происходит метаболизм фруктозы. Это означает, что выявленная на групповом уровне тема отражает реальный биохимический процесс.

ALDOB занимает центральное положение в этом процессе, катализируя расщепление фруктозо-1-фосфата — вторую ключевую реакцию метаболизма фруктозы. Совпадение независимых источников — статистического (DAVID) и экспериментального (НРА) — повышает надёжность вывода о функциональной согласованности списка.