

Базы данных KEGG и GO

1. Вступление

В рамках задания данного практикума мне был предоставлен [список из 59 ID генов](#). Моя задача – выяснить с помощью специальных баз данных и сервисов, что это за гены и что их может объединять.

2. Групповой анализ полного списка ID

GO enrichment analysis

В качестве сервиса для анализа обогащения был использован [Enrichr](#), позволяющий проводить GO enrichment analysis для списка генов человека. GO (Gene Ontology) представляет собой систему функциональной аннотации генов и белков, включающую три основные категории: biological process, molecular function, cellular component.

- Сервис GO enrichment analysis позволяет выявлять биологические процессы, статистически обогащенные в списке генов.
- Сервис GO enrichment analysis позволяет определять метаболические пути, связанные с исследуемым набором генов.
- Сервис GO enrichment analysis позволяет визуализировать функциональные связи между генами и аннотациями, реконструировать общую биологическую роль исследуемого набора.

Статистическая значимость определялась с использованием Fisher exact test. Для оценки статистической значимости использовалась поправка Benjamini–Hochberg на множественное тестирование. Анализ проводился по базе GO Biological Process 2026.

Для исследуемого списка генов был получен 501 статистически значимый GO-термины, где наиболее значимыми оказались 10 находок, показанные на Рис. 1.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Icosanoid Metabolic Process (GO:0006690)	1.248e-73	6.251e-71	1099.15	184512.13
2	Arachidonate Metabolic Process (GO:0019369)	1.206e-60	3.022e-58	782.19	107917.33
3	Long-chain Fatty Acid Metabolic Process (GO:0001676)	1.370e-54	2.287e-52	325.75	40401.21
4	Unsaturated Fatty Acid Metabolic Process (GO:0033559)	6.236e-41	7.810e-39	333.39	30863.59
5	Icosanoid Biosynthetic Process (GO:0046456)	1.829e-35	1.832e-33	741.62	59319.81
6	Prostaglandin Metabolic Process (GO:0006693)	5.819e-32	4.859e-30	522.59	37585.29
7	Leukotriene Metabolic Process (GO:0006691)	8.125e-32	5.815e-30	885.96	63423.60
8	Omega-hydroxylase P450 Pathway (GO:0097267)	1.310e-31	8.206e-30	239292.00	17016019.40
9	Fatty Acid Metabolic Process (GO:0006631)	2.144e-31	1.194e-29	96.11	6787.34
10	Prostanoid Biosynthetic Process (GO:0046457)	2.365e-28	1.185e-26	1272.57	80950.34

Рис. 1. Наиболее значимые находки для анализа, отсортированные по значению p-value.

Полный список из 501 значимого GO-термина представлен в [таблице выдачи обогащения](#).

Вывод

Полученные результаты GO enrichment analysis показывают, что исследуемый список генов представляет собой функционально связанную группу, участвующую преимущественно в метаболизме арахидоновой кислоты и воспалительных процессах. Наиболее значимые GO-термины были связаны с метаболизмом арахидоновой кислоты, эйкозаноидов, простагландинов и лейкотриенов. Интересно, что анализ выявил последовательные этапы синтеза (Рис. 2.)

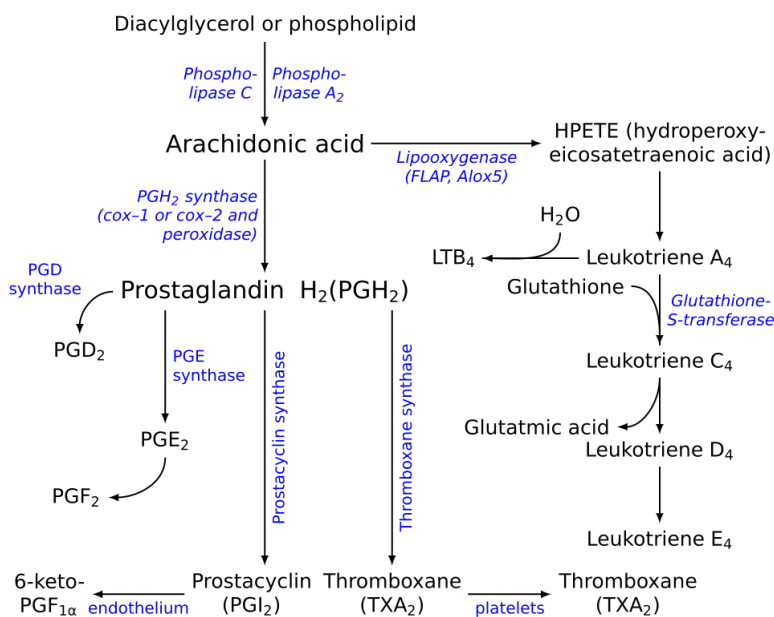


Рис. 2. Пути биосинтеза эйкозаноидов из арахидоновой кислоты.

Интересным также стало значимое обогащение терминов, связанных с метаболизмом длинноцепочечных и ненасыщенных жирных кислот. Это круто, так как арахидоновая кислота - полиненасыщенная жирная кислота и служит центральным субстратом для синтеза эйкозаноидов.

Кроме того, есть много генов семейства цитохромов P450, которые участвуют в детоксикации, в метаболизме сигнальных липидов, влияющих на воспаление и клеточный ответ на стресс. Также в списке есть гены антиоксидантной защиты.

3. Индивидуальный анализ одного гена из списка

Human Protein Atlas

Human Protein Atlas представляет собой базу данных, содержащую информацию об экспрессии и локализации белков человека в различных тканях и клетках организма. С помощью [Human Protein Atlas](#) можно решать следующие задачи:

1. Анализировать тканеспецифичную экспрессию генов и белков человека.
2. Исследовать внутриклеточную локализацию белков.
3. Изучать связь экспрессии генов с физиологическими процессами и заболеваниями.

Для индивидуального анализа я выбрала ген PTGS2. Он кодирует индуцируемую циклооксигеназу COX-2, участвующую в синтезе простагландинов из арахидоновой кислоты. Экспрессия PTGS2 повышается при воспалении и иммунном ответе.

Ген экспрессируется во множестве различных тканей, но особенно высокая экспрессия наблюдается в мочевом пузыре (151.3 nTPM), мужской репродуктивной системе (семенные пузырьки: 120.2 nTPM; простата: 23,4 nTPM), костном мозге (110.4 nTPM), шейке матки (58.4 nTPM), также можно отметить экспрессию в легких (34.5 nTPM) и желчном пузыре (29.6 nTPM) (Рис. 3 и Рис. 4).

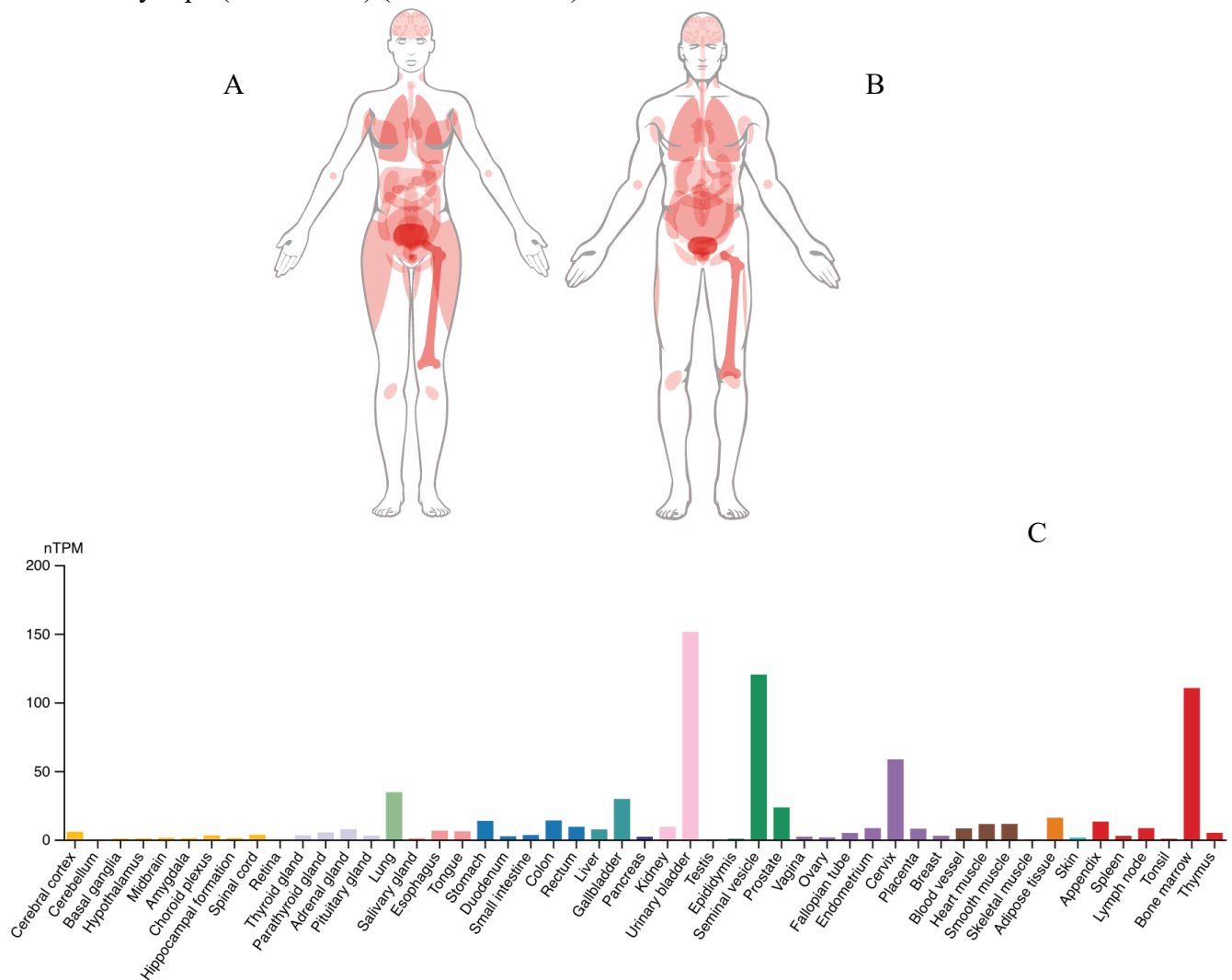


Рис. 3. С: Данные по экспрессии гена PTGS2 в различных тканях и органах человека, приведенные в базе данных The Human Protein Atlas; А: Районы экспрессии данного гена в женском организме; В: Районы экспрессии данного гена в мужском организме.

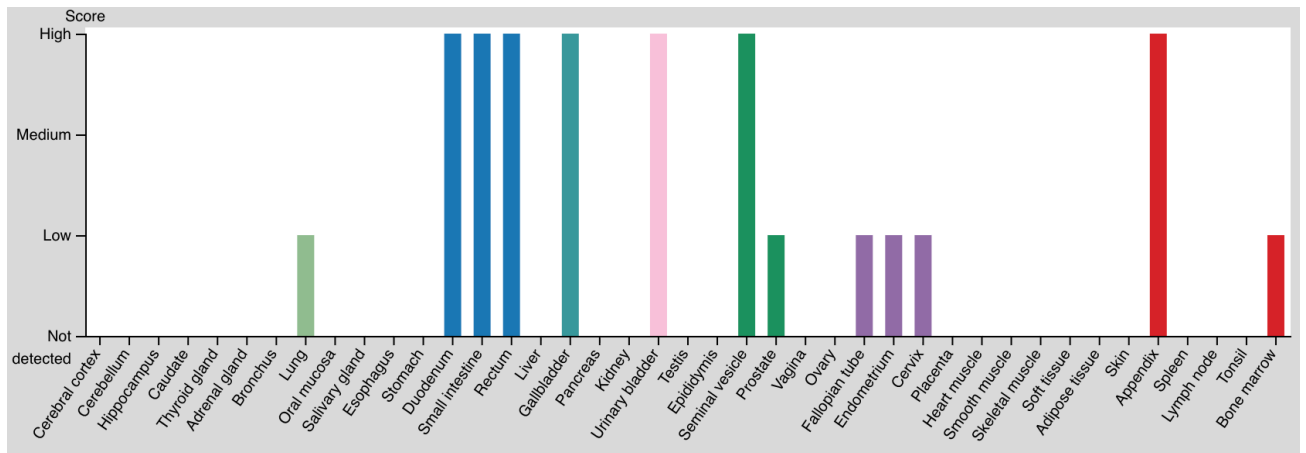


Рис. 4. Данные по экспрессии белка PTGS2 в различных тканях и органах человека, приведенные в базе данных The Human Protein Atlas.

Интересно, что PTGS2 экспрессируется в широком спектре тканей организма, что может свидетельствовать о важной роли простагландинов не только в воспалении, но и в регуляции физиологических процессов различных органов.

Субклеточная локализация: белок преимущественно локализован в везикулах, а также в цитозоле (Рис. 5). Имеет 2 транскрипта, то есть сам ген PTGS2 кодирует 2 изоформы мРНК.

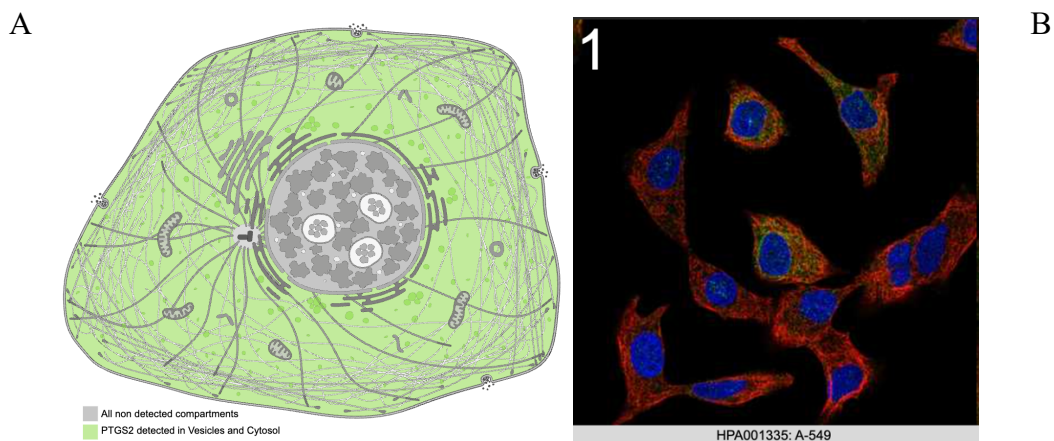


Рис. 5. Изображение локализации белка PTGS2 в клетке. А: Основные субклеточные локализации; В: Субклеточная локализация кодируемого белка (белков) в клеточных линиях человека (зеленый - целевой белок, красный - микротрубочки, синий - ядро).

В головном мозге человека экспрессия гена повышена в коре головного мозга (6.4 nTPM), хориодном сплетении (3.9 nTPM) и белом веществе (3.6 nTPM) (Рис. 6).

Наличие экспрессии в тканях головного мозга также может указывать на участие PTGS2 в нейровоспалительных процессах и регуляции функций нервной системы.

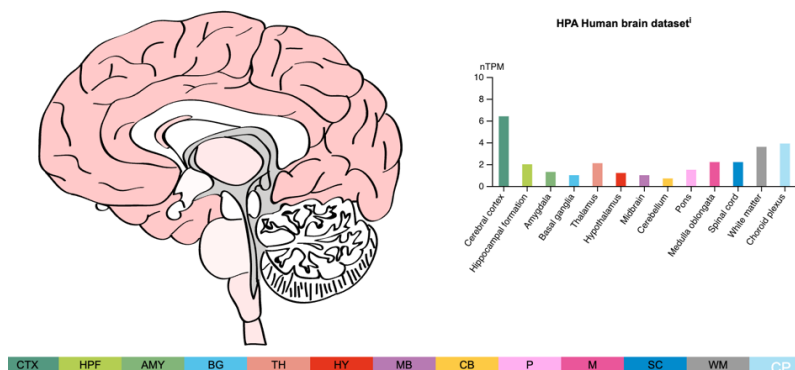


Рис. 6. Отделы головного мозга человека с повышенной экспрессией белка PTGS2.

В таблицу на Рис. 7 указывается, что ген PTGS2 связан с 21 реакцией, куда входят: биосинтез простагландинов, метаболизм омега-3 жирных кислот, метаболизм эйкозаноидов, метаболизм линолеиновой кислоты, транспортные реакции, метаболизм арахидоновой кислоты:

Pathway / Subsystem	Subsystem-associated compartments	# proteins	# metabolites	# reactions for this protein
Prostaglandin biosynthesis	Cytosol, Endoplasmic reticulum	49	80	6
Omega-3 fatty acid metabolism	Cytosol, Peroxisome, Nucleus, Endoplasmic reticulum	52	73	6
Eicosanoid metabolism	Cytosol, Mitochondria, Peroxisome, Endoplasmic reticulum, Golgi apparatus, Nucleus	42	82	3
Linoleate metabolism	Cytosol, Endoplasmic reticulum	63	42	1
Transport reactions	Cytosol, Extracellular, Lysosome, Peroxisome, Mitochondria, Golgi apparatus, Endoplasmic reticulum, Nucleus, Inner mitochondria	563	1599	2
Arachidonic acid metabolism	Cytosol, Peroxisome, Endoplasmic reticulum, Extracellular, Nucleus, Mitochondria, Golgi apparatus	107	88	3

Рис. 7. Сводка метаболических процессов, связанных с продуктом гена PTGS2.

Вывод

Таким образом, индивидуальный анализ PTGS2 с использованием Human Protein Atlas подтвердил результаты группового GO enrichment analysis и позволил более подробно охарактеризовать роль данного гена в воспалительном липидном метаболизме. Высокая экспрессия PTGS2 в тканях, связанных с иммунным ответом и активным метаболизмом, а также участие гена в многочисленных метаболических реакциях, включая метаболизм эйкозаноидов, простагландинов, омега-3 жирных кислот и арахидоновой кислоты, хорошо согласуются с результатами enrichment analysis, где наиболее значимыми оказались термины, связанные с воспалением и липидным обменом.

4. Заключение

В ходе выполнения практикума были исследованы 59 генов человека с использованием методов функционального анализа и специализированных баз данных. Проведенный GO enrichment analysis показал, что исследуемый список генов представляет собой функционально связанную группу, участвующую преимущественно в метаболизме арахидоновой кислоты, синтезе эйкозаноидов и регуляции воспалительных процессов. Наиболее значимыми оказались GO-термины, связанные с метаболизмом простагландинов, лейкотриенов. Это значит, что большинство исследуемых генов участвует в последовательных этапах единого процесса. Кроме того, в списке были выявлены гены цитохромов P450 и антиоксидантной защиты, что указывает на тесную связь воспалительных процессов с окислительным стрессом и регуляцией клеточного ответа на повреждение.

Индивидуальный анализ гена PTGS2 с использованием Human Protein Atlas подтвердил результаты группового анализа и позволил более подробно рассмотреть роль одного из ключевых участников. Было показано, что PTGS2 активно участвует в синтезе простагландинов, экспрессируется в различных тканях организма и связан с многочисленными метаболическими и воспалительными процессами.

Таким образом, использование GO enrichment analysis и Human Protein Atlas позволило не только определить общую биологическую функцию исследуемого набора генов, но и проследить вклад конкретного гена в работу всей системы. Выполненный анализ показал, что исследуемые гены образуют согласованную сеть ферментов и регуляторных белков, обеспечивающих синтез и метаболизм биологически активных липидных молекул, участвующих в воспалении, иммунном ответе и клеточной защите.