

1. Описание входных данных

В качестве входных данных у нас имеется текстовый файл со списком из 11 генов человека.

Ссылка: https://kodomo.fbb.msu.ru/FBB/year_24/lectures/lists_go/list14.txt

Вот наши гены с описанием получающихся с них белков:

VCAN - белок входит в состав внеклеточного матрикса тканей, включая глаза, кожу и сердце.

BCAN - специфический для головного мозга протеогликан хондроитинсульфата, играющий решающую роль в стабилизации структуры внеклеточного матрикса в центральной нервной системе

DSE - фермент, который превращает D-глюкуроновую кислоту в L-идуроновую кислоту, необходимую для синтеза компонента внеклеточного матрикса.

UST - белок, связан с заболеваниями молочной железы и острым некрозом сетчатки.

DSEL - связан с синдромом Эйкена и дисплазией Шнекенбекена.

NCAN - протеогликан хондроитинсульфата, находящийся во внеклеточном матриксе головного мозга и необходимый для адгезии нейронных клеток, роста нейритов и синаптической пластичности.

BGN - бигликан, небольшой богатый лейцином протеогликан, необходимый для структуры внеклеточного матрикса, сборки коллагеновых фибрилл, роста костей и клеточной сигнализации.

CHST14 - фермент, который необходим для образования важнейшего гликозаминогликана в соединительной ткани.

CSPG5 - специфический для головного мозга трансмембранный белок, имеющий решающее значение для развития нервной системы, в частности для роста нейритов, ветвления дендритов и синаптической пластичности.

CSPG4 - трансмембранный протеогликан, который выступает в качестве ключевого регулятора пролиферации, миграции и ангиогенеза клеток.

DCN - протеогликан, богатый лейцином, имеющий решающее значение для сборки коллагеновых фибрилл и структурной целостности внеклеточного матрикса.

Думаю все эти гены связаны между собой внеклеточным матриксом. А еще ассоциированы с разными видами рака.

2. Групповой сервис STRING

Данным сервисом можно решать следующие задачи:

- Анализировать белок-белковые взаимодействия
- Анализировать коэкспрессию генов из исследуемого списка
- Исследовать обогащение по GO терминам и KEGG путям

Покопавшись в интернете, нашли, что сервис использует тестирования Фишера и Колмогорова-Смирнова. Сервис также делает поправку на множественное тестирование, что следует из наличия **False Discovery Rate (FDR)**, которая оценивает долю ошибок первого рода.

Таблицу находок скачать не получилось. Вот ссылка на нее:

<https://string-db.org/cgi/network?taskId=bclG12ktsEzu&sessionId=bx4iUL7Rvo4h>

Если говорить о проведенном GO-анализе, то по биологическим процессам было получено шесть терминов. По молекулярным функциям пять терминов, а по клеточным компонентам шесть. KEGG путь был получен всего один - биосинтез гликозаминогликанов. Картинки ниже:

Biological Process (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0019800	Peptide cross-linking via chondroitin 4-sulfate glycosaminoglycan	2 of 6	2.78	1.19	0.0069
GO:0006790	Sulfur compound metabolic process	6 of 336	1.5	1.79	1.60e-05
GO:0030204	Chondroitin sulfate metabolic process	4 of 31	2.36	2.76	4.75e-06
GO:0030208	Dermatan sulfate biosynthetic process	4 of 5	3.16	4.05	4.11e-08
GO:0006029	Proteoglycan metabolic process	6 of 80	2.13	3.49	4.11e-08
GO:1903510	Mucopolysaccharide metabolic process	6 of 84	2.11	3.47	4.11e-08

Molecular Function (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0047757	Chondroitin-glucuronate 5-epimerase activity	2 of 3	3.08	1.36	0.0035
GO:0008146	Sulfotransferase activity	3 of 53	2.01	1.27	0.0035
GO:0005540	Hyaluronic acid binding	3 of 30	2.25	1.57	0.0011
GO:0030021	Extracellular matrix structural constituent conferring compression r...	3 of 15	2.55	1.78	0.00052
GO:0005539	Glycosaminoglycan binding	5 of 245	1.56	1.42	0.00052

Cellular Component (Gene Ontology)					
GO-term	description	count in network	strength	signal	false discovery rate
GO:0000139	Golgi membrane	5 of 664	1.13	0.87	0.0030
GO:0062023	Collagen-containing extracellular matrix	5 of 407	1.34	1.29	0.00030
GO:0031012	Extracellular matrix	6 of 552	1.29	1.4	4.67e-05
GO:0005794	Golgi apparatus	11 of 1650	1.08	1.45	1.01e-09
GO:0043202	Lysosomal lumen	7 of 97	2.11	4.53	6.16e-11
GO:0005796	Golgi lumen	7 of 106	2.07	4.44	6.16e-11

Рис1. Находки GO анализа с помощью сервиса STRING

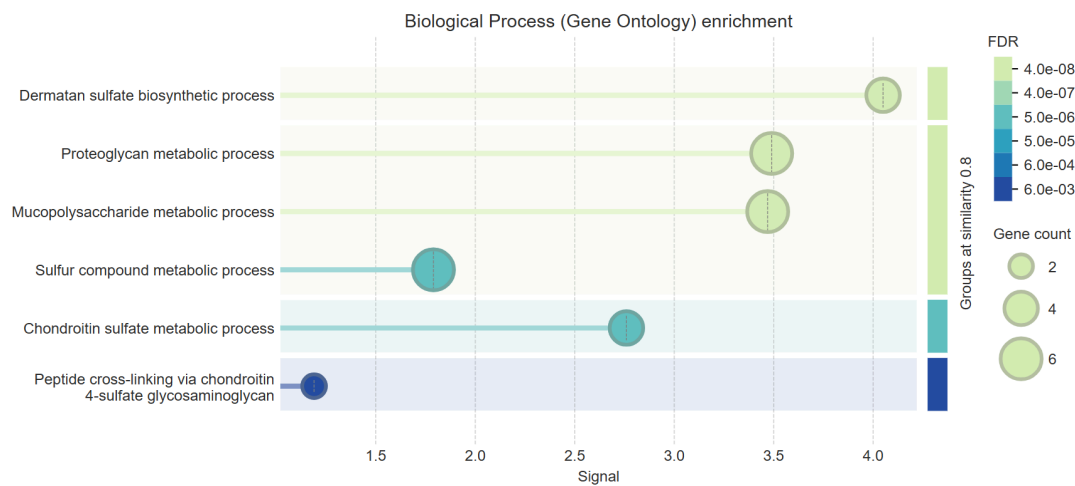


Рис2. Визуализация находок GO анализа по биологическим процессам.

Смотрим на FDR. Самые хорошие значения у дерматан-сульфат биосинтетического, протеингликанового и микополисахаридного метаболических процессов. Молекулярные функции соответствующие: поддержание структуры внеклеточного матрикса и связывание гликозаминогликана. Представлены наши белки в люмине клеток, что логично.

GLYCOSAMINOGLYCAN BIOSYNTHESIS - CHONDROITIN SULFATE / DERMATAN SULFATE

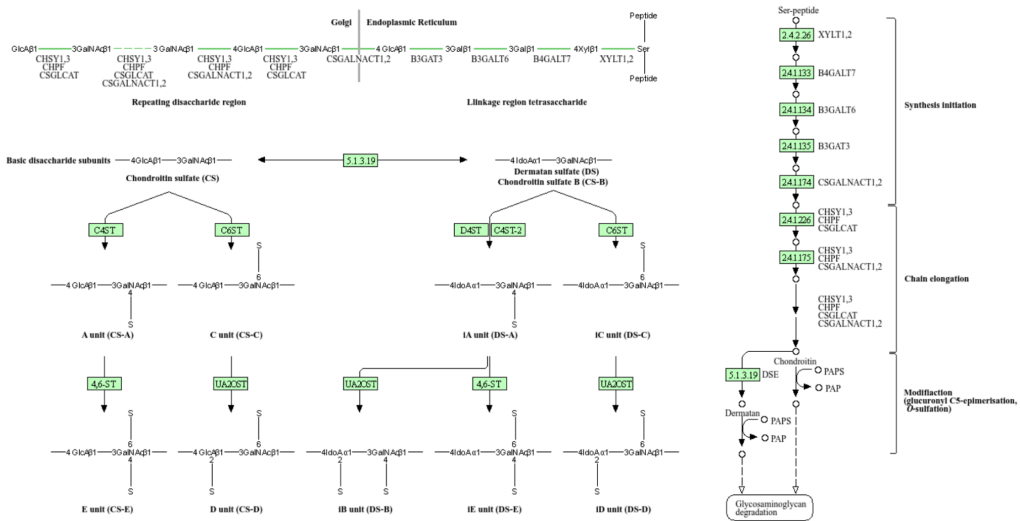


Рис 2. Единственный KEGG путь, полученный сервисом STRING

Обогащение пути биосинтеза гликозаминогликанов (chondroitin sulfate / dermatan sulfate) обусловлено присутствием генов, кодирующих протеогликаны внеклеточного матрикса (VCAN, BCAN, NCAN, CSPG4/5, BGN, DCN), а также ферменты их постсинтетической модификации (DSE, DSEL, CHST14, UST).

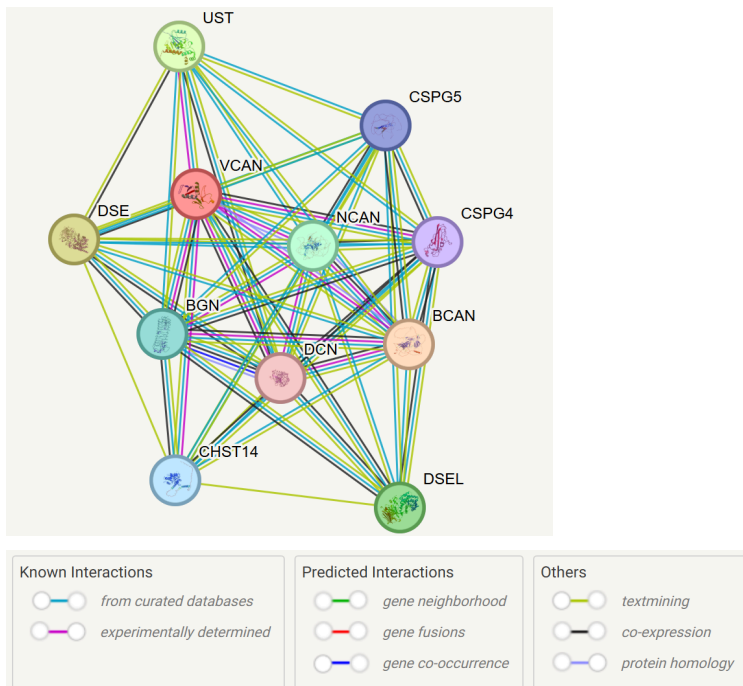


Рис 3. Визуализация взаимодействий внутри нашей выборки белков.

На рис3 представлена плотная сеть взаимодействий. Большое количество розовых и голубых линий свидетельствует об экспериментальной подтвержденности взаимодействий. Салатовые линии показывают, что белки совместно упоминаются в статьях достаточно часто, чтобы это не было случайной находкой.

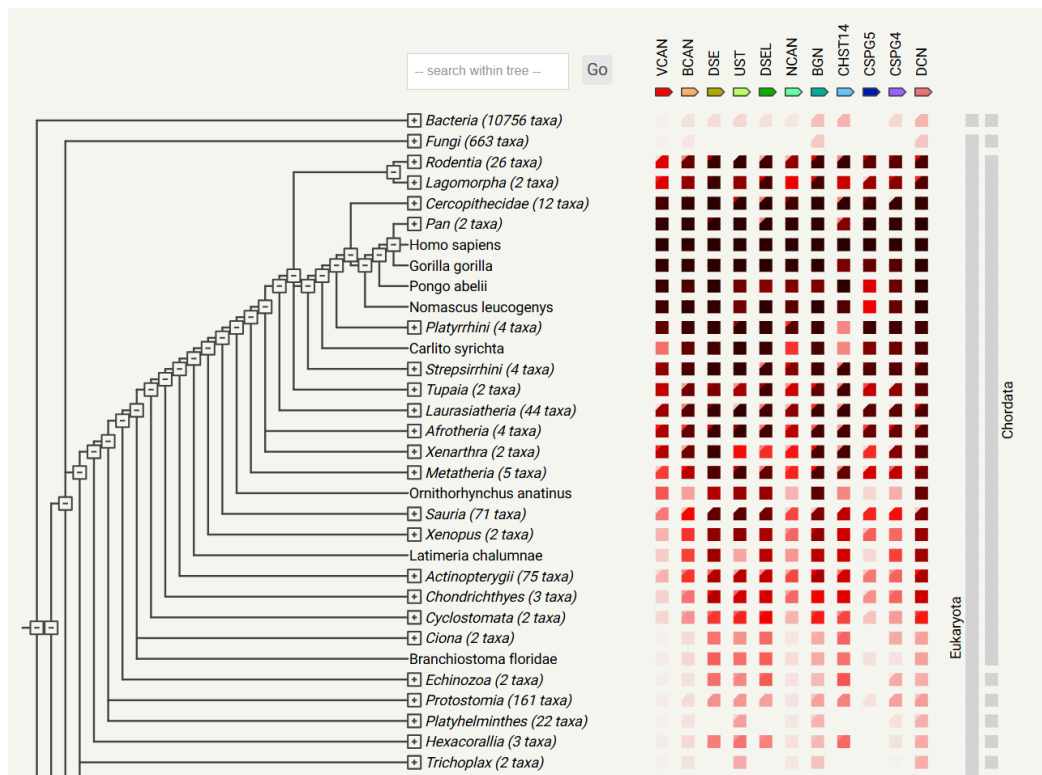


Рис 4. Семейства генов, закономерности встречаемости которых в разных геномах демонстрируют сходство

Исследуемые гены консервативны и представлены у всех млекопитающих (интерсно, что похожие по последовательности находки есть даже у грибов). В наибольшей степени они сходны у приматов, что видно по черной окраске. Это отражает эволюционную важность метаболического пути, в котором данные гены имеют ключевую роль.

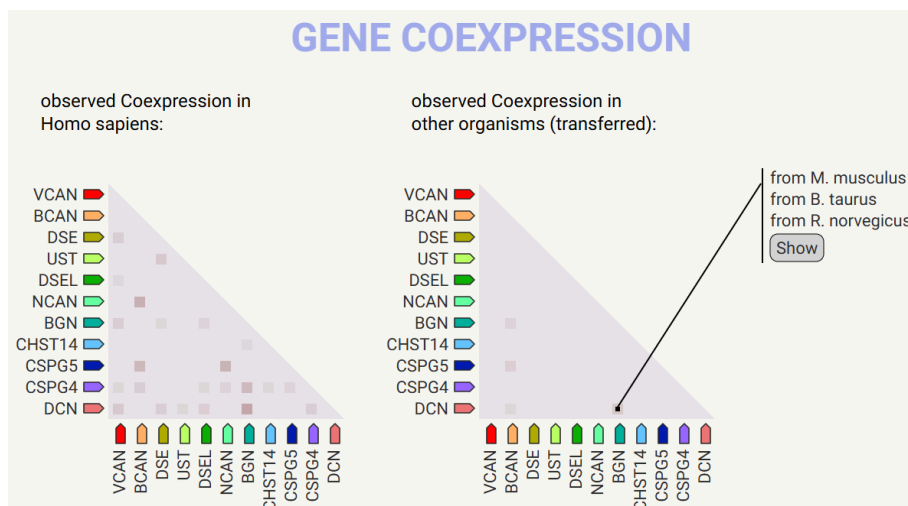


Рис 5. Козэкспрессия наших генов в людях и других организмах

При том, что наши гены участвуют в одном метаболическом пути, уровень их коэкспрессии низок. Не уверена, что так должно быть, возможно сервис по этой задаче обладает слишком маленьким количеством информации.

3.Индивидуальный сервис Human Protein Atlas

Данным сервисом можно решать следующие задачи:

- Определять локализацию белков во всех основных органах и тканях человека
- Изучать влияние уровней экспрессии белков на выживаемость пациентов с различными видами рака
- Проводить сингл селл анализ

Будем работать с BCAN. Из результатов видно, что этот ген кодирует протеогликан внеклеточного матрикса. В пути биосинтеза гликозаминогликанов, полученном в прошлом пункте, он влияет на связывание воды, обеспечение структурной жесткости, участвует в регуляции воспалений.

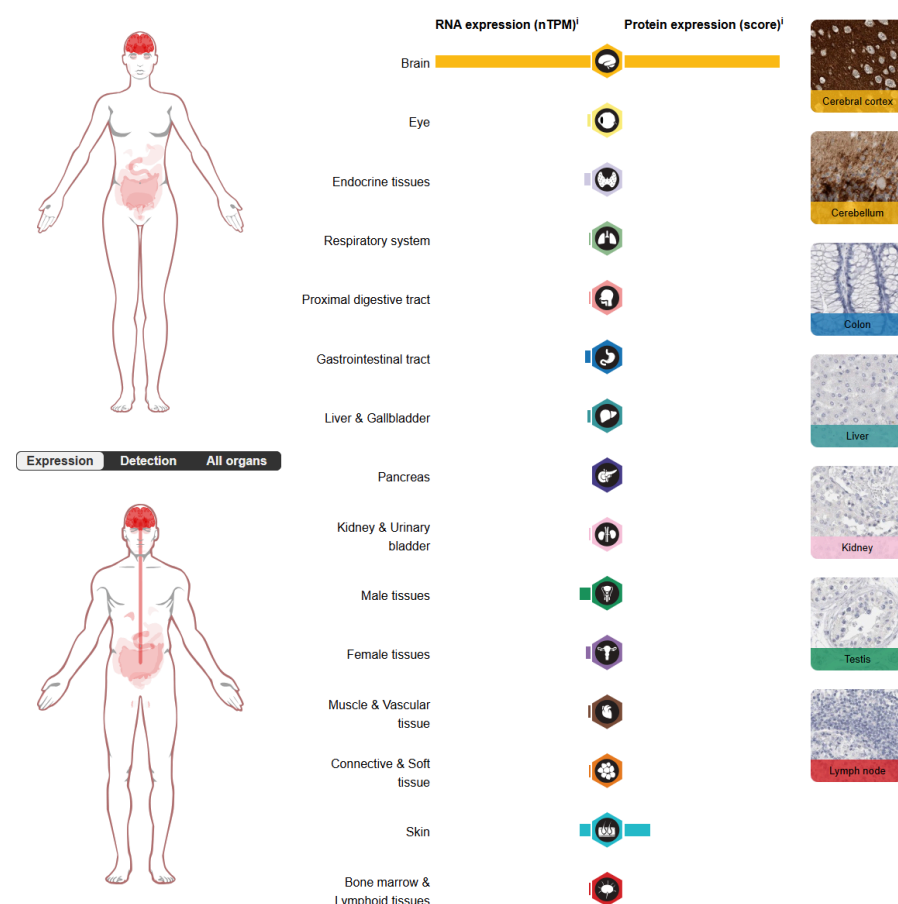


Рис5. РНК и белковая коэкспрессия BCAN в разных частях тела человека.

Бесспорный чемпион по РНК и белковой экспрессии - головной мозг. Для мужчин еще и спинной мозг, но тут по классике - женщин всегда изучают меньше, чем мужчин, и данных по ним всегда меньше. Небольшой уровень экспрессии РНК в тестикулах, яичниках и РНК+белков в коже. Кожа, честно говоря, прям таки удивляет. Объяснить это я себе не могу.

Single cell type groups

RNA single cell type group specificity: Group enriched (Glial cells, Pigment cells)

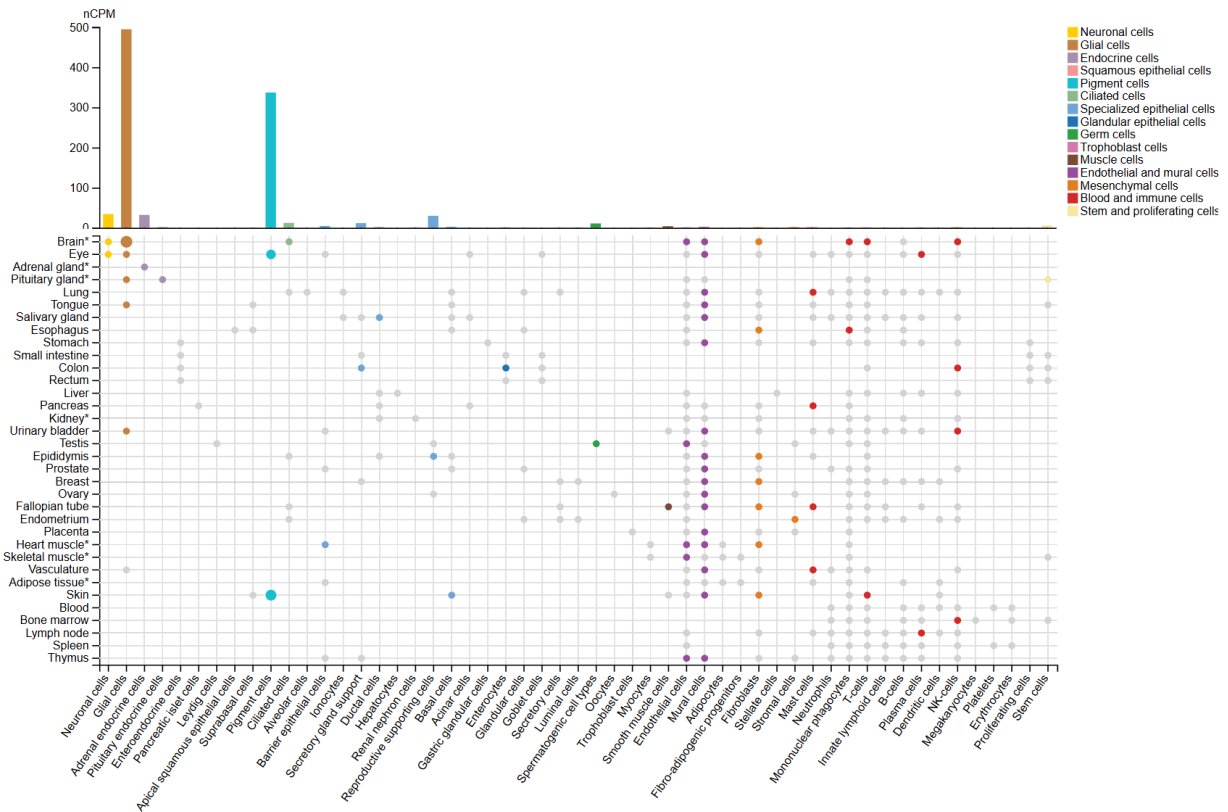
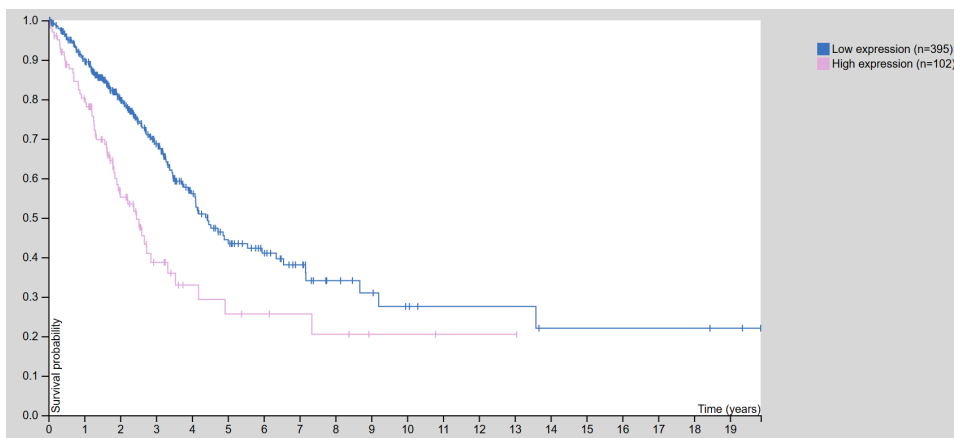


Рис6. Обогащение сингсельных групп

Обогащены глиальные и пигментные клетки. Ну тут понятно: глия потому что мозг, пигментные потому что кожа. Еще точек много для mural cells(выстилают поверхность кровяных сосудов), но они размером поменьше, поэтому насыщение низкое.



BCAN is potential prognostic, high expression is unfavorable in Lung Adenocarcinoma (TCGA)

Рис7. Вероятность выжить с одним из видов рака легких в зависимости от уровня экспрессии BCAN

Это мой любимый график в этом практикуме. Видим, что вероятность подольше пожить с раком при высокой экспрессии нашего гена значительно ниже. Все потому что он весь такой важный и влияет на синтез гликозаминогликанов. Гликозаминогликаны - это высокогидрофильные углеводные компоненты межклеточного матрикса, которые удерживают воду, обеспечивая тургор тканей, формируют структурный каркас

соединительной ткани, выступают смазочным материалом в суставах, регулируют ионный обмен, свертывание крови и клеточную пролиферацию. Понятное дело, отход от нормальной экспрессии *VCAN* все портит. Но для меня остается вопросом, почему выживаемость снижается с повышением экспрессии, а не наоборот. Интересно еще, что ген экспрессируется в мозге, а влияние наибольшее на течение рака легких.

4. Выводы

Обе рассмотренных базы очень удобны в использовании и позволяют выполнять широкий ряд задач. Про наш набор генов узнали, что они влияют на структуру внеклеточного матрикса, что в свою очередь влияет на широкий спектр всевозможных процессов в организме. Данные, полученные при анализе одного из генов с помощью Human Protein Atlas, перекликается с общей информацией, предоставленной STRING для полного списка.