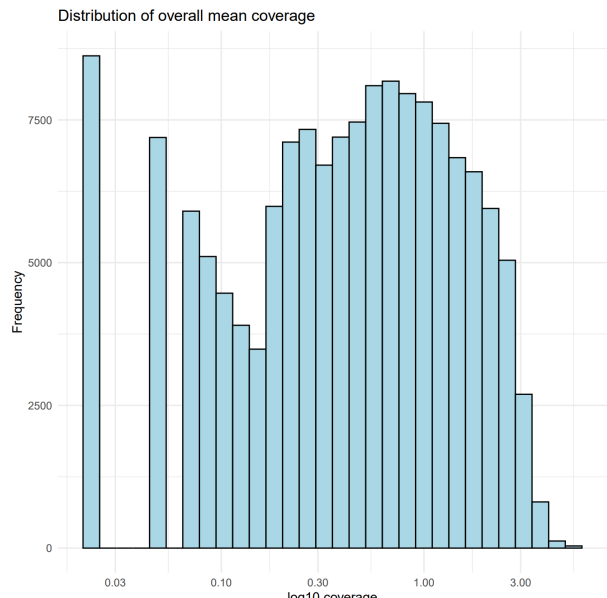


- 17 файлов из kallisto (распределяю в таблице дизайна, распределяя по 3 группам: Диабетики 2 типа, Активные и с ожирением:

```
design <- tibble("Sample" = colnames(counts),
               "group" = if_else(colnames(counts) %in% c('66', '74', '59', '72', '61'), "T2D",
                                if_else(colnames(counts) %in% c('63', '64', '60', '58', '62', '65'),
                                "Obese", "Active"))))
```

Самосоогласованность образцов:

Большое количество генов имеет нулевое покрытие на общем графике:

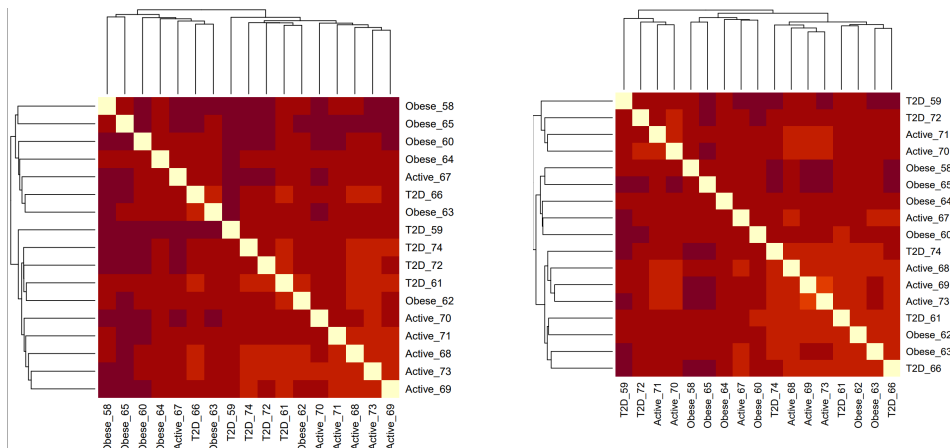


Всвязи с этим провожу фильтрацию всех файлов, делая отсечку в минимум 10 ридов на ген:

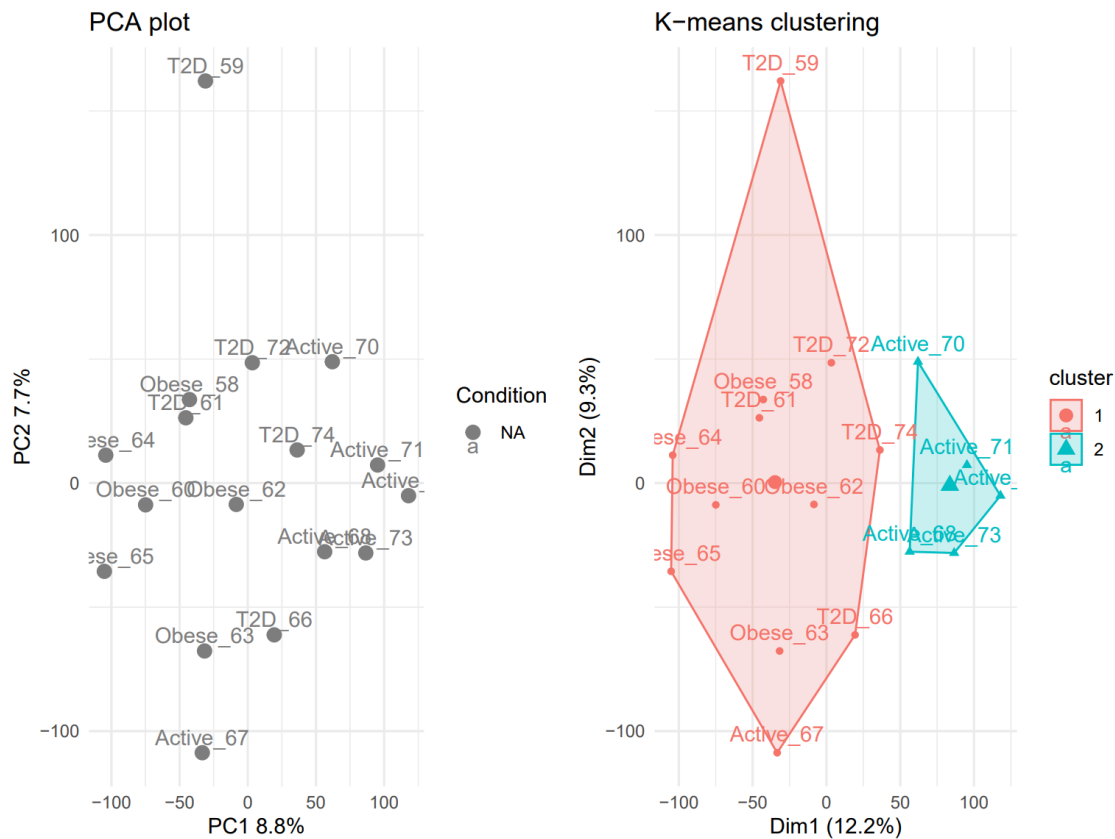
```
filtered_counts <- counts[apply(counts, 1, mean) >= 10, ]
write_tsv(rownames_to_column(filtered_counts, 'gene.id'),
          'filtered_star_counts.tsv')
```

Корреляция между образцами:

Слева график корреляции с неотфильтрованными рядами, а справа с отфильтрованными: (светло-красный указывает на наличие корреляции, а темно-красный на её отсутствие)



Видно, что некоторые образцы из разных групп хорошо коррелируют между собой. Однако не видно общего кластера образцов. Поэтому провожу поиск главной компоненты, чтобы посмотреть, кластеризуются ли образцы по какому-либо признаку друг с другом:



Здесь дополнительно помогает метод кластеризации по К-средним. Он ищет такие точки(центроиды), расстояние до которых от всех точек, принадлежащих этому кластеру будет минимально. Соответственно, наши образцы образовали два кластера. В одном находятся образцы с ожирением и диабетом, а во втором с активным образом жизни.

DESeq Obese_vs_Active:

Для расчёта использовал отсечку padj в 0.05. Изменение уровня экспрессии ($\log_2\text{FoldChange}$) должно было быть как минимум в $2^{1.3} \approx 2.5$ раза

```
dge$upreg <- dplyr::filter(res, log2FoldChange > 1.3 & padj < 0.05)
dge$downreg <- dplyr::filter(res, log2FoldChange < -1.3 & padj < 0.05)
```

Количество ап и даун регулированных генов:

```
> nrow(dge$upreg)
[1] 212
> nrow(dge$downreg)
[1] 124
```

DESeq T2D_vs_Active:

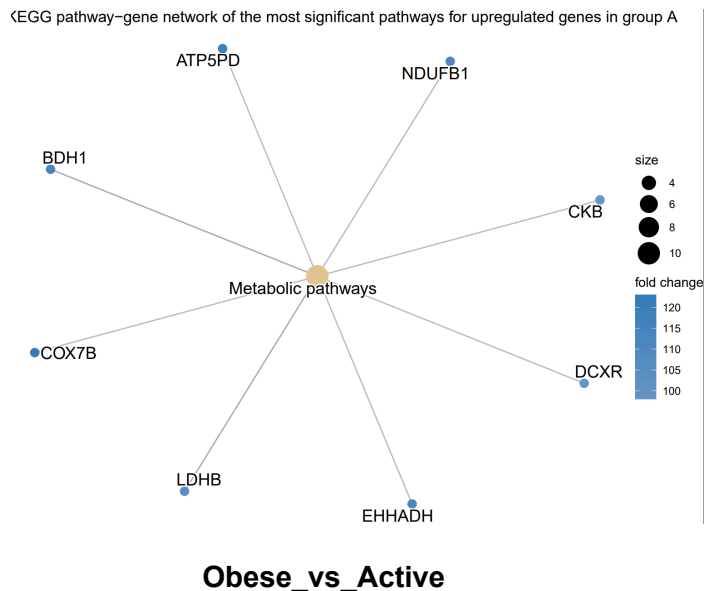
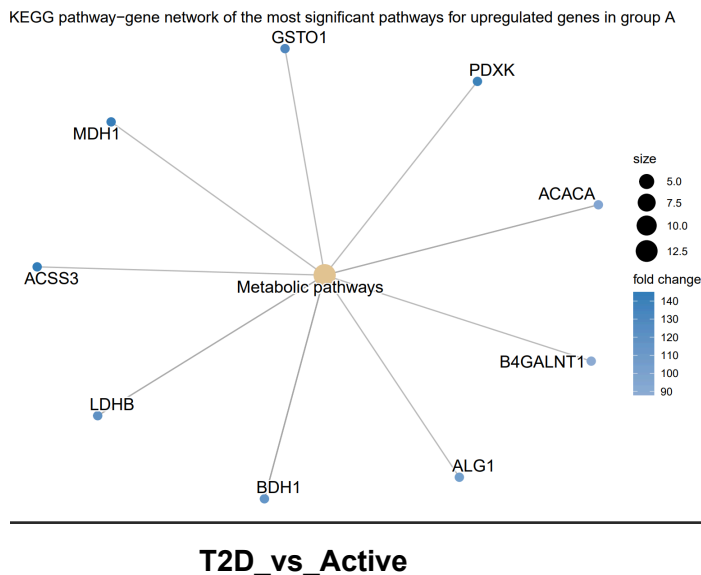
Аналогичный анализ, количество ап и даун-регулируемых генов:

```
> nrow(dge$upreg)
[1] 159
> nrow(dge$downreg)
[1] 145
```

Количество генов большое, поэтому не стал понижать log2FoldChange и перешёл к функциональной аннотации.

Функциональная аннотация

- Даунрегулированные гены в сравнении обеих групп с группой Active:



T2D_vs_Active:

Интересно, что в этой группе менее экспрессируются гены, вовлеченные в цикл трикарбоновых кислот и в глюконеогенез. Например:

ACACA (Ацил-КоА-карбоксилаза альфа)

ACSS3 (Ацил-КоА-синтетаза короткой цепи 3)

MDH1 (Мальатедегидрогеназа 1)

Это гены, напрямую участвующие в ЦТК

PDXK (Пиридоксаль-фосфат зависимая оксидоредуктаза)

LDHB (Лактатдегидрогеназа B)

А эти два гена участвуют в глюконеогенезе, что логично

То, что эти гены оказались даунрегулированными звучит логично. Так как они все косвенно вовлечены в метаболизм глюкозы, который при диабете понижен.

Мне кажется, что такое снижение экспрессии генов возможно из-за двух причин:

1. В общем пути метаболизма глюкозы очень много положительно-обратных связей. Т.е чем больше глюкозы поступает, тем интенсивнее экспрессируются гены её метаболизма. На фоне общего снижения уровня глюкозы мы видим в первую очередь понижение экспрессии именно генов ЦТК, т.к для его функционирования нужен постоянный и стабильный приток веществ из гликолиза
2. Возможно, такое снижение экспрессии это результат общего патологического состояния клеток, из-за которого “заглушаются” гены метаболизма именно ЦТК. Такая же ситуация наблюдается, например, в раковых клетках или в клетках при гипоксии (тоже патологическое состояние). То есть снижение экспрессии этих генов обусловлено не просто низким уровнем глюкозы, а именно патологическим состоянием, который этот низкий уровень глюкозы вызывает

Obese_vs_Active

Здесь так же даунрегулированы гены из разных метаболических путей

Некоторые гены, вовлеченные в бета-окисление жирных кислот и метаболизм кетоновых тел:

BDH1 (3-Hydroxybutyryl-CoA dehydrogenase 1)

EHADH (Enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase)

Здесь, из-за большого присутствия жирных кислот в клетках действует отрицательно-обратная связь, которая сдерживает их переработку

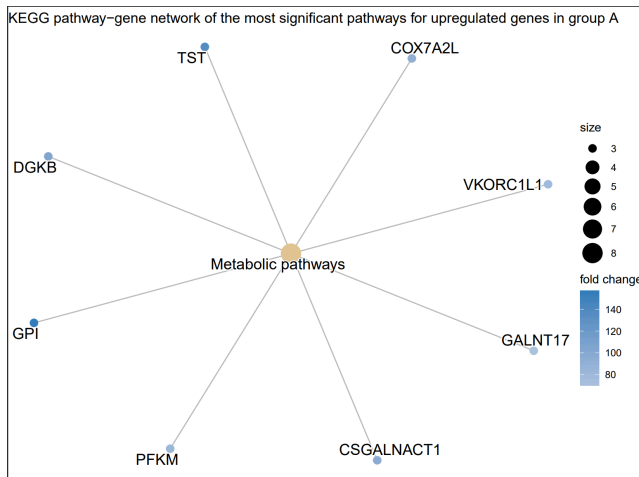
В то же время снижена экспрессия генов электрон-транспортной цепи:

NDUFB1 (NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 1)

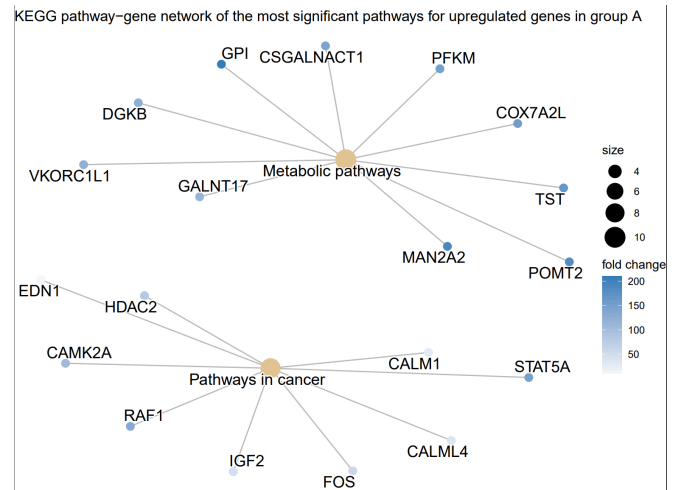
COX7B (Cytochrome c oxidase subunit 7B)

Это скорее всего необходимо из-за огромного притока NADH из ЦТК. Так как Ацетил-КоА при бета-окислении жирных кислот образуется очень много, необходимо как-то регулировать уровень образования АТФ

- **Апрегулированные гены в сравнении обеих групп с группой Active:**



T2D_vs_Active



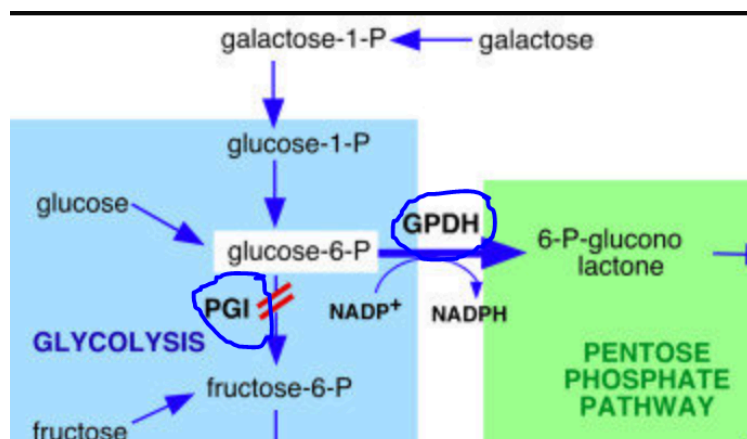
Obese_vs_Active

T2D_vs_Active:

- Здесь повысили экспрессию ген первого этапа гликолиза:

GPI (Glucose-6-phosphate isomerase). Он повысился в обеих группах, но, как мне кажется, по разным причинам. Этот фермент стоит “на перепутье” между гликолизом и пентозофосфатным путём. Но именно GPI запускает гликолиз, а другой фермент GPDH(Glycerol-3-phosphate dehydrogenase) “перетягивает одеяло” в сторону пентозофосфатного пути.

Повышенная экспрессия GPI в диабетических клетках наверное нужна, чтобы “экстренно” забрать появившуюся глюкозу в гликолиз, не теряя молекулы в пентозофосфатный путь, который гораздо менее энергетически выгоден.



- Второй интересный ген, это:

TST (Thiosulfate sulfurtransferase), который так же повысил экспрессию в обеих группах.

Про него есть одна интересная статья в Nature:

“Thiosulfate sulfurtransferase prevents hyperglycemic damage to the zebrafish pronephros in an experimental model for diabetes”

<https://doi.org/10.1038/s41598-022-16320-1>

В ней говорится, что этот белок позволяет справиться с гипергликемическим стрессом, помогая справляться с окислительным стрессом, “предоставляя электроны в антиокислительную систему”. И так же проводится связь с ожирением, что объясняет повышение его экспрессии и во второй группе: *“Since inflammation appears to play a major role in diabetes, enhanced antioxidant activity would be crucial for counteracting the metabolic abnormalities associated with obesity”*

- Ещё два гена, которые повысили экспрессию, связаны с гликозилированием:

GALNT17 (Polypeptide N-acetylgalactosaminyltransferase 17)

CSGALNACT1 (Chondroitin sulfate N-acetylgalactosaminyltransferase 1)

Это тоже очень логично и интересно, потому что основным клиническим признаком диабета является гликозилированный гемоглобин. Видимо это ответ клеток на гипергликемию, когда они пытаются хоть как-то снизить уровень глюкозы в крови, пришивая её хоть куда то, где она будет не так вредна.

Obese_vs_Active

- В этой группе также сверхэкспрессирован GPI. Здесь я так же не очень понимаю, где причина, а где следствие. Возможно у людей с ожирением этот в целом более активен, что приводит к БОльшему образованию Ацетил-КоА и БОльшему образованию жирных кислот из него. Или же это повышение экспрессии вызвано тем патологическим состоянием, которое вызывает ожирение.
- Ещё повысил экспрессию ген PFKM (Phosphofructokinase, muscle). Это буквально следующий фермент в гликолизе после GPI и он тоже

является регуляторным. Этот фермент производит первую молекулу АТФ в гликолизе и является общим регулятором скорости гликолиза.

- Ещё несколько генов с повышенной экспрессией так или иначе вовлечены в гликозилирование и синтез олигосахаридов (видимо для гликокаликса):

GALNT17 (Polypeptide N-acetylgalactosaminyltransferase 17)- involved in the initiation of O-linked glycosylation

MAN2A2 (Mannosidase alpha class 2A member 2) - involved in the processing of N-linked oligosaccharides

POMT2 (Protein O-mannosyltransferase 2) - involved in the O-mannosylation of proteins

- Именно в этой группе появился новый метаболический путь, который связан с раковыми клетками. Но нормальный fold change там есть только у трёх генов:

RAF1 (Raf-1 proto-oncogene, serine/threonine kinase) - involved in the MAPK/ERK signaling pathway

CAMK2A (Calcium/calmodulin dependent protein kinase II alpha)

STAT5A (Signal transducer and activator of transcription 5A) - involved in the JAK/STAT signaling pathway

Все эти гены - регуляторные в путях, которые отвечают за пролиферацию и деление клеток. Поэтому мне кажется, что они являются не признаками потенциального развития рака, а просто их базовый уровень экспрессии при ожирении выше, т.к. жировая ткань постоянно и очень активно делится. Однако в базах данных эти регуляторные белки ассоциированы именно с опухолевыми клетками, что привело к образованию в cnetplot этого “ракового” метаболического островка

Вывод

При менее строгих отсечках (pvalue вместо padj, выше порог отбора) количество ап и даун генов значительно вырастает и можно увидеть новые метаболические пути. Однако не смотря на то, что в моих результатах не так много биологических путей и

генов, все они очень логично объясняют состояния пациентов. И после анализа самих генов кажется, что по-другому быть и не может. Очень понравилось, интересно и круто смотреть на клетку вот так “изнутри”