

Отчёт Решетниковой Г.Д., студентки 4-ого курса Биологического ф-та СПбГУ.  
Анализ данных qRT-PCR для генов антоцианового синтеза *F3H*, *DFR*, *ANS*.

#### Введение:

Посевная рожь (*Secale cereale* L.) – ценная продовольственная и кормовая культура, которая сохранила своё генетическое и фенотипическое разнообразие, в том числе и антоциановую окраску различных частей растений.

Известны мутации, влияющие на антоциановую окраску ржи. В настоящее время в Петергофской коллекции поддерживаются 6 мутаций безантоциановости *vi1-vi6* (от лат. *var. viride* - "зеленый") (Войлоков, 2014). Мутации *vi1*, *vi2*, *vi3* и *vi6* затрагивают разные структурные гены биосинтеза антоцианов, а мутации *vi4* и *vi5* аллельны и затрагивают регуляторный ген биосинтеза антоцианов *Vs* (неопубликованные данные).

Для некоторых генов известны механизмы, лежащие в основе мутаций, связанных с нарушением биосинтеза антоцианов. Так, нонсенс мутация в гене *ANS* (антоцианидинсинтаза) приводит к отсутствию антоциановой пигментации у безантоциановой линии *vi2* (Андреева и др., 2018). Выяснено, что некоторые из этих мутаций (*vi4* и *vi5* в частности) затрагивают регуляторные гены биосинтеза антоцианов, другие - структурные, кодирующие ферменты. (Войлоков и др., 2014). Путь биосинтеза антоцианов включает десятки ферментов, но основными регуляторными из них являются *F3H*, *DFR*, *ANS* (Winkel-Shirley, 2001).

Особенности работы структурных генов биосинтеза антоцианов у ржи не изучены, в связи с чем представляет интерес провести анализ экспрессии основных структурных генов, участвующих в позднем этапе синтеза антоцианов: *F3H*, *DFR*, *ANS*.

Поэтому, цель работы:

1. Изучить экспрессию основных генов биосинтеза антоцианов у линий ржи с различной окраской зерна.

Для выполнения цели поставлены следующие задачи:

1. Получение зерновок у семи линий ржи с различной окраской зерна (желтый, коричневый);
2. Выделение тотальной РНК из перикарпа зерновок;
3. Проведение ОТ-ПЦР в реальном времени для анализа экспрессии генов биосинтеза антоцианов ржи *F3H*, *DFR*, *ANS*;
4. Статистический анализ полученных данных.

Гипотезы для статистического анализа:

1. Существуют статистически значимые различия в экспрессии генов антоцианового пути у растений разных линий.
2. Существуют статистически значимые различия в экспрессии генов антоцианового пути у растений с перикарпами разных цветов.

Материалы и методы:

Растительный материал:

Линия (в скобках указано обозначение линий в графиках)	Генотип	Фенотип перикарпа
vi1(line1)	<i>vi1vi1</i>	Коричневый
vi2(line2)	<i>vi2vi2</i>	Жёлтый
vi2(line3)	<i>vi2vi2</i>	Коричневый
vi3(line4)	<i>vi3vi3</i>	Жёлтый
vi6(line5)	<i>vi6vi6</i>	Жёлтый
Ф-5/63(line6)	<i>Vs<sup>b</sup>Vs<sup>b</sup></i>	Коричневый
7(line7)	<i>vsvs</i>	Жёлтый

1. После выделения тотальной РНК с последующей обратной транскрипцией получаем кДНК, с которой будем проводить ПЦР-РВ в шести повторностях для каждой из линий. По разнице значений пороговых циклов для генов интереса и гена домашнего хозяйства (в настоящей работе будет использован ген гистона H2) будем оценивать степень экспрессии генов интереса;
2. Полученные данные о пороговых значениях циклов собираем в таблицу ('data.tsv') и проводим соответствующие статистические анализы.

Методы и ход работы:

Для проверки гипотез будет проведено 2 дисперсионных анализа:

1. Дисперсионный анализ экспрессии генов для семи линий ржи;
2. Анализ экспрессии генов растений в группах сформированных по окраске зерна (желтый, коричневый);

Сперва получали дельтаCt между генами интереса и референсным геном. А так же создавали факторы для дальнейшей работы.

В случае нормально распределенных данных и при равных дисперсиях сравниваемых совокупностей следует использовать параметрические тесты; если хотя бы одно из этих условий не выполнено – то непараметрические тесты.

1. Дисперсионный анализ экспрессии генов для семи линий ржи;

Сперва проводим проверку на нормальность данных для генов *F3H*, *DFR*, *ANS* в зависимости от линии (shapiro.test). Затем используем непараметрические тесты Краскала-Уоллиса и U-тест Манна-Уитни (см. "Результаты и обсуждения"). Формируем графики значений  $\Delta Cq$  для генов интереса.

2. Анализ экспрессии генов растений в группах сформированных по окраске зерна (желтый, коричневый);

Сперва проводим проверку на нормальность данных для генов *F3H*, *DFR*, *ANS* в зависимости от цвета (shapiro.test). Затем используем непараметрический тест Вилкоксона и поправки на множественные сравнения (см. “Результаты и обсуждения”). Формируем графики значений экспрессий, нормированных по референсному гену.

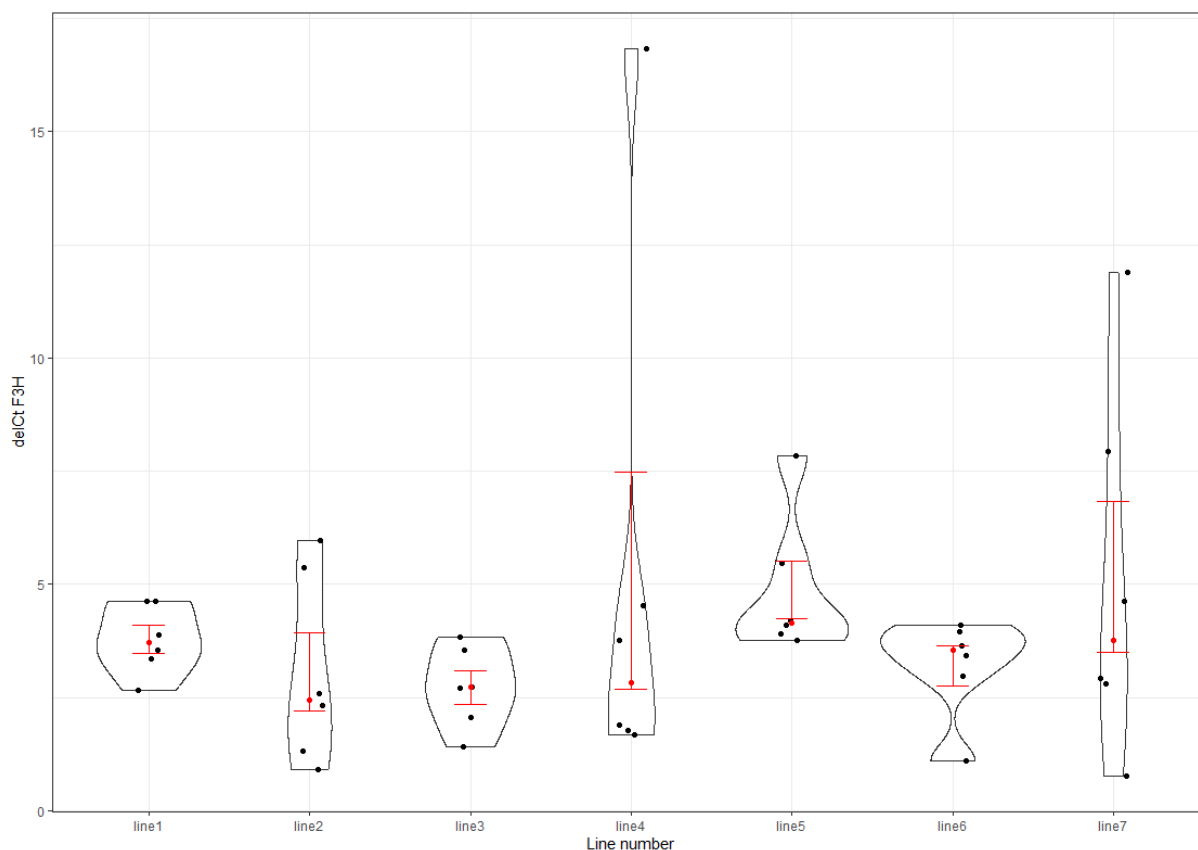
Соответствующие графики будут представлены в разделе “Результаты и обсуждения”.

Результаты и обсуждение:

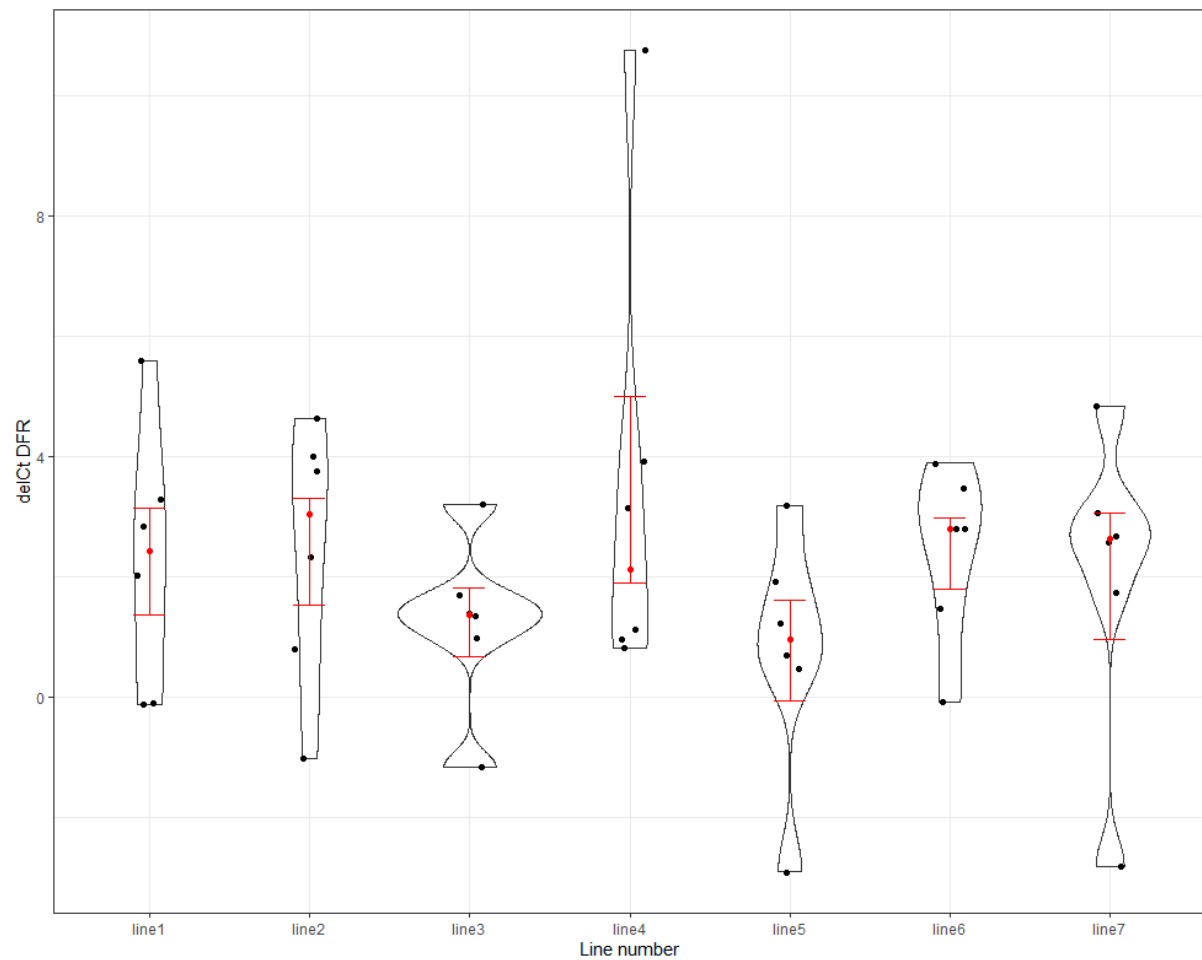
1. Дисперсионный анализ экспрессии генов для семи линий ржи.

Рассчитываем  $\Delta Ct$  относительно референсного гена. Формируем графики значений  $\Delta Ct$  в зависимости от линии для каждого из генов:

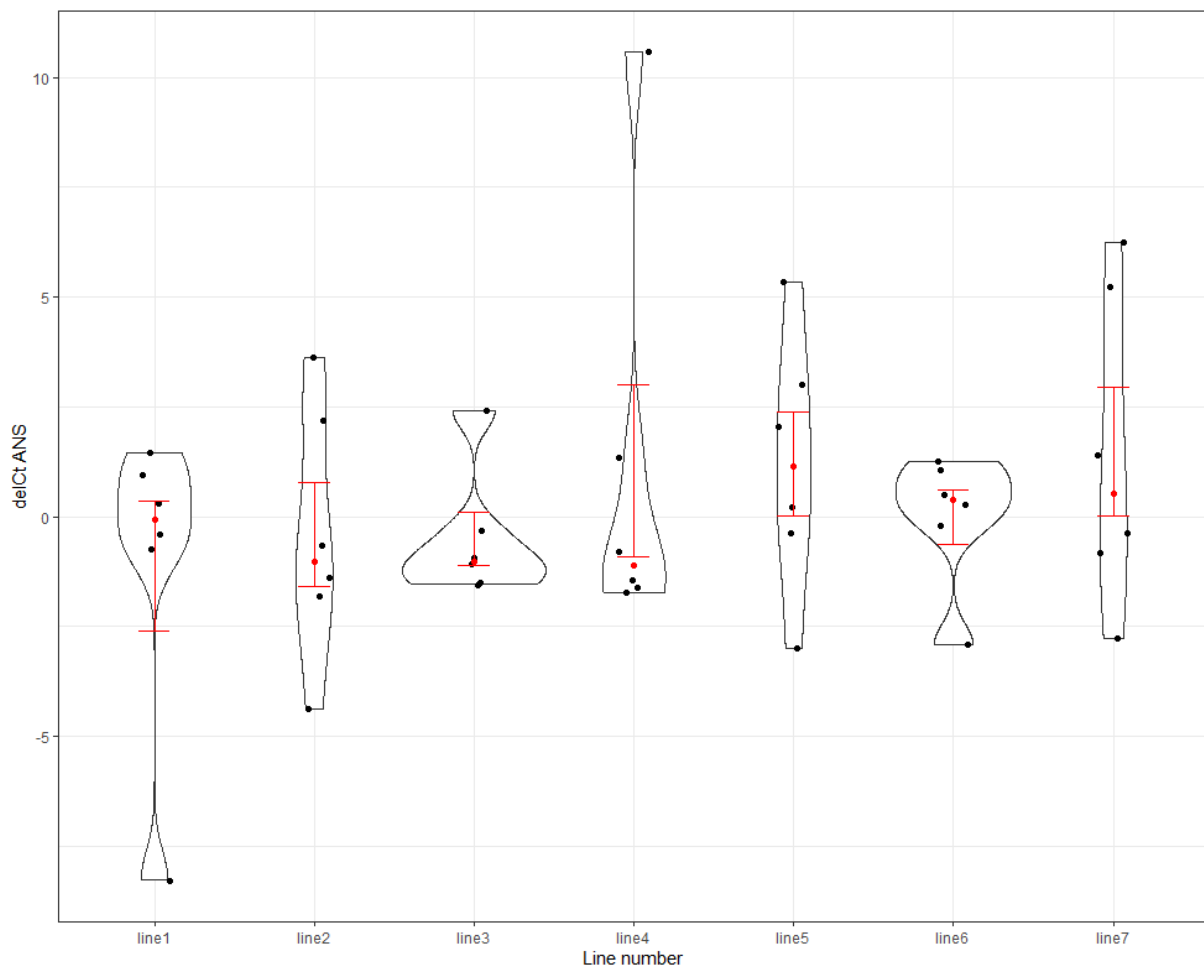
- *F3H*. Несмотря на то, что на графике отчетливо видно, что у line4 есть ощутимый выброс, соответствующая медиана и ошибка не отличаются от других линий существенно. Для теста пары 3-5 мы имели наименьшее p.value (0.091), что объясняется наибольшим отличием для ошибки и одним из наибольших отличий для значения медианы. Можно, считать, что отличия между 3й и 5й линией для экспрессии гена *F3H* - наибольшие за весь проведенный нами анализ в рамках этой работы, но все же не настолько существенные, чтобы полагать существенные различия.



- *DFR*. Несмотря на то, что на графике отчетливо видно, что у 4-ой линии есть ощутимый выброс, соответствующая медиана и ошибка не отличаются от других линий существенно. Имеем, что все группы не отличаются между собой статистически значимо.



- *ANS*. Несмотря на то, что на графике отчетливо видно, что у 4-ой линии есть ощутимый выброс, соответствующая медиана и ошибка не отличаются от других линий существенно. Имеем, что все группы не отличаются между собой статистически значимо.



Проверка на нормальность во всех трёх случаях показала, что распределения ненормальные. (p-value > 0.05 – распределение нормальное, p-value < 0.05 – распределение ненормальное):

```
# Для гена F3H
#   line      F3H
# 1 line1 0.599464724 – норм. распределение
# 2 line2 0.258073980 – норм. распределение
# 3 line3 0.817154285 – норм. распределение
# 4 line4 0.002298943 – не норм. распределение
# 5 line5 0.026194917 – не норм. распределение
# 6 line6 0.069972115 – норм. распределение
# 7 line7 0.494993564 – норм. распределение
```

```
# Для гена DFR
#   line      DFR
# 1 line1 0.55567002 – норм. распределение
# 2 line2 0.49814106 – норм. распределение
# 3 line3 0.45967333 – норм. распределение
# 4 line4 0.02415555 – ненорм. распределение
# 5 line5 0.44348075 – норм. распределение
# 6 line6 0.42001400 – норм. распределение
# 7 line7 0.14412558 – норм. распределение
```

```
# Для гена ANS
#   line      ANS
# 1 line1 0.008772804 - ненорм. распределение
# 2 line2 0.819011083 - норм. распределение
# 3 line3 0.018959032 - ненорм. распределение
# 4 line4 0.003010515 - ненорм. распределение
# 5 line5 0.994697754 - норм. распределение
# 6 line6 0.070845590 - норм. распределение
# 7 line7 0.512757904 - норм. распределение
```

Дополнительно провели тест для оценки равенства дисперсий – дисперсии различаются для гена *F3H*, для *DFR* и *ANS* дисперсии были равны. Условия для использования параметрических тестов не были выполнены, поэтому используем непараметрические тесты.

Непараметрический тест Краскала-Уоллиса. Во всех трёх случаях не наблюдалось статистически значимых различий в значениях  $\Delta C_q$  между линиями.

Помимо теста Краскала-Уоллиса провели U-тест Манна-Уитни.

Результатом теста является таблица с попарными сравнениями результатов каждой линии (например, пересечение line1 и line2 даёт информацию о степени различий между линией 1 и линией 2).

Стандартно в R используется поправка Холма, однако, я использовала поправку Бендамини-Хохмана, так как она позволяет дать более точные результаты в случае моих данных. Несмотря на поправку U-тест Манна-Уитни так же показал, что не наблюдается статистически значимых различий в значениях  $\Delta C_q$  между линиями.

#### 1. *F3H*

```
      line1 line2 line3 line4 line5 line6
line2 0.827 -     -     -     -     -
line3 0.755 0.904 -     -     -     -
line4 0.904 0.904 0.904 -     -     -
line5 0.755 0.755 0.091 0.812 -     -
line6 0.904 0.904 0.812 0.937 0.432 -
line7 0.937 0.827 0.812 0.904 0.904 0.904

P value adjustment method: BH
```

P-value > 0.05 во всех случаях. Но между 3 и 5 линией возможно есть различия (p-value = 0.091, но они статистически не значимые).

#### 2. *DFR*

```
      line1 line2 line3 line4 line5 line6
line2 0.94 -     -     -     -     -
line3 0.92 0.92 -     -     -     -
line4 0.94 0.94 0.94 -     -     -
line5 0.92 0.84 0.94 0.84 -     -
line6 0.94 0.94 0.84 0.94 0.84 -
line7 0.94 0.94 0.84 0.94 0.84 0.94

P value adjustment method: BH
```

P-value > 0.05 во всех случаях.

#### 3. *ANS*

```

      line1 line2 line3 line4 line5 line6
line2 0.98  -    -    -    -    -
line3 0.98  0.98  -    -    -    -
line4 0.98  0.98  0.98  -    -    -
line5 0.98  0.98  0.98  0.98  -    -
line6 0.98  0.98  0.98  0.98  0.98  -
line7 0.98  0.98  0.98  0.98  1.00  0.98

P value adjustment method: BH

```

P-value > 0.05 во всех случаях.

2. Анализ экспрессии генов растений в группах сформированных по окраске зерна (желтый, коричневый);

После проверки экспрессии генов между линиями возникла мысль оценить экспрессию генов антоцианового пути у растений разных цветов – жёлтого и коричневого.

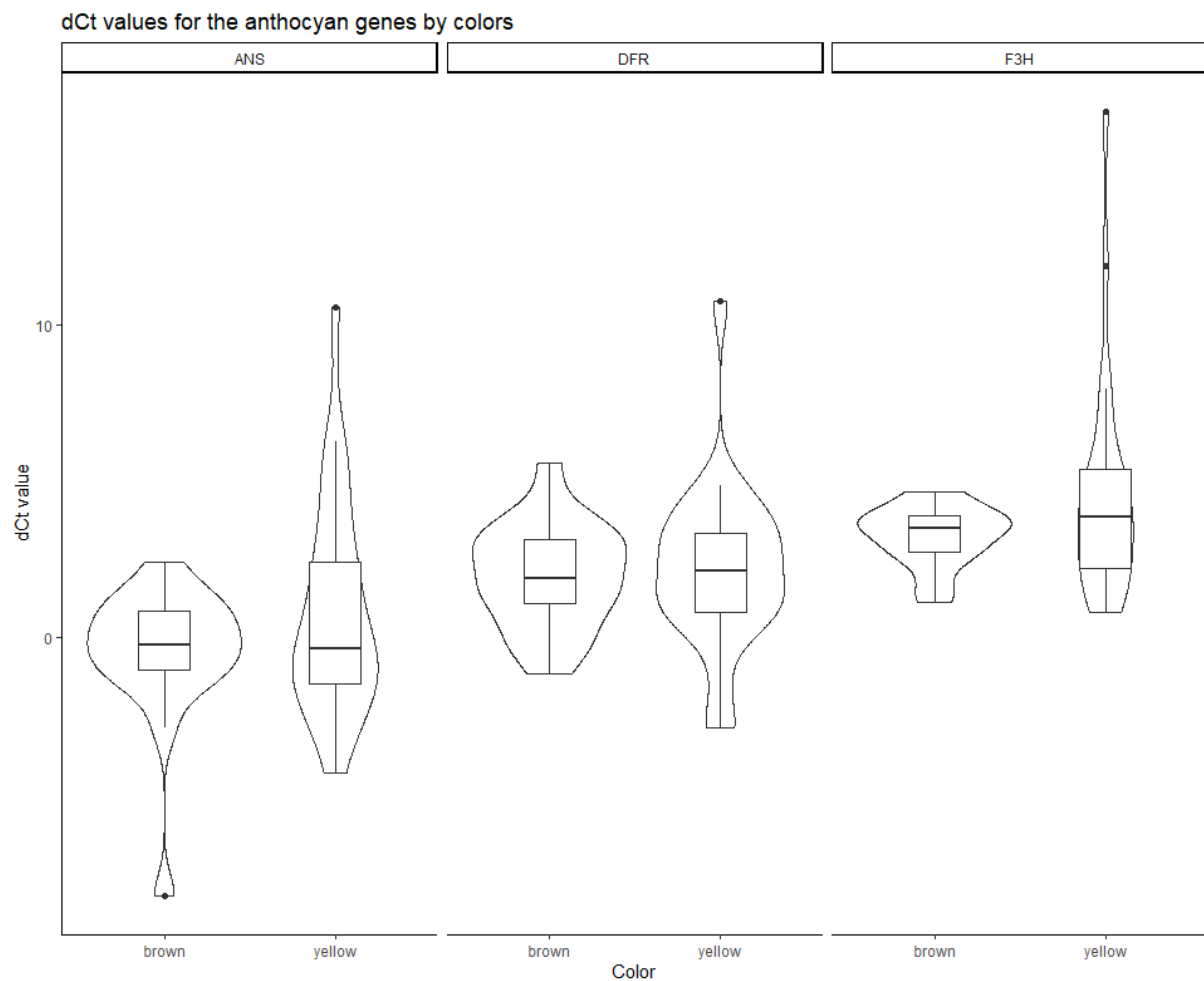
Сперва проверяли данные на нормальность. Во всех трёх случаях распределение не было нормальным, а значит, дальнейший анализ следует проводить, используя непараметрический тест – тест Вилкоксона, так как имеем 2 группы (желтый и коричневый) с поправкой на множественные сравнения (так как сравниваем 3 гена) – сперва с поправкой Холма, затем с поправкой Бендамини-Хохмана. В результате статистически значимых различий в экспрессии генов антоцианового пути в зерновках ржи разного цвета не наблюдались:

```

  name  p.value p.adj.holm p.adj.BH
1 ANS    0.606      1      0.909
2 DFR    0.930      1      0.930
3 F3H    0.357      1      0.909

```

Графики значений  $\Delta C_t$ , нормированных по референсному гену:



На графике наглядно видно, что данные не имеют статистически значимых различий в рамках групп, сформированных по цвету. Значения медиан почти равны в группах, квантили отличаются не существенно.

Однако все же, можно отметить небольшие наблюдения по значениям данных экспрессий в соответствии с группой и цветом. Мы имеем выбросы для значений экспрессии генов растений с желтым фенотипом перикарпа зерновки в 4й и 7й группах.

#### Обсуждение:

Не наблюдали статистически значимых изменений в экспрессии генов антоцианового пути от линии и цвета перикарпа зерновки. Возможно, данные стоит исследовать большее число линий и другие фенотипы перикарпов зерен и взять большую выборку. Более того, следует проверить экспрессию транскрипционных факторов этих генов (или структурных генов поздних этапов биосинтеза антоцианов), так как фенотипы зерна различаются.

Однако все же, можно отметить небольшие наблюдения по значениям данных экспрессий в соответствии с группой и цветом. Мы имеем выбросы для значений экспрессии генов растений с желтым фенотипом перикарпа зерновки в 4-ой и 7-ой группах.



Уровень экспрессии генов не меняется в рамках линии и цвета, что подтверждается и графиками. В будущем имеет интерес сделать корреляционный анализ зависимости экспрессии генов друг от друга, однако, это не является целью настоящей работы.

**Вывод:**

Обе гипотезы в ходе работы не были подтверждены. Статистически значимых изменений в экспрессии генов антоцианового пути у растений ржи разных линий и у растений с разной окраской перикарпа зерна — нет.