

Сборка и функциональный анализ транскриптома пшеницы сорта Саратовская 29

Александр В. Вихорев^{1,2}

Научный руководитель: Олеся Ю. Шоева ²

¹Новосибирский национальный исследовательский государственный университет (НГУ)

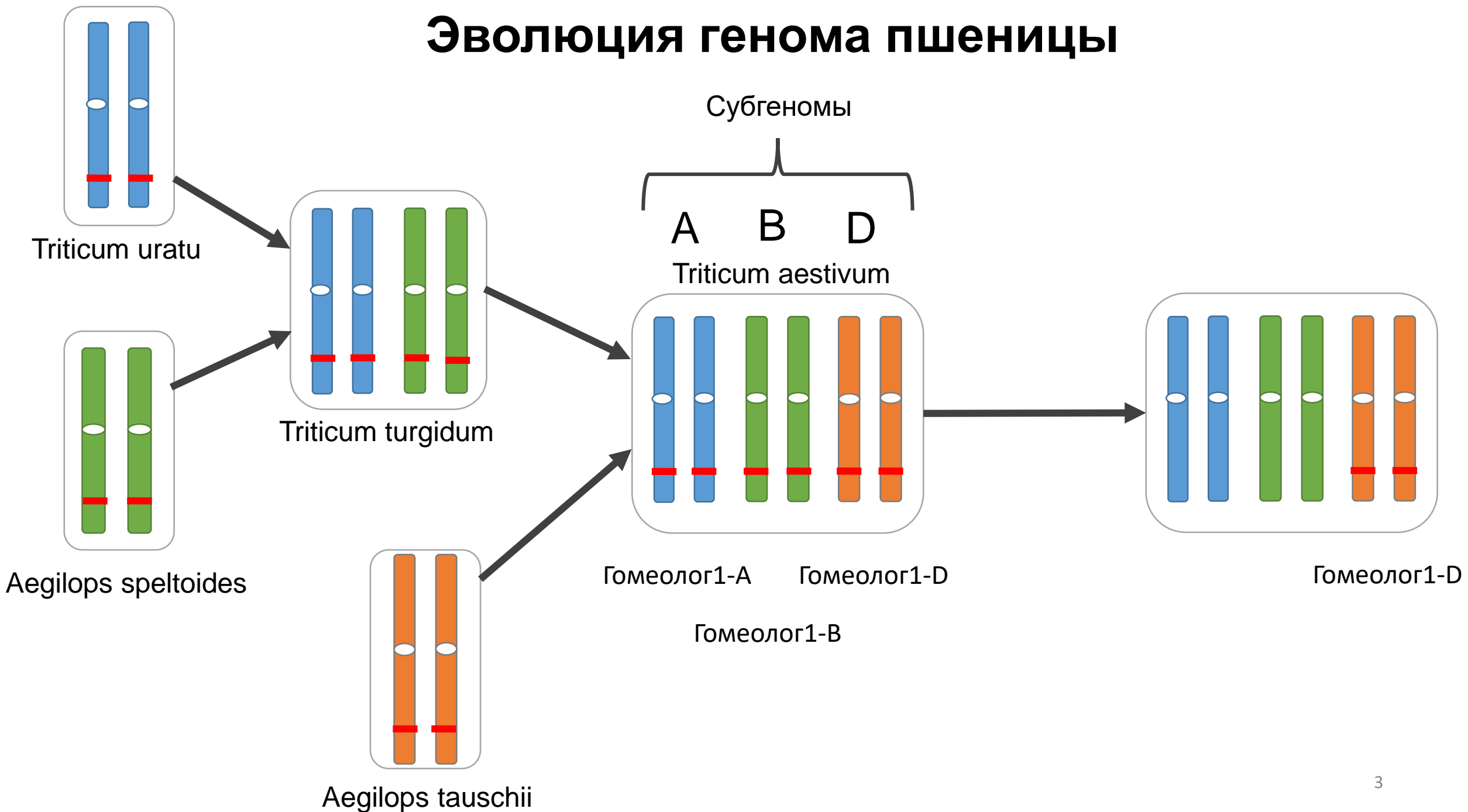
²Институт Цитологии и Генетики СО РАН (ИЦиГ), сектор функциональной генетики злаков
e-mail: vikhorev@bionet.nsc.ru

Мягкая пшеница (*Triticum aestivum*)

- Одно из самых важных растений
- Обеспечивает потребление до 20% от всех калорий
- Аллогексаплоид (AABBDD), 42 хромосомы, ~120 000 генов

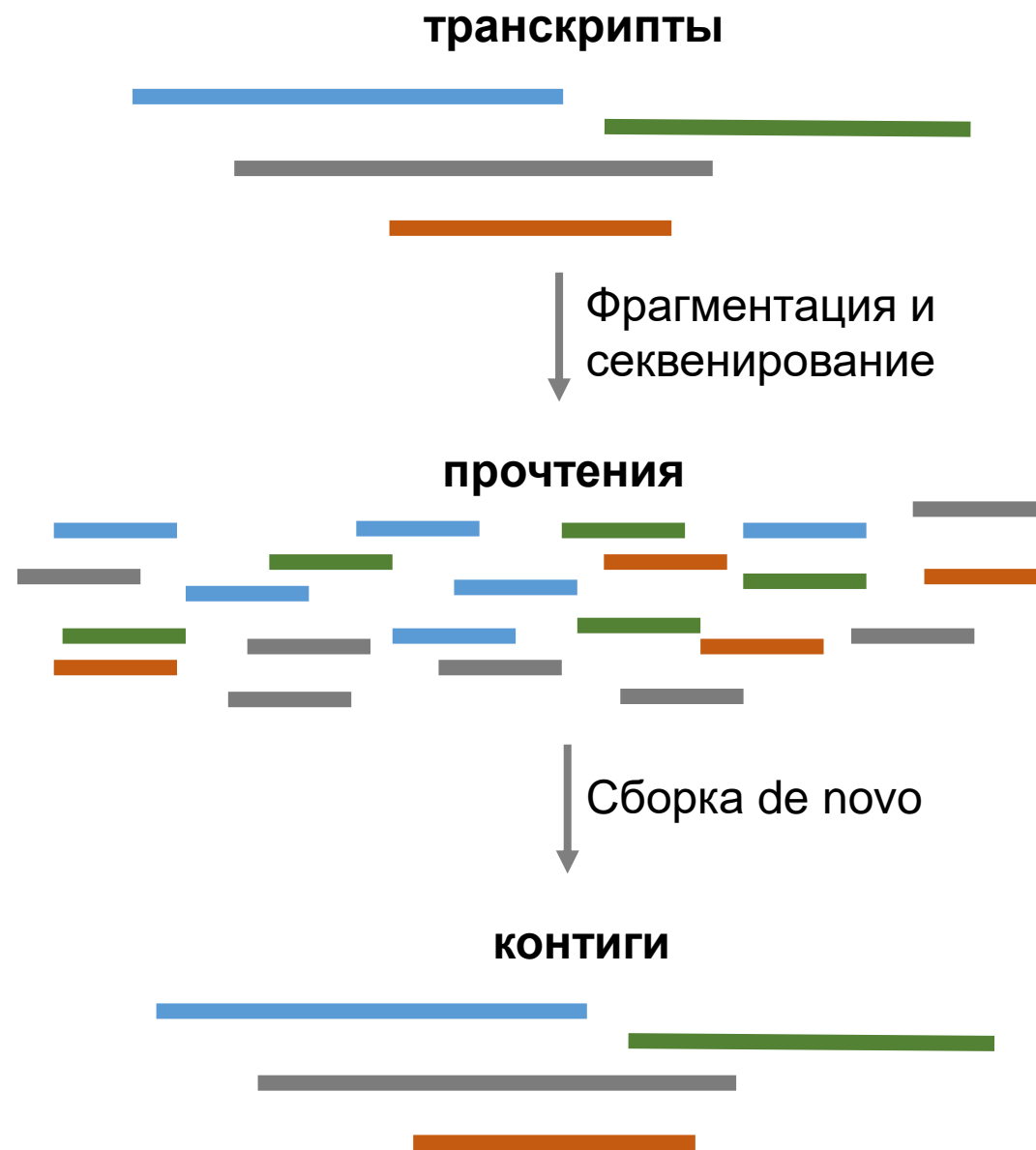


Эволюция генома пшеницы



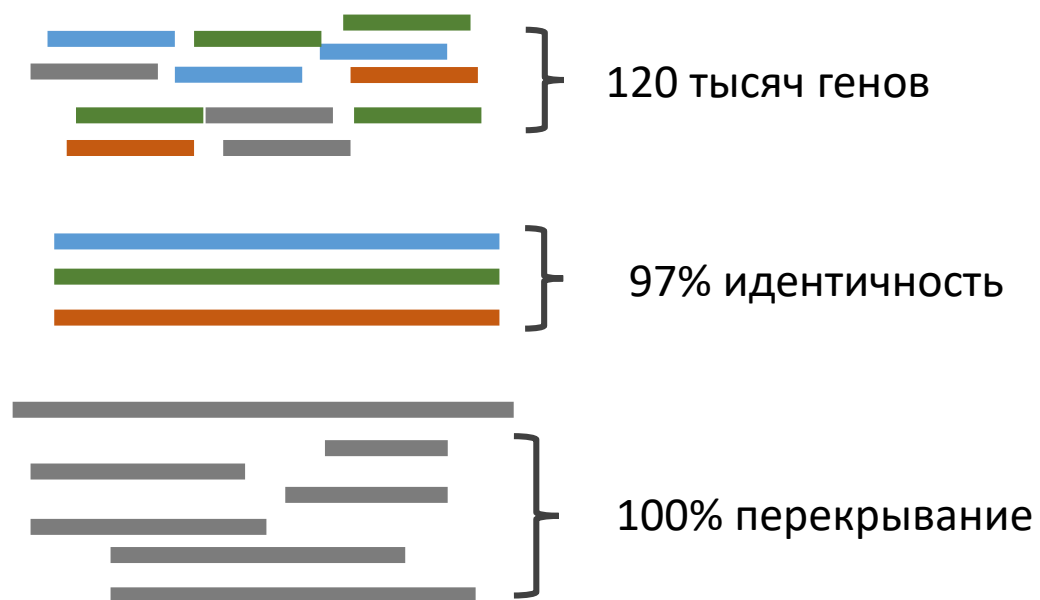
Изучение транскриптома при помощи RNA-seq

- Подходит для работы с любыми организмами
- Позволяет находить новые транскрипты и изоформы генов
- Собран референсный геном и транскриптом модельного сорта Chinese Spring в ходе работы консорциума IWGSC.
- Ресеквенирование местных сортов является актуальной задачей



Сборка транскриптома пшеницы – нетривиальная задача

- Огромный размер генома и транскриптома
- Большое количество высоко гомологичных последовательностей
- Большое количество избыточных контигов



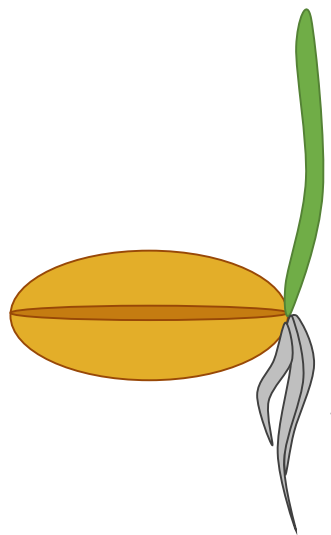
Цель работы – изучение транскриптома аллогексаплоидной пшеницы сорта Саратовская 29 на стадии проростков с помощью RNA-seq анализа.

Задачи:

1. Разработать конвейер для сборки транскриптома пшеницы;
2. Собрать de novo транскриптом пшеницы сорта Саратовская 29 на стадии проростков;
3. Функционально аннотировать полученную de novo сборку транскриптома;
4. Осуществить поиск и функциональный анализ дифференциальной экспрессии генов между корнем и coleoptile;
5. Оценить роль субгеномов пшеницы в развитии корня и coleoptile.

Объект исследования

Сорт – Саратовская 29



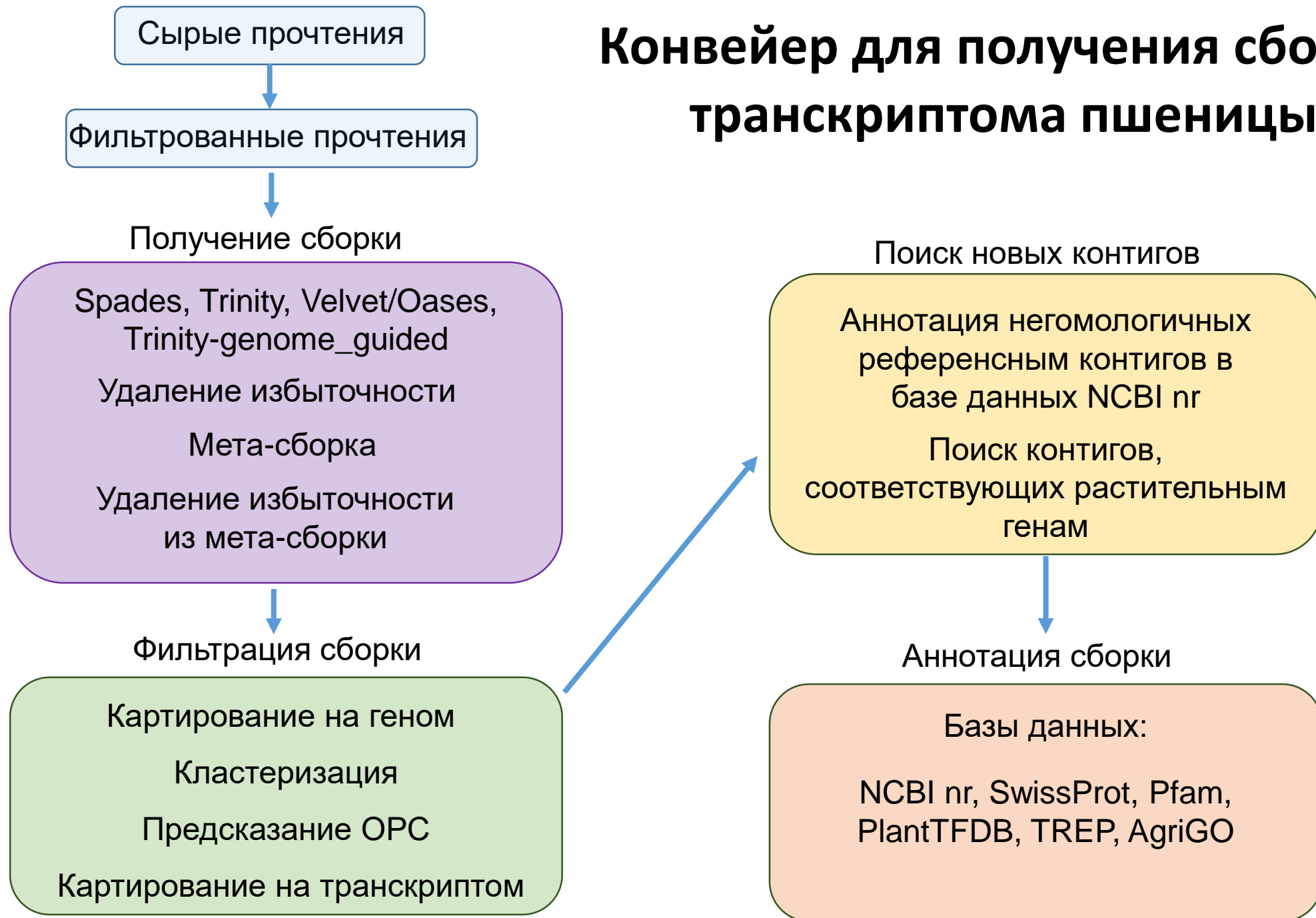
3 библиотеки из колеоптиле
C1_S1, C2_S2, C3_S3

3 библиотеки из корней
R1_S4, R2_S5, R3_S6

Парные прочтения длиной 75 п.н.
были получены на аппарате
Illumina NextSeq 550



Конвейер для получения сборки транскриптома пшеницы



Оценка сборки транскриптома de novo

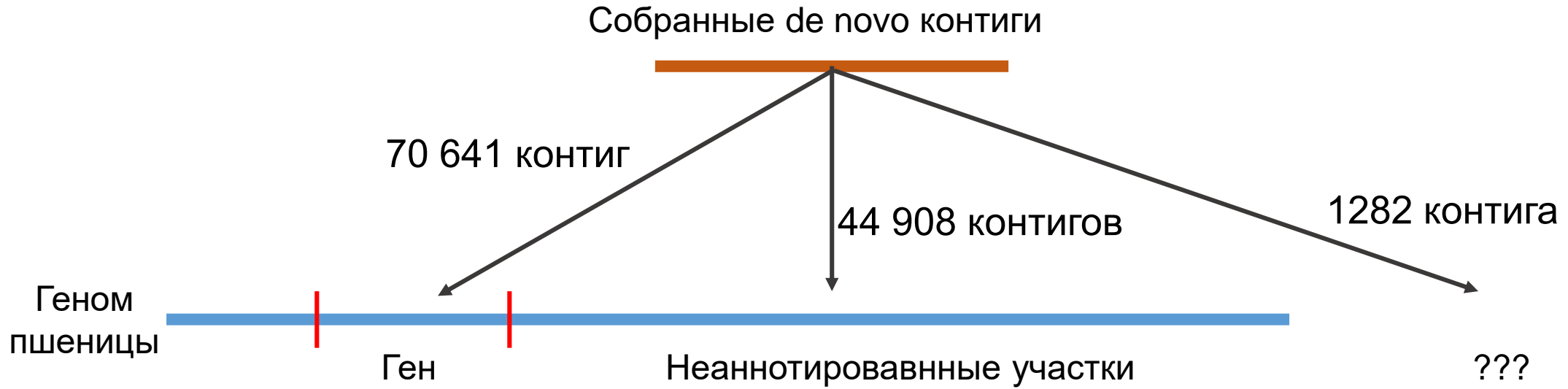
Характеристики транскриптома	Spades	Trinity	Velvet/Oases	Trinity-GG	Мета-сборка
N50	2010	1542	1980	1530	1305
ГЦ-состав, %	48	49	49	51	53
Количество контигов	196 343	331 095	708 744	217 574	585 991
Количество контигов короче 1000 п.н.	98 772	110 648	316 226	78 460	177 305
BUSCO, %	94.4	94.2	94.3	94.3	95.9
Покрытие референсного транскриптома, %	18	16	17	30	53
Сумма нормализованных характеристик	1.3588	1.0877	1.5642	2.7242	5

Формула для нормализации характеристик:

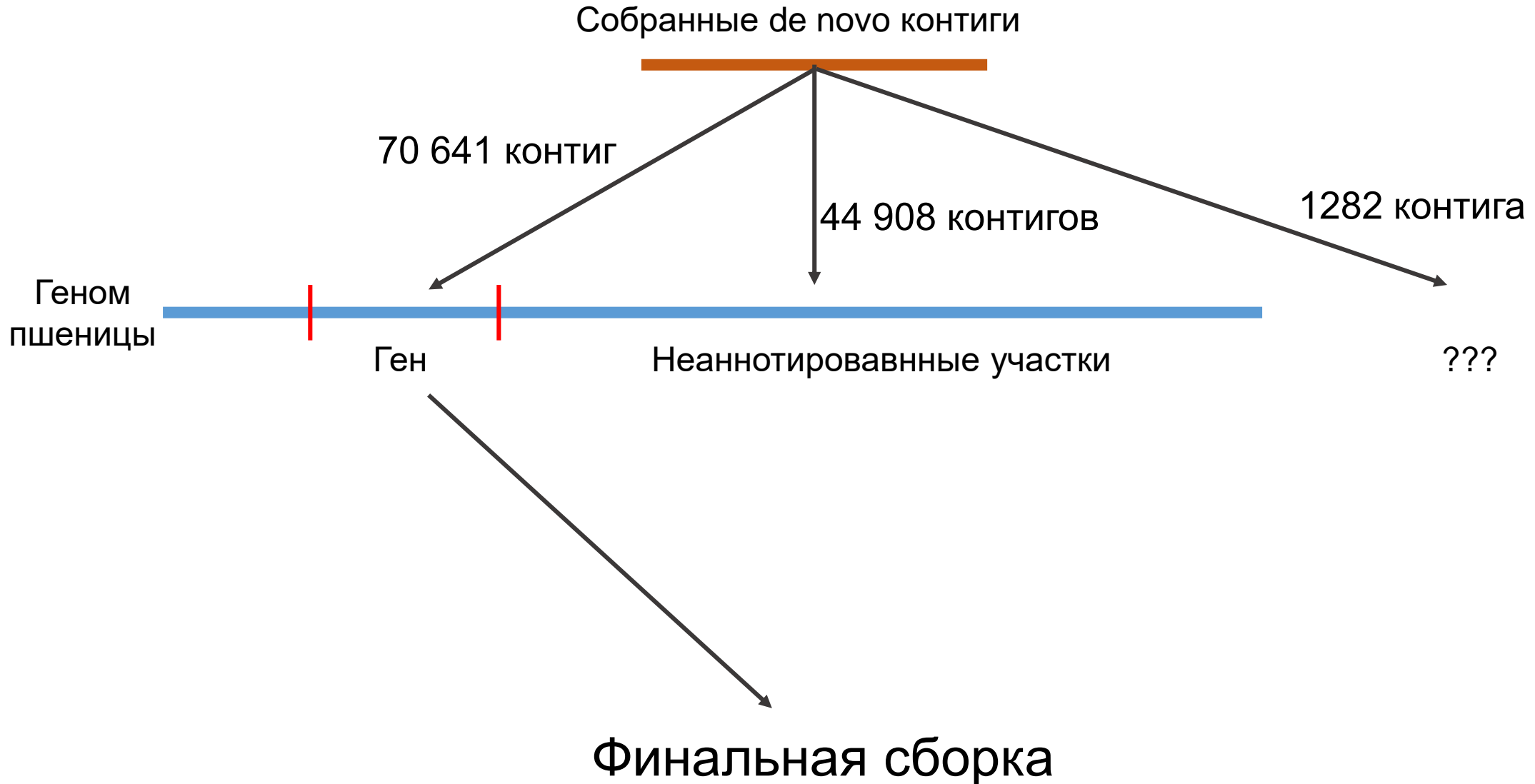
$$\sum N_{ij} = \sum \frac{R_{ij} - \min(R_i)}{\max(R_i) - \min(R_i)} \quad \text{Holzner, Marz, 2018}$$

Мета-сборка является наиболее полной

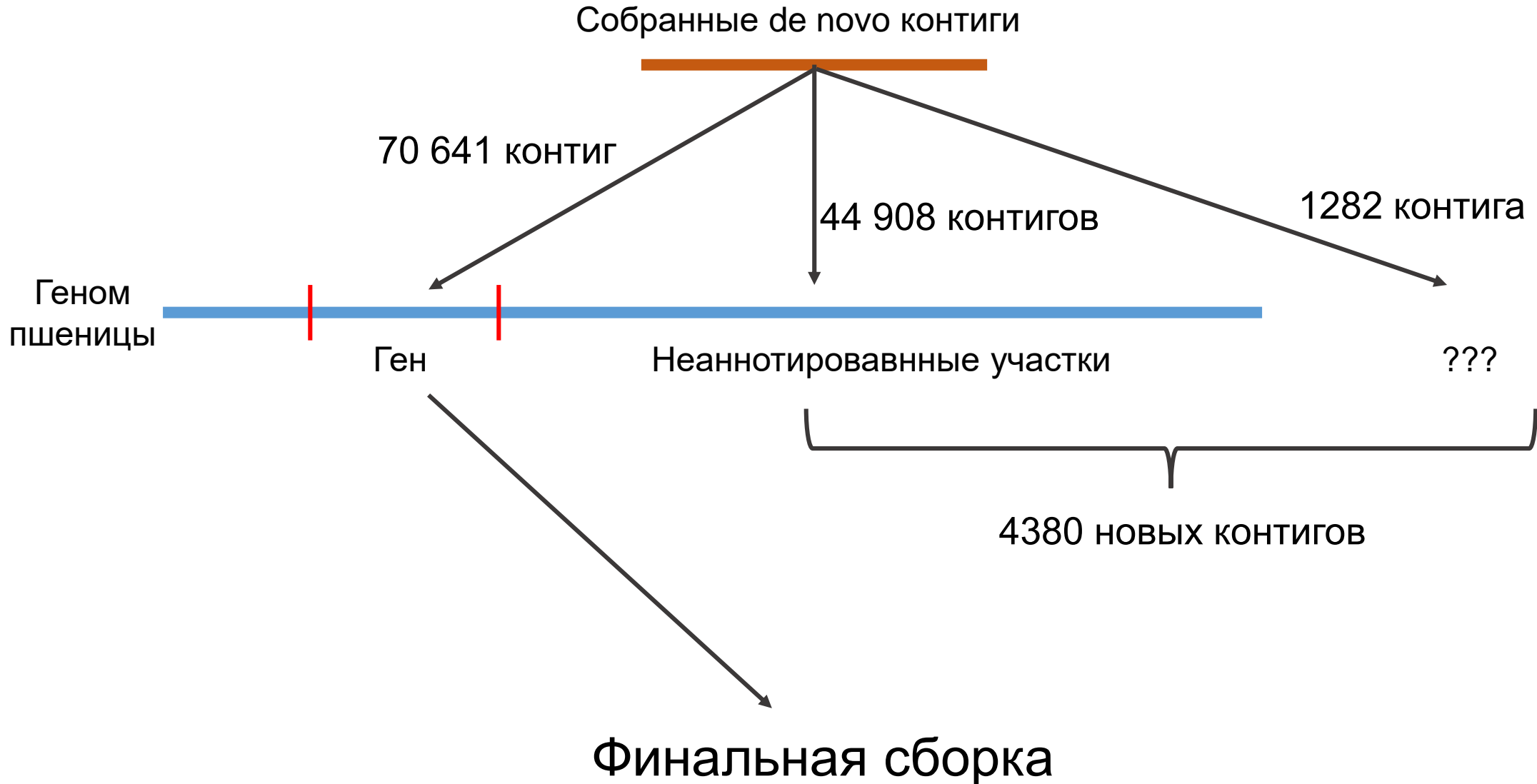
Получение финальной сборки транскриптома



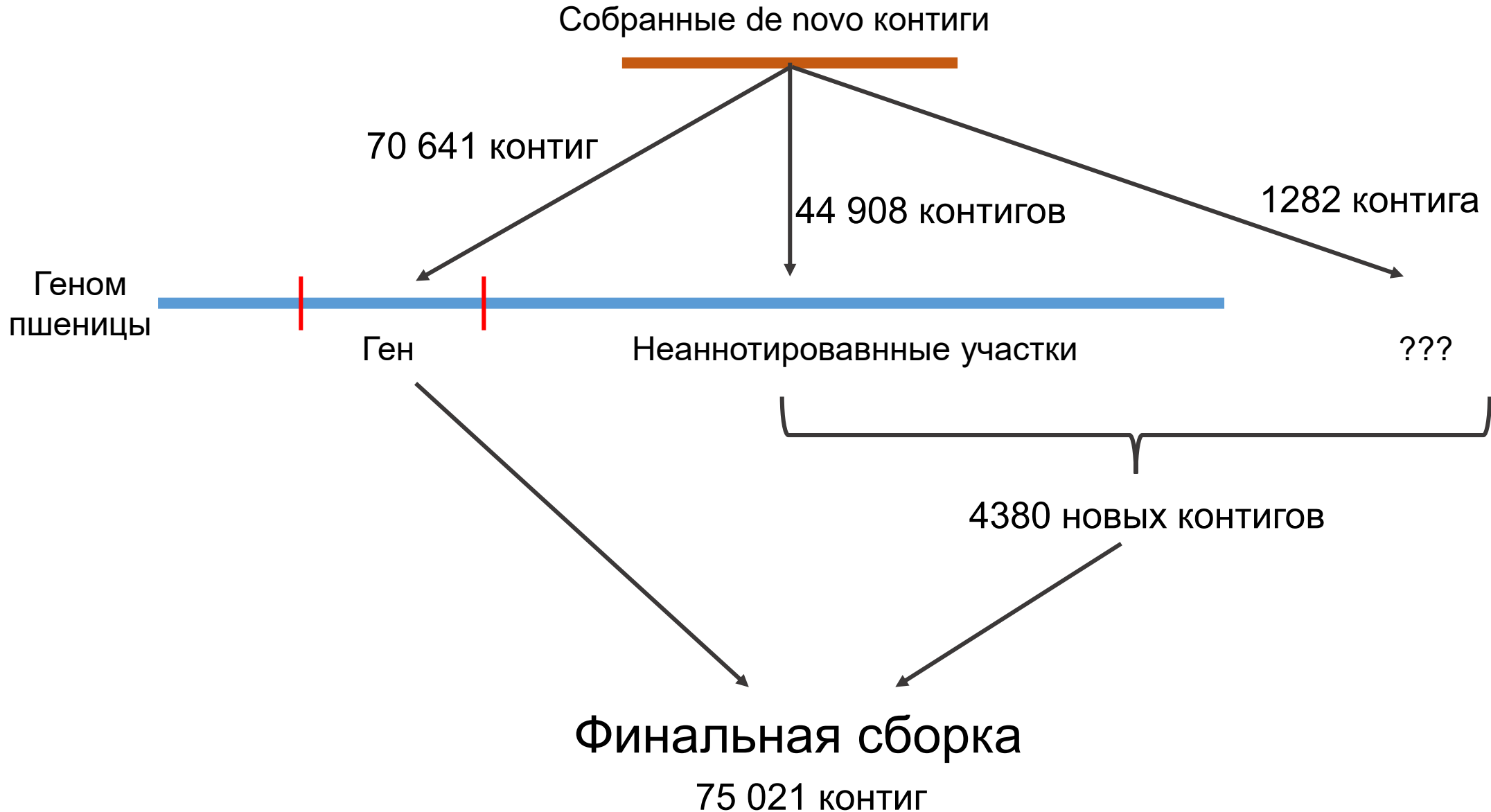
Получение финальной сборки транскриптома



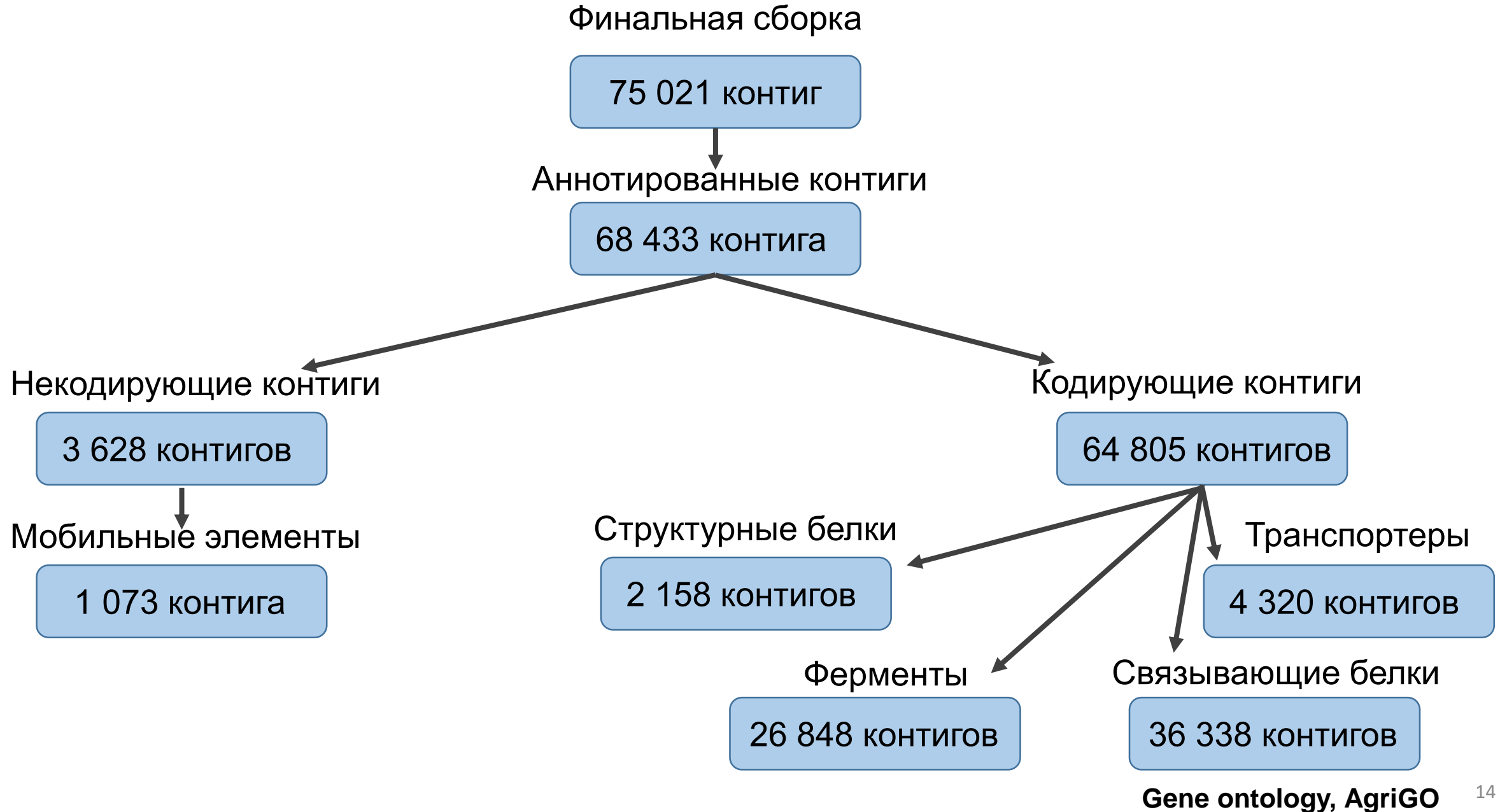
Получение финальной сборки транскриптома



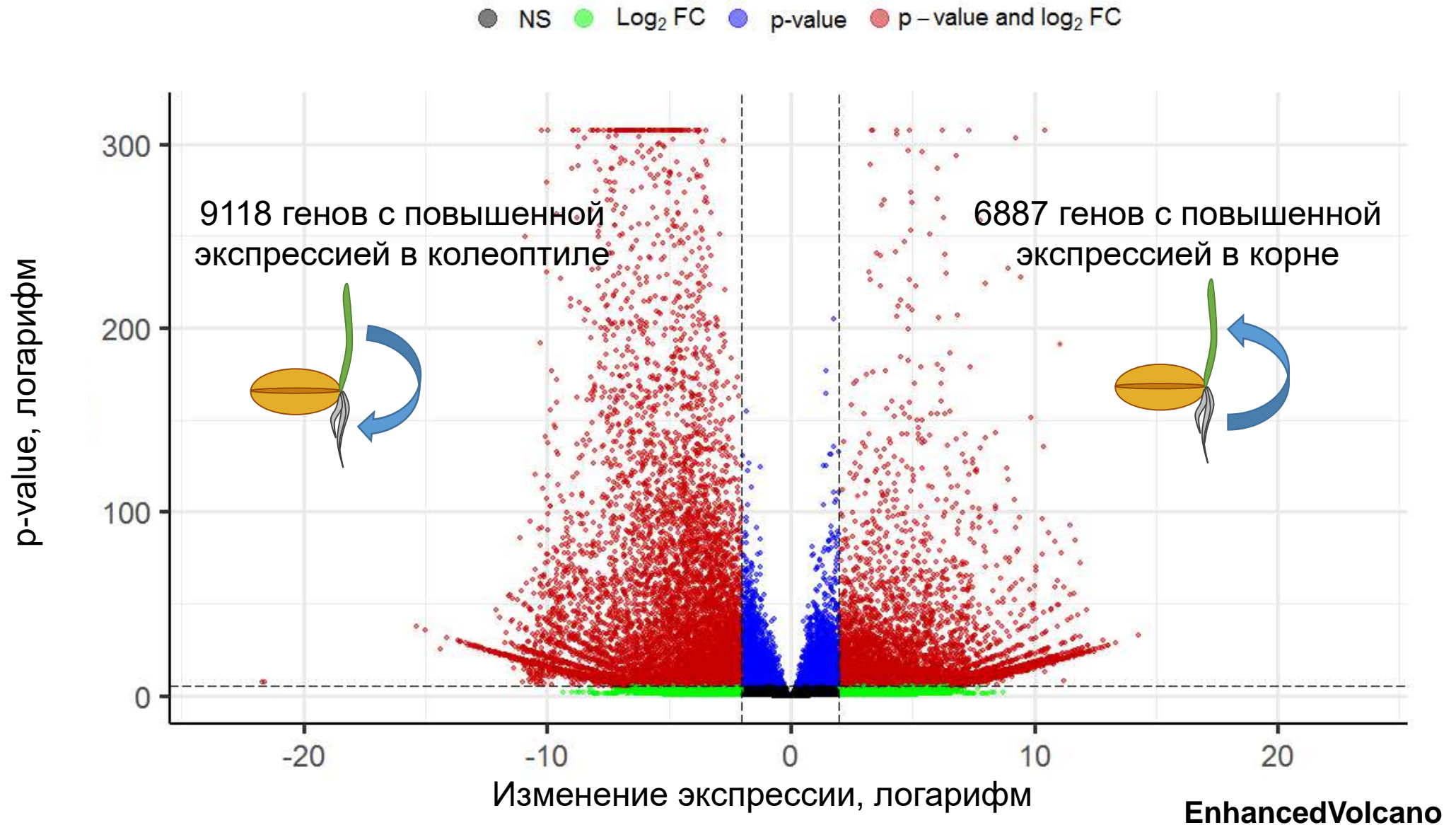
Получение финальной сборки транскриптома



Аннотация сборки транскриптома



Анализ диф. экспрессии генов



Аннотация ДЭГ с повышенной экспрессией в корне



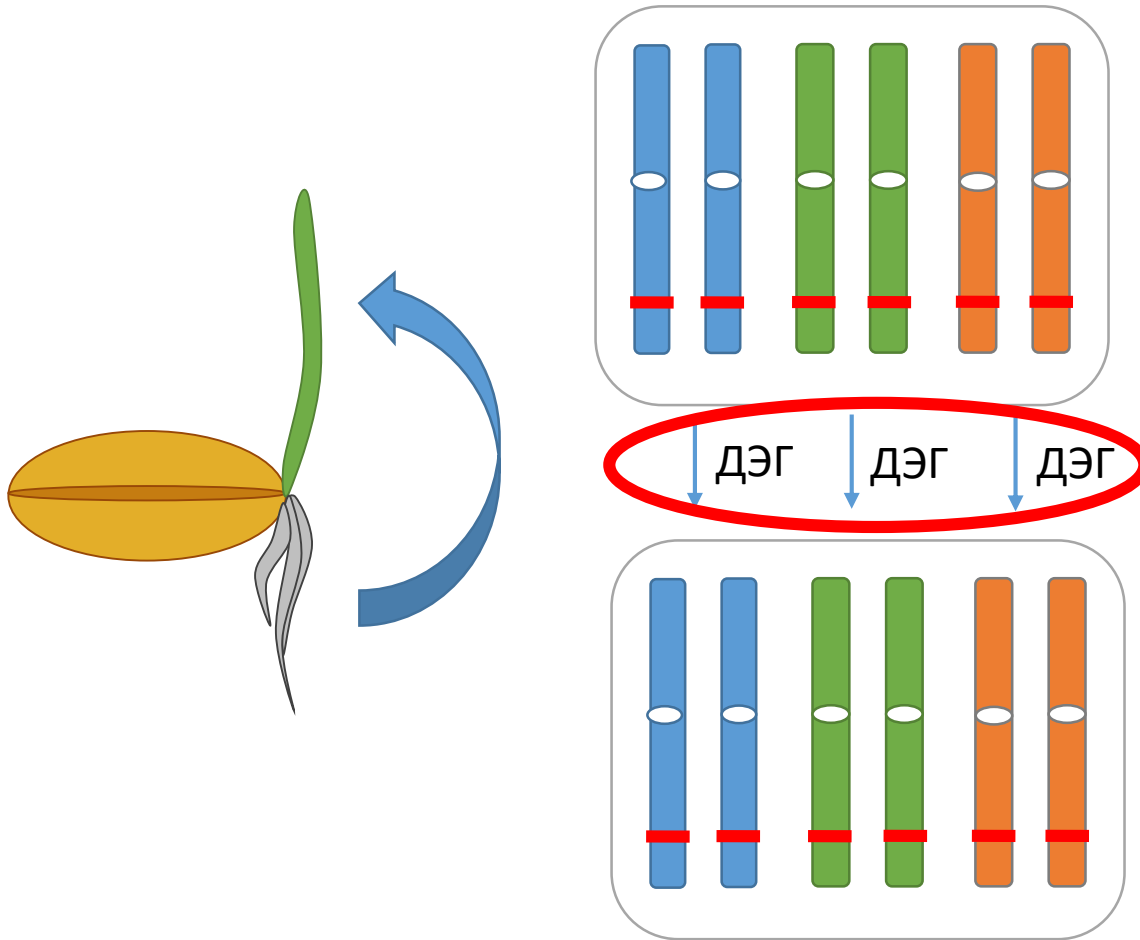
Среди генов, вовлеченных в ответ на стресс и развитие корней, могут быть найдены важные маркеры для селекции

Аннотация ДЭГ с повышенной экспрессией в колеоптиле



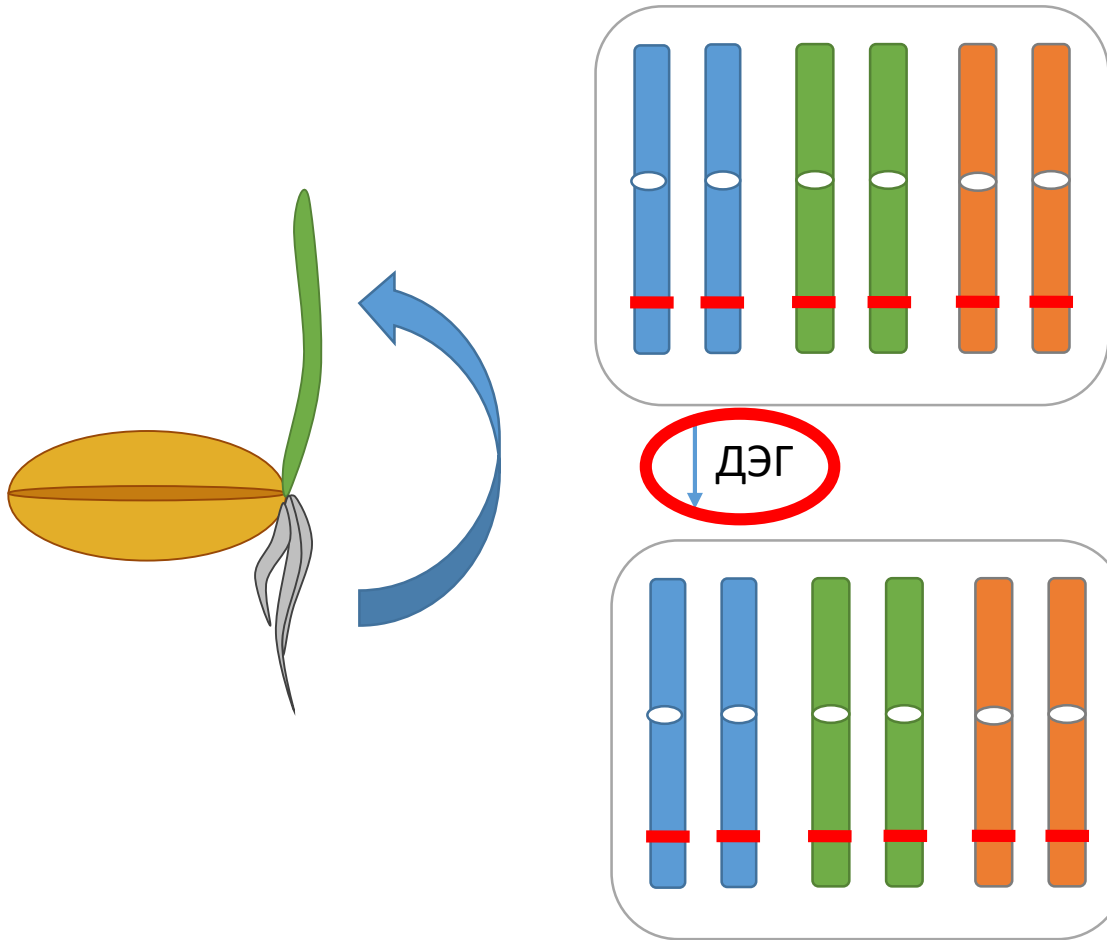
Большая часть генов относится к фотосинтезу

Анализ диф. экспрессии гомеологов



H₀: все три гомеолога изменяют экспрессию

Анализ диф. экспрессии гомеологов

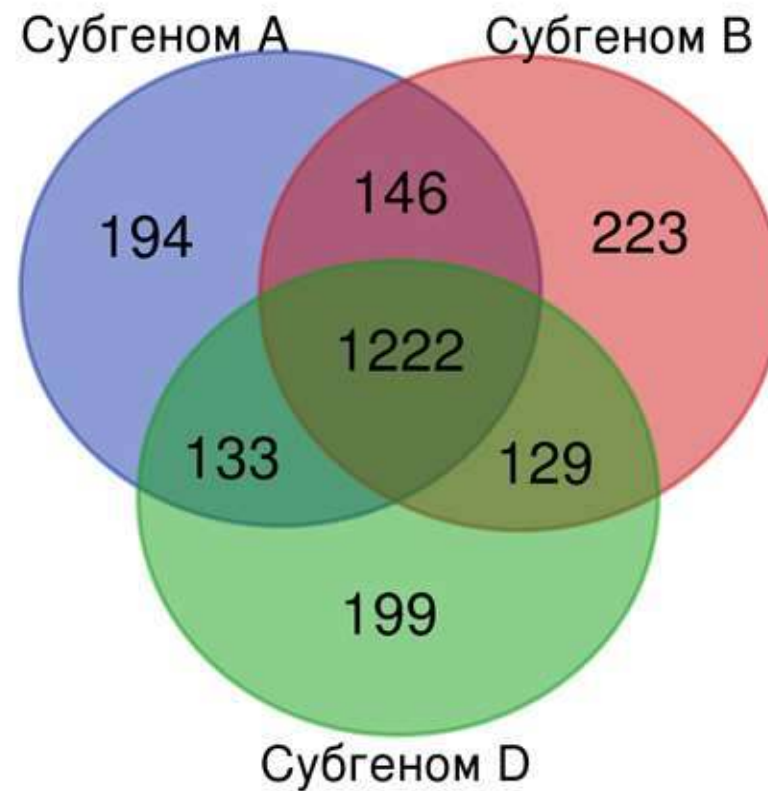
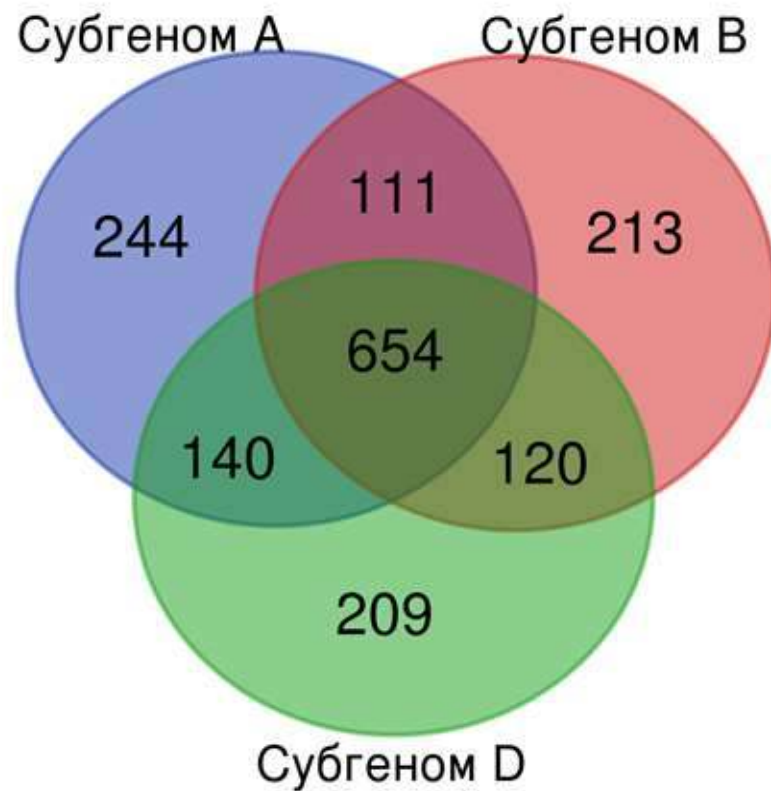


H0: все три гомеолога изменяют экспрессию

H1: экспрессию могут изменять только два или один гомеолог

Анализ диф. экспрессии гомеологов

Повышенная экспрессия в корне Повышенная экспрессия в колеоптиле

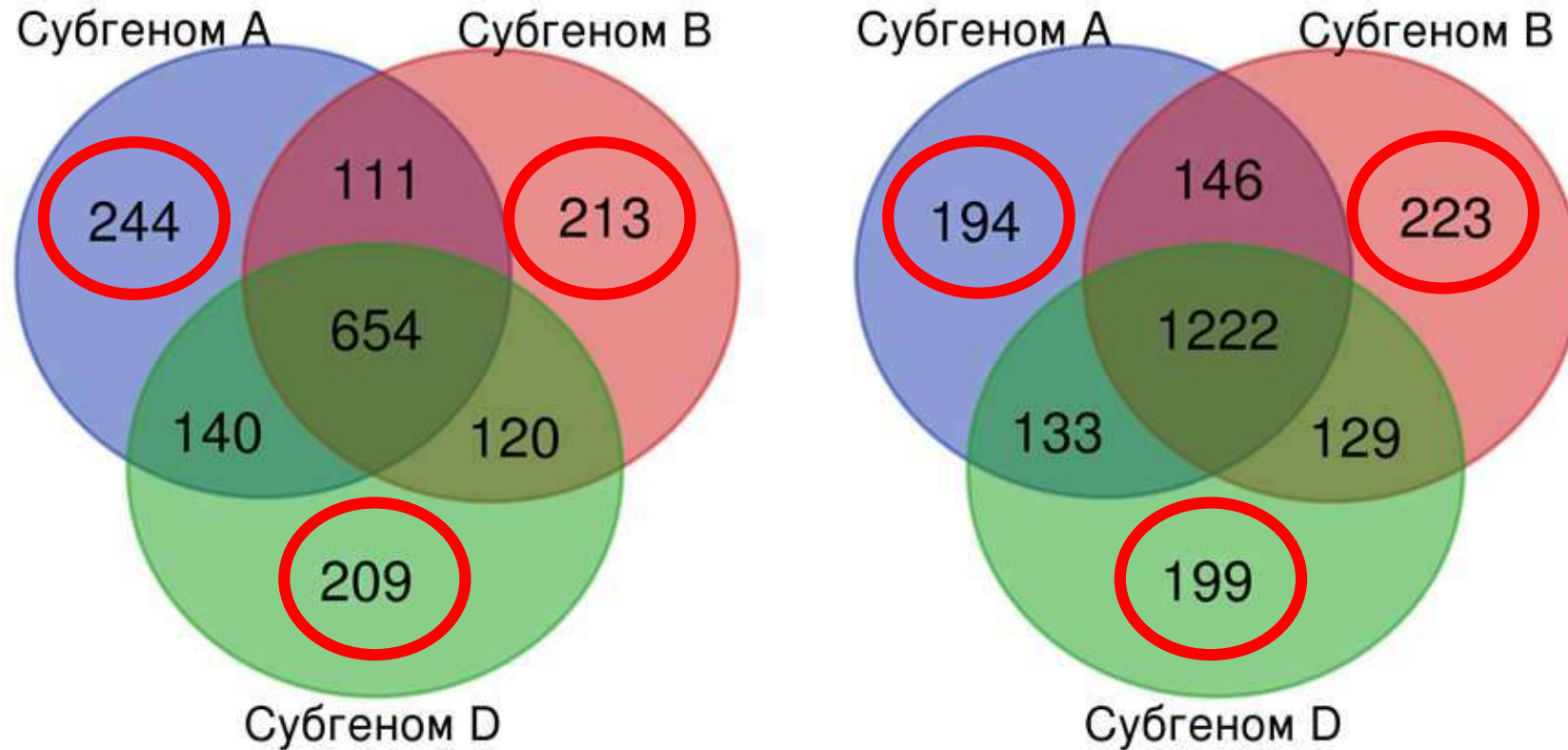


$\chi^2: p < 0.01$

Гомеологи из разных субгеномов экспрессируются неравномерно

Аннотация субгеном-специфичных ДЭГ

Повышенная экспрессия в корне Повышенная экспрессия в колеоптиле



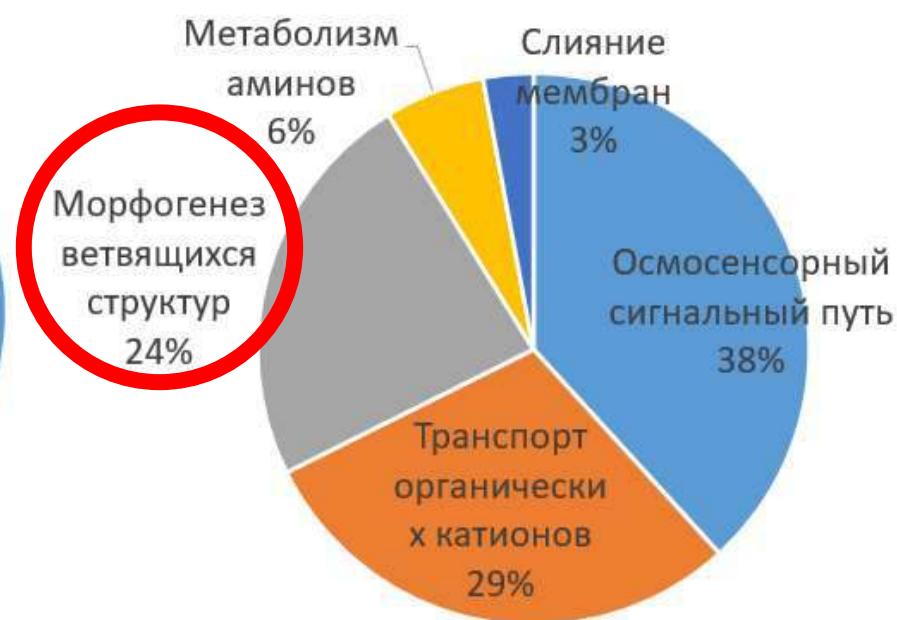
Были выявлены термины ГО, обогащённые для каждого из субгеномов

Аннотация субгеном-специфичных ДЭГ с повышенной экспрессией в корне

Субгеном А



Субгеном В



Субгеном D



Для каждого из субгеномов были выделены специфические термины, функционально связанные с развитием корня

Выводы:

1. Был разработан конвейер для сборки de novo транскриптома пшеницы, включающий использование индивидуальных сборщиков, объединение индивидуальных сборок и удаление избыточности. Полученная с помощью данного конвейера сборка транскриптома превосходит индивидуальные сборки по качеству и полноте;
2. Была получена сборка транскриптома пшеницы сорта Саратовская 29 на стадии проростков, состоящая из 75 021 контига, из которых 4380 (5.8%) не представлены в референсной сборке транскриптома пшеницы сорта Chinese spring;
3. Было аннотировано 68 433 собранных de novo контига, из которых белок-кодирующими являются 64 805 (94,7 %) контигов, а 3 628 (5,3 %) – некодирующими контигами;
4. Было обнаружено 6887 генов с повышенной экспрессией в корне и 9118 генов с повышенной экспрессией в coleoptile. Среди генов с повышенной экспрессией в корне преобладают гены, вовлеченные во вторичный метаболизм и в ответ на факторы окружающей среды, а среди генов с повышенной экспрессией в coleoptile преобладают гены, ответственные за фотосинтез;
5. Была обнаружена функциональная специализация субгеномов в ходе развития корня пшеницы на стадии проростка: субгеном А отвечает за элонгацию корневых волосков, субгеном В отвечает за морфогенез ветвящихся структур, а субгеном D отвечает за регуляцию развития корней. Не было выявлено функциональной специализации субгеномов в развитии coleoptile.

Благодарности:

- Шмаков Николай Александрович
- Глаголева Анастасия Юрьевна
- Васильев Геннадий Владимирович

Оценка качества данных

До предобработки

	GC, %	Q20, %	Q30, %	Кол-во п.н., Gb	Кол-во ридов, M
C1_S1_R1	51.5481	89.868	86.001	3.9744	52.2986
C2_S2_R1	53.1476	89.6397	85.772	12.4686	164.0737
C3_S3_R1	51.7196	89.8057	85.93	3.2590	42.8847
R1_S4_R1	51.5059	90.0609	86.2494	3.8455	50.6028
R2_S5_R1	51.0072	90.2337	86.4458	4.3988	57.8842
R3_S6_R1	51.0977	90.4154	86.657	3.5325	46.4841

Среднее количество прочтений с качеством Q20 больше 90%

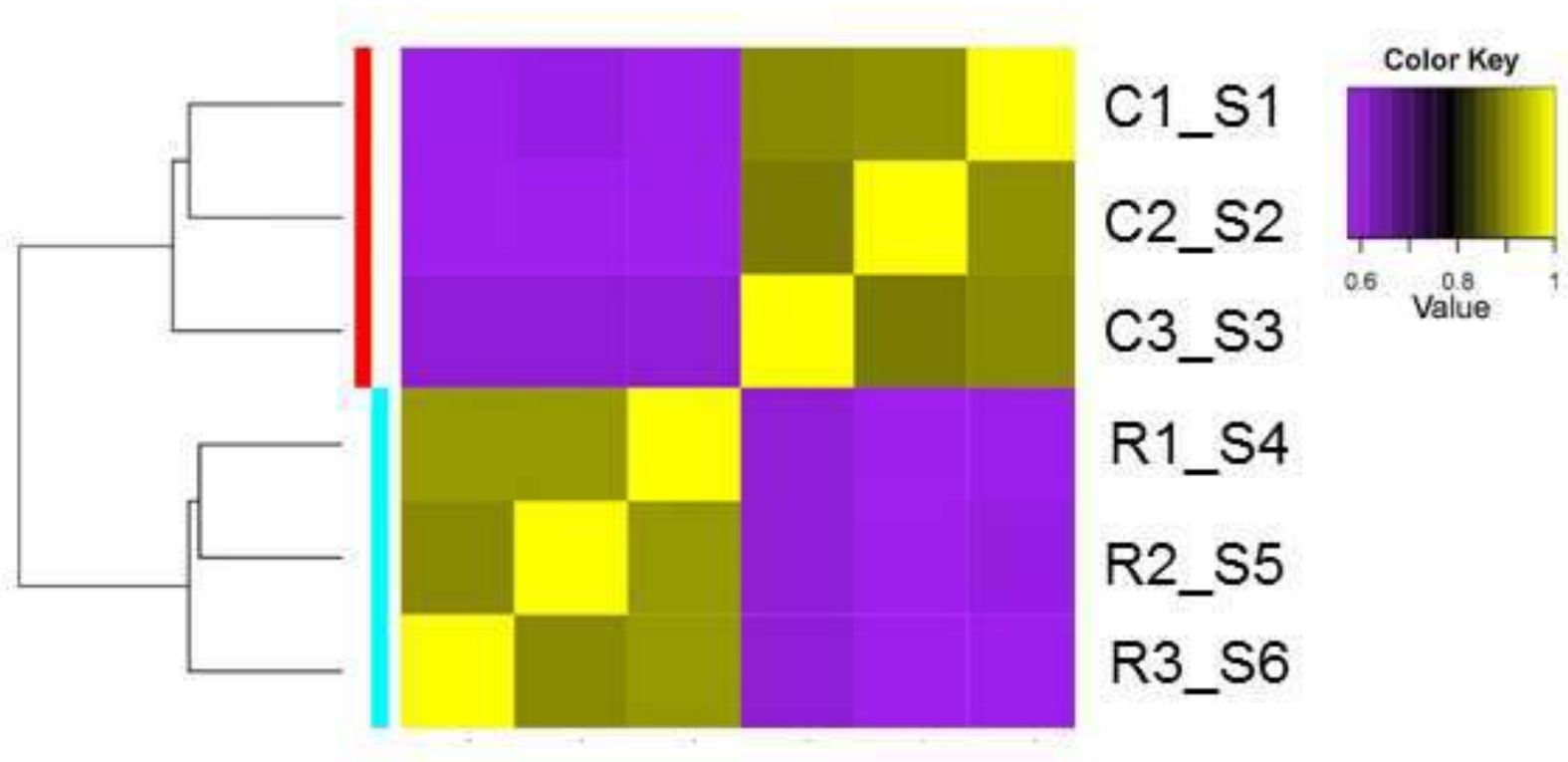
Предобработка данных

После предобработки

	GC, %	Q20, %	Q30, %	Кол-во п.н., Gb	Кол-во ридов, Mb	кол-во прошедших
C1_S1_R1	51.4075	91.6044	87.7964	3.8173	50.6783	0.9690
C2_S2_R1	53.009	91.5056	87.7066	11.933	158.4755	0.9659
C3_S3_R1	51.6372	91.4668	87.6542	3.1390	41.6893	0.9721
R1_S4_R1	51.3585	91.7454	87.9949	3.6986	49.0999	0.9703
R2_S5_R1	50.9244	91.8109	88.0807	4.2442	56.3409	0.9733
R3_S6_R1	50.9616	92.0035	88.2963	3.4039	45.1636	0.9716

97% прочтений прошли предобработку

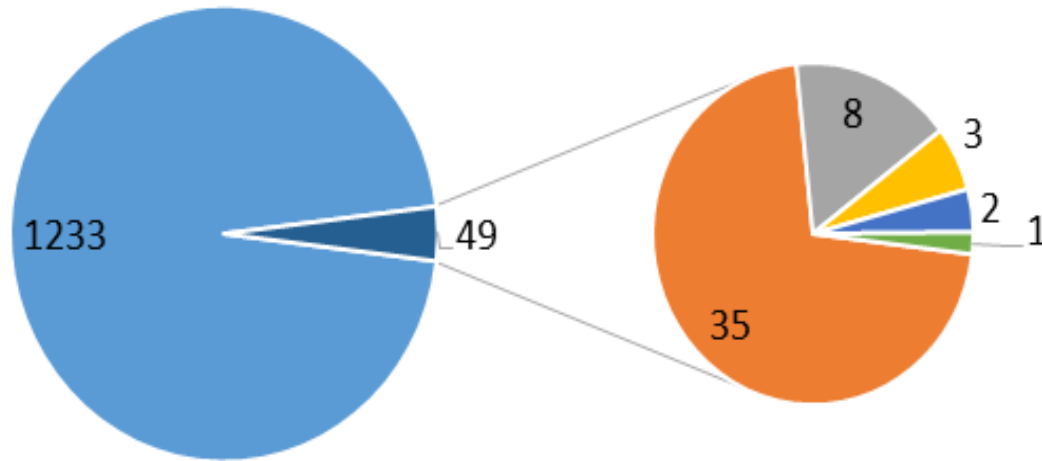
Анализ корреляции между библиотеками



Библиотеки из одних органов коррелируют друг с другом, значит они подходят для дальнейшего анализа.

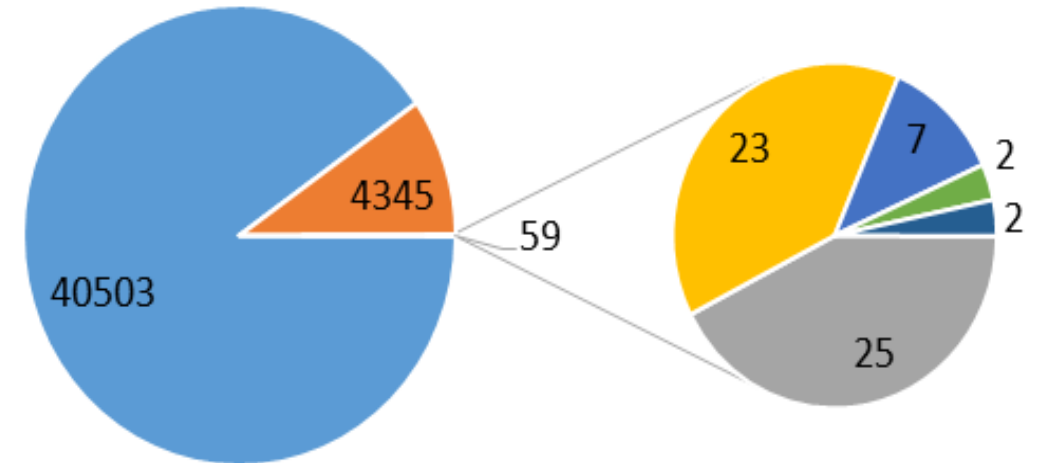
Аннотация негомологичных референсным контигов

А. Не гомологичные референсному геному контиги



- Не аннотированные
- Цветковые растения
- Бактерии
- Грибы
- Вирусы
- Животные

Б. Не гомологичные реф. транскриптому контиги

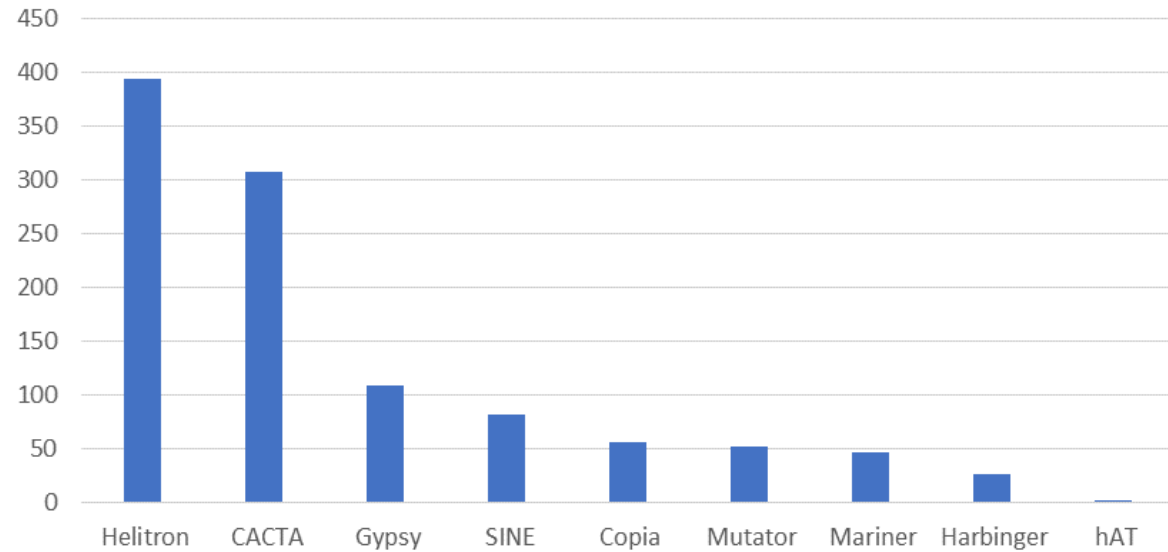


- Не аннотированные
- Цветковые растения
- Бактерии
- Животные
- Грибы
- Хлорелла
- Мхи

4380 контигов, относящиеся к цветковым растениям, были включены в финальную сборку

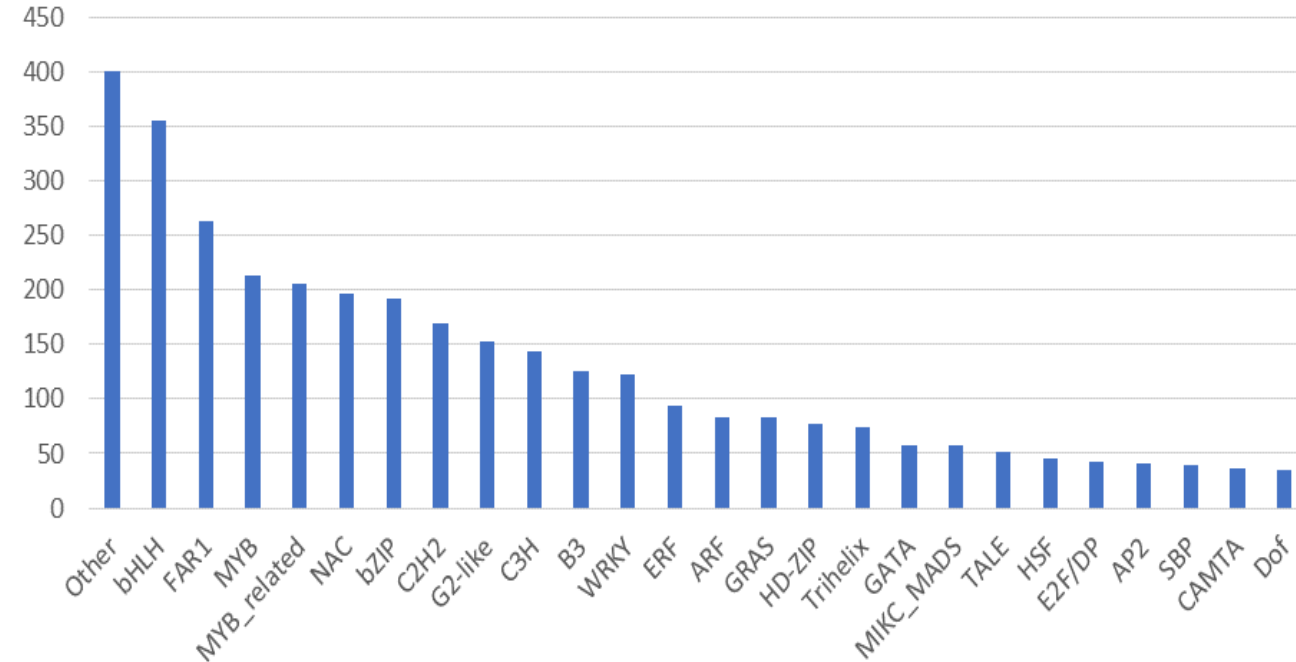
Аннотация сборки транскриптома

Мобильные элементы



1073 контига соответствуют мобильным генетическим элементам

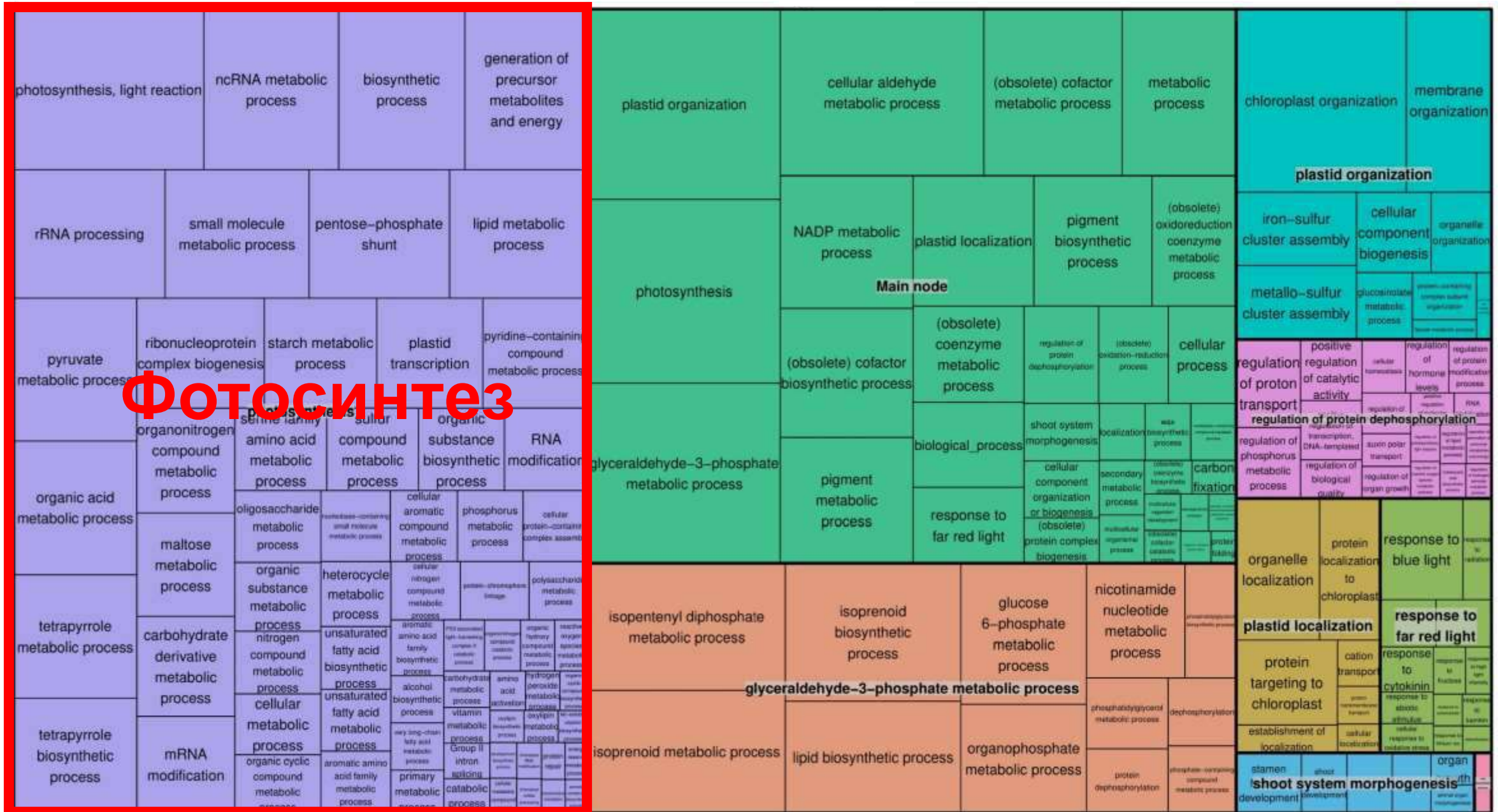
Транскрипционные факторы



3358 контигов соответствуют генам транскрипционных факторов

В базах данных NCBI nr, SwissProt и Pfam было аннотировано 91% собранных контигов

Аннотация ДЭГ с повышенной экспрессией в колеоптиле

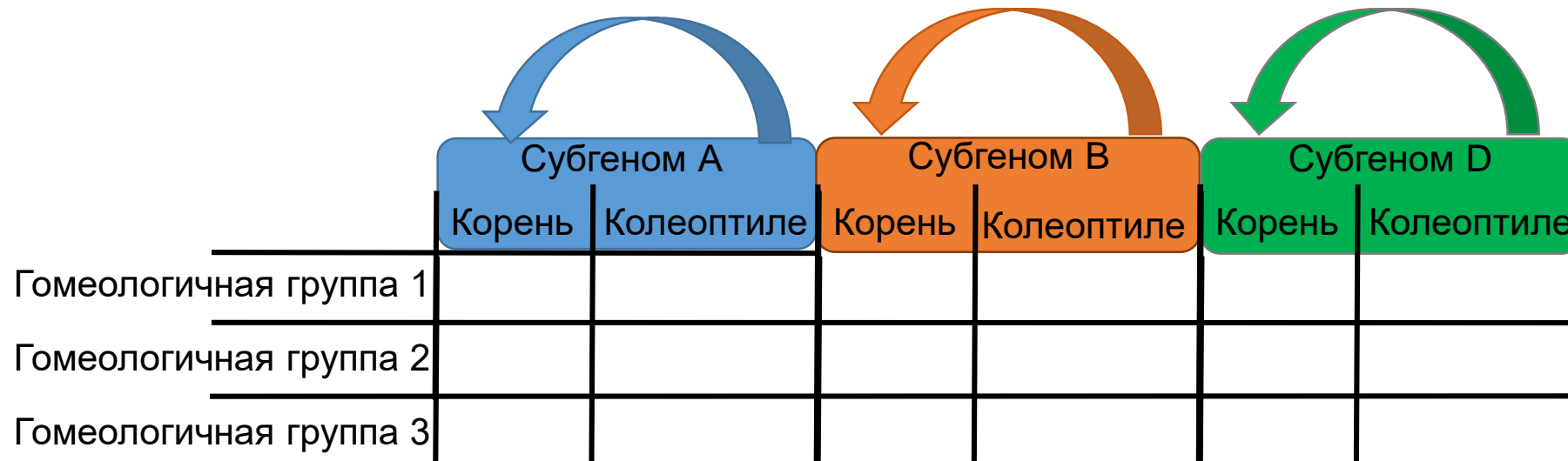


Анализ диф. экспрессии гомеологов



Гомеологичные гены были объединены в гомеологичную группу и была составлена таблица их экспрессии в субгеномах А, В и D

Анализ диф. экспрессии гомеологов



Дифференциальная экспрессия была посчитана между корнем и колеоптиле для гомеологической группы внутри каждого субгенома