

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ  
«ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ ИМ. А.А. ХАРКЕВИЧА  
РОССИЙСКОЙ АКАДЕМИИ НАУК»

*На правах рукописи*

Клепикова Анна Владимировна

**Создание и анализ экспрессионных карт *Arabidopsis thaliana* и  
*Capsella bursa-pastoris***

03.02.07 – генетика

03.01.09 – математическая биология, биоинформатика

Диссертация на соискание ученой степени кандидата биологических наук

Научный руководитель:  
кандидат биологических наук  
Алексей Александрович Пенин

Москва – 2020

## ОГЛАВЛЕНИЕ

<b>Введение</b> .....	<b>5</b>
1. Актуальность темы исследования .....	5
2. Цель и задачи исследования .....	5
3. Методология и методы исследования .....	6
4. Научная новизна .....	6
5. Теоретическая и практическая значимость .....	7
6. Положения, выносимые на защиту .....	7
7. Степень достоверности и апробация результатов .....	8
8. Структура и объем работы .....	8
9. Список публикаций по теме диссертации .....	9
<b>Список сокращений</b> .....	<b>10</b>
<b>Глава 1. Создание и анализ транскриптомной карты <i>A. thaliana</i></b> .....	<b>11</b>
1.1. Введение .....	11
1.2. Обзор литературы .....	13
1.2.1. История создания транскриптомных карт растений .....	13
1.2.2. Технические характеристики транскриптомных карт .....	14
1.2.3. Представленность таксонов среди видов, для которых созданы транскриптомные карты .....	17
1.2.4. Степень детализации транскриптомных карт .....	19
1.2.5. Кластеризация органов и тканей растений по их транскрипционному профилю .....	21
1.2.6. Общее описание транскриптома растений .....	22
1.2.7. Транскрипционные профили образцов .....	24
1.2.8. Поиск стабильно экспрессирующихся генов .....	25
1.2.9. Анализ тканеспецифичной экспрессии генов .....	26
1.2.10. Детальное изучение экспрессии транскрипционных факторов .....	29
1.2.11. Экспрессионные базы данных .....	30
1.2.12. Проблемы и перспективы изучения транскриптомных карт .....	31
1.3. Результаты .....	33
1.3.1. Выбор образцов транскриптомной карты .....	33
1.3.2. Секвенирование и первичная обработка данных .....	34
1.3.3. Интегральные параметры экспрессии генов .....	35
1.3.4. Сравнение параметров экспрессии генов в образцах .....	39
1.3.5. Анализ дифференциальной экспрессии .....	40
1.3.6. Стабильность экспрессии генов .....	42
1.3.7. Специфичность экспрессии генов .....	43
1.3.8. Паттерны экспрессии семейств транскрипционных факторов .....	44
1.4. Обсуждение .....	46
1.4.1 Структура транскриптомной карты .....	46
1.4.2. Интегральные параметры транскриптома <i>A. thaliana</i> .....	48
1.4.3. Схожесть профилей экспрессии органов и тканей .....	50
1.4.4. Различия профилей экспрессии образцов и разнообразие экспрессии генов .....	51
1.4.5. Стабильно экспрессирующиеся гены .....	52
1.4.6. Паттерны экспрессии генов .....	54
1.4.7. Разнообразие паттернов экспрессии транскрипционных факторов .....	55
1.4.8. Создание экспрессионной базы данных <i>TraVA</i> .....	56
1.5. Заключение .....	57
<b>Глава 2. Временная серия апикальных меристем <i>A. thaliana</i></b> .....	<b>59</b>
2.1. Введение .....	59
2.2. Обзор литературы .....	60
2.2.1. Важность перехода к цветению для жизни растений .....	60
2.2.2. Фотопериодический путь инициации цветения .....	60

2.2.3. Передача индуктивного сигнала в апикальную меристему .....	62
2.2.4. Основные события, происходящие в апикальной меристеме .....	63
2.2.5. Другие пути инициации цветения .....	64
2.2.6. Степень изученности глобальных изменений транскриптома апикальной меристемы при переходе к цветению .....	65
2.3. Результаты .....	67
2.3.1. Сбор образцов временной серии, секвенирование и первичная обработка данных .....	67
2.3.2. Динамика ключевых генов – регуляторов цветения .....	69
2.3.3. Общий анализ дифференциально экспрессирующихся генов .....	71
2.3.4. Кластеризация генов согласно их профилям экспрессии .....	72
2.3.5. Анализ профилей экспрессии отдельных классов генов .....	74
2.3.6. Верификация ключевых стадий временной серии меристем с помощью секвенирования независимо собранных образцов .....	75
2.4. Обсуждение .....	77
2.4.1. Морфология растений <i>A. thaliana</i> при переходе к цветению .....	77
2.4.2. Основные паттерны экспрессии генов на стадии перехода .....	77
2.4.3. Паттерны экспрессии генов, связанных с S-фазой .....	78
2.4.4. Паттерны экспрессии генов, связанных с митозом .....	78
2.4.5. Предлагаемая модель, описывающая изменение экспрессии генов в точке перехода к цветению .....	79
2.4.6. Расширение предложенной модели .....	81
2.4.7. Соответствие наблюдаемых изменений в экспрессии генов данным гистологического анализа .....	82
2.5. Заключение .....	84
<b>Глава 3. Органоспецифический ответ на холодостресс у <i>A. thaliana</i>.....</b>	<b>85</b>
3.1. Введение .....	85
3.2. Обзор литературы .....	86
3.2.1. Приспособление растений к холодострессу .....	86
3.2.2. Механизмы восприятия пониженной температуры .....	87
3.2.3. Пути передачи холодострессового сигнала .....	87
3.2.4. Транскрипционная регуляция холодострессового ответа .....	88
3.2.5. Регулон генов СВФ .....	90
3.2.6. Баланс между развитием и защитой при холодострессовой адаптации .....	91
3.2.7. Исследованность ответа на холод .....	92
3.3. Результаты .....	94
3.3.1. Выбор образцов и получение первичных данных .....	94
3.3.2. Анализ дифференциальной экспрессии .....	95
3.3.3. Изменение экспрессии генов регулона СВФ в различных органах .....	95
3.3.4. Определение общих и органоспецифических ДЭ генов .....	96
3.3.5. Поведение известных генов, участвующих в ответе на холод .....	97
3.3.6. Экспрессионные характеристики генов, отвечающих на холодострессовое воздействие - Энтропия Шеннона .....	99
3.3.7. Экспрессионные характеристики генов, отвечающих на холодострессовое воздействие - ДЭ-значение .....	101
3.3.8. Анализ регуляторных элементов в промоторных областях ДЭ генов .....	102
3.4. Обсуждение .....	104
3.4.1 Предварительный анализ данных .....	104
3.4.2. Анализ общих и органоспецифических ДЭ генов .....	105
3.4.3. Неравномерная изученность органов <i>A. thaliana</i> .....	107
3.4.4. Паттерны экспрессии генов, отвечающих на холодострессовое воздействие .....	109
3.4.5. Дифференциальное участие регулируемых холодом генов в нормальном развитии растения .....	112

3.4.6. Привлечение разных транскрипционных факторов в органах <i>A. thaliana</i> .....	113
3.5. Заключение .....	114
<b>Глава 4. Транскриптомная карта <i>Capsella bursa-pastoris</i> .....</b>	<b>115</b>
4.1. Введение .....	115
4.2. Обзор литературы .....	116
4.2.1. Полиплоидизация - основной путь эволюции растительных геномов .....	116
4.2.2. Последствия полиплоидизации для транскриптома .....	117
4.2.3. <i>Capsella bursa-pastoris</i> – модельный объект для изучения ранней эволюции .....	118
4.3. Результаты .....	120
4.3.1. Сбор образцов, секвенирование и первичная обработка данных .....	120
4.3.2. Общие экспрессионные характеристики транскриптома <i>C. bursa-pastoris</i> .....	120
4.3.3. Анализ дифференциальной экспрессии .....	121
4.3.4. Анализ дифференциальной экспрессии гомеологичных генов .....	123
4.3.5. Изучение промоторных областей гомеологичных генов .....	124
4.4. Обсуждение .....	126
4.4.1 Сбор образцов, секвенирование и первичная обработка данных .....	126
4.4.2. Отсутствие отличий между субгеномами по общим экспрессионным характеристикам .....	126
4.4.3. Различия во вкладе в дифференциальную экспрессию двух субгеномов .....	127
4.4.4. Различия в паттернах экспрессии гомеологичных генов .....	128
4.4.5. Интеграция транскриптомных данных в базу <i>TraVA</i> .....	129
4.5. Заключение .....	130
<b>Материалы и методы.....</b>	<b>131</b>
<b>Выводы.....</b>	<b>136</b>
<b>Список литературы .....</b>	<b>137</b>
<b>Приложения.....</b>	<b>160</b>

## ВВЕДЕНИЕ

### 1. Актуальность темы исследования

Несмотря на большой прогресс генетики растений в последние годы, функции многих генов остаются неизвестными или малоизученными. Даже для *Arabidopsis thaliana* (L.) Heynh. – модельного объекта генетики растений, детальные эксперименты, направленные на определение функций проведены не более чем для 30% генов. Для остальных функции до сих пор постулируются на основании массовых экспериментов с использованием данных по экспрессии генов и белков, в результате чего генам присваиваются категории формальной системы аннотации – ГО (Генная Онтология).

Однако, несмотря на большое количество данных по *A. thaliana*, включая подробный атлас экспрессии (Schmid et al., 2005), около 30% генов до сих пор не имеют даже такого варианта аннотации или их описания слабо детализированы. Это связано как с низкой чувствительностью методов, которые использовались при получении данных для создания классификации, так и с недостаточно высоким разнообразием образцов. Создание атласа экспрессии генов *A. thaliana* с использованием новых методов с более высоким разрешением не проводилось, хотя это позволило бы получить принципиально новую информацию по уровням и паттернам экспрессии. Это приводит к негативным последствиям, особенно в случае модельного объекта, так как результаты, полученные для него, служат основой для функциональной аннотации генов других видов, для которых проведение долгосрочных комплексных исследований невозможно.

Это делает актуальным создание транскриптомной карты этого объекта с использованием максимально доступных методов анализа, а также расширение спектра анализируемых образцов. При этом одним из важнейших критериев выбора в таком эксперименте должна являться возможность изучения процессов, анализ которых ранее не проводился. Помимо этого, так как паттерн экспрессии может опосредованно свидетельствовать о функции генов (Assis and Bachtrog, 2013), эту информацию можно использовать для изучения таких явлений, как ранние стадии субфункционализации генов после процессов полиплоидизации. Ранее проведение таких исследований было затруднительно из-за сложности различения экспрессии гомеологичных генов.

### 2. Цель и задачи исследования

Целью настоящего исследования было создание транскриптомных карт модельного объекта генетики растений *Arabidopsis thaliana* и близкого к нему аллотетраплоида *Capsella bursa-pastoris*, а также оценка возможности использования полученных данных для анализа биологических процессов.

В рамках поставленной цели решались следующие задачи:

- 1) создание и анализ детализированной транскриптомной карты *A. thaliana* с использованием метода RNA-seq;
- 2) изучение динамики экспрессии генов в меристеме *A. thaliana* при переходе к цветению;
- 3) изучение тканевой специфичности ответа на холодостресс у *A. thaliana*;
- 4) создание транскриптомной карты раннего аллотетраплоида *C. bursa-pastoris* (L.) Medik. с использованием метода RNA-seq.
- 5) анализ механизмов эволюции экспрессии гомеологичных генов на ранних этапах эволюции аллотетраплоида *C. bursa-pastoris*.

### 3. Методология и методы исследования

Все проведенные в нашей работе исследования были выполнены по стандартному плану: выбор изучаемых образцов, сбор образцов, выделение РНК и приготовление библиотек кДНК, секвенирование, первичная обработка полученных чтений, экспрессионный анализ. При сборе биологических образцов для получения более однородных результатов применялся основанный на морфологических маркерах метод синхронизации стадии развития растений. Для конструирования библиотек, секвенирования транскриптомов и первичной обработки чтений использовались стандартные для платформы Illumina методы. В арсенал методов транскриптомики входили анализ дифференциальной экспрессии и обогащения списков генов категориями ГО, кластерный анализ, расчет меры стабильности, вычисление паттерна экспрессии с помощью энтропии Шеннона и другие биоинформатические методы.

### 4. Научная новизна

Наша работа была посвящена созданию и анализу транскриптомной карты *Arabidopsis thaliana*, а также рассмотрению отдельных ее аспектов: временной серии апикальных меристем при переходе к цветению и органоспецифического ответа на холодостресс. Кроме того, высокая детализация атласа экспрессии *A. thaliana* была использована для планирования транскриптомной карты *Capsella bursa-pastoris*. В работе была создана транскриптомная карта *A. thaliana*, вошедшая в число самых крупных атласов экспрессии генов у растений. Ее анализ позволил подробно охарактеризовать транскриптом всего растения, определить паттерны экспрессии генов, в том числе ранее не изучавшихся транскрипционных факторов, и определить гены стабильно экспрессирующиеся в разных органах и тканях. Анализ входящей в ее состав временной серии апикальных меристем побега впервые показал согласованные изменения экспрессии генов, участвующих в прохождении клеточного цикла и вероятно связанные с изменением его длительности при переходе к цветению. Был продемонстрирован органоспецифический ответ на действие пониженной температуры.

## 5. Теоретическая и практическая значимость

Созданная на основе полученных данных база Transcriptome Variation Analysis (TraVA) начиная с 2016 года используется ежегодно более чем четырьмя тысячами исследователей по всему миру. Группой профессора Provar'ta (<https://csb.utoronto.ca/nicholas-provar't/>) была проведена интеграция данных в графическую оболочку efp browser. В этом виде они умолчанию приводятся в основной базе данных по *A. thaliana* TAIR ([arabidopsis.org](http://arabidopsis.org)) для демонстрации паттерна экспрессии генов.

Основанные на транскриптомной карте профили экспрессии генов из баз данных TraVA и TAIR используются в многочисленных исследованиях функций генов, проводимых другими научными группами; среди них можно назвать изучение ответа на дальний красный свет (Oh et al., 2020), регуляции состояния хроматина (Zhao et al., 2020), а также создание сетей генетической регуляции (Kulkarni et al., 2018). С использованием базы TraVA проводится учебный курс «Plant Bioinformatics» на платформе Coursera (<https://ru.coursera.org/learn/plant-bioinformatics>).

С помощью транскриптомной карты *C. bursa-pastoris* было подтверждено отсутствие геномного доминирования на уровне экспрессии и показано, что различия в уровнях экспрессии гомеологичных генов связаны с возникновением новых сайтов связывания транскрипционных факторов в промоторе одного из гомеологов. Данные транскриптомной карты *C. bursa-pastoris* и результаты анализа экспрессии в различных органах под воздействием низкой температуры также были интегрированы в базу данных TraVA и используются как публичный ресурс.

Полученные результаты имеют также практическую значимость. Найденные с помощью транскриптомной карты *A. thaliana* стабильно экспрессирующиеся гены могут быть использованы для корректного проведения широкого круга экспрессионных анализов с помощью метода ОТ-ПЦР в реальном времени.

## 6. Положения, выносимые на защиту

1. С использованием методов высокопроизводительного секвенирования созданы транскриптомные карты модельного объекта биологии растений *Arabidopsis thaliana* и близкого к нему аллотетраплоида *Capsella bursa-pastoris*, позволяющие упростить функциональный анализ генов.
2. Момент перехода к цветению в апикальной меристеме побега *Arabidopsis thaliana* характеризуется изменением динамики клеточных делений, вероятно связанным с сокращением длительности клеточного цикла за счет фаз G<sub>1</sub> и G<sub>2</sub>.
3. Развитие холодостресса у *Arabidopsis thaliana* сочетает универсальные для всего растения процессы с тканеспецифичными.

4. Различия паттернов экспрессии гомеологичных генов могут возникать из-за возникновения сайтов посадки транскрипционных факторов у одного из гомеологов.

### **7. Степень достоверности и апробация результатов**

По материалам диссертации опубликовано четыре статьи в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях и семинарах:

1. The 26<sup>th</sup> International Conference on Arabidopsis Research (ICAR) 2015, Париж, Франция;
2. The 29<sup>th</sup> International Conference on Arabidopsis Research (ICAR) 2018, Турку, Финляндия;
3. Life of Genomes 2018, Казань, Россия;
4. The 44<sup>th</sup> the Federation of European Biochemical Societies Congress (FEBS) 2018, Прага, Чехия;
5. The 11<sup>th</sup> International Conference on Bioinformatics of Genome Regulation and Structure\ Systems Biology 2018, Новосибирск Россия;
6. The International Conference on Polyploidy 2019, Гент, Бельгия;
7. VII съезд Вавиловского общества генетиков и селекционеров (ВОГиС), 2019, Санкт-Петербург, Россия.

### **8. Структура и объем работы**

Диссертация изложена на 179 страницах машинописного текста и содержит следующие разделы: введение; список сокращений; главы 1-4, каждая из которых содержит введение, обзор литературы, результаты, обсуждение и заключение; выводы, список литературы и приложения. Материал включает 24 рисунка, одну таблицу и список литературы, содержащий 407 ссылок.

## 9. Список публикаций по теме диссертации

По теме диссертации опубликовано четыре статьи в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (PubMed, WoS и Scopus):

1. **Klepikova AV**, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 2016 Dec;88(6):1058-1070. doi: 10.1111/tpj.13312.

2. **Klepikova AV**, Penin AA. Gene Expression Maps in Plants: Current State and Prospects. *Plants* 2019 Aug;8(9):309. doi: 10.3390/plants8090309.

3. **Klepikova AV**, Logacheva MD, Dmitriev SE, Penin AA. RNA-seq analysis of an apical meristem time series reveals a critical point in *Arabidopsis thaliana* flower initiation. *BMC Genomics* 2015 Jun 18;16:466. doi: 10.1186/s12864-015-1688-9.

4. **Klepikova AV**, Kulakovskiy IV, Kasianov AS, Logacheva MD, Penin AA. An update to database TraVA: organ-specific cold stress response in *Arabidopsis thaliana*. *BMC Plant Biol.* 2019 Feb 15;19(Suppl 1):49. doi: 10.1186/s12870-019-1636-y.

5. Kasianov AS, **Klepikova AV**, Kulakovskiy IV, Gerasimov ES, Fedotova AV, Besedina EG, Kondrashov AS, Logacheva MD, Penin AA. High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. *Plant J.* 2017 Jul;91(2):278-291. doi: 10.1111/tpj.13563.

**СПИСОК СОКРАЩЕНИЙ**

ГО – Генная Онтология (Gene Ontology)

ДЭ – дифференциально экспрессирующийся

ДЭАВ – дифференциально экспрессирующийся в паре гомеологичных генов из субгеномов А и В

ОТ-ПЦР – полимеразная цепная реакция с обратной транскрипцией

п.н. – пар нуклеотидов

полиА – полиаденилированная

ТФ – транскрипционный фактор

COR – гены, отвечающие на холод (cold-response)

CV – коэффициент вариации (coefficient of variation)

ФРKM – число фрагментов на тысячу п.н. гена на миллион картированных чтений (Fragments Per Kilobase per Million mapped reads)

РРKM – число чтений на тысячу п.н. гена на миллион картированных чтений (Reads Per Kilobase per Million mapped reads)

# ГЛАВА 1. СОЗДАНИЕ И АНАЛИЗ ТРАНСКРИПТОМНОЙ КАРТЫ *ARABIDOPSIS THALIANA*

## 1.1. Введение

Одной из центральных задач биологии является изучение функций генов и их взаимодействия в процессе развития и жизнедеятельности организма с помощью развитого инструментария генетических, физиологических, биохимических и других методов. В подавляющем большинстве случаев эксперименты, направленные на определение функций генов, структуры и динамики генетических сетей, проводятся на модельных объектах, таких как *Arabidopsis thaliana* (L.) Heynh. или *Drosophila melanogaster* Meigen. Затем полученные данные экстраполируются на другие виды. Однако вплоть до нынешнего времени лишь для небольшой доли генов были проведены эксперименты, позволяющие непосредственно изучить их функции. Подобные исследования обычно низкопроизводительны, занимают много времени и требуют значительных усилий, поэтому прирост числа исследованных генов происходит медленно. Это делает необходимым поиск массовых подходов, позволяющих хотя бы косвенно судить о функции генов. Таким подходом может быть изучение профилей экспрессии генов (Assis and Bachtrog, 2013).

Экспрессией гена называется количество его мРНК в образце; из-за ограниченных возможностей прямого измерения, количество мРНК обычно считается пропорциональным некоторой измеримой величине (уровню флуоресценции в случае микрочипов или числу чтений при секвенировании РНК). Профилем, или паттерном, экспрессии гена называется соотношение его уровней экспрессии в органах и тканях организма; для полноты представления о нем требуется исследование транскриптомов как можно большего числа различных клеток, тканей или органов. Такой подход называется транскриптомной картой, транскриптомным атласом или атласом экспрессии генов. В работе Schmid et al. транскриптомная карта определена как оценка глобальной экспрессии генов во всех органах и тканях в ходе всего онтогенеза организма от эмбриогенеза до старения (Schmid et al., 2005). Соответственно, полномасштабный транскриптомный атлас требует существенных материальных затрат для сбора и обработки образцов. Тем не менее, в последние два десятилетия появились технологии получения экспрессионных данных в масштабе всего генома, что облегчает создание транскриптомных карт разной степени детализации. С их помощью можно изучать функционирование транскриптома в целом, определять функции отдельных генов и выявлять тканеспецифичные группы генов, детерминирующих специализацию развития клеток (Su et al., 2004; Chikina et al., 2009; Wang et al., 2010).

Как было сказано выше, в настоящее время основным способом определения функции генов немодельных объектов является экстраполяция данных с модельных организмов на основании сходства нуклеотидных последовательностей. Однако хорошо известны случаи, когда продукты генов с высоким сходством нуклеотидных последовательностей обладают противоположными биологическими функциями. Например, гены *A. thaliana* *TERMINAL FLOWER* (*TFL*) и *FLOWERING LOCUS T* (*FT*) принадлежат к одному семейству и обладают 60%-ым сходством нуклеотидных последовательностей, при этом их роли в инициации цветения противоположны. Более того, замена одной аминокислоты в белке TFL, который в норме препятствует переходу к цветению, изменяет его функцию на аналогичную продукту гена *FT*, являющегося одним из основных активаторов цветения (Hanzawa et al., 2005). Соответственно, перенос функции исключительно по сходству нуклеотидной последовательности может приводить к ошибочным заключениям. Сравнение профилей экспрессии ортологичных генов у модельных и немодельных объектов в таких случаях может уточнить возможность переноса функции. Так, в приведенном выше примере профили экспрессии генов *TFL* и *FT* различаются и при этом эволюционно консервативны, то есть могут служить одним из косвенных свидетельств функции (Hanzawa et al., 2005).

Таким образом, для правильного функционального переноса на немодельные виды необходима подробная транскриптомная карта модельного объекта, каким в биологии растений является *A. thaliana*. В обзоре литературы Главы 1 мы обсудим современное состояние и перспективы развития области транскриптомики, посвященной созданию атласов экспрессии генов; в разделе 1.3 приведены результаты, полученные нами при создании и анализе транскриптомной карты *A. thaliana* с помощью секвенирования РНК; основные выводы из полученных результатов приведены в разделе 1.4.

## 1.2. Обзор литературы

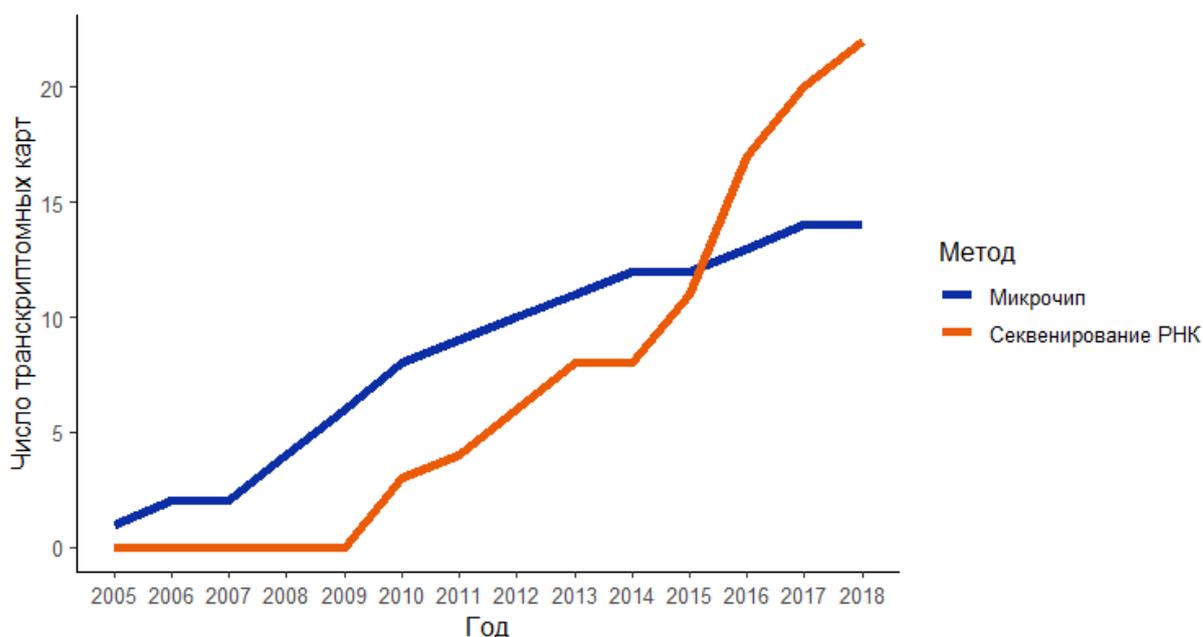
### 1.2.1. История создания транскриптомных карт растений

Появление атласов экспрессии генов стало возможным только после существенного развития транскриптомики. Начиная с проведенного в 1995 году изучения экспрессии генов *A. thaliana* с помощью ДНК-микрочипа (Schena et al., 1995), массовый анализ транскрипции, основанный на гибридизации меченной флуорофорами кДНК с нанесенными на микрочип комплементарными последовательностями ДНК, нашел широкое применение в разных областях растительной биологии (Rossel, 2002; Becker, 2003; Wellmer et al., 2004). В первой половине 2000-ых годов уже существовали работы со значительным для того времени числом образцов, среди которых следует упомянуть изучение транскриптомов 15 зон корня *A. thaliana* (Birnbaum, 2003), анализ профилей транскрипции генов, изменяющих экспрессию в ходе клеточного цикла, (Menges et al., 2002) и исследование циркадной регуляции экспрессии генов (Harmer et al., 2000). Однако в этих работах рассматривались отдельные органы и ткани или определенные условия выращивания растений (такие, как повышенная интенсивность света, (Rossel, 2002). Кроме того, в большинстве из них была представлена только часть (до трети) генов из-за технических ограничений производства микрочипов.

Первая детализированная транскриптомная карта растения была создана в 2005-ом году для *A. thaliana* (Schmid et al., 2005). Арабидопсис – классический объект биологии растений и первое растение с секвенированным геномом (Meinke et al., 1998; Arabidopsis Genome Initiative, 2000), благодаря чему стало возможным создание экспрессионного микрочипа Affymetrix ATH1, содержащего пробы к приблизительно 22 тысячам генов. Это составляло более 80% известных к тому моменту генов и около 70% генов из аннотации TAIR10 (Schmid et al., 2005; Lamesch et al., 2012). Транскриптомная карта Schmid et al. стала знаковой работой в области растительной транскриптомики и во многом определила ключевые вопросы, изучавшиеся в последующих атласах.

В течение последних полутора десятилетий прогресс в технологиях секвенирования и сборки геномов приводит к неуклонному росту числа полных последовательностей растительных геномов. Среди видов, чей геном был секвенирован, есть как модельные объекты, такие как коротконожка двуколосковая (*Brachypodium distachyon* (L.) P.Beauv.) (International Brachypodium Initiative, 2010) и пастушья сумка (*Capsella bursa-pastoris* (L.) Medik.) (Kasianov et al., 2017), так и хозяйственно-ценные растения, в том числе кукуруза (Schnable et al., 2009), рис (Goff et al., 2002; Yu et al., 2002), люцерна (Young et al., 2011) и соя (Schmutz et al., 2010), а также многие другие растения. Благодаря развитию геномики, новые детализированные транскриптомные карты создаются ежегодно (Рисунок 1.1). В обзоре литературы нами будут

рассмотрены тридцать семь атласов экспрессии генов, список которых приведен в Приложении, Таблица 1.1.



**Рисунок 1.1.** Рост числа опубликованных транскриптомных карт растений.

### 1.2.2. Технические характеристики транскриптомных карт

Традиционно для создания транскриптомных атласов применяют два метода широкомасштабного анализа экспрессии: микрочипы и секвенирование РНК. За исключением работы Nobuta et al., 2007, где с помощью технологии масштабного параллельного секвенирования сигнатур (MPSS, (Brenner et al., 2000) было изучено 18 образцов риса (12 различных тканей растения и 6 образцов под воздействием абиотических стрессов) (Nobuta et al., 2007), все рассматриваемые нами транскриптомные карты были созданы с помощью этих двух технологий.

Из рассмотренных нами тридцати семи атласов экспрессии генов 14 было создано с помощью технологии микрочипов. Длительная история разработки и применения микрочипов делает их надежной технологией с подробными протоколами использования и последующей обработки данных (Brazma et al., 2001). Однако этот метод имеет ряд существенных ограничений. К ним относятся фиксированность набора генов, пробы к которым присутствуют на чипе, и вытекающая из этого невозможность обнаружения новых альтернативных изоформ РНК и неаннотированных ранее генов, а также неточная оценка экспрессии генов с близкими последовательностями (Sekhon et al., 2013). Особенно существенно эти ограничения затрудняют анализ экспрессии генов полиплоидных растений, к которым относятся многие сельскохозяйственные культуры. Это может приводить к некорректной оценке экспрессионного

профиля образца, особенно если его транскриптом ранее не анализировался, и таким образом влиять на биологические выводы. Следует, однако, отметить, что степень детализации микрочипа напрямую зависит от качества сборки и аннотации генома, на основе которого он создан, поэтому прогресс геномных технологий позволяет создавать все более полные микрочипы.

Параметром, отражающим глубину анализа с помощью микрочипов, является число генов, пробы к которым присутствуют на микрочипе. В рассматриваемых нами транскриптомных картах это число распределялось от 21 439 (ячмень, (Druka et al., 2006) до 61 115 (пшеница, (Schreiber et al., 2009), и, по оценке авторов, составляло от 72% аннотированных на тот момент генов (кукуруза (Sekhon et al., 2011) до 100% (пшеница (Schreiber et al., 2009) и сорго (Shakoor et al., 2014). В подавляющем большинстве работ использовались ранее сконструированные микрочипы, однако в некоторых случаях чипы создавались самими исследователями. К таким работам относятся транскриптомные карты табака (Edwards et al., 2010), кукурузы (Sekhon et al., 2011), сорго (Shakoor et al., 2014) и мха *Physcomitrella patens* (Hedw.) Bruch & Schimp. (Ortiz-Ramírez et al., 2016).

Атласы экспрессии генов, созданные с помощью микрочипов, продолжают появляться, однако рост их числа ощутимо замедляется, и на смену им приходят транскриптомные карты, основанные на секвенировании РНК (Рисунок 1.1). Среди рассмотренных нами атласов 22 было получено с помощью этого метода. Согласно нашим сведениям, самыми ранними среди них были карта риса (Zhang et al., 2010), содержащая 8 образцов, и две карты сои (Libault et al., 2010b; Severin et al., 2010) – 9 и 14 образцов соответственно.

Сравнению точности и производительности микрочипов и секвенирования РНК посвящено значительное количество работ (Zhao et al., 2014; Rai et al., 2018); в отношении транскриптомных карт такое сравнение было проведено в работе Sekhon et al. для кукурузы (Sekhon et al., 2013). Использование одной и той же РНК для анализа экспрессии с помощью микрочипов (Sekhon et al., 2011) и секвенирования (Sekhon et al., 2013) позволило избежать биологической вариабельности и сравнить технические характеристики методов. Экспрессионные профили образцов, полученные двумя технологиями, были схожи между собой (коэффициент корреляции Пирсона для разных образцов лежал в диапазоне от 0,70 до 0,83). В то же время секвенирование РНК позволяло обнаружить транскрипты существенно большего числа генов (82,1% от всех аннотированных генов против 56,5% у чипов), включая тканеспецифические, а также лучше различало экспрессию паралогичных генов.

Техническое качество транскриптомных карт, полученных с помощью секвенирования, можно оценить по таким параметрам как технология секвенирования, длина чтений и глубина секвенирования.

Наиболее часто для создания атласов экспрессии генов применялась платформа Illumina (Illumina, США) – от секвенатора GAII до HiSeq2500, в зависимости от года публикации. Помимо Illumina, использовались платформы SOLiD (версии 3.0 и 4.0, Applied Biosystems, США) (Meyer et al., 2012; Ibarra-Laclette et al., 2013) и GS-FLX+ (Roche Applied Sciences, США) (Cañas et al., 2017).

Длина чтения является важным параметром, влияющим на успешность картирования и, таким образом, на объем получаемых данных. Большая длина чтения обеспечивает лучшее картирование, что не столь важно для видов с хорошо собранными и аннотированными простыми геномами, однако необходимо для полиплоидов и других вариантов сложных геномов. Длины чтений, полученные в разных работах, напрямую зависят от года публикации статьи (что отражает развитие технологий секвенирования): так, если в ранних работах длины чтений в основном не превышали 36 п.н. (пар нуклеотидов) (Libault et al., 2010b; Severin et al., 2010), то в дальнейшем в основном использовались чтения длиной 100 п.н. (Vlasova et al., 2016; Walley et al., 2016), а максимальная длина чтения для платформы Illumina составила 150 п.н. (McCormick et al., 2018). Следует упомянуть так же о структуре чтений: парные чтения увеличивают разрешающую способность картирования, поэтому в большинстве работ использовали именно их, однако для объектов, обладающих хорошей сборкой генома (как, например, кукуруза) с успехом были использованы одиночные чтения (Walley et al., 2016). Такой дизайн эксперимента снижает стоимость секвенирования по сравнению с парными чтениями, что позволяет проанализировать большее число образцов и/или увеличить глубину секвенирования.

Увеличение глубины секвенирования позволяет проводить анализ с более высоким разрешением и, соответственно, делать более надежные выводы об экспрессии. Глубина секвенирования оценивалась нами как суммарное по всему атласу число картированных чтений (для тех работ, в которых было приведено это число). Минимальное суммарное число чтений было представлено в атласе экспрессии генов розы (Dubois et al., 2012) – 9,3 млн, максимальное – кукурузы, 4,5 млрд (Stelpflug et al., 2016). Следует отметить, что более значимая характеристика полноты описания транскриптома – число чтений, приходящихся на каждый образец в отдельности, так как в пределах одной транскриптомной карты эти значения могут существенно различаться. Необходимый размер отдельной библиотеки зависит от целей исследования: для изучения транскриптома отдельной клетки требуется несколько миллионов чтений, в то время как для анализа событий сплайсинга в отдельном органе необходимо около 100 млн чтений (Conesa et al., 2016).

Помимо глубины секвенирования, на результаты анализа транскриптомной карты большое влияние имеют наличие референсного генома и его качество. В подавляющем большинстве случаев изучение экспрессии в отсутствие референсного генома не позволяет

сделать надежные выводы об исследуемых процессах. Это связано с крайней сложностью сборки транскриптома: число контигов или скаффолдов зачастую в разы превышает число генов организма, что делает невозможным корректный анализ экспрессии даже в случае такого простого с точки зрения организации генома и числа генов объекта как *A. thaliana* (Ono et al., 2015). Тем не менее, картирование чтений на транскриптом в настоящее время широко применяется для немодельных объектов с большими геномами. Среди рассматриваемых нами транскриптомных карт в девяти проводились секвенирование и сборка референсного транскриптома; в их число входят роза, авокадо, арахис (Dubois et al., 2012; Ibarra-Laclette et al., 2013; Clevenger et al., 2016).

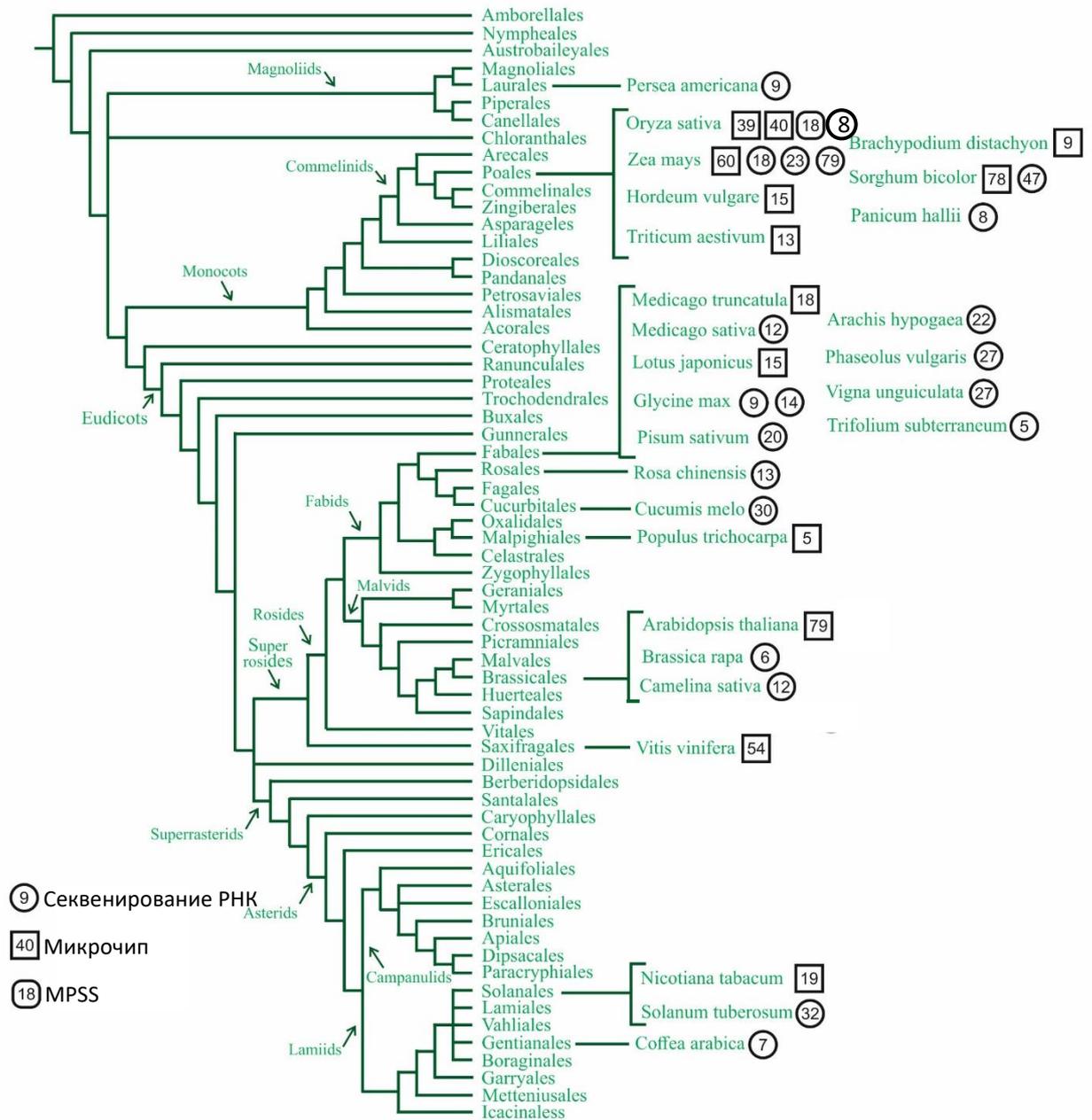
Стремительное развитие геномных технологий и использование в сборке геномов таких методов как синтетические длинные чтения или определение конформации хромосом (Hi-C) в будущем позволят получать полные последовательности геномов большого размера и сложности. В этом случае повторный анализ атласов экспрессии генов, в настоящий момент использующих транскриптомы, предоставит возможность проверить и уточнить сделанные выводы.

### 1.2.3. Представленность таксонов среди видов, для которых созданы транскриптомные карты

Покрытосеменные (Magnoliophyta) составляют большую часть современных растений: к ним относится не менее 350 тысяч видов (The Plant List, 2013), а также абсолютное большинство сельскохозяйственных растений. Эти факты определяют систематическое положение видов, для которых созданы транскриптомные карты; разнообразие изученных растений было оценено нами с помощью последнего варианта классификации APG IV (The Angiosperm Phylogeny Group, 2016) (Рисунок 1.2).

Только три атласа экспрессии из тридцати семи созданы для нецветковых растений: мха *P. patens* (Ortiz-Ramírez et al., 2016; Perroud et al., 2018) и сосны приморской (*Pinus pinaster* Aiton) (Cañas et al., 2017). Базальные Цветковые представлены только авокадо (*Persea americana* Mill.) из семейства Lauraceae надпорядка Magnoliidae (Ibarra-Laclette et al., 2013).

Больше трети транскриптомных атласов созданы для Однодольных, при этом единственное представленное семейство – Poaceae, что ожидаемо, так как злаки традиционно составляют основную часть сельскохозяйственной продукции: в 2017 году Продовольственная и сельскохозяйственная организация ООН оценила объем собранного урожая злаков в 2 642 миллионов тонн (FAO, 2018). Среди изученных растений присутствовали такие важные культуры как рис (Nobuta et al., 2007; Wang et al., 2010), кукуруза (Sekhon et al., 2011; Stelplflug et al., 2016), сорго (McCormick et al., 2018) и ячмень (Druka et al., 2006).



**Рисунок 1.2.** Таксономия видов, для которых созданы транскриптомные карты. Филогенетическое дерево Покрытосеменных в системе APGIV адаптировано из (The Angiosperm Phylogeny Group, 2016).

Более представлены двудольные растения: транскриптомные карты были созданы для 22 растений. Из них к астеридам относится только *Coffea arabica* L. из Rubiaceae (Ivamoto et al., 2017), а также *Nicotiana tabacum* L. (Edwards et al., 2010) и картофель *Solanum tuberosum* L. (Massa et al., 2011) из семейства Solanaceae. Кроме того, база данных eFP browser (Winter et al., 2007) содержит экспрессионные профили томата и картофеля (Solanaceae), однако эти транскриптомные данные использовались в основном в целях аннотации геномов и не анализировались отдельно, поэтому мы не рассматривали их в нашем обзоре (Xu et al., 2011; Sato et al., 2012).

Несколько лучше изучены розиды, хотя представленность семейств также невелика: согласно нашим сведениям, существуют транскриптомные атласы видов из семейств Brassicaceae, Cucurbitaceae, Fabaceae, Rosaceae и Salicaceae. В семейство Fabaceae (десять транскриптомных карт) входит большое количество хозяйственно-ценных растений, таких, как арахис *Arachis hypogaea* L. (Clevenger et al., 2016), люцерна (O'Rourke et al., 2015) и соя (Libault et al., 2010b; Severin et al., 2010), а к Brassicaceae (три атласа) относятся классический модельный объект *A. thaliana* (Schmid et al., 2005) и сельскохозяйственные растения рода *Brassica*.

Для ряда растений было создано несколько транскриптомных карт – из-за появления новых методов (Shakoog et al., 2014; McCormick et al., 2018) или в случае разных направлений исследования объекта (Jiao et al., 2009; Wang et al., 2010).

Таким образом, в настоящее время только отдельные таксоны изучены с точки зрения глобального анализа экспрессии. Известно, что недостаточная и смещенная выборка анализируемых видов может привести к ошибочным выводам, например, неправильный набор таксонов при анализе эволюции Покрытосеменных (Goremykin, 2003) привел к парадоксальному выводу, что амборелла (*Amborella trichopoda* Baill.) не является базальным цветковым. В дальнейшем использование сбалансированного набора таксонов показало ошибочность этой гипотезы (Stefanović et al., 2004). Поэтому дальнейшее изучение транскриптомики видов, находящихся в разных частях эволюционного древа, необходимо для получения более полного и точного представления о структуре и вариативности растительного транскриптома.

#### 1.2.4. Степень детализации транскриптомных карт

Рассмотренные нами атласы экспрессии генов существенно различаются по своей детализации, то есть по числу образцов: наименее детализированная транскриптомная карта содержала пять образцов (клевер подземный, *Trifolium subterraneum* L. (Kaur et al., 2017), а две наиболее подробные – 79 (*A. thaliana* и *Z. mays*) (Schmid et al., 2005; Stelpflug et al., 2016) (Рисунок 1.2). Изучаемые образцы отражают общую анатомию растения, и их базовый набор практически одинаков для всех карт. В их число, как правило, входят: созревающие и зрелые семена и/или плоды и/или эмбрионы, прорастающее семя, проросток (или его части), стебель, лист (чаще более, чем на одной стадии развития, и в виде отдельных частей), цветок (как правило, отдельные его органы) и корень.

Для хозяйственно ценных однодольных детализация экспрессии склоняется в сторону важных для продуктивности образцов, то есть соцветий и органов цветка (пестик, тычинка, цветочные чешуи) (Druka et al., 2006; Zhang et al., 2010) и семени, в особенности эндосперма (Wang et al., 2010; Sekhon et al., 2011, 2013; Stelpflug et al., 2016).

Многие бобовые растения являются важными сельскохозяйственными культурами и используются для обогащения почвы азотом, что необходимо для эффективного земледелия. Поэтому в транскриптомной карте сои (*Glycine max* (L.) Merr.) в больших подробностях изучены разные части корня и клубеньки, в которых обитают симбиотические азотфиксирующие бактерии (Libault et al., 2010b), а для люцерны усеченной (*Medicago truncatula* Gaertn.), модельного объекта исследований симбиоза растения и азотфиксирующих бактерий, создана серия развития клубеньков (Benedito et al., 2008). Кроме того, так как плоды сои широко используются в пищевой промышленности, второй атлас экспрессии генов *G. max* содержит семь стадий развития цветка (Severin et al., 2010).

Традиционная масличная культура рыжик посевной (*Camelina sativa* (L.) Crantz) обладает рядом экономических преимуществ, таких, как короткий вегетационный период и способность расти в разнообразных условиях, и может быть использована для получения биотоплива, что объясняет возвращение интереса к ней. Развитие семян и цветка – основных для продуктивности органов *C. sativa* – подробно рассмотрено в транскриптомном атласе (Kagale et al., 2016).

Особняком стоит работа Jiao et al., представляющая собой клеточный атлас риса – исследование экспрессии генов в 40 типах клеток, выделенных с помощью лазерной микродиссекции из трех зон корней, примордиев листьев на разных стадиях развития и прорастающего семени трех возрастов (Jiao et al., 2009).

В ряде работ, помимо основного набора данных, представляющих развитие растения в нормальных условиях, используют образцы, подвергавшиеся стрессовому воздействию, биотическому или абиотическому (Nobuta et al., 2007; Massa et al., 2011; Dubois et al., 2012). В транскриптомной карте розы отдельное внимание уделено образцам, подвергшимся заражению грибным патогеном *Botrytis cinerea*, который снижает качество собираемых цветов, уменьшая продуктивность (Dubois et al., 2012). Для сорго (злака, культивируемого на большом разнообразии почв) была проанализирована экспрессия в образцах, полученных при выращивании растений на различающихся по содержанию азота почвах (McCormick et al., 2018). В транскриптомную карту риса, помимо стандартных, входило восемь образцов каллусов в процессе генетической трансформации (Wang et al., 2010).

В атласе экспрессии генов *A. thaliana* подчеркивается важность сбора отдельных частей одного органа: по сравнению с профилем экспрессии целого цветка, число экспрессирующихся генов хотя бы в одном из четырех органов цветка (чашелистике, лепестке, тычинке или плодолистике) было выше на 11% (Schmid et al., 2005). То же наблюдение было сделано в клеточном атласе риса, где в многочисленных случаях экспрессия гена, наблюдаемая в отдельном типе клеток, не выявлялась при анализе органа, в состав которого он входит (Jiao et al., 2009).

Этот факт необходимо принимать во внимание при создании новых транскриптомных карт, и включать в них как можно более детально разобранные части органов. Кроме того, для получения данных, сравнимых с существующими атласами экспрессии, стандартный набор образцов, описанный выше, должен быть основой транскриптомной карты.

#### *1.2.5. Кластеризация органов и тканей растений по их транскрипционному профилю*

Общая оценка сходства и различий изучаемых образцов – один из первых вопросов, ставящихся исследователями в посвященных транскриптомным картам работах. Как правило, этот вопрос решается с помощью кластерного анализа тем или иным математическим методом. Основой для кластеризации служат транскрипционные профили образцов: как в виде исходных данных (уровней флуоресценции для микрочипов и числа чтений, приходящихся на ген, для секвенирования кДНК), так и различным образом процессированные (Pirim et al., 2012). Кластеризация образцов, с одной стороны, применяется для проверки правильности сбора образцов (биологические повторности одного образца должны находиться на близком расстоянии), а с другой – является источником ценной информации о схожести различных органов и тканей растения на уровне транскриптома.

Несмотря на различия, обусловленные особенностями строения конкретных растений, кластеризации образцов разных транскриптомных карт довольно схожи и воспроизводят морфологическое и физиологическое сходство органов. В основном образцы кластеризуются по функциональным особенностям, то есть в группы объединяются скорее разные стадии развития одного органа, нежели разные органы или ткани, собранные с растения определенного возраста. Так, образцы объединяются по наличию хлорофилла (что было главным дискриминирующим фактором для ячменя (Druka et al., 2006), по содержанию характерной ткани (хорошо выделяются образцы, содержащие меристематическую ткань, в атласе *Z. mays* (Stelpflug et al., 2016), а также, хотя в целом кластеризация по возрасту не является частой, образцы различных стареющих органов группируются вместе, что может объясняться преобладанием процессов высыхания и распада тканей над определяющими тканеспецифичность (Schmid et al., 2005).

На примере карт с высокой степенью детализации можно увидеть, что отдельные клады образуют пыльники, различные образцы листьев и зеленые междоузлия, все типы корней, семена, цветки, органы цветков, меристемы (Wang et al., 2010; Shakoор et al., 2014; Stelpflug et al., 2016; McCormick et al., 2018). Это подтверждается и картами меньшего разрешения, где, как правило, группировались корни, цветки, семена и зеленые части растения (Benedito et al., 2008; Libault et al., 2010b; Severin et al., 2010; Alves-Carvalho et al., 2015).

Анализируя кластеризации образцов различных транскриптомных карт, можно отметить ряд объединяющих их особенностей. Во-первых, у растений, принадлежащих к разным таксонам,

наиболее отличным от остальных органом является пыльник (или, в зависимости от собранного образца, тычинка или пыльца). Это было отмечено в атласах экспрессии риса и кукурузы (Nobuta et al., 2007; Wang et al., 2010; Stelpflug et al., 2016). Кроме того, анализ транскриптомных карт, содержащих подробные серии развития семян, показывает, что, хотя семена кластеризуются вместе, между ними существуют значительные расстояния на дереве кластеризации. Это говорит о больших изменениях транскриптома, происходящих в онтогенезе семян, и необходимости точно выбирать стадию развития семени при изучении специфических процессов, что особенно важно для сельскохозяйственных культур (Sekhon et al., 2011). Как уже было сказано в разделе 1.2.4, профили экспрессии отдельных тканей или органов могут значительно отличаться от транскриптома целой гетерогенной структуры, что отражается в их отдельной кластеризации (Schmid et al., 2005).

Таким образом, изучение кластеризации образцов транскриптомной карты дает представление о разнообразии входящих в нее органов и тканей растения и чрезвычайно полезно для выбора образцов при создании нового атласа экспрессии, сравнимого с уже существующими.

#### *1.2.6. Общее описание транскриптома растений*

Самой простой характеристикой экспрессионного профиля как отдельных образцов, так и транскриптомной карты в целом является число экспрессирующихся генов – в каждом образце, во всех образцах или хотя бы в одном образце. Несмотря на внешнюю простоту этой задачи, ее решение довольно нетривиально.

Во-первых, выбор порога, отделяющего экспрессирующиеся гены от неэкспрессирующихся, в настоящее время не имеет строгой формализации и произвольно определяется исследователями. К тому же, в случае секвенирования РНК наблюдаемый уровень экспрессии гена зависит от числа полученных при секвенировании чтений: в случае недостаточной глубины секвенирования транскрипты низкоэкспрессирующихся генов не будут обнаружены.

Во-вторых, при оценке общего числа экспрессирующихся генов большое значение имеет качество аннотации генома. Как правило, для определения положения генов в геноме используется картирование секвенированной РНК, поэтому на аннотацию влияет выбор образцов, из которых выделялась РНК. Целое растение плохо подходит для этой цели (см. раздел 1.2.4), и обычно используется смесь РНК из простых с точки зрения сбора материала образцов. Соответственно, при анализе экспрессии в других образцах для части тканеспецифических генов могут отсутствовать аннотация в геноме и комплементарные зонды на микрочипе. В таком случае число экспрессируемых генов в образце будет недооценено, что может существенно повлиять на биологические выводы. С другой стороны, аннотация низкого качества,

фрагментированная сборка генома или анализ на основе сборки транскриптома могут привести к переоценке числа транскрибирующихся генов. Известно, что оценка числа генов конкретного вида зависит от качества сборки генома и методов аннотации и может значительно меняться со временем (Sterck et al., 2007), причем эта проблема не решена до конца даже для такого хорошо изученного генома как геном *Homo sapiens* (Willyard, 2018).

Учитывая вышесказанное, необходимо принимать во внимание метод, которым создана транскриптомная карта, а также качество генома вида, для которого она сделана. При дальнейшем описании доли экспрессируемых генов мы опирались на приведенные в статьях данные; следует учитывать, что они получены с применением разных порогов по уровню экспрессии.

Число генов, экспрессирующихся хотя бы в одном образце транскриптомной карты достаточного разрешения, дает представление о доле генов генома, в принципе способных к транскрипции. Это значение оказалось довольно стабильным: за небольшим исключением, доля генов, экспрессирующихся хотя бы в одном образце, составляла 75-85% вне зависимости от таксона, к которому принадлежало растение. Отличающиеся случаи могут объясняться методом (MPSS, (Nobuta et al., 2007), анализом только подвыборки генов (максимально надежных моделей генов, (Sekhon et al., 2011) или собственной сборкой и аннотацией транскриптома (O'Rourke et al., 2015).

Для транскриптомной карты сои была проведена проверка полноты обнаружения экспрессирующихся генов: 51 529 (74,5%) генов экспрессировалось хотя бы в одном из 9 образцов, а после добавления 14 проб корней из другой работы авторов (Libault et al., 2010a), число экспрессируемых генов возросло до 52 947 (Libault et al., 2010b). Таким образом, правильный выбор образцов и достаточное их количество позволяет обнаружить экспрессию большинства потенциально способных к транскрипции генов.

В большинстве транскриптомных карт число экспрессирующихся генов в каждом образце также было схоже. Как правило, разница между образцами с максимальным и минимальным числом экспрессирующихся генов составляла 10-20%, а доля от общего числа генов в разных образцах – 50-70%. В редких случаях доля экспрессирующихся генов в образцах была выше: в органах кукурузы она составила от 73 до 81% (Sekhon et al., 2011), а в образцах проса Холла (*Panicum hallii* Vasey) – 77-90% (Meyer et al., 2012). Следует отметить, что в первом случае авторами был сконструирован свой микрочип, а во втором была сделана сборка транскриптома.

В целом для различных объектов прослеживается тенденция к экспрессии большинства генов в большом числе образцов: транскрипты от 50% до 73% генов были обнаружены в большинстве проб (Schmid et al., 2005; Kagale et al., 2016). Доли генов, экспрессирующихся во всех образцах, были довольно схожи для большинства карт, где было приведено это значение, и

составляли 42-50% вне зависимости от степени детализации атласов, которая варьировалась от 6 (Tong et al., 2013) до 60 образцов (Sekhon et al., 2011). В транскриптомной карте кукурузы было показано участие этих генов в процессах, связанных с трансляцией, модификацией и транспортом белков, а также передачей сигналов, то есть с базовыми биологическими процессами (Sekhon et al., 2011).

Анализ неэкспрессирующихся генов имеет определенные сложности, так как транскрипты могут не детектироваться из-за недостаточной глубины секвенирования или по причине отсутствия в выборке образца, специфического для данного гена. Нам удалось обнаружить два характерных значения для доли неэкспрессируемых генов: примерно 10% (8,6% для атласа кукурузы и 12% для *A. thaliana* и *C. sativa*) (Schmid et al., 2005; Sekhon et al., 2011; Kagale et al., 2016) и 22-23% (Benedito et al., 2008; Libault et al., 2010b). Эти гены зачастую были связаны с ответом на абиотические стимулы (что скорее всего вызвано отсутствием образцов, подвергнутых стрессовым воздействиям (Sekhon et al., 2011), а также кодировали транскрипционные факторы, в особенности содержащие домен F-box (Schmid et al., 2005; Sekhon et al., 2011).

Общий обзор экспрессируемости генов позволяет сделать вывод о сходстве глобальных характеристик транскриптомов растений из самых разных таксонов и лишь небольших их различиях, которые могут быть вызваны методическими особенностями экспериментов.

#### 1.2.7. Транскрипционные профили образцов

Как было сказано выше (раздел 1.2.6), в отношении числа экспрессирующихся генов разные образцы одной транскриптомной карты демонстрируют большую степень сходства. Однако профили экспрессии, то есть соотношение уровней экспрессии всех генов в образце, значительно различаются между органами и тканями (Schmid et al., 2005). В транскриптомном атласе *A. thaliana* была предложена удобная мера сравнения транскрипционных профилей образцов – параметр *Z* (Schmid et al., 2005); в дальнейшем она широко использовалась для самых разных объектов. Параметр *Z* описывает отклонение экспрессии в данном образце от «среднего» по карте образца. Следует отметить, что структура транскриптомной карты (состав и число проб) напрямую влияет на расчет профиля усредненного образца; это может привести к отличающимся профилям экспрессии одного и того же органа в разных атласах.

Так, в транскриптомных картах *A. thaliana*, *M. truncatula* и кукурузы листья наиболее похожи на усредненный образец, то есть имеют распределение параметра *Z* со средним в нуле и с небольшой дисперсией (Schmid et al., 2005; Benedito et al., 2008; Sekhon et al., 2011), однако в транскриптомной карте сои таким образом были семена (Severin et al., 2010), в других атласах отличающихся от среднего сильнее всего (Benedito et al., 2008). Для некоторых образцов

наблюдались бимодальные распределения параметра  $Z$ : такими образцами были цветки *A. thaliana* (Schmid et al., 2005), корни и клубеньки сои (Severin et al., 2010), семена *M. truncatula* (Benedito et al., 2008), что говорит об экспрессии в этих частях растения большого количества органоспецифичных генов. В большинстве тканей *P. hallii* параметр  $Z$  распределялся симметрично относительно нуля (Meyer et al., 2012), а ткани сорго, напротив, в основном характеризовались смещением среднего параметра  $Z$ , что отражает их широкое разнообразие (Shakoor et al., 2014).

### 1.2.8. Поиск стабильно экспрессирующихся генов

Определение генов, экспрессия которых находится на одном уровне в разных тканях, органах и условиях – стабильно экспрессирующихся генов (далее «стабильных генов»), крайне важно для решения задачи нормализации экспрессии генов между разными образцами. Стабильные гены находят широкое применение в качестве внутренних референсов при сравнении экспрессии методом ОТ-ПЦР (полимеразная цепная реакция с обратной транскрипцией) в реальном времени (Czechowski, 2005). Поиск стабильно экспрессирующихся генов был проведен для большого количества растительных объектов (Imai et al., 2014; Liu et al., 2014; Niu et al., 2014; Saha and Blumwald, 2014; Long et al., 2015) (для обзора см. (Kozera and Rapacz, 2013), и транскриптомные карты являются важной его частью, предоставляя информацию об экспрессии генов в большом количестве тканей и органов, а так же в разнообразных условиях выращивания.

В классической работе Czechowski et al. по определению референсных генов для *A. thaliana* было показано, что традиционно используемые для нормализации экспрессии так называемые «гены домашнего хозяйства» – *ACT2* (кодирующий актин), *TUB6* (тубулин), *EF-1a* (фактор элонгации 1), *UBQ10* (убиквитин) и *GAPDH* (глицеральдегид-3-Р-дегидрогеназа) – имеют очень большой разброс уровней экспрессии в данных микрочипов по различным органам, стадиям и условиям выращивания *A. thaliana* (Czechowski, 2005). Вместо нестабильных генов домашнего хозяйства в этой работе были отобраны и валидированы наиболее равномерно экспрессирующиеся гены.

В большинстве транскриптомных карт в качестве меры стабильности экспрессии гена использовался коэффициент вариации (coefficient of variation, CV) – стандартное отклонение, деленное на среднюю экспрессию по всем образцам. Необходимо учитывать, что чем больше образцов взято в анализ и чем выше их разнообразие, тем меньшим будет количество стабильных генов, кроме того, всегда остается возможность обнаружения образца, где экспрессия гена, стабильного в остальных пробах, будет значимо отличаться.

Из традиционных референсов только *GAPDH* и убиквитины стабильно экспрессировались в транскриптомных картах (Benedito et al., 2008; Wang et al., 2010; Meyer et al., 2012; Verdier et al., 2013); большинство классических «генов домашнего хозяйства» не входило в число самых стабильных генов и обладало большой дисперсией экспрессии вплоть до CV 0,96 (Wang et al., 2010; Sekhon et al., 2011).

В зависимости от применяемых порогов по значению CV, которые в основном составляли от 0,1 до 0,3, число стабильно экспрессирующихся генов различалось от нескольких десятков (Meyer et al., 2012; O'Rourke et al., 2015) до почти тысячи (Jiao et al., 2009; Vlasova et al., 2016). Во всех работах функции большей части этих генов не были известны (Benedito et al., 2008; Massa et al., 2011), а анализ аннотации Генной Онтологии (подробнее см. раздел 1.3.3) показал их вовлеченность в необходимые для жизнедеятельности клеток и тканей процессы. Так, гены, стабильные в разных типах клеток риса, чаще кодировали убиквитины и белки теплового шока, а так же были вовлечены в контроль глюконеогенеза и клеточного дыхания (Jiao et al., 2009); в образцах кукурузы гены ассоциировались с киназами, связыванием нуклеотидов и модификацией белков (Sekhon et al., 2011); в органах и тканях риса – с трансляцией, транспортом белков, убиквитином, клеточным циклом (Wang et al., 2010); а гены, стабильные в образцах *P. hallii*, кодировали 3' экзонуклеазу гистонов, гистон H3 и регуляторные киназы (Meyer et al., 2012).

Таким образом, изучение стабильно экспрессирующихся генов позволяет не только выбрать референсы для нормализации экспрессии, но и дает ценную информацию о самых базовых для жизнедеятельности организма генах, транскрипты которых необходимы во всех клетках, тканях и органах.

### 1.2.9. Анализ тканеспецифичной экспрессии генов

Рассмотрение стабильно экспрессирующихся генов позволяет дать ответ на вопрос об универсальных для всего организма процессах; обратной задачей является поиск генов, экспрессирующихся в определенных тканях и органах. Анализ ткане- и органоспецифических генов помогает выяснить, какие молекулярные механизмы лежат в основе формирования уникальных морфологических и физиологических свойств определенных тканей и органов в ходе онтогенеза.

Для определения тканеспецифичных генов используются два основных подхода. Первый – это анализ дифференциальной экспрессии между всеми возможными парами образцов с последующим выявлением генов, экспрессирующихся преимущественно в одном образце. Такой подход, помимо идентификации тканеспецифичных генов, позволяет изучать общее разнообразие транскриптома различных частей растения. Этот метод был применен для

большого числа объектов (Quesada et al., 2008; Meyer et al., 2012; Alves-Carvalho et al., 2015; O'Rourke et al., 2015; Ortiz-Ramírez et al., 2016; Vlasova et al., 2016). Количество дифференциально экспрессирующихся (ДЭ) генов значительно различалось как внутри, так и между транскриптомными картами: минимальное число ДЭ генов варьировало от 83 (междоузлия и клубеньки, (Quesada et al., 2008) до 2 653 (последовательные стадии развития початка кукурузы, (Davidson et al., 2011), а максимальное – от 4 989 (плоды кофе на разных стадиях созревания, (Ivamoto et al., 2017) до 27 945 (между семенем и цветком сои, (Severin et al., 2010). Как и общий анализ профилей экспрессии с помощью параметра Z (см. раздел 1.2.7), изучение разницы в уровнях экспрессии генов показывает значительные отличия транскриптомов тканей и органов. Уровни изменения экспрессии ДЭ генов могут быть использованы для кластеризации и выявления тканеспецифичных генов, что было сделано для *P. hallii* (Meyer et al., 2012), тополя волосистоплодного (*Populus trichocarpa* Torr. & A. Gray, (Quesada et al., 2008), гороха (Alves-Carvalho et al., 2015), *C. sativa* (Kagale et al., 2016).

Вторым подходом к определению тканеспецифичных генов является использование численных мер ширины паттерна экспрессии гена. Паттерном экспрессии мы далее будем называть транскрипционный профиль гена, то есть соотношение его уровней экспрессии во всех образцах набора данных. В самом простом случае используется бинарная классификация: для каждого образца определяется, экспрессируется ли в нем данный ген, и тот ген, для которого наличие экспрессии определено только в одном образце, называется тканеспецифическим. Очевидным минусом такого подхода является необходимость выбора порога экспрессируемости (подробно обсуждалось в разделе 1.2.6). С помощью этого метода для разных объектов было обнаружено около тысячи тканеспецифических генов. Образцы, характеризующиеся наибольшим числом уникальных генов, были разными у разных растений. Так, в клубеньках *M. truncatula* специфично экспрессировалось 332 гена (Benedito et al., 2008), в корнях фасоли – около 300 (Vlasova et al., 2016), в прорастающих семенах табака – 211 (Edwards et al., 2010), в семенах коровьего гороха *Vigna unguiculata* (L.) Walp. – 684 (Yao et al., 2016). Для кукурузы наибольшее число тканеспецифических генов (334) характеризовало лист, тогда как междоузлия и ось початка имели крайне мало уникальных генов – факт, который необходимо учитывать при создании специфичных трансгенных конструкций, направленных на увеличение продуктивности этих органов, которые составляют 60% кормовой массы (Hansey et al., 2010; Sekhon et al., 2011). Небольшое количество обнаруживаемых тканеспецифических генов затрудняет анализ функций с помощью Генной Онтологии, однако в табаке уникальные для прорастающих семян гены были связаны с развитием семян, запасанием питательных веществ, системным ответом на избыток воды и свет высокой интенсивности (Edwards et al., 2010); среди специфических для гипокотыля

сосны генов чаще встречались гены, ассоциированные с метаболизмом ДНК (Cañas et al., 2017), а в семенах *V. unguiculata* – с запасанием липидов (Yao et al., 2016).

Параметр  $Z$ , использующийся для общей оценки транскрипционного профиля образца (см. раздел 1.2.7), также применялся для поиска тканеспецифических генов. Максимальные значения параметра  $Z$  свидетельствуют о наибольшем отклонении экспрессии гена в данной пробе от условного «среднего образца»; таким образом, можно отобрать гены, значимо отличающиеся от среднего только в одном образце. С его помощью были идентифицированы уникальные для клубеньков и семян гены сои (Severin et al., 2010), а для еще одного представителя семейства Fabaceae, лядвенца японского (*Lotus japonicus* L.), наибольшее число генов, обнаруженных с помощью параметра  $Z$ , также характеризовало клубеньки (Verdier et al., 2013).

Для поиска уникальных для органов и тканей генов в *B. distachyon* была применена мера специфичности SPM (SPecificity Measure), принимающая значения от 0, если ген не экспрессируется в данном образце, до 1, если ген экспрессируется исключительно в данном образце (Xiao et al., 2010; Sibout et al., 2017). Это позволило обнаружить значительно большее количество тканеспецифических генов, нежели параметр  $Z$  и бинарное наличие/отсутствие экспрессии: в семенах их число составило 3 966, в листьях – 1 303 и в междоузлиях – 1 176.

Энтропия Шеннона ( $H$ ) является широко используемой мерой тканевой специфичности (Schug et al., 2005; Lin et al., 2014). Распределение энтропии Шеннона показывает, что большая часть генов экспрессируется во всех или практически во всех образцах, кроме того, выраженный пик на низких значения  $H$  демонстрирует наличие тканеспецифических генов (Sekhon et al., 2013). С использованием энтропии Шеннона были обнаружены гены риса, уникальные для цветка, эндосперма, корней и каллуса (Wang et al., 2010).

Таким образом, к настоящему времени создано немало методов определения тканеспецифических генов. Исследование таких генов необходимо для уточнения механизмов, обеспечивающих уникальность каждого органа и ткани. Кроме того, составление списков тканеспецифических генов имеет важную практическую цель: определение генов-мишеней воздействия при генетических манипуляциях с целью увеличения экономической продуктивности (Sekhon et al., 2011). Промоторы тканеспецифических генов также могут быть использованы для создания трансгенных конструкций, экспрессирующихся в определенном органе на определенной стадии развития, что может снизить негативные последствия конститутивной экспрессии трансгенов (Yao et al., 2016).

### 1.2.10. Детальное изучение экспрессии транскрипционных факторов

Из всех тканеспецифических генов особое значение для определения уникальных особенностей ткани или органа имеют транскрипционные факторы (ТФ) – ДНК-связывающие белки, регулирующие активность других генов. Координированная регуляция экспрессии генов необходима для правильного протекания биологических процессов, и ТФ являются ключевыми узлами в сети генетической регуляции (Kagale et al., 2016). Семейства ТФ в большой степени консервативны среди эукариот (Jin et al., 2014), что позволяет сравнивать результаты, полученные с помощью транскриптомных карт растений разных таксонов.

Анализ 120 тканеспецифических ТФ сои позволил обнаружить предпочтительную экспрессию определенных семейств ТФ в определенных органах и тканях: в клубеньках присутствовали ТФ из семейств C2H2 (Zn) и NIN-like, в апикальной меристеме были перепредставлены гомеодоменные (HD) ТФ, а в цветках – LIM, MADS и NAC. Участие ТФ семейства HD в развитии меристемы (Laux et al., 1996), а MADS – в развитии цветка (Pelaz et al., 2001) *A. thaliana* известно, поэтому схожие паттерны экспрессии ортологов этих ТФ у сои подтвердили их возможное участие в тех же процессах (Libault et al., 2010b).

Особенности экспрессии ТФ табака и сосны были продемонстрированы на отдельных семействах: Ethylene Response Factors (ERF) и Myb-R2R3 соответственно (Edwards et al., 2010; Cañas et al., 2017). В случае *N. tabacum* значимой тканевой специфичности обнаружено не было, хотя гены одного из подсемейств ERF, вовлеченного в ответ на метилжасмонаты, обладают более высоким уровнем экспрессии в корнях, что, вероятно, связано с регуляцией роста корней (Edwards et al., 2010). Из 58 генов *P. pinaster*, относящихся к семейству Myb-R2R3, ранее изучалось только восемь. С помощью кластеризации по профилям экспрессии в работе Cañas et al. было выявлено одно подсемейство ТФ, равномерно экспрессирующееся во всех образцах, остальные же обладали тканеспецифичностью. Наличие известных Myb-R2R3 ТФ в каждом из кластеров может помочь в изучении функций остальных ТФ этого семейства (Cañas et al., 2017).

Детальное изучение ТФ *P. patens* было сосредоточено на регуляторах развития спорофита: их паттерн экспрессии в транскриптомном атласе был ограничен апикальной меристемой, причем гены, кодирующие ТФ из одного белкового комплекса, экспрессировались на схожем уровне (Ortiz-Ramírez et al., 2016).

Гексапloid *C. sativa* предоставляет возможность для изучения поведения триплицированных в ходе полиплоидизации ТФ. В геноме *C. sativa* было обнаружено 4 527 генов, кодирующих ТФ, из которых 82% сохранялись во всех трех копиях. Представленность ТФ в разных органах заметно различалась; анализ тканевой специфичности показал разделение ТФ на

13 кластеров. Примечательно, что паралогичные тройки только в 36% случаев оказывались не в одном кластере, что говорит о сохранении паттернов экспрессии (Kagale et al., 2016).

В клеточном атласе риса, имеющем самое большое разрешение относительно тканевой специфичности, была поставлена интересная задача: поиск регуляторных мотивов в промоторах генов, уникально или преимущественно экспрессирующихся в определенных типах клеток. Такой анализ представляет собой следующий шаг после простого определения тканеспецифических генов, и, вместе с улучшающейся аннотацией некодирующих участков генома, вносит существенный вклад в построение сетей генетической регуляции – так, в атласе риса была показана связь одного из мотивов в промоторах тканеспецифичных генов с ответом на абиотический стресс и заражение патогенами (Jiao et al., 2009).

### 1.2.11. Экспрессионные базы данных

Результаты транскриптомных экспериментов представлены в крупных базах данных, таких, как NCBI (National Center for Biotechnology Information, Национальный центр Биотехнологической Информации, <https://www.ncbi.nlm.nih.gov>) или EBI (European Bioinformatics Institute, Европейский Биоинформатический Институт, <https://www.ebi.ac.uk>), где они в основном хранятся в виде так называемых «сырых данных» (исходных необработанных чтений для секвенирования РНК и ненормализованных значений уровней флуоресценции для микрочипов). В настоящее время для некоторых видов растений в публичном доступе находится значительное число экспериментов: так, в декабре 2018 года для *A. thaliana* в базе данных NCBI (GEO) содержалось более 2 000 исследований, проведенных с помощью микрочипов, и более 1 000 – путем секвенирования РНК; каждый проект включает от одного до сотен образцов. Такой объем данных чрезвычайно тяжело поддается повторному анализу из-за значительных компьютерных ресурсов, требующихся для предварительной обработки. Кроме того, во многих случаях метаданные – формализованное описание образца – предоставлены не полностью, что также затрудняет и иногда делает невозможным использование публичных данных.

Частично эта проблема решена с помощью баз данных отдельных лабораторий и исследовательских групп, представляющих результаты масштабных экспрессионных экспериментов, таких, как транскриптомные карты, в удобной для пользователя форме. Число образцов в таких базах ограничивается десятками и сотнями образцов, зачастую принадлежащих к одному проекту; данные в них уже прошли всю необходимую предобработку и представляются в виде финальных значений, отражающих уровень экспрессии. Существующие экспрессионные базы данных можно разделить на две категории, различающиеся интерфейсом и инструментарием.

Базы, созданные для удобного отображения транскриптомных карт, отличаются наглядностью представления. Примером может служить широко используемая база данных «eFP browser» (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>), в которой транскриптомные карты изображаются в виде схематического рисунка соответствующего растения, на котором цветом обозначен относительный уровень экспрессии гена (Winter et al., 2007). Расширенная версия eFP browser, база данных ePlant (<http://bar.utoronto.ca/eplant>), в дополнение к экспрессии включает в себя иные типы данных, такие как координаты гена на хромосоме или филогения растения, однако не предоставляет возможности проводить какие-либо анализы экспрессии (Waese et al., 2017).

Базы данных второго типа обладают разнообразными аналитическими инструментами, но визуализация результатов, как правило, сводится к разного рода таблицам из-за сложности одновременного представления большого количества данных. С помощью таких баз возможно проведение дополнительного анализа, благодаря чему повышается эффективность использования данных. Так, в базе данных PLEXdb, (<http://www.plexdb.org>), охватывающей транскриптомные карты ячменя, кукурузы и риса (Druka et al., 2006; Wang et al., 2010; Sekhon et al., 2011), пользователь может выбирать разные типы нормализации чтений, такие как RPKM/FPKM (число чтений на тысячу п.н. гена на миллион картированных чтений или число фрагментов на тысячу п.н. гена на миллион картированных чтений, Reads Per Kilobase per Million mapped reads или Fragments Per Kilobase per Million mapped reads, метрики для одно- и парноконцевых чтений соответственно) и другие; в Melonet-DB (<http://gene.melonet-db.jp/cgi-bin/top.cgi>) возможно одновременное представление экспрессии нескольких генов (Yano et al., 2018); другие базы данных позволяют проводить анализ дифференциальной экспрессии (<https://mtgea.noble.org>, (Benedito et al., 2008; He et al., 2009) или поиск коэкспрессирующихся генов (<https://mtgea.noble.org>, PLEXdb, TomExpress <http://tomexpress.toulouse.inra.fr> (Zouine et al., 2017).

### *1.2.12. Проблемы и перспективы изучения транскриптомных карт*

Рассмотрение транскриптомных карт растений показывает, что к настоящему времени широкий спектр растительных видов изучен с применением передовых технологий транскриптомики. Однако нами был выявлен и ряд проблем в области создания и анализа транскриптомных карт. В первую очередь, среди атласов экспрессии генов крайне неравномерно представлены таксоны растений (Рисунок 1.2), что обуславливается созданием транскриптомных карт в первую очередь для хозяйственно-ценных видов и может приводить к ошибочным биологическим выводам, особенно в области эволюционных исследований (подробнее см. раздел 1.2.3). Мета-анализ атласов экспрессии также затруднен отсутствием координированных усилий

по созданию транскриптомных карт; международный консорциум, объединяющий разные научные группы и продолжающий геномный проект «1000 растений» (1000 Plants, <https://sites.google.com/a/uAlberta.ca/onekp/>), мог бы помочь с решением этой проблемы. В рамках консорциума возможно создание стандартов получения и анализа транскриптомных данных, аналогичных MIQE и MIAME для ОТ-ПЦР в реальном времени и микрочипов, соответственно (Brazma et al., 2001; Bustin et al., 2009), и, учитывая продолжающееся снижение цены секвенирования, определить обязательный минимальный набор образцов, таким образом значительно повысив возможности сравнительного анализа.

Далее, существующие эксперименты сконцентрированы на анализе белок-кодирующих генов; только в нескольких работах из рассмотренных нами анализировались малые РНК или длинные некодирующие РНК (Nobuta et al., 2007; Clevenger et al., 2016). Для животных уже существуют специализированные атласы некодирующих РНК (Liu et al., 2017a), и в ближайшее время следует ожидать появления аналогичных карт для растений. Дальнейшее развитие методов определения нуклеотидной последовательности и появление таких подходов, как секвенирование РНК одиночных клеток и одномолекулярное секвенирование в реальном времени, будут способствовать созданию транскриптомных карт лучшего разрешения. Эти методы уже доказали свою полезность при создании атласов экспрессии генов: с их помощью были созданы карта мозга дрозофилы (секвенирование РНК одиночных клеток, (Davie et al., 2018) и основанная на одномолекулярном секвенировании транскриптомная карта кролика (Chen et al., 2017). В дальнейшем эти методы могут быть адаптированы для растений (хотя применение некоторых из них затрудняется жесткой клеточной стенкой, для обзора см. (Yuan et al., 2018).

Как было описано в разделе 1.2.1, первая транскриптомная карта растения *A. thaliana* во многом определила направления анализа атласов экспрессии (Schmid et al., 2005). Однако при ее создании применялась технология микрочипов, имеющая существенные ограничения (подробнее см. раздел 1.2.2), а вся область транскриптомики в то время только начинала развиваться. Следует отметить, что эти транскриптомные данные широко использовались при анализе профилей разнообразных генов (для примера см. (Mara et al., 2010; Kunieda et al., 2013; Yamaguchi et al., 2018), а также для создания Генной Онтологии *A. thaliana*, вследствие чего все последующие анализы могли унаследовать имеющиеся в ней технические недостатки. Соответственно, представляется целесообразным создание атласа экспрессии *A. thaliana* с помощью современного метода секвенирования РНК для обновления и уточнения данных о самом исследованном объекте среди растений, а также для получения информации о ранее неизученных генах, тканях, органах и стадиях развития.

### 1.3. Результаты

#### 1.3.1. Выбор образцов транскриптомной карты

Полная транскриптомная карта организма, согласно Schmid *et al.*, должна охватывать как можно большее число его органов, тканей и типов клеток на разных стадиях онтогенеза, от эмбрионального развития до старения, а также содержать образцы, собранные в разных условиях жизни организма, например, под воздействием абиотических или биотических стрессов (Schmid *et al.*, 2005). Наша работа сосредоточена на создании подробной транскриптомной карты *A. thaliana* в «нормальных» условиях выращивания, то есть при типичной для *Arabidopsis* температуре +22°C (Arabidopsis Biological Resource Center, 2015) и при длительной освещенности в течении суток (т.н. условия «длинного дня», 16 часов света и 8 часов темноты, индуцирующие цветение у *A. thaliana* (Koornneef *et al.*, 1991); подробнее об условиях выращивания см. Материалы и Методы). Основой транскриптомной карты *A. thaliana* послужили 79 образцов разных частей и органов на различных стадиях развития, от примордиев до стареющих и засыхающих структур; в их число вошли корни, меристемы, листья, органы цветка, целые цветки, семена, стручки, стебли. Изучение функционирования транскриптома в динамике представляет отдельный интерес, и для этой цели хорошо подходят временные серии отдельных органов (Bar-Joseph *et al.*, 2012), то есть сбор образцов последовательно, день за днем, или, что возможно в случае растений благодаря их модульной структуре (Schmid *et al.*, 2005), одновременный сбор органов, находящихся на разных стадиях развития. В наш атлас экспрессии вошли следующие временные серии: меристемы, цветки, семена и листья. Все образцы транскриптомной карты были собраны в двух биологических повторностях, каждая из которых представляла собой пул из 15 растений, что позволяет уменьшить число необходимых биологических повторностей. Подробное описание собранных нами образцов находится в Приложении, Таблица 1.2.

Известно, что некоторые гены не экспрессируются в нормальных условиях выращивания; особенно это касается генов, участвующих в ответе на стрессовые воздействия. Примером может служить ген *DOF1.3* (*AT1G26790*), один из ранних генов холодового ответа *A. thaliana*, экспрессия которого в норме не обнаруживается в таких частях растения, как листья, стебли или корни, однако активируется под действием низких температур (Park *et al.*, 2015). Чтобы оценить, насколько полно созданный нами атлас экспрессии описывает возможные состояния транскриптома *A. thaliana*, мы использовали собранные в лаборатории временные серии листовых проб, подвергнутых воздействию пониженной температуры (+4°C), повышенной температуры (+42°C) и механическим повреждениям (протыкание листьев иглой).

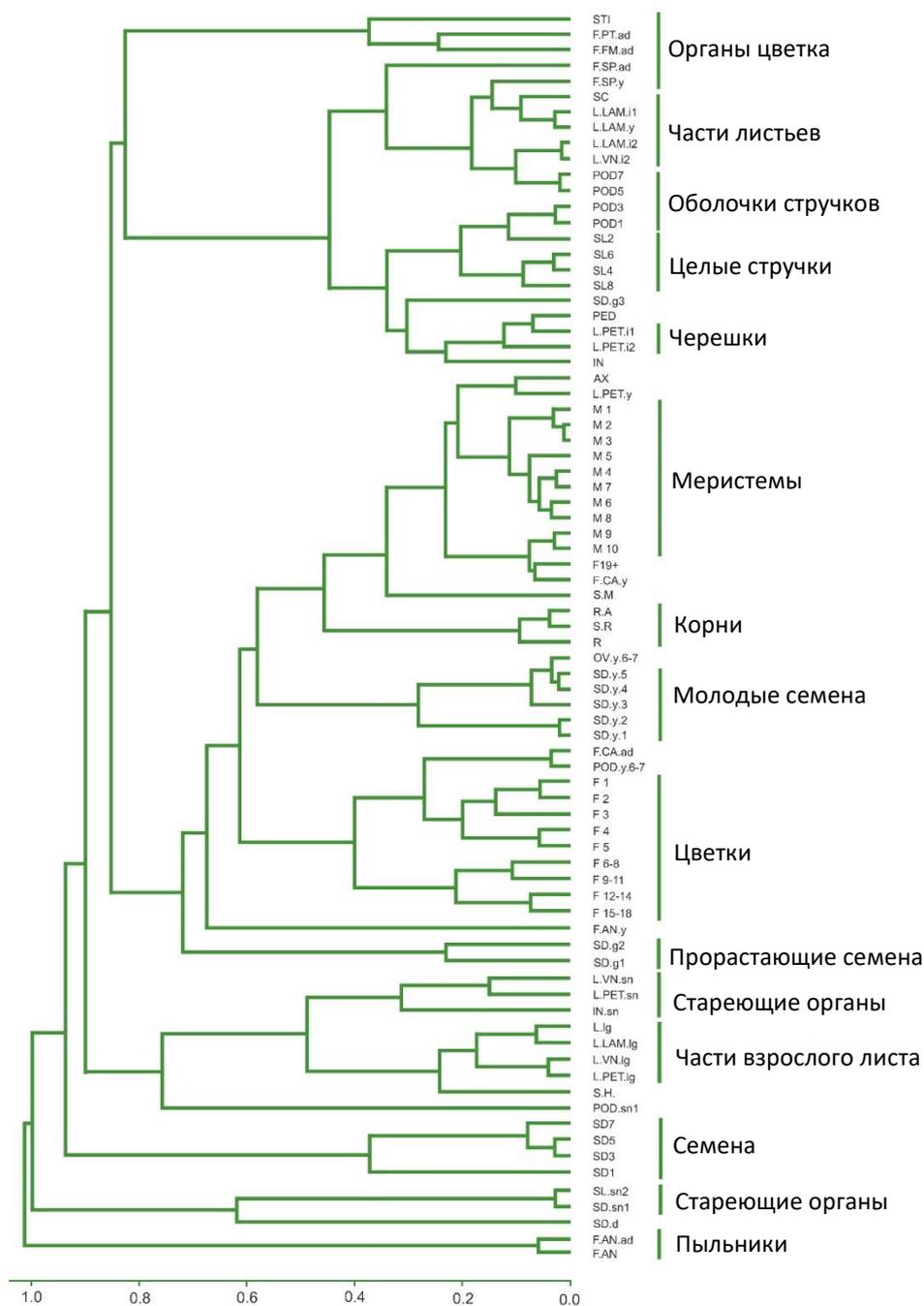
### 1.3.2. Секвенирование и первичная обработка данных

В Лаборатории эволюционной геномики из собранных нами образцов была выделена РНК, сконструированы библиотеки кДНК для секвенирования на платформе Illumina и проведено само секвенирование с помощью прибора HiSeq2000 (Illumina, США). Затем полученные данные были предобработаны для удаления технических последовательностей и чтений низкого качества, уникально картированы на геном *A. thaliana*, и для каждого гена было определено суммарное число чтений, приходящихся на ген (см. Материалы и Методы). Это значение является исходным для всех дальнейших анализов.

В результате в среднем для каждого образца было получено 22,7 млн уникально картированных высококачественных чтений (Клепикова et al., 2016), Таблица S2 Приложения), а общий объем данных транскриптомной карты составил 3,6 млрд чтений. Упомянутые выше образцы, подвергнутые стрессовым воздействиям, были ранее секвенированы в той же лаборатории с получением в среднем 18,9 млн чтений и общим объемом 606 млн чтений.

Следует отметить, что первая стадия конструирования библиотек кДНК заключается в селекции фракции мРНК, что может быть сделано двумя основными методами: (1) с помощью гибридизации мРНК на микрочастицах, несущих олиго-дТ последовательности, и удаления этих частиц из раствора, где остаются не полиаденилированные (полиА) РНК, в основном рРНК; (2) с помощью гибридизации рибосомальной РНК на микрочастицах, несущих зонды, комплементарные к участкам рРНК данного вида, и удаления из раствора, где остаются все остальные типы РНК, частиц с рРНК. Второй способ позволяет проанализировать большее разнообразие РНК, однако требует значительного количества исходного материала, что невозможно в случае приготовления библиотек из таких структур растения как, например, меристемы, поэтому для транскриптомной карты *A. thaliana* был применен протокол селекции фракции полиА мРНК. По этой причине основные анализы проводились для белок-кодирующих генов.

Полученные данные в первую очередь были проверены на пригодность к дальнейшему анализу в отношении сходства биологических повторностей. Для этого был использован квадрат коэффициента корреляции Пирсона ( $r^2$ ). Его значения для всех пар повторностей лежали в диапазоне от 0,83 до 1 со средним 0,97 и медианой 0,98 (Приложение, Таблица 1.2). Для анализа и наглядного представления сходства образцов между собой было построено основанное на  $r^2$  иерархическое дерево кластеризации (Рисунок 1.3).



**Рисунок 1.3.** Дерево кластеризации образцов транскриптомной карты. В качестве меры расстояния между парой образцов (ось X) использовался вычтенный из единицы квадрат корреляции Пирсона.

### 1.3.3. Интегральные параметры экспрессии генов

Первым этапом анализа атласа экспрессии *A. thaliana* было общее описание транскриптома в терминах числа экспрессирующихся генов как в отдельных образцах, так и по всей транскриптомной карте в целом.

В аннотации генома *A. thaliana* (TAIR10) представлено 33 323 гена, из них 27 201 описаны как кодирующие мРНК. Как было сказано выше, из-за особенностей протокола приготовления

библиотек для секвенирования анализ экспрессируемости проводился только для таких генов. Абсолютный уровень экспрессии гена зависит от глубины секвенирования (размера библиотеки) образца: разница в размерах библиотек между двумя образцами в два раза, может привести двукратному отличию в уровне экспрессии данного гена, даже если реальная его экспрессия (количество мРНК) в этих образцах одинакова. Поэтому перед проведением поиска экспрессирующихся генов в образцах транскриптомной карты значения суммарного числа чтений, приходящихся на ген, для всех образцов были нормализованы на размер библиотеки с помощью метода медианы, реализованного в пакете «DESeq» статистической среды R (Anders and Huber, 2010), см. Материалы и Методы).

Выбор порога, отделяющего экспрессирующиеся гены от неэкспрессирующихся, может непосредственно повлиять на результаты анализа (Hart et al., 2013). Так как в нашей работе в качестве численного измерения экспрессии использовалось общее число чтений, приходящихся на данный ген в данном образце, а не RPKM или FPKM, часть методов определения экспрессируемости гена оказались не применимы (например, предложенный в работе Hart *et al.* метод zFPKM (Hart et al., 2013). Для снижения числа ложноположительных результатов нами был использован «строгий» порог экспрессируемости: на ген должно приходиться 16 или более нормализованных чтений в каждой из двух биологических повторностей образца (Su et al., 2014).

Из 27 201 гена *A. thaliana*, кодирующих мРНК, хотя бы в одном образце экспрессировался 24 621 ген. Во всех 79 образцах экспрессировалось 10 654 генов, что составляет 40% от белок-кодирующих генов или же треть от всех генов. Число экспрессирующихся генов в каждом образце в отдельности различалось немногим более, чем на 4 000: меньше всего транскрибирующихся генов было обнаружено в образце M1 (15 525 генов, вегетативная меристема), а больше всего – в F3 (19 613 генов, третий цветок на стадии распускания первого цветка) (Приложение, Рисунок 1.1).

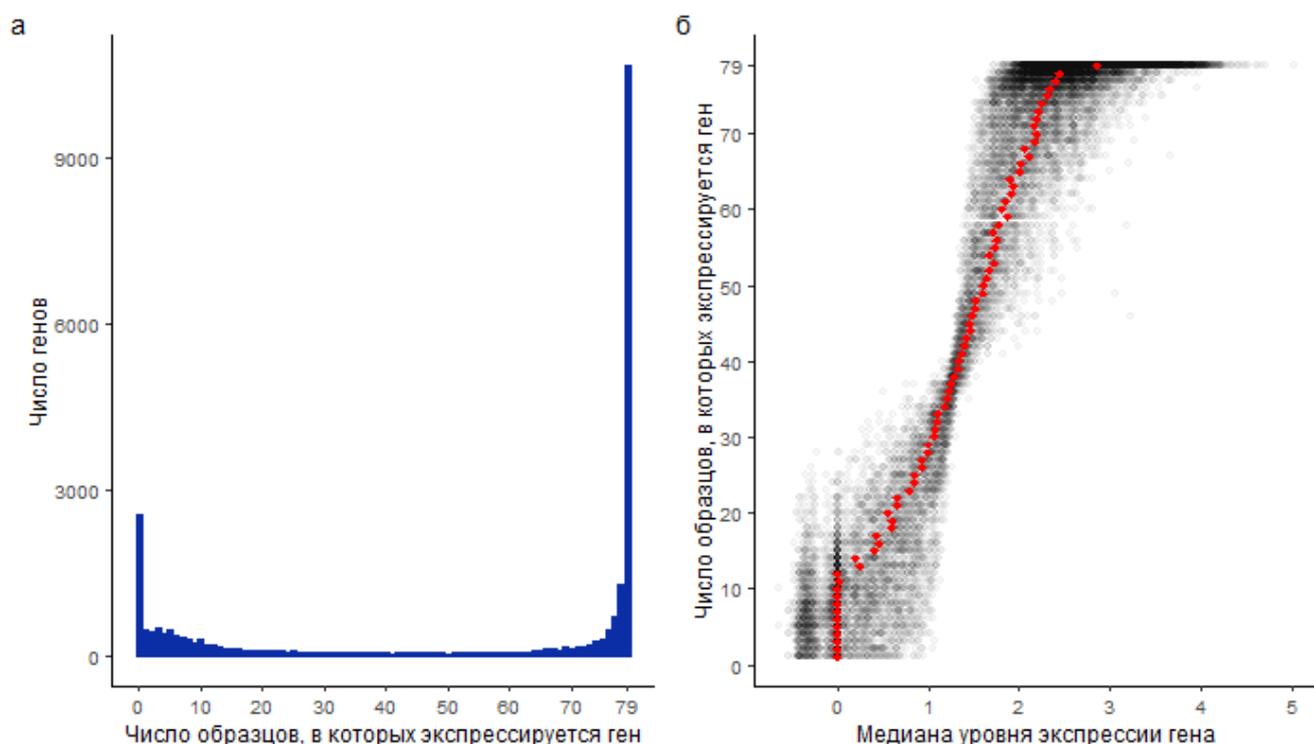
Собранные в нашей работе 79 образцов в значительной мере описывают анатомию растения, однако остается вероятность, что транскрипты некоторых генов не были обнаружены, так как эти гены экспрессируются либо в частях и/или стадиях развития растения, не вошедших в созданную нами транскриптомную карту, либо в иных условиях выращивания. Для проверки полноты атласа экспрессии мы использовали набор образцов, подвергнутых воздействию абиотических стрессов (см. раздел 1.3.1). Транскрипты 26 генов, кодирующих мРНК (96 генов, если принимать в расчет все 33 323 аннотированных гена *A. thaliana*), обнаруживались только под воздействием стрессов, но не в основном атласе экспрессии. Из всех присутствующих в геноме *A. thaliana* генов 7 617 не экспрессировалось ни в транскриптомной карте, ни в стрессовых образцах (из них белок-кодирующих – 2 580, а основную часть составляют псевдогены и некодирующие РНК).

Для получения общего представления о функциях изучаемых генов широко используется классификация «Генная Онтология» (Blake, 2013). Генная Онтология (ГО, Gene Ontology) – это система универсального описания генов и их продуктов для всех биологических видов (Ashburner et al., 2000). Описания разделены на три группы (словаря): молекулярная функция, которую выполняет продукт гена, биологический процесс, в котором он участвует, и клеточный компонент, в котором продукт гена локализуется. Внутри словаря существует система из вложенных категорий (терминов), описывающих ген с разной степенью точности: например, ген может быть описан как «участвующий в клеточном метаболизме кетонов» (GO:0042180), что является самым нижним и точным уровнем описания, и этот термин входит в более крупную категорию «метаболизм малых молекул» (GO:0044281), и так далее вплоть до самого общего термина «биологический процесс» (GO:0008150). Каждый ген с той или иной точностью описан терминами из трех словарей.

Для списка неэкспрессирующихся генов был проведен анализ обогащения категориями ГО и других баз данных (таких, как InterPro, (Mitchell et al., 2019): количество генов, аннотированных каждым термином в списке неэкспрессирующихся генов, сравнивалось с теоретически ожидаемым, исходя из общего числа таких генов в геноме (см. Материалы и Методы). С его помощью была показана перепредставленность таких терминов, как «защитная реакция в ответ на грибы» (GO:0050832), «убийство клеток других организмов» (GO:0031640), «РНК-направленная активность ДНК-полимеразы» (GO:0003964) (Klepikova et al., 2016), Таблица Приложения S9).

Как было сказано в разделе 1.2.1, в 2005-ом году был создан первый атлас экспрессии генов растения, модельного объекта *A. thaliana* (Schmid et al., 2005). Сравнение нашей транскриптомной карты с ним служит дополнительным способом проверки состоятельности полученных нами данных. Атлас Schmid *et al.* был получен с помощью технологии микрочипов и содержит информацию о 21 150 генах, пробы к которым присутствовали на микрочипе, в 79 биологических образцах. Транскрипты 94% этих генов были обнаружены в наших данных. Кроме того, экспрессия 5 877 генов была выявлена в созданной нами транскриптомной карте, но не в данных Schmid *et al.*

Самой простой интегральной характеристикой экспрессии гена может служить ширина паттерна экспрессии, посчитанная как число образцов, в которых этот ген транскрибируется (Рисунок 1.4 а, см. раздел 1.2.7). Распределение генов по числу образцов, в которых они экспрессируются, имеет два пика: транскрипты основной части генов, кодирующих полиА РНК, наблюдаются почти во всех образцах (15 296 генов – в 65 и более образцах), второй пик приходится на небольшое число образцов (4 920 генов экспрессируется в 15 и менее образцах), а в промежутке от 16 до 64 образцов находится 4 405 генов.



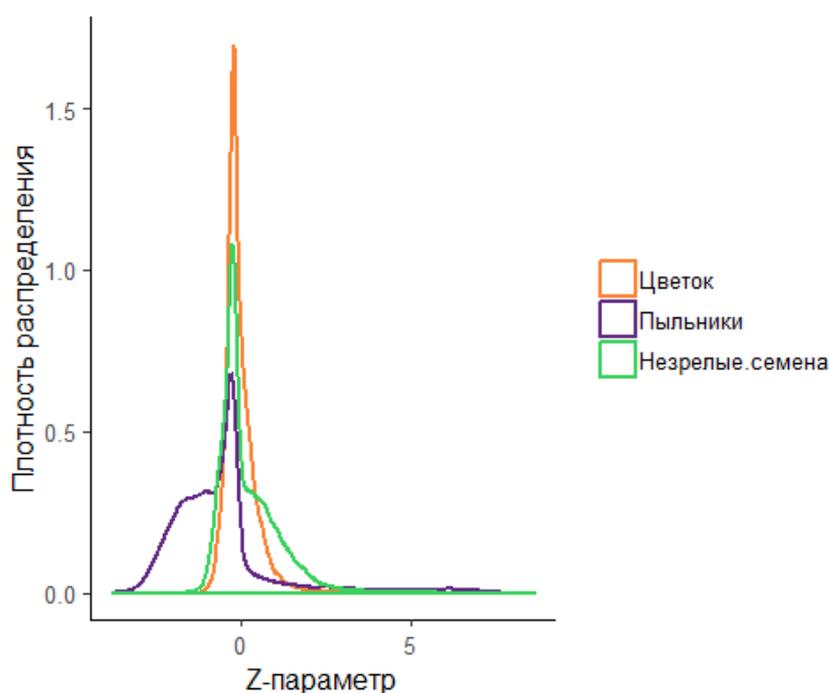
**Рисунок 1.4.** Экспрессионные характеристики генов. (а) Распределение генов по числу образцов, в которых ген экспрессируется (порог экспрессируемости составляет 16 и более чтений в каждой из биологических повторностей); (б) Распределение числа образцов, в которых экспрессируется ген, в зависимости от медианного уровня экспрессии гена (шкала X представляет собой десятичный логарифм медианы экспрессии гена).

Кроме того, нами была продемонстрирована связь между числом образцов, в которых ген экспрессируется, и уровнем экспрессии (Рисунок 1.4 б, Приложение, Рисунок 1.2). Средняя экспрессия, как и медианная, оказалась выше для генов, обладающих более широким паттерном (то есть экспрессирующихся в большем числе образцов). Этот же тренд прослеживается также для минимума и максимума экспрессии, хотя и в меньшей степени (Приложение, Рисунок 1.2).

При анализе экспрессируемости генов *A. thaliana* следует учитывать некоторые недостатки существующей сборки и аннотации генома (TAIR10). Так, ряд соседних на хромосоме генов (например, *AT3G30385* и *AT3G30387*, *AT3G28290* и *AT3G28300*, *AT5G50580* и *AT5G50680*) имеют идентичные нуклеотидные последовательности, что, скорее всего, является следствием ошибок сборки или аннотации. Это препятствует картированию чтений на геном при требовании уникального картирования (то есть для каждого чтения в геноме должно находиться только одно место с совпадающей последовательностью), и, таким образом, эти гены будут иметь нулевое число чтений во всех образцах, на самом деле являясь экспрессируемыми. Для оценки количества таких генов мы провели неуникальное картирование, что добавило в список экспрессирующихся хотя бы в одном образце 213 белок-кодирующих (334 всего) генов. Из них во всех образцах экспрессировалось 143 (159) генов.

### 1.3.4. Сравнение параметров экспрессии генов в образцах

Помимо общего описания транскриптома организма, интерес представляет сравнение профилей экспрессии отдельных образцов. Для представления распределения уровней экспрессии мы использовали параметр  $Z$  ( $Z$ -score), который представляет собой разницу числа чтений на ген в данном образце и среднего по всем образцам числа чтений, нормированную на стандартное отклонение по всем образцам (см. Материалы и Методы). Распределение параметра  $Z$  позволяет судить о том, насколько экспрессия генов в конкретном образце отличается от усредненной по всем образцам; при большой близости образца к «усредненному» параметр  $Z$  будет нормально распределен со средним 0.



**Рисунок 1.5.** Распределение параметра  $Z$  в трех образцах транскриптомной карты: полностью распустившегося цветка (Цветок, F1), вскрывшихся пыльников полностью распустившегося цветка (Пыльники, F.AN), незрелых семян, собранных с полностью сбросившего околоцветник стручка (Незрелые семена, SD.y1).

В целом распределения параметра  $Z$  во всех образцах обладали большой схожестью. Однако в ряде проб распределение было смещено вправо, что говорит о превышении среднего уровня транскрипции некоторыми генами (SD.y1, молодые семена; OV6-7, семяпочки до оплодотворения), а в других образцах – влево (F.AN, открытые пыльники; F.AN.ad, зрелые пыльники до вскрытия) (Рисунок 1.5, Приложение, Рисунок 1.3).

Другим использованным нами способом сравнения образцов было изучение соотношений терминов ГО, описывающих экспрессирующиеся в этих образцах гены. Традиционно используются два варианта аннотации генов терминами ГО: основная, наиболее полная, где генам по возможности присваивается самое точное описание молекулярной функции,

биологического процесса и клеточного компонента; и вторая, так называемая аннотация «slim», которая представляет собой срез сети онтологии на более высоком уровне и, таким образом, дает более общее описание гена. Однако в случае наших данных даже аннотация «slim» оказалась недостаточно общей для получения сравнимой картины по разным образцам из-за большого количества терминов ГО и слишком малого числа генов в каждом из них. Чтобы получить более крупные категории, нами был сделан срез сети аннотации на третьем уровне, считая первым уровнем начальные категории «молекулярная функция» (GO:0003674), «биологический процесс» (GO:0008150) и «клеточный компонент» (GO:0005575). Соотношение терминов третьего уровня ГО было практически одинаковым во всех образцах; средний уровень экспрессии и среднее число образцов, в которых экспрессируются гены различных ГО категорий, показывало различия между необходимыми для функционирования всех тканей генами (например, входящих в категории GO:0048856 – «развитие анатомических структур» и GO:0044424 – «внутриклеточные компоненты») и органоспецифическими категориями (GO:0070505 – «оболочка пыльцевого зерна»).

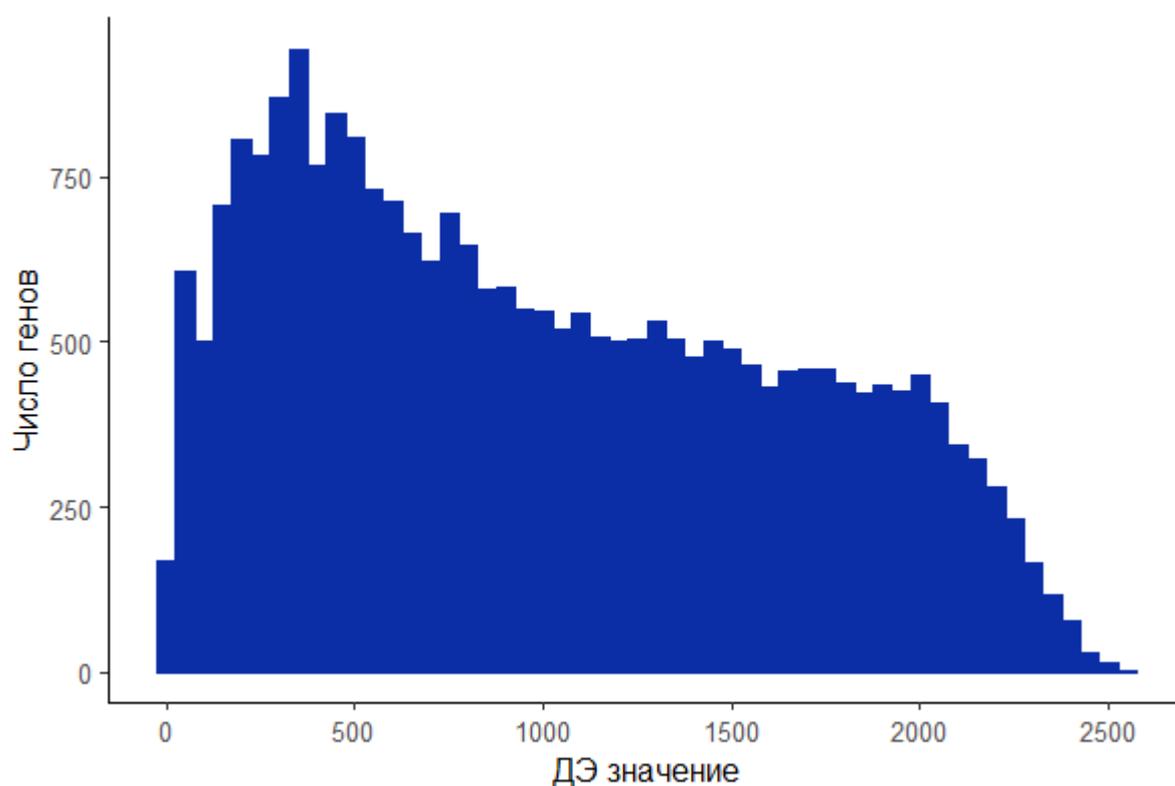
### *1.3.5. Анализ дифференциальной экспрессии*

Изучение различий в уровнях экспрессии генов между образцами является широко используемым способом получения информации как о функциональной разнице между пробами (например, контрольной и подверженной стрессовому воздействию), так и о поведении исследуемого гена в различных условиях или тканях (Soneson and Delorenzi, 2013). Для общей характеристики разнообразия собранной нами транскриптомной карты мы проанализировали дифференциальную экспрессию генов между всеми возможными парами образцов, что дало 3 081 попарное сравнение. Анализ был проведен с использованием пакета статистической среды R DESeq (Anders and Huber, 2010).

Число дифференциально экспрессирующихся генов между разными парами образцов варьировалось в широких пределах: наиболее схожими оказались образцы M1 и M2 (меристемы до перехода к цветению, 14 ДЭ генов), а наиболее различающимися – F3 и SD.d (третий цветок при распускании первого и сухие семена, 15 149 ДЭ генов) (Клерикова et al., 2016), Таблицы S12 и S13 Приложения). Число ДЭ генов использовалось в качестве меры расстояния между образцами для построения иерархической кластеризации (Приложение, Рисунок 1.4).

Гены, экспрессия которых возрастает в одном образце по сравнению с другим, и гены, которые имеют сниженную экспрессию в первом образце, могут быть вовлечены в разные процессы, протекание которых отличается между образцами. По этой причине такие гены следует рассматривать отдельно. Чтобы получить общее представление о генах с различным уровнем транскрипции в разных частях и органах растения, мы проанализировали

перепредставленность категорий ГО в списках генов, ДЭ между каждой парой образцов; гены, экспрессия которых возрастает или снижается в одной из проб, рассматривалось отдельно. Всего было обнаружено 1 528 категорий ГО, обогащенных более чем в два раза хотя бы в одном списке генов, а также 1 531 терминов из других баз данных. Многие термины, относящиеся к фотосинтетическим процессам, были перепредставлены в списки ДЭ генов; к ним относятся такие категории, как «фотосинтетические мембраны» (GO:0034357, 2 169 сравнений) и «темновая стадия фотосинтеза» (GO:0019685, 1 510 сравнений) (Клерікова et al., 2016), Таблица S14 Приложения). Кроме того, часто оказывались дифференциально экспрессирующимися гены, вовлеченные в регуляцию хроматина и деления клеток, такие как «коровые гистоны» (IPR007125, 2013 сравнений), «хроматин» (GO:0000785, 1 490 сравнений), «фазы клеточного цикла» (GO:0022403, 1 445 сравнений), «циклин, С-конец» (IPR004367, 1 525 сравнений).



**Рисунок 1.6.** Распределение ДЭ-значений генов, дифференциально экспрессирующихся хотя бы в одном сравнении образцов.

Для описания разнообразия экспрессии в отношении генов мы ввели интегральную меру «ДЭ-значение гена» – число парных сравнений, в которых данный ген являлся ДЭ. ДЭ-значение отражает, во-первых, широту паттерна экспрессии (гены, экспрессирующиеся в большем числе образцов, потенциально могут иметь более высокое ДЭ-значение) и, во-вторых, статистически достоверную разницу в уровнях экспрессии в этих образцах. ДЭ-значение позволяет определить гены, которые вовлечены в орган-специфические либо в базовые процессы организма. Распределение ДЭ-значений транскриптомной карты представлено на Рисунке 1.6.

Хотя теоретически максимально возможное ДЭ-значение равнялось числу парных сравнений образцов (3 081), наибольшее наблюдаемое ДЭ-значение составило 2 533. Для 331 гена ДЭ-значения превышали 2 300. Анализ перепредставленности категорий ГО показал обогащение списка этих генов терминами, относящимися к фотосинтезу и связанному с ним метаболизму (такими, как «фотосинтетическая мембрана», GO:0034357, и «метаболизм глюкозинолатов», GO:0019757) (Приложение, Таблица 1.3).

852 гена (из числа экспрессирующихся хотя бы в одном образце) имели ДЭ-значения от 100 и меньше и были обогащены категориями, связанными с процессингом РНК (GO:0006397), локализацией белков GO:0034613) и транспортом (GO:0016192) (Приложение, Таблица 1.4). Распределение числа образцов, в которых экспрессируются гены с минимальным ДЭ-значением, оказалось бимодальным с пиками на одном и 79 образцах (Приложение, Рисунок 1.5).

### 1.3.6. Стабильность экспрессии генов

Поиск генов, стабильно экспрессирующихся в различных органах и условиях, является важной задачей в сравнительном анализе экспрессии. Как и во многих других работах, посвященных анализу транскриптомных карт (Wang et al., 2010; Meyer et al., 2012; O'Rourke et al., 2015), в качестве меры равномерности экспрессии мы использовали коэффициент вариации, то есть отношение стандартного отклонения уровня экспрессии гена по всем образцам к среднему. Коэффициент вариации у 47 генов был меньше 0,2, у 339 генов – меньше 0,25 и у 970 – меньше 0,3; ДЭ-значение у этих генов находилось в диапазоне от 0 до 488 (при максимуме 3 081). Анализ обогащения списков этих генов терминами ГО показал перепредставленность категорий, связанных с транспортом нуклеиновых кислот и белков (GO:0050658, GO:0006886), мембранами (GO:0030662, GO:0030117) и процессингом РНК (GO:0008380, GO:0016071) (Klerikova et al., 2016), Таблица Приложения S20).

Чтобы проверить, сохраняет ли экспрессия этих генов стабильность в других условиях выращивания, мы добавили результаты секвенирования РНК растений, подвергнутых воздействию абиотических стрессов (см. раздел 1.3.1), к транскриптомной карте и пересчитали коэффициент вариации. При пороговом значении CV 0,2 стабильными оказались 34 гена, 274 гена имели CV меньше 0,25, 792 гена – меньше 0,3. Для дальнейшего отбора наиболее стабильно экспрессирующихся генов *A. thaliana*, список генов, стабильных в образцах транскриптомной карты, сравнивался со списком генов, стабильных как в стрессовых, так и в нормальных условиях, и в финальный анализ вошли только гены, присутствующие в обоих списках (Рисунок 1.7). Отношение стандартного отклонения уровня экспрессии к среднему по всем образцам у этих генов различалось от 0,15 до 0,2, а ДЭ-значение – от 0 до 117, тогда как энтропия Шеннона варьировалась от 4,56 до 4,57. Перепредставленность категорий ГО для списка самых

стабильных генов в целом совпадала с обогащением генов, стабильных в транскриптомной карте *A. thaliana*.

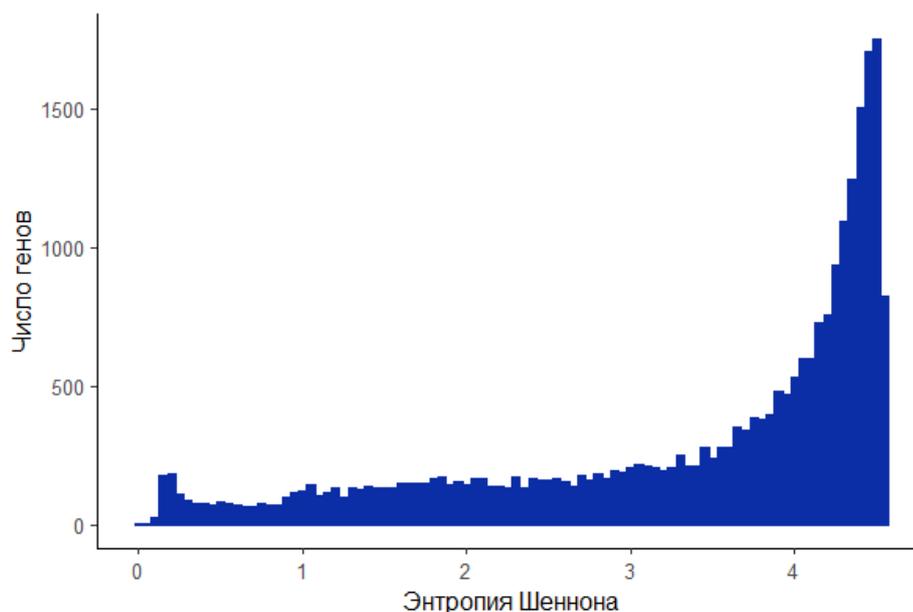


**Рисунок 1.7.** Диаграмма Венна для генов, стабильно экспрессирующихся в транскриптомной карте, и генов, стабильных в объединенных данных транскриптомной карты и стрессовых воздействий. Приведены количества стабильных генов при трех порогах:  $CV \leq 0,20$ ,  $CV \leq 0,25$ ,  $CV \leq 0,30$ .

### 1.3.7. Специфичность экспрессии генов

Еще одним способом изучения глобального транскриптома объекта является описание паттернов экспрессии генов и поиск среди них специфично экспрессирующихся в разных органах, тканях и типах клеток. В качестве метрики тканевой специфичности гена в ряде работ использовалось число образцов, в которых экспрессируется ген (Ponger et al., 2001; Huminiecki et al., 2003; Vinogradov, 2003); этот способ был применен нами для общего описания экспрессии генов (см. раздел 1.3.3; подробнее о методах определения тканеспецифичных генов см. раздел 1.2.7). Однако такая оценка не учитывает возможные отличия в уровнях экспрессии гена в разных образцах, что может внести погрешность в анализ; кроме того, как обсуждалось в раздел 1.2.7, этот способ подразумевает произвольный выбор порога по уровню экспрессии, что затрудняет проведение сравнений между транскриптомными картами.

Этих недостатков лишена такая мера ширины паттерна экспрессии, как энтропия Шеннона (H). Максимально возможная энтропия определяется числом образцов, исходя из полностью равномерной экспрессии в них; в нашей транскриптомной карте наибольшая энтропия составила 4,57. Распределение энтропии H смещено в область высоких значений, и, таким образом, показывает, что большинство генов обладают широким паттерном экспрессии (Рисунок 1.8). Второй локальный максимум, значительно меньших размеров (1,8%), находится в области значений H 0,1-0,25 и образуется генами, специфично экспрессирующимися в определенных образцах.



**Рисунок 1.8.** Распределение энтропии Шеннона экспрессирующихся в транскриптомной карте генов.

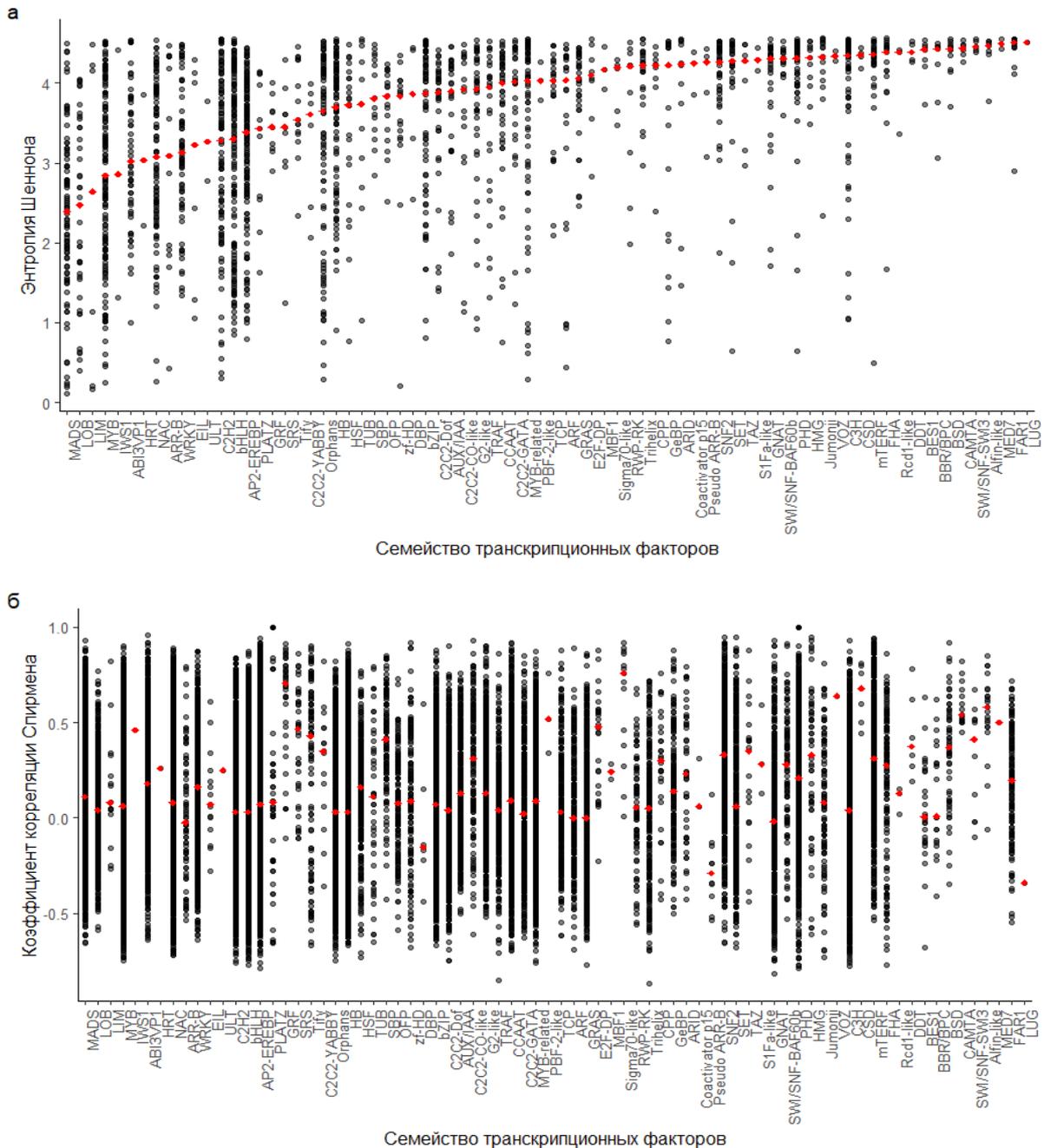
Анализ перепредставленности категорий ГО и других баз данных в списке генов с энтропией, превышающей 4,53, показал связь этих генов с такими процессами, как транспорт нуклеиновых кислот, процессинг РНК (GO:0006397), и такими клеточными компонентами, как мембраны (GO:0031965) (Klerikova et al., 2016), Таблица Приложения S17). Гены, энтропия которых была ниже 0,15, характеризовались категориями «межклеточный сигналинг» (GO:0007267), «пектинэстеразная активность» (GO:0030599), «клеточная стенка» (GO:0005618) (Klerikova et al., 2016), Таблица Приложения S16).

### 1.3.8. Паттерны экспрессии семейств транскрипционных факторов

Изучение генов, кодирующих транскрипционные факторы, необходимо для понимания функционирования сетей генетической регуляции, контролирующей разные аспекты развития и жизнедеятельности организма. Согласно базе данных PlantTFDB (Jin et al., 2014), в геноме *A. thaliana* аннотировано 2 757 ТФ, относящихся к 82 семействам (Jin et al., 2015), из них 2 148 генов экспрессируется хотя бы в одном образце нашей транскриптомной карты.

Для изучения паттерна экспрессии регуляторных генов мы проанализировали распределения значений энтропии Шеннона для всех семейств ТФ (Рисунок 1.9 а). Большинство семейств характеризовалось большой дисперсией  $H$ : среди принадлежащих к ним ТФ встречаются как тканеспецифичные, так и экспрессирующиеся во всех образцах гены. Тем не менее, медианы энтропии разных семейств ТФ варьировались от 2,39 до 4,51; наиболее узкий паттерн экспрессии имели такие классы ТФ, как MADS, LOB, LIM и MYB, тогда как самые широко экспрессирующиеся гены, кодирующие регуляторы транскрипции, входили в семейства SWI/SNFSWI3, SNF2, CAMTA, DDT и FAR.

Во многих случаях продукты генов, входящих в семейства ТФ, формируют белковые комплексы, как, например, субъединицы рибосом или компоненты фотосистем, что позволяет предположить высокую медиану корреляции экспрессии внутри семейства. Однако, согласно нашим данным, коэффициенты корреляции Спирмена между всеми парами генов, относящихся к одному классу ТФ, варьировались в широких пределах, а медиана корреляций в большинстве из них была близка к 0 (Рисунок 1.9 б). Кроме того, медианы корреляций семейств ТФ не зависели от медианы энтропии Шеннона.



**Рисунок 1.9.** Распределение (а) энтропии Шеннона в семействах ТФ и (б) коэффициента корреляции Спирмена между всеми парами ТФ в каждом семействе.

## 1.4. Обсуждение

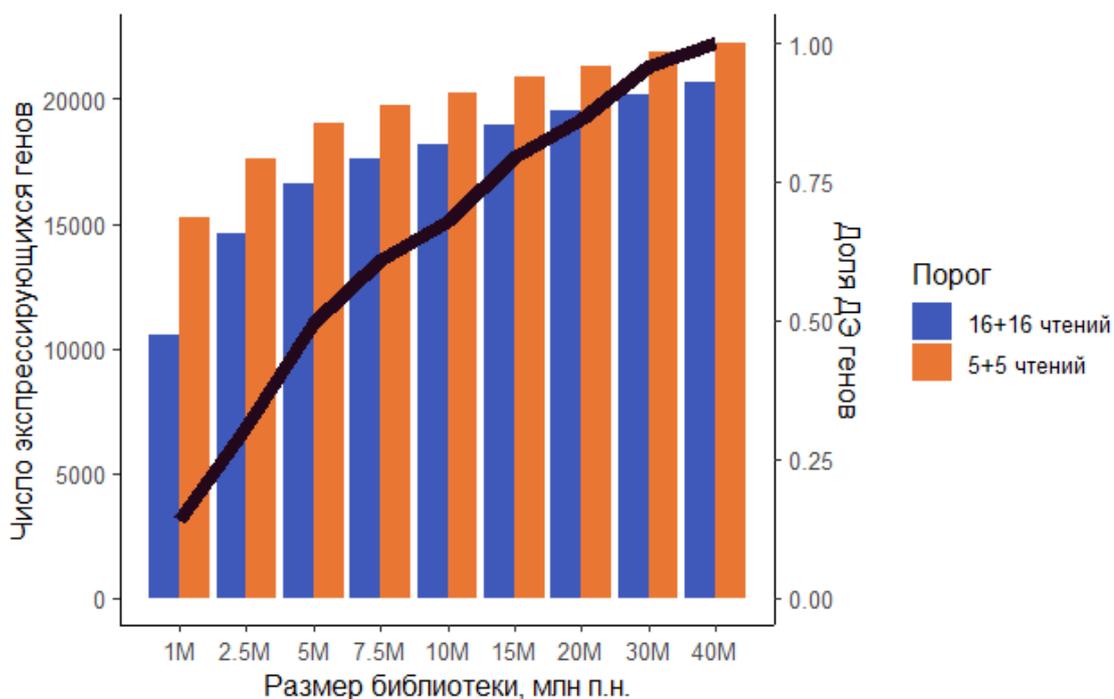
### 1.4.1. Структура транскриптомной карты

Целью нашей работы было создание высокоразрешенной транскриптомной карты *A. thaliana* с помощью современного метода анализа экспрессии – секвенирования РНК. С 2005-го года область транскриптомики, посвященная созданию и анализу атласов экспрессии генов, неуклонно развивалась, благодаря чему мы имели возможность при выборе образцов опираться на ранее созданные транскриптомные карты, среди которых особенно важна карта *A. thaliana* (Schmid et al., 2005). Анализ существовавших к тому времени атласов позволил определить минимальный необходимый набор образцов (подробнее см. раздел 1.2.4); в их число вошли семена, цветки, листья, корни. Чтобы наиболее эффективно расширить этот список и включить в него менее изученные органы *A. thaliana*, мы в подробностях рассмотрели выбор биологических проб в атласе экспрессии, созданном с применением микрочипов (Schmid et al., 2005). Изучение кластеризации образцов этой транскриптомной карты показало, что разные части и стадии развития корней, хотя и характеризуются значительным числом специфичных для этого органа генов, в большой степени схожи между собой; то же касается последовательно собранных листьев розетки (см. (Schmid et al., 2005), Рисунок 1С). Соответственно, повторное исследование этих образцов с той же степенью детализации не представляется целесообразным. Вместо них нами были собраны три образца корней (апекс корня, корень без апекса и корень проростка) и серия разных стадий развития третьего розеточного листа (Приложение, Таблица 1.2). С другой стороны, чрезвычайно важным представляется включение в транскриптомную карту органов и тканей, которым было уделено мало внимания в предыдущих широкомасштабных исследованиях, поэтому мы собрали такие образцы, как апикальная меристема побега, органы цветка на стадии 9 (Smyth et al., 1990), желтеющий лист, междоузлие и стручок.

В анализе экспрессии дополнительный интерес представляют временные серии органов и тканей, позволяющие изучать динамику транскриптома и конструировать сети генетической регуляции (Bar-Joseph et al., 2012). Временные серии могут собираться последовательно день за днем или единомоментно, если органы на разных стадиях развития присутствуют на растении одновременно, как в случае цветков. К первым в нашей транскриптомной карте относились части третьего розеточного листа, а также серия апикальных меристем побега от вегетативной стадии семидневного проростка до развития примордиев цветков (подробнее см. Главу 2). Одновременно собирались такие органы, как цветки, семена разной степени созревания, стручки вместе с семенами и оболочки стручков без семян.

Выделение тотальной РНК, приготовление библиотек, секвенирование и первичная обработка данных были проведены в Лаборатории эволюционной геномики (см. раздел 1.3.2 и

Материалы и Методы). Нами было показано, что размера библиотеки в 20 млн чтений достаточно для обнаружения более 95% экспрессирующихся генов и 86% дифференциально экспрессирующихся генов по сравнению с результатами, получаемыми при 40 млн чтений (Рисунок 1.10, неопубликованные данные). Поэтому для образцов транскриптомной карты было получено не менее 17 млн (23 млн в среднем) уникально картированных чтений, что позволяет проводить анализ экспрессии подавляющего большинства генов.



**Рисунок 1.10.** Влияние размера библиотеки на число генов, транскрипты которых обнаруживаются в образце при двух порогах (5 и более или 16 и более чтений в каждой из биологической повторностей), и на количество ДЭ генов. Для проведения этой проверки нами были собраны в двух биологических повторностях образцы соцветий дикого типа *A. thaliana* и мутанта *ap3-6* на стадии распускания первого, базального цветка. Выделенная из образцов РНК была использована для приготовления библиотек, секвенированных на платформе Illumina. Для каждого образца было получено 40 миллионов чтений, из которых были сделаны случайные подвыборки раннего размера. Для каждой подвыборки были определены число экспрессирующихся генов и найдены ДЭ гены между мутантом и диким типом.

Квадрат коэффициента корреляции Пирсона подтвердил пригодность биологических повторностей для дальнейшего анализа; кроме того,  $r^2$  использовался для построения дерева кластеризации образцов. Как видно на Рисунке 1.3, кластеризация отражает анатомию растения, и образцы преимущественно группируются по морфологическому и физиологическому сходству. Так, отдельные клады образуют органы цветков, части листьев, оболочки стручков без семян, меристемы, корни, цветки, прорастающие семена, незрелые семена, созревающие семена и стареющие органы. Преимущественная кластеризация образцов по органам, а не по возрасту растения, с которого были собраны образцы, была показана и для других растений, например, ячменя, кукурузы и риса (Druka et al., 2006; Jiao et al., 2009; Wang et al., 2010; Stelpflug et al., 2016).

Единственным исключением было объединение жилки и черешка стареющего (желтеющего и засыхающего) листа со стареющим междоузлем, что также отмечалось в атласе *A. thaliana*, созданном с помощью микрочипов (Schmid et al., 2005); этот факт объясняется преобладанием процессов, связанных угасанием жизнедеятельности, над органоспецифическими.

Среди собранных нами образцов дальше всех отстояли зрелые пыльники после вскрытия (F.AN, собранные с первого, базального, цветка в момент полного распускания) и зрелые пыльники перед вскрытием (F.AN.ad, собранные с третьего цветка на оси соцветия в момент полного распускания первого цветка). Аналогичный результат был получен при анализе транскриптомной карты таких эволюционно далеких от *A. thaliana* растений, как рис и кукуруза (Nobuta et al., 2007; Wang et al., 2010; Stelpflug et al., 2016); кроме того, уникальные транскриптомные характеристики пыльцы отмечались в ранних работах, посвященных широкомасштабному анализу экспрессии этого органа (Becker, 2003).

Ряд образцов нашей транскриптомной карты обладал чрезвычайно схожими экспрессионными профилями: минимальный  $r^2$  соседних образцов временной серии апикальных меристем (M1, M2, M3 и M6, M7, M8) составил 0,9, а в серии незрелых семян – 0,92 (SD.y1, SD.y2 и SD.y4, SD.y5, OV.y6-7) (Klepikova et al., 2016), Таблица S3 Приложения). Соответственно, такие образцы дублируют друг друга в отношении большинства транскриптомных анализов, и, при отсутствии задачи создания как можно более полной транскриптомной карты или изучения временных серий, могут быть без потерь исключены из плана сбора.

#### 1.4.2. Интегральные параметры транскриптома *A. thaliana*

Прежде всего следует отметить, что описанные выше (см. раздел 1.3.2) технические ограничения метода – отбор фракции полиаденилированной РНК при приготовлении библиотек для секвенирования – позволяют проводить точную оценку экспрессии только белок-кодирующих генов, поэтому большинство анализов в этом разделе ограничиваются ими.

Самым простым способом сравнительного описания транскриптомов является подсчет числа экспрессирующихся генов. Так как наши данные обладали достаточным размером библиотек (см. раздел 1.4.1), мы использовали «строгий» порог экспрессируемости в 16 нормализованных чтений в каждой из биологических повторностей для снижения числа ложноположительных результатов (Su et al., 2014).

90% (24 621 из 27 201) генов, кодирующих мРНК, экспрессировались хотя бы в одном из образцов, что оказалось несколько выше, нежели наиболее характерные для других транскриптомных карт значения в 75-85%, однако если рассматривать все гены *A. thaliana*, то это число снизится до 77% (25 706 из 33 323 генов). Во всех 79 образцах нашей транскриптомной карты экспрессировалось 10 654 гена, что составляет 40% от белок-кодирующих генов или же

треть от всех генов, что практически не отличается от значений, полученных в других атласах (42-50%). Анализ количества генов, экспрессирующихся в каждом из образцов в отдельности, показал, что разница между образцами с минимальным и максимальным числом экспрессирующихся генов составляет 20% (4 088 генов), что согласуется с результатами, полученными при изучении других видов. Доля экспрессирующихся генов составляла от 57% в образце M1 (15 525 генов, вегетативная меристема) до 72% в F3 (19 613 генов, третий цветок при полном распускании первого цветка) (Приложение, Рисунок 1.1), что также близко к диапазону 50-70%, полученному при анализе опубликованных транскриптомных карт. Следует отметить, что в некоторых работах отмечены наименее транскрипционно активные образцы: стручок сои (52,4%, (Libault et al., 2010b), семена *M. truncatula* (55,2% и 50,8%, (Benedito et al., 2008; Verdier et al., 2013), эндосперм риса (51,8%, (Wang et al., 2010). В нашей работе число экспрессирующихся генов в большинстве образцов незрелых и созревающих семян не отличалось от среднего (17 906), однако сухие, желтеющие и зрелые зеленые семена обладали минимальным числом экспрессирующихся генов (57,8-60%), уступая только меристемам (Приложение, Рисунок 1.1).

Наиболее общим описанием тканеспецифичности экспрессии гена может служить число образцов, в которых экспрессируется ген. Большинство генов (15 296) экспрессируются практически во всех образцах, относясь, таким образом, к необходимым для базовой жизнедеятельности клеток и тканей, в то время как вторая по размеру группа генов транскрибируется в 15 и менее образцах, являясь тканеспецифичной (Рисунок 1.4 а). Схожее наблюдение о преобладании широко экспрессирующихся генов было сделано в предыдущей карте *A. thaliana*, а также в атласе *C. sativa* (Schmid et al., 2005; Kagale et al., 2016). Гены, экспрессирующиеся в большем числе образцов, характеризуются более высокой экспрессией (Рисунок 1.4 б), тогда как тканеспецифические гены имеют меньшую медиану уровня экспрессии, что также свидетельствует о структурной роли первых и регуляторной – вторых.

Для проверки полноты обнаружения генов, способных к транскрипции, мы использовали полученные ранее в лаборатории три временные серии листьев *A. thaliana*, подвергнутых воздействию абиотических стрессов: холодового (+4°C), теплового (+42°C) и механического (повреждение листьев с помощью иглы). Нами были отобраны гены, экспрессирующиеся хотя бы в одном из образцов проверочного набора данных (все три временные серии рассматривались совместно); сравнение списка этих генов с экспрессирующимися в транскриптомной карте показало, что только 26 генов, кодирующих мРНК (и 96 генов всего), экспрессируются исключительно в стрессовых данных. Таким образом, 79 образцов нашей транскриптомной карты достаточно для детектирования подавляющего большинства (99,9%) генов.

Транскрипты 7 617 генов не обнаруживались ни в транскриптомной карте, ни в дополнительных стрессовых данных; в их число входили 4 084 псевдогена, 2 580 белок-кодирующих генов, 628 – тРНК и 325 генов с иной аннотацией (такие, как микроРНК и рРНК). Гены, кодирующие мРНК, обогащены таким категориями ГО, как «защитный ответ на грибы» (GO:0050832), «убийство клеток других организмов» (GO:0031640), «РНК-зависимая активность ДНК-полимеразы» (GO:0003964) (Klerikova et al., 2016), Таблица Приложения S9). Перепредставленность таких категорий у неэкспрессирующихся генов может быть связана с отсутствием среди использованных нами данных образцов, подвергнутых воздействию биотического стресса.

Сравнение нашей транскриптомной карты *A. thaliana*, созданной с помощью секвенирования РНК, с более ранним атласом, основанном на результатах микрочипов (Schmid et al., 2005), показывает, что секвенирование позволяет обнаружить большинство (94%) экспрессирующихся генов. При этом транскрипты 5 877 генов детектируются исключительно в нашем атласе, что, как и в случае атласов кукурузы (Sekhon et al., 2013), говорит о важности создания транскриптомных карт современными методами анализа экспрессии даже для видов, уже изучавшихся ранее.

#### 1.4.3. Схожесть профилей экспрессии органов и тканей

В предыдущем разделе мы отмечали сходство числа экспрессирующихся генов во всех образцах транскриптомной карты. Однако уровни транскрипции многих генов существенно различаются между образцами (подробнее см. раздел 1.4.4); таким образом, профили экспрессии различных образцов, то есть соотношение уровней экспрессии всех генов в образце, имеют свои особенности (Schmid et al., 2005). Для описания профиля экспрессии традиционно используется параметр  $Z$ , имеющий стандартное распределение. Если изучаемый образец похож на «усредненный», то его распределение параметра  $Z$  стремится к стандартному. Большинство образцов нашей транскриптомной карты имело симметричное или в небольшой степени смещенное влево распределение параметра  $Z$  с медианой, близкой к 0 (Рисунок 1.5 и Приложение, Рисунок 1.3). Существенное смещение распределения влево объясняется низким уровнем экспрессии большинства генов в образце (как в случае желтеющих семян, SD.sn1) или присутствием высокоэкспрессируемых органоспецифических генов (как гены *ABCG20* и *ATEXP24* в пыльниках F.AN и F.AN.ad) (Рисунок 1.5). Следует также отметить, что мода параметра  $Z$  в пыльниках, несмотря на асимметрию и смещение распределения, близка к нулю. Однако в результатах, полученных с помощью микрочипов, пыльца имела практически равномерное распределение (Schmid et al., 2005), что, как и низкое число экспрессирующихся генов, является следствием отсутствия проб к пыльцеспецифическим генам. Более

транскрипционно активные образцы, такие как меристемы (M1-10), семязпочки до оплодотворения (OV6-7), молодые семена (SD.y1), семязпочки, имеют смещенное вправо распределение параметра Z (Приложение, Рисунок 1.3). В нашей транскриптомной карте цветки обладали наиболее симметричным распределением, схожим со стандартным, как и в атласе *M. truncatula* (Benedito et al., 2008), однако в карте экспрессии сои цветки имели бимодальное распределение параметра Z (Severin et al., 2010), что говорит о видоспецифичности профилей экспрессии образцов.

#### 1.4.4. Различия профилей экспрессии образцов и разнообразие экспрессии генов

Для дальнейшего изучения разнообразия тканей и органов *A. thaliana* мы провели поиск генов, экспрессия которых статистически значимо различается между образцами, то есть проанализировали дифференциальную экспрессию во всех возможных парах образцов (3 081 сравнений).

По числу дифференциально экспрессирующихся генов образцы M1 и M2 (меристемы до перехода к цветению) оказались наиболее похожими друг на друга (14 ДЭ генов), тогда как уровни экспрессии 15 149 генов отличались между F3 и SD.d (третий цветок при распускании первого и сухие семена), то есть число ДЭ генов варьировало в очень широких пределах. Как и ожидалось, между числом ДЭ генов и корреляцией  $r^2$  прослеживалось соответствие: чем выше число ДЭ генов, тем ниже корреляция между этими образцами. Как и коэффициент корреляции, число ДЭ генов использовалось в качестве меры расстояния между образцами для построения иерархической кластеризации (Приложение, Рисунок 1.4). Несмотря на ряд отличий в структуре дерева кластеризации, группировка образцов по морфологическому и физиологическому сходству сохраняется и даже становится более отчетливой: клады формируют пыльники, стареющие органы, меристемы и обогащенные меристематическим тканями образцы (такие, как F.SA.y - пестик цветка на стадии 9 по Smyth et al. (Smyth et al., 1990), семена, корни, цветки, листья разных стадий).

Гены, дифференциально экспрессирующиеся между двумя образцами, отражают разницу протекания в них биологических процессов. Изучение аннотации ДЭ генов категориями Генной Онтологии и других баз данных позволяет получить общее представление о функциях этих генов, а также о процессах, в которых они участвуют. Мы проанализировали перепредставленность терминов ГО во всех парных сравнениях, отдельно рассматривая гены, экспрессия которых повышена или понижена в первом образце по сравнению со вторым. Большое количество обогащенных категорий (1 528 – ГО и 1 531 – других баз данных) свидетельствует о разнообразии включенных в транскриптомную карту образцов (Klepikova et al., 2016), Таблица S11 Приложения). Наиболее часто встречались термины, связанные с фотосинтетическими

процессами, что отражает разницу профилей экспрессии зеленых и незеленых частей растения. Также были перепредставлены категории, ассоциированные с регуляцией хроматина и делением клеток, что вызвано отличиями между быстро развивающимися и зрелыми органами.

Помимо разницы между образцами, анализ дифференциальной экспрессии позволяет изучать профили экспрессии генов, выделяя те пробы, между которыми число транскриптов генов значимо различаются. Чтобы оценить разнообразие экспрессии генов в транскриптомной карте, мы использовали меру «ДЭ-значение гена» – число парных сравнений, в которых данный ген являлся ДЭ. В соответствии с числом парных сравнений, наибольшее ДЭ-значение может быть 3 081, однако максимальное ДЭ-значение в наших данных составило 2 533. Распределение ДЭ-значений (Рисунок 1.6) показывает, что 50% генов находятся в диапазоне от 439 до 1 528, при этом мода распределения составляет 229.

Гены, экспрессирующиеся в небольшом числе образцов  $N$ , потенциально могут иметь ДЭ-значение, равное  $(79 - N) * N$ . Например, ген, специфически экспрессирующийся в одном образце, может иметь ДЭ-значение 78 (ген *ATIG33550*, имеющий ДЭ-значение 78, экспрессируется уникально в зрелых пыльниках). Таким образом, гены с узким паттерном экспрессии имеют небольшие ДЭ-значения, и составляют, наряду со стабильно экспрессирующимися генами, основную часть распределения с ДЭ-значением ниже 100 (Приложение, Рисунок 1.5). Эти гены имели обогащение категориями ГО, связанными с базовыми биологическими процессами, такими, как процессинг РНК, транспорт и локализация белков (Приложение, Таблица 1.4).

Гены с высокими ДЭ-значениями должны экспрессироваться в большом числе образцов и одновременно иметь статистически значимую разницу в уровнях транскрипции между ними; ДЭ-значения выше 2 300 характеризовали 331 ген. Следует отметить, что большинство этих генов имели тривиальные названия (64% по сравнению с 30% среди всех генов) и, таким образом, были ранее изучены в работах, посвященных функциональному анализу, или входили в известные семейства генов. Многие из генов с высокими ДЭ-значениями были вовлечены в регуляцию и функционирование фотосинтеза (Приложение, Таблица 1.3), что соответствует разнице между активно фотосинтезирующими (листья и междоузлия) и незелеными (семена, корни) образцами транскриптомной карты.

#### *1.4.5. Стабильно экспрессирующиеся гены*

Поиск генов, обладающих равномерным уровнем экспрессии в разных органах и тканях растения, необходим для корректного сравнения экспрессии генов между образцами с помощью ОТ-ПЦР в реальном времени (Kozera and Rapacz, 2013). Первоначально предполагалось, что такие гены должны быть вовлечены в необходимые для жизнедеятельности клетки процессы и, соответственно, не подвергаться тканеспецифической регуляции, благодаря чему их назвали

генами домашнего хозяйства. Первые работы, посвященные поиску референсов, рассматривали гены, кодирующие актины, тубулины, 18S РНК, глицеральдегид-3-фосфат дегидрогеназу (Thellin et al., 1999), однако в дальнейшем с применением массовых транскриптомных данных было показано, что эти гены имеют большую дисперсию экспрессии в разных тканях и органах (Czechowski, 2005). С другой стороны, функции наиболее стабильных генов, идентифицированных в работе Czechowski et al., по-прежнему остаются не до конца исследованными (как в случае генов *AT4G33380* и *AT4G34270*), также остается неясным какие клеточные процессы требуют равномерной экспрессии участвующих в них генов во всем организме.

Для поиска стабильно экспрессирующихся генов мы использовали коэффициент вариации; с его помощью были отобраны 970 наиболее стабильно экспрессирующихся генов ( $CV < 0,3$ ). Стабильность экспрессии этих генов также подтверждается низкими (от 0 до 488) ДЭ-значениями, а присутствие их транскриптов в всех образцах – высокими  $H$  (4,56-4,57). Обогащение терминами ГО у стабильных, наименее дифференциально и наиболее широко экспрессирующихся генов частично перекрывается (Приложение, Таблицы 1.6, 1.5, 1.7); в число перепредставленных категорий входят «процессинг мРНК», «внутриклеточный транспорт» и «транспорт с помощью везикул».

Таким образом, с помощью транскриптомной карты мы отобрали гены, имеющие равномерный уровень экспрессии в широком спектре процессов развития; чтобы проверить, изменяется ли экспрессия этих генов в других условиях выращивания, мы проанализировали набор данных, объединяющий атлас экспрессии и три временные серии воздействия абиотических стрессов. Добавление этих данных привело к падению числа стабильных генов до 792 (при пороге  $CV \leq 0,3$ ). Используя как исключительно транскриптомную карту, так и объединенные данные, мы отобрали 27 наиболее стабильных генов ( $CV \leq 0,2$ , Рисунок 1.7). ДЭ-значения этих генов лежали в диапазоне от 0 до 117 (с медианой 15), показывая минимальные значимые различия в экспрессии, при этом следует отметить, что изменение уровней экспрессии между пробами для них не превышало четырех раз (медиана – 2,3), а энтропия Шеннона равнялась 4,56 или 4,57 (максимальная для наших данных), что подтверждает равномерность их экспрессии в транскриптомной карте.

Описание этих генов в базе данных TAIR ([arabidopsis.org](http://arabidopsis.org)) показывает, что они связаны с самыми разными процессами, включая цветение, транспорт лизосом, конденсацию хромосом и адаптацию к стрессу, и, таким образом, могут служить референсными генами в широком спектре экспериментов по измерению экспрессии генов с помощью ОТ-ПЦР в реальном времени. В работе Czechowski et al. был определен набор генов *A. thaliana*, стабильно экспрессирующихся в существовавших на тот момент транскриптомных данных (Czechowski, 2005). Один из этих генов

(AT4G34270) присутствует среди найденных нами 27 самых стабильных генов, а остальные характеризуются коэффициентом вариации меньше 0,3 в объединенном наборе данных. Что касается «традиционных» референсных генов, только ген *UBC* (AT5G25760, определенный также как самый стабильный классический референс в работе Czechowski et al. и в других видах растений (Benedito et al., 2008; Wang et al., 2010; Meyer et al., 2012; Verdier et al., 2013) имел CV меньше 0,3, для остальных же мера стабильности лежала в диапазоне от 0,41 до 0,96.

#### 1.4.6. Паттерны экспрессии генов

Гены, экспрессирующиеся во всех органах и тканях организма, в том числе стабильные гены, составляют структурную и регуляторную основу необходимых для жизнедеятельности процессов (Eisenberg and Levanon, 2013). С другой стороны, развитие из тотипотентной зиготы всего многообразия тканей и органов требует пространственно ограниченной, тканеспецифичной экспрессии генов, определяющих уникальные характеристики типов клеток. В соответствии с этим, изучение тканевой специфичности экспрессии генов является одной из важных задач транскриптомики, решение которой ведет к пониманию онтогенеза тканей и органов. Кроме того, клеточно- или тканеспецифические гены применяются в ряде прикладных задач, таких как клеточный сортинг (Bargmann and Birnbaum, 2010) или создание узконаправленных трансгенных конструкций (Sekhon et al., 2011).

Подробные транскриптомные карты служат мощным инструментом для выявления генов, экспрессия которых ограничивается одним типом клеток или тканей. Для определения ширины паттерна транскрипции гена мы использовали энтропию Шеннона. Распределение  $H$  (Рисунок 1.8) показывает, что транскрипты большинства генов обнаруживаются во всех или почти во всех образцах (пик в области высоких значений энтропии). Схожий результат был получен в работе Li et al., где использовался анализ коэффициента эксцесса распределения экспрессии для идентификации генов, экспрессирующихся в двух и более, но не во всех тканях *A. thaliana* и других организмов (Li et al., 2012). Эти гены являются основой жизнедеятельности всех типов клеток и участвуют в таких процессах, как транспорт нуклеиновых кислот, метаболизм липопротеинов и мембран.

Распределение энтропии Шеннона, полученное на основе наших данных, повторяет форму распределения  $H$  кукурузы (Sekhon et al., 2013), что свидетельствует о межвидовой универсальности соотношения широко экспрессирующихся и тканеспецифичных генов. Так, оба распределения обладают вторым локальным максимумом в области низких значений  $H$ . У *A. thaliana* гены, составляющие этот пик, участвуют в клеточном сигналинге, формировании эндомембран и клеточной стенки; кроме того, продукты многих из них имеют пектинэстеразную

активность, что соответствует большому числу генов, уникально экспрессирующихся в пыльниках.

#### 1.4.7. Разнообразие паттернов экспрессии транскрипционных факторов

Как среди универсально экспрессирующихся генов, так и тканеспецифических особое место занимают гены, кодирующие транскрипционные факторы – белки, имеющие ДНК-связывающий домен и способные, благодаря связыванию с определенными участками в промоторах других генов, регулировать их экспрессию. Изучение паттернов экспрессии ТФ необходимо для построения сетей генетической регуляции разнообразных процессов, протекающих в организме, а для ранее неизученных регуляторных генов такой анализ может дать новую информацию об их функции.

Геном *A. thaliana* содержит 2 757 ТФ, относящихся к 82 семействам (Jin et al., 2014, 2015). Распределение энтропии Шеннона для 2 148 ТФ, транскрипты которых были обнаружены в наших данных, представлено на Рисунке 1.9 а.

Наименьшая медиана энтропии характеризовала такие классы ТФ, как MADS, LOB, LIM и MYB; узкий паттерн экспрессии, ограниченный некоторыми органами или стадиями развития, ранее был показан для некоторых из этих ТФ, среди которых есть гены, участвующие в контроле развития органов цветка, листьев и перехода к цветению (Ng and Yanofsky, 2001). На другом конце распределения находятся семейства SWI/SNF SWI3, SNF2, CAMTA, DDT и FAR, которые участвуют в таких универсальных клеточных процессах, как ремоделирование хроматина, репарация ДНК, сигналинг и ответ на стресс (Jerzmanowski, 2007; Lin et al., 2007).

MADS-box-содержащие гены, вероятно, самое изученное семейство ТФ растений; для определения их функций и эволюции было использовано множество подходов, включая характеристику мутантов и трансгенных растений, полногеномное секвенирование и анализ экспрессии и репортерных генов (Ng and Yanofsky, 2001). Однако даже для такого хорошо изученного семейства детализированная транскриптомная карта способна дать новую информацию о паттернах экспрессии генов, таких как *AGL97* (*AT1G46408*) и *AGL52* (*AT4G11250*), специфичных для пыльцы, или *AGL51*, который экспрессируется в черешках и междоузлиях.

Другое семейство с низкой энтропией – LBD (содержащее LATERAL ORGAN BOUNDARIES (LOB)-домен), в состав которого входят 45 генов. Среди них наиболее изучен ген *AS2*, участвующий в контроле развития листа и определяющий его полярность (Semiarti et al., 2001; Xu et al., 2003) путем подавления гомеобоксных генов семейства *KNOX* (Lin, 2003). Продукт гена *AS2* также действует в органах цветка, где его функция частично дублируется геном *ASL1* (*AT5G66870*) из семейства LOB (Chalfun-Junior et al., 2005). Взаимодублирование функций

генов этого семейства предполагалось ранее, хотя анализ белков LBD показал отсутствие взаимозаменяемости между доменом LOB белка AS2 и других белков, то есть уровень дублирования по крайней мере между этим геном и остальными ограничен (Matsumura et al., 2009). В соответствии с этим результатом паттерны экспрессии генов LBD значительно отличались: некоторые из них экспрессировались в большинстве образцов транскриптомной карты (например, *LBD39*, *LBD37* и *LBD11*), тогда как другие демонстрировали узкий паттерн (*LBD10*, *LBD2* и *LBD20*), вплоть до единственного органа: так, экспрессия *AT2G31310* (*LBD14*) ограничена корнями, а *AT3G50510* (*LBD28*) и *AT3G13850* (*LBD22*) – пыльцой.

Экспрессия ТФ одного семейства может коррелировать в разных тканях и органах, если их продукты формируют белковые комплексы, как компоненты фотосистем, белки субъединиц рибосом или «флоральные квартеты» – комплексы MADS-box белков, регулирующих тип органа цветка (Honma and Goto, 2001). Однако распределение корреляции Спирмена внутри классов генов (Рисунок 1.9 б) не показывает наличия координированной экспрессии: корреляция большинства ТФ в пределах одного семейства близка к нулю. Низкая координированность экспрессии ТФ одного класса ранее наблюдалась в условиях стресса (Chen et al., 2002) и, с помощью наших данных, была подтверждена для развития различных тканей и органов *A. thaliana*.

#### 1.4.8. Создание экспрессионной базы данных TraVA

Созданная нами транскриптомная карта *A. thaliana* послужила основой для публично доступной базы данных Transcriptome Variation Analysis (TraVA, [travadb.org](http://travadb.org)). В ней представлены уровни экспрессии всех аннотированных генов *A. thaliana* в 79 образцах нашего атласа, а также для каждого гена приведены результаты анализа дифференциальной экспрессии между всеми парами образцов. Визуализация уровней экспрессии поддерживается как для одного гена, так и в мультигенном формате, кроме того, пользователь может выбрать тип нормализации на размер библиотеки, настроить цветовую кодировку уровней экспрессии и выбрать инструмент для анализа дифференциальной экспрессии.

К марту 2019-го года посещаемость базы данных TraVA достигла 4 000 уникальных пользователей в год; на ее основе организован ряд заданий учебного курса «Plant Bioinformatics» (<https://ru.coursera.org/learn/plant-bioinformatics>); TraVA интегрирована в один из основных интернет-ресурсов генетики и геномики *A. thaliana* – базу данных TAIR.

## 1.5. Заключение

Наша работа была посвящена созданию транскриптомной карты *A. thaliana* с помощью секвенирования РНК. В состав карты вошло 79 образцов различных частей и стадий развития корней, листьев, цветков, семян и стручков, что делает ее наравне с атласом экспрессии генов кукурузы (Stelpflug et al., 2016) и первой транскриптомной картой растения (Schmid et al., 2005) самым высокоразрешенным набором экспрессионных данных. Большое разнообразие исследованных органов и тканей позволило детектировать транскрипты подавляющего большинства белок-кодирующих генов; следует отметить, что добавление к транскриптомной карте образцов, подвергнутых действию абиотических стрессов, не добавило большого количества экспрессирующихся генов. Гены, экспрессия которых не была обнаружена, чаще были связаны с реакцией на биотические стрессы и регулировали защиту от грибов и вирусов, что отражает отсутствие биотического стресса среди изучаемых нами данных и говорит о крайней специфичности функции таких генов.

Сравнение общих экспрессионных характеристик (числа экспрессирующихся генов во всех образцах и хотя бы в одном образце, профили экспрессии образцов) транскриптомной карты *A. thaliana* и других растений показало высокую консервативность глобального состояния транскриптома в разных таксонах. Мы определили гены, наиболее стабильно экспрессирующиеся как в органах и тканях *A. thaliana*, так и под действием абиотических стрессов; эти гены были обогащены категориями ГО, связанными с транспортом нуклеиновых кислот и мембранами.

Чтобы найти новые транскрипционные регуляторы, мы проанализировали экспрессионные паттерны генов, кодирующих ТФ, с помощью энтропии Шеннона. Гены, регулирующие «местные» биологические процессы – переход к цветению, определение места закладки латеральных органов, развитие корней – имели наименьшую медианную энтропию, что было характерно для семейств транскрипционных факторов MADS, LBD, MYB; напротив, SWI/SNF-SWI3 и SNF2, регулирующие такие универсальные процессы, как упаковка хроматина, обладали близкой к максимальной энтропией. Изучение паттернов транскрипции ранее неисследованных ТФ позволило обнаружить тканеспецифические функции некоторых из них, что было показано нами на примере гена *LBD14*, уникально экспрессирующегося в корнях, и *LBD22* и *LBD28*, специфичных для пыльцы.

Все образцы транскриптомной карты имели как схожие профили экспрессии, так и одинаковое распределение категорий ГО, что подчеркивает универсальность транскриптома во всех органах и тканях растения и позволяет предположить, что их функциональная

специализация возникает скорее под действием относительно небольшого числа генов, нежели по причине радикальных изменений во всем транскриптоме.

Экспрессионные данные созданной нами транскриптомной карты представляют собой важный ресурс для научного сообщества и доступны в широко используемой публичной базе данных TraVA (<http://travadb.org/>).

## ГЛАВА 2. ВРЕМЕННАЯ СЕРИЯ АПИКАЛЬНЫХ МЕРИСТЕМ *ARABIDOPSIS THALIANA*

### 2.1. Введение

Рассматриваемая нами в Главе 1 транскриптомная карта *A. thaliana* включает в себя 79 образцов различных органов и частей растения; отдельное внимание мы уделяли сбору образцов, организованных во временные серии (цветки, листья, семена), что позволяет анализировать динамику процесса развития органа. Особенно примечательна временная серия апикальной меристемы побега, так как в этой структуре растения, содержащей пул стволовых клеток, происходят основные события перехода к цветению. Зацветание *A. thaliana* инициируется внешними и внутренними сигналами, приводящими к активации экспрессии регуляторных генов в апикальной меристеме, после чего меристема чрезвычайно быстро переходит из вегетативной стадии (закладывание примордиев листьев) в репродуктивное (формирование примордиев цветков). Переход к цветению у разнообразных растений подробно изучался с помощью методов традиционной генетики, что позволило определить ключевые активаторы и репрессоры, построить сеть их генетических взаимодействий и описать разные пути регуляции их функции. С другой стороны, изменения транскриптома апикальной меристемы, следующие за активацией положительных регуляторов зацветания, описаны недостаточно (Torti et al., 2012; Zhang et al., 2015).

Временная серия апикальных меристем собиралась нами с седьмого дня после прорастания, когда розетка *A. thaliana* содержит только шесть листьев, до появления бутонов цветков (стадия 5 по (Smyth et al., 1990) и, таким образом, покрывала момент перехода цветения. Общий обзор профилей экспрессии генов, проведенный при анализе транскриптомной карты, показал, что транскриптомы апикальных меристем на разных стадиях существенно отличаются. Исходя из этого, мы подробно проанализировали экспрессию генов во временной серии апикальных меристем в процессе перехода к цветению, чему посвящена Глава 2. Обзор литературы Главы 2 сосредоточен на фотопериодической регуляции зацветания и кратко затрагивает другие способы его инициации; в разделе 2.3 описаны результаты экспрессионного анализа временной серии АМ; сделанные на их основании выводы обсуждаются в разделе 2.4.

## 2.2. Обзор литературы

### 2.2.1. Важность перехода к цветению для жизни растений

Зацветание – переход всех или части апикальных меристем побега к формированию цветков – является крайне важной стадией в жизни растения, определяя его репродуктивный успех. Для однолетних растений, таких, как *A. thaliana*, своевременный переход к цветению особенно важен, потому что переключение типа развития апикальной меристемы (АМ) с вегетативного на репродуктивное происходит однократно и необратимо. Таким образом, цветение должно прийтись на наиболее благоприятный период года, и выбор времени зацветания происходит с учетом множества факторов окружающей среды. Среди них длина светового дня (фотопериод) – наиболее стабильный, не меняющийся год от года признак наступления необходимых для растения условий (Song et al., 2013b), тогда как температура окружающей среды и внутренние физиологические сигналы растения отражают события конкретного сезона (Cheng et al., 2017; Antoniou-Kourouniotti et al., 2018).

Генетический контроль зацветания образует сложную интегрированную сеть, в которой действие различных внешних и внутренних стимулов на соответствующие рецепторы приводит к активации ключевых генов-регуляторов и обеспечивает своевременный переход к репродукции (Andrés and Coupland, 2012).

### 2.2.2. Фотопериодический путь инициации цветения

Для многих видов растений долгота дня (продолжительность суточного светового периода) имеет решающее значение для определения времени перехода к цветению. Для зацветания так называемых растений длинного дня (ДД) продолжительность светового периода должна превысить пороговое значение. Напротив, растения короткого дня (КД) зацветают при уменьшении продолжительности дня. В то же время существуют нейтральные к фотопериоду растения, на которых долгота дня не оказывает влияния. Способ индукции цветения с помощью фотопериода различается как между видами, так и между расами одного вида растений (Shrestha et al., 2014). Следует отметить, что изменение долготы светового периода в первую очередь служит стабильным свидетельством наступления благоприятных условий в отношении температуры и влажности (лета в высоких и средних широтах и зимы в тропическом климате); таким образом, фотопериод является глобальным регулятором времени инициации цветения, точная настройка которого происходит с помощью других стимулов, таких как температура или гормональная регуляция.

*A. thaliana* относится к факультативным растениям ДД, то есть переходит к цветению при превышении длиной дня порогового значения, но может зацвести и в условиях КД, хотя это требует большего времени (Turck et al., 2008; Giakountis et al., 2010).

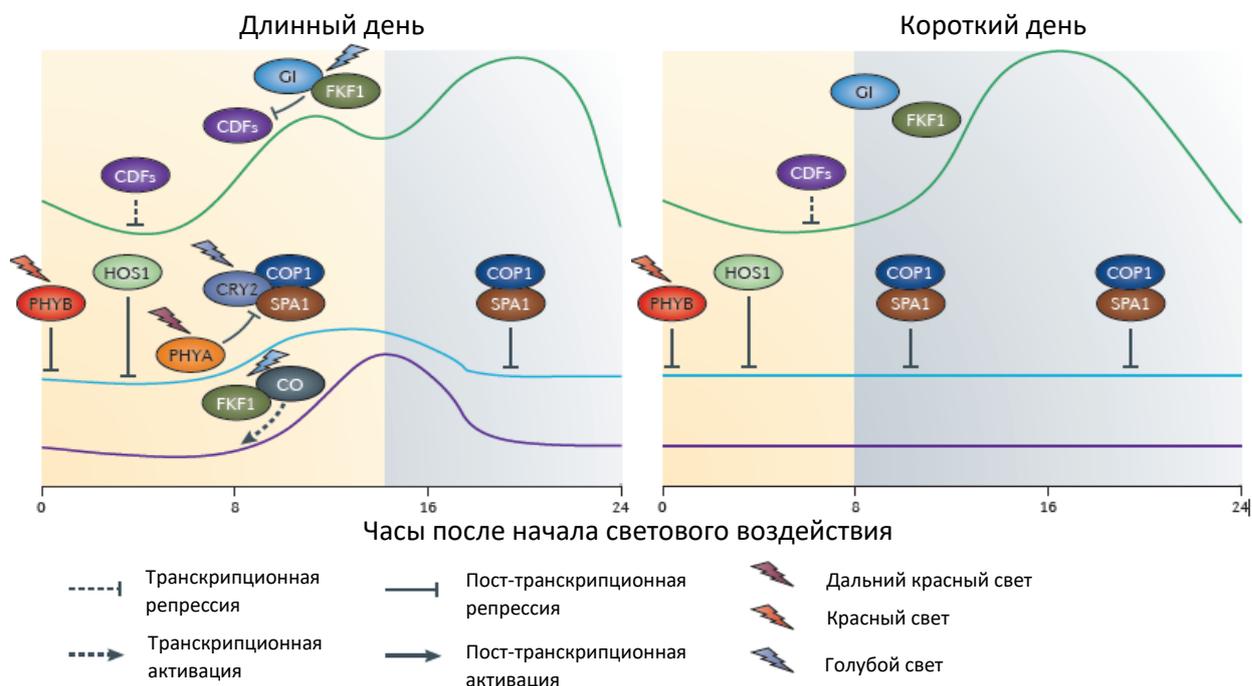
Определение изменений в долготе дня и развитие соответствующей реакции начинаются с рецепторных белков, воспринимающих свет. Растения обладают большим репертуаром фоторецепторов, поглощающих разные длины волн; в их число входят пять рецепторов к красному (630-760 нм) и дальнему красному (710-800 нм) свету, называемых фитохромами (PhyA-E) (Quail, 2002), семь рецепторов голубого света – два фототропина PHOT1 и PHOT2 (Briggs and Christie, 2002), два криптохрома CRY1 и CRY2) (Lin, 2002) и три белка, содержащих домен LOV/F-box/Kelch, FKF1 (FLAVIN BINDING, KELCH REPEAT, F-BOX PROTEIN 1), ZTL (ZEITLUPE), LKP2 (LOV KELCH PROTEIN 2) (Demarsy and Fankhauser, 2009), и один рецептор ультрафиолета-В UVR8 (Rizzini et al., 2011). Дифференциальная активация фоторецепторов в зависимости от «качества» света (соотношения длин волн) вызывает физиологический ответ растения – переход к цветению.

Главную роль в фотопериодической активации цветения играют гены *FKF1*, *GIGANTEA (GI)*, *CONSTANS (CO)* и *FLOWERING LOCUS T (FT)*, преимущественно экспрессирующиеся в сосудах листьев (Putterill et al., 1995; Fowler et al., 1999; Kardailsky et al., 1999; Kobayashi et al., 1999; Imaizumi et al., 2003). Специфический для растений белок GI и убиквитин-лигаза FKF1, являющаяся рецептором голубого света, входят в число компонентов циркадных часов, то есть обладают суточной периодичностью экспрессии с пиком во второй половине светового дня (Baudry et al., 2010; de Montaigu et al., 2010). После индукции голубым светом в течении ДД FKF1 взаимодействует с GI, что приводит к деградации репрессоров транскрипции гена *CO*, белков CYCLING DOF FACTOR (CDF) (Sawa et al., 2007; Song et al., 2012). Таким образом, в конце ДД, но не КД, экспрессия *CO* возрастает (Рисунок 2.1).

Помимо регуляции экспрессии, *CO* подвергается посттрансляционным модификациям. Белок *CO* убиквитинируется убиквитин-лигазным комплексом, включающим в себя CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) и SUPPRESSOR OF PHYTOCHROME A (SPA1), что приводит к деградации *CO* в 26S протеасоме (Laubinger et al., 2006; Liu et al., 2008b). Активность этого комплекса подавляется на свету с помощью FKF1-зависимого механизма (Lee et al., 2017), то есть деградация *CO* с помощью комплекса COP1-SPA1 в основном происходит в темноте. Таким образом, только пик количества мРНК гена *CO* в конце ДД после деградации репрессоров CDF с помощью GI-FKF1 приводит к накоплению белка *CO* (Рисунок 2.1).

Стабилизации *CO* в конце светового периода ДД также способствуют фоторецепторы PHYA и CRY2 (Briggs, 2001; Zuo et al., 2011). Светозависимое физическое взаимодействие CRY2 со SPA1 и COP1 уменьшает каталитическую активность COP1-SPA1 и препятствует деградации *CO* (Рисунок 2.1) (Zuo et al., 2011). Фитохромы синтезируются в неактивной форме (Pr) и под действием света переходят в активную форму (Pfr), которая способна к транспорту в ядро. PHYA также участвует в стабилизации *CO* с помощью подавления активности COP1-SPA1. С другой

стороны, PHYB с помощью белка PHL (PHYTOCHROME-DEPENDENT LATE-FLOWERING) связывает CO и репрессирует его активность (Endo et al., 2013).



**Рисунок 2.1.** Сеть транскрипционной и посттрансляционной регуляции экспрессии *CONSTANS*, контролирующая фотопериодический путь перехода к цветению. Адаптировано из (Andrés and Coupland, 2012).

Помимо комплекса COP1-SPA1, убиквитин-лигаза HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1 (HOS1) способствует снижению стабильности CO в начале светового дня, а также при понижении температуры окружающей среды (Рисунок 2.1, см. также раздел 3.2.4) (Lazaro et al., 2012).

### 2.2.3. Передача индуктивного сигнала в апикальную меристему

Фотопериодический путь активации цветения начинается в листьях, однако главные события перехода к репродуктивному развитию происходят в апикальной меристеме побега (Andrés and Coupland, 2012). Существование некоторого соединения, «флоригена», передающего сигнал из листа в апекс растения, предполагалось еще в начале XX века (Чайлахян, 1937). Таким агентом оказался белок FT, входящий в широко распространенное семейство CETS (CENTRORADIALIS, TERMINAL FLOWER 1 и FT) (Kardailsky et al., 1999; Kobayashi et al., 1999; An, 2004). Ген *FT* экспрессируется в клетках-спутницах ситовидных элементов флоэмы, а небольшой белок FT транспортируется из листа в АМ через флоэму при участии компонента плазмодесм клеток-спутниц FT-INTERACTING PROTEIN 1 (FTIP1) (Liu et al., 2012; Endo et al., 2018).

CO, транскрипционный фактор, содержащий два домена «цинковые пальцы» (типа В-box) и домен ССТ (CO, CO-LIKE и TOC1), активирует экспрессию гена *FT* (Tiwari et al., 2010). CO связывается с проксимальной частью промотора *FT* (Tiwari et al., 2010; Song et al., 2012); в CO-зависимой регуляции *FT* принимают участие и другие ТФ, такие как NUCLEAR FACTOR Y (Сао et al., 2014) или PICKLE (Jing et al., 2019). Сложная система транскрипционной и посттрансляционной регуляции CO (см. раздел 2.2.2) обеспечивает активацию транскрипции *FT* только в условиях ДД (Рисунок 2.1).

Помимо CO экспрессия *FT* напрямую инициируется компонентами фотопериодического пути GI, FKF1 и CDF1 (Sawa et al., 2007; Sawa and Kay, 2011; Song et al., 2012), ТФ семейства bHLH PHYTOCHROME-INTERACTING FACTOR 4 (PIF4), интегрирующим температурную активацию и фотопериод (Kumar et al., 2012), и CIB1, связывающий проксимальную часть промотора *FT* в ответ на голубой свет (Liu et al., 2008a).

Репрессия транскрипции *FT* также вовлекает большое количество белков. Хромодоменный белок LIKE HETEROCHROMATIN PROTEIN1 (LHP1), связывается с гистонем H3, несущим метку репрессированного хроматина (триметилирование по лизину 27, H3K27me3), в локусе *FT*, что приводит к его инактивации (Turck et al., 2007). Схожий механизм задействован в репрессии *FT* с помощью H3K27me3 деметилазы JUMONJI 13 (Zheng et al., 2019). MADS-доменные белки SHORT VEGETATIVE PHASE (SVP) и FLOWERING LOCUS C (FLC), участвующие в температурной регуляции зацветания (см. раздел 2.2.5), связываются с последовательностями CArG в промоторе и первом интроне *FT* (Helliwell et al., 2006; Lee et al., 2007), негативно регулируя его транскрипцию. ТФ семейства AP2 TEMPRANILLO1 (TEM1) и TEM2, TARGET OF EAT 1 (TOE1), TOE2, SCHNARCHZAPFEN (SNZ) и SCHLAFMÜTZE (SMZ) также функционируют как репрессоры (Castillejo and Pelaz, 2008; Mathieu et al., 2009).

Таким образом, разветвленная генетическая сеть регуляции фотопериодического пути инициации цветения обладает двумя основными точками конвергенции сигнальных путей – генами *CO* и *FT* и обеспечивает точную физиологическую реакцию на изменение долготы дня.

#### 2.2.4. Основные события, происходящие в апикальной меристеме

После транспорта в АМ *FT* взаимодействует с FLOWERING LOCUS D (FD), ТФ из семейства basic leucine zipper (bZIP), что приводит к активации экспрессии генов, обеспечивающих переход к репродуктивному развитию (Andrés and Coupland, 2012). Первым после переноса фотопериодического сигнала в АМ активируется ген *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), кодирующий MADS-доменный ТФ (Borner et al., 2000; Samach et al., 2000). Экспрессия гена *SOC1* и его гомолога *FRUITFULL* (*FUL*) абсолютно необходима для зацветания (Melzer et al., 2008). *SOC1* совместно с еще одним ТФ семейства

MADS – AGL24, активирует экспрессию *LEAFY (LFY)*, гена, определяющего репродуктивное состояние АМ и вовлеченного в инициацию примордиев цветков (Lee et al., 2008a). Транскрипция генов *SOC1*, *FUL*, *LFY* и *APETALA1 (API)*, флоральный регулятор, частично дублирующий функцию *LFY* также индуцируется ТФ SQUAMOSA BINDING PROTEIN LIKE (SPL), являющихся мишенями FD (Yamaguchi et al., 2009; Jung et al., 2012). Гены *API* и *LFY* и ген *TERMINAL FLOWER1 (TFL1)* взаимно негативно регулируют друг друга: *TFL1* подавляет экспрессию *API* и *LFY* в центре АМ, где находится пул стволовых клеток, тогда как последние в свою очередь негативно регулируют экспрессию *TFL1* в примордиях цветков (Shannon, 1993; Liljgren et al., 1999). Примечательно, что *TFL1* является гомологом FT, а сходство их последовательностей достигает 60%, причем одна аминокислотная замена превращает функцию *TFL1* в FT и наоборот (Hanzawa et al., 2005; Ho and Weigel, 2014). FD в АМ выполняет двойную роль: взаимодействуя с *TFL1*, тормозит переход к цветению, а совместно с FT – активирует (Hanano and Goto, 2011).

#### 2.2.5. Другие пути инициации цветения

Инициация цветения с помощью изменения длины суточного светового периода занимает важное место в переходе к репродуктивному развитию, однако не является единственным способом его активации. Среди других путей можно назвать автономный, гормональный, вернализационный и температуро-зависимый.

Индукция цветения с помощью продолжительной обработки низкими температурами называется вернализацией. Разные виды растений, как и разные расы в пределах одного вида, различаются по требованиям к холодному воздействию: некоторым растениям необходим период пониженных температур для перехода к цветению при достижении оптимальной длины дня, зацветание других видов и рас не зависит от холодного периода (Wollenberg and Amasino, 2012). Основной механизм вернализации – эпигенетический сайленсинг экспрессии гена *FLC*, который кодирует MADS-доменный репрессор цветения (Michaels and Amasino, 1999; Sheldon et al., 1999). *FLC* напрямую связывается с промоторами генов фотопериодического пути – *FT*, *FD* и *SOC1*, а также с первым интроном FT, блокируя их транскрипцию (Helliwell et al., 2006; Searle et al., 2006).

Для активации экспрессии *FLC* необходима модификация хроматина по лизину 4 и лизину 36 гистона H3 (Crevillén and Dean, 2011). В процессе длительного холодного воздействия в нуклеосомах локуса *FLC* с помощью Polycomb repressive complex 2 (PRC2) и длинных некодирующих РНК *COOLAIR* и *COLD AIR* накапливается метилирование H3K27me3, что приводит к репрессии транскрипции *FLC* (Heo and Sung, 2011; Song et al., 2013a). Уровень

метиляции гистона H3 поддерживается в дальнейшем и после повышения температуры, обеспечивая стабильный сайленсинг *FLC* (Csorba et al., 2014).

Автономный путь инициации цветения также направлен на регуляцию экспрессии *FLC* с помощью пост-транскрипционных механизмов и эпигенетической модификации хроматина (Quesada et al., 2005; Bäurle and Dean, 2008). Основные регуляторы автономного пути – белки FCA, FLK, FPA, FY, PAPS1, PAPS2, PAPS4, PCFS4 и PEP (для обзора см. (Cheng et al., 2017)). Эти гены образуют сеть взаимной регуляции; основная активность их продуктов заключается в процессинге 3'-конца РНК (Rataj and Simpson, 2014). Благодаря этому белки автономного пути регулируют альтернативное полиаденилирование антисмысловых РНК и формирование 3'-конца РНК *FLC* (Liu et al., 2007).

Температура окружающей среды имеет крайне важное влияние как на рост и развитие растения, так и на его репродуктивный успех, поэтому в определении оптимального времени зацветания участвует термочувствительный путь (Carovilla et al., 2015). Температурная регуляция зацветания достигается с помощью включения в нуклеосомы гистонового варианта H2A.Z хроматин-ремоделирующими комплексами, в состав которых входят белки ACTIN-RELATED PROTEIN6 (ARP6) и PHOTOPERIOD-INSENSITIVE EARLY FLOWERING1 (PIE1) (Talbert and Henikoff, 2014). Гистон H2A.Z делает ДНК менее доступной для ТФ и замедляет работу РНК-полимеразы II, что ограничивает экспрессию генов при низкой температуре. При потеплении количество H2A.Z в промоторах генов, в том числе гена *FT*, снижается, что делает его доступным для ТФ (Kumar et al., 2012; Talbert and Henikoff, 2014). Кроме того, важную роль в температурно-зависимом зацветании играет репрессор транскрипции SVP, количество мРНК которого снижается при росте температуры (Fernández et al., 2016).

На переход к цветению влияет также гиббереллиновый (ГК) сигнальный путь. Повышение уровня ГК приводит к деградации транскрипционных репрессоров группы DELLA и, таким образом, активирует зацветание (Yu et al., 2012; Galvão et al., 2015). В листьях ГК участвуют в активации транскрипции *FT*, что происходит путем снижения количества белков DELLA, которые связываются с ТФ SPL и CO и блокируют их способность индуцировать экспрессию *FT* (Yu et al., 2012; Xu et al., 2016). В апикальной меристеме биологически активные ГК также снижают количество белков DELLA, что через микроРНК159 и ТФ MYB33 приводит к активации экспрессии *SOC1* и *LFY* (Achard et al., 2004).

#### 2.2.6. Степень изученности глобальных изменений транскриптома апикальной меристемы при переходе к цветению

В предыдущих разделах обзора литературы мы кратко осветили разные аспекты процесса перехода растения к зацветанию. Разнообразные пути активации цветения *A. thaliana* подробно

охарактеризованы: известны сенсоры разнообразных внешних и внутренних сигналов, влияющих на определение времени перехода к репродуктивному развитию; путь передачи этих сигналов в АМ; ключевые гены, регулирующие зацветание. Однако физиологические события, происходящие после активации регуляторных генов (*LFY*, *API*) и обеспечивающие изменение состояния меристемы, остаются неизвестными. Изучить этот вопрос может помочь исследование динамики всего транскриптома АМ и поиск в ней определенных паттернов.

До сих пор число исследований, посвященных широкомасштабному анализу экспрессии в АМ, остается ограниченным. Среди них можно назвать работу Torti et al., в которой был проведен анализ профилей экспрессии генов в трех образцах АМ в процессе перехода к цветению, что позволяет проследить общую динамику транскриптома, однако уровень детализации такой временной серии невысок (Torti et al., 2012). Еще одно исследование рассматривает транскриптом АМ соцветия в контексте формирования и развития цветков и не затрагивает процесс перехода вегетативной меристемы в репродуктивную (Zhang et al., 2015). Кроме того, с помощью флуоресцентного сортирования клеток были изучены транскриптомы различных компартментов АМ *A. thaliana* (Yadav et al., 2014), а с использованием иммунопреципитации помеченных определенным эпитопом рибосом – транслирующиеся РНК (Tian et al., 2019).

Таким образом, не смотря на очевидную важность такого исследования, динамика транскриптома АМ *A. thaliana* в процессе перехода к цветению оставалась неизученной.

## 2.3. Результаты

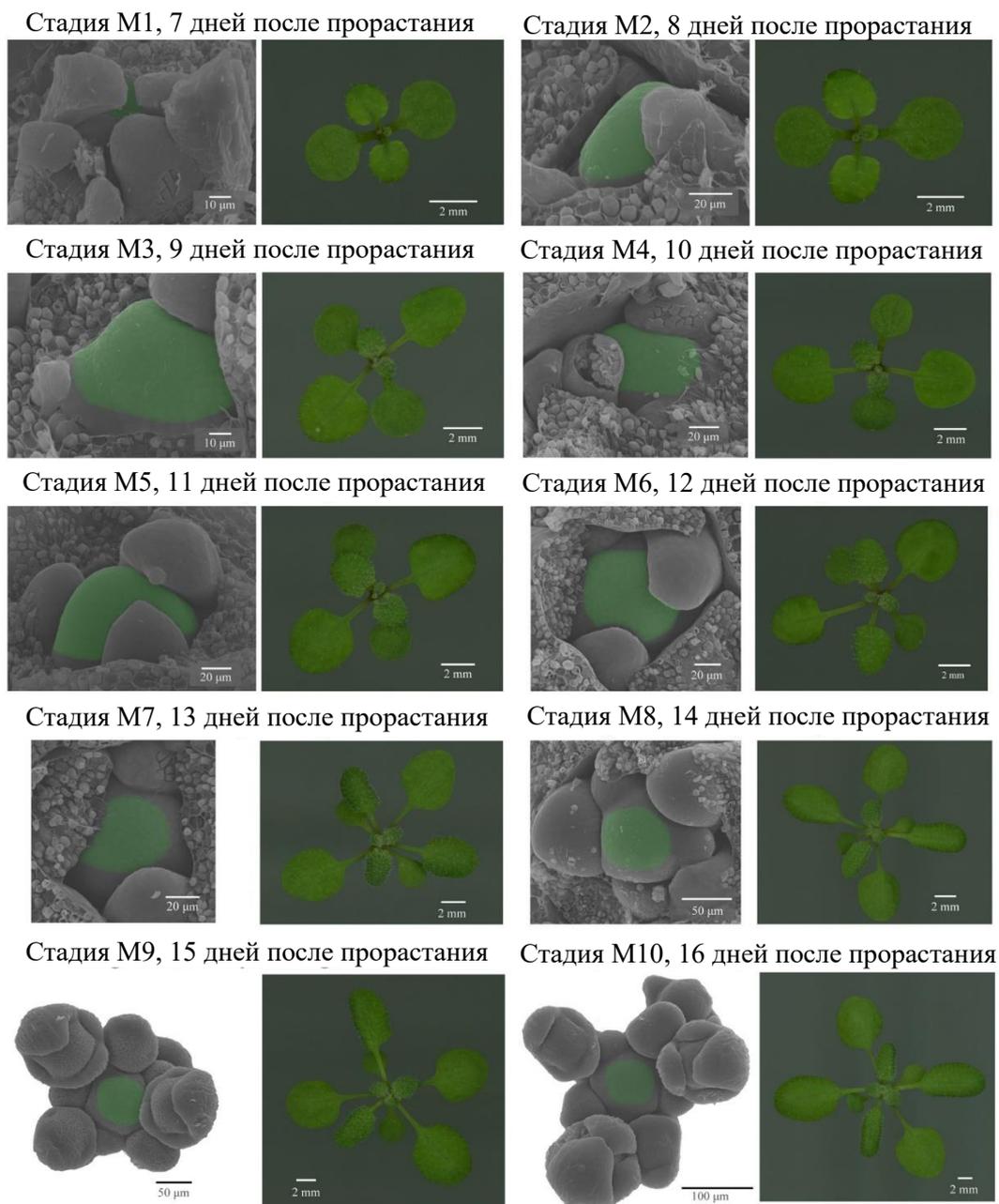
### 2.3.1. Сбор образцов временной серии, секвенирование и первичная обработка данных

В процессе перехода к цветению морфология и физиология апикальной меристемы побега претерпевает существенные изменения (Jacqard et al., 2003), в том числе отражающиеся на экспрессии генов, что делает АМ одной из интереснейших с точки зрения транскриптомики структур растения (Andrés and Coupland, 2012). Для детального анализа этих транскриптомных изменений удобным инструментом служат высокоразрешенные временные серии АМ. Однако, как было сказано выше (см. раздел 2.2.6), к моменту публикации нашей работы (Klerikova et al., 2015) существовало только одно исследование динамики транскриптома АМ с помощью секвенирования РНК, включающее в себя три образца, что явно недостаточно для полного отражения происходящих в АМ процессов. Поэтому нашей целью было создание более разрешенной временной серии АМ в процессе зацветания.

Для этого нами ежедневно собирались пулы из 15 АМ в двух биологических повторностях; сбор проводился в течении десяти дней, с 7 по 16 день после прорастания растений (стадии М1-М10). Хотя используемая в работе раса *A. thaliana* Col-0 (CS70000) в высокой степени гомозиготна (см. Материалы и Методы), неоднородность скорости роста сделала необходимой разработку и применение морфологических маркеров для синхронизации стадий развития собираемых растений.

Общий вид розеток *A. thaliana* и микрофотографии меристем с прилежащими тканями приведены на Рисунке 2.2. Растения характеризовались по таким параметрам, как видимое невооруженным глазом число листьев, морфология листьев (длина, наличие трихом, развитие центральной жилки), число цветков. Описание каждой стадии в соответствии с выбранными морфологическими маркерами приведено в Приложении, Таблица 2.1.

Так как временная серия меристем является частью транскриптомной карты (см. Главу 1), сбор образцов, выделение тотальной РНК, приготовление библиотек, секвенирование и первичная обработка данных проводились как описано в разделе 1.3.2, а число уникально картированных чтений для каждого образца приведено в (Klerikova et al., 2016), Таблица S2 Приложения). Как в случае других образцов атласа экспрессии *A. thaliana*, биологические повторности АМ были схожи между собой – квадрат коэффициента корреляции Пирсона распределялся от 0,96 до 0,99 (Приложение, Таблица 1.2).



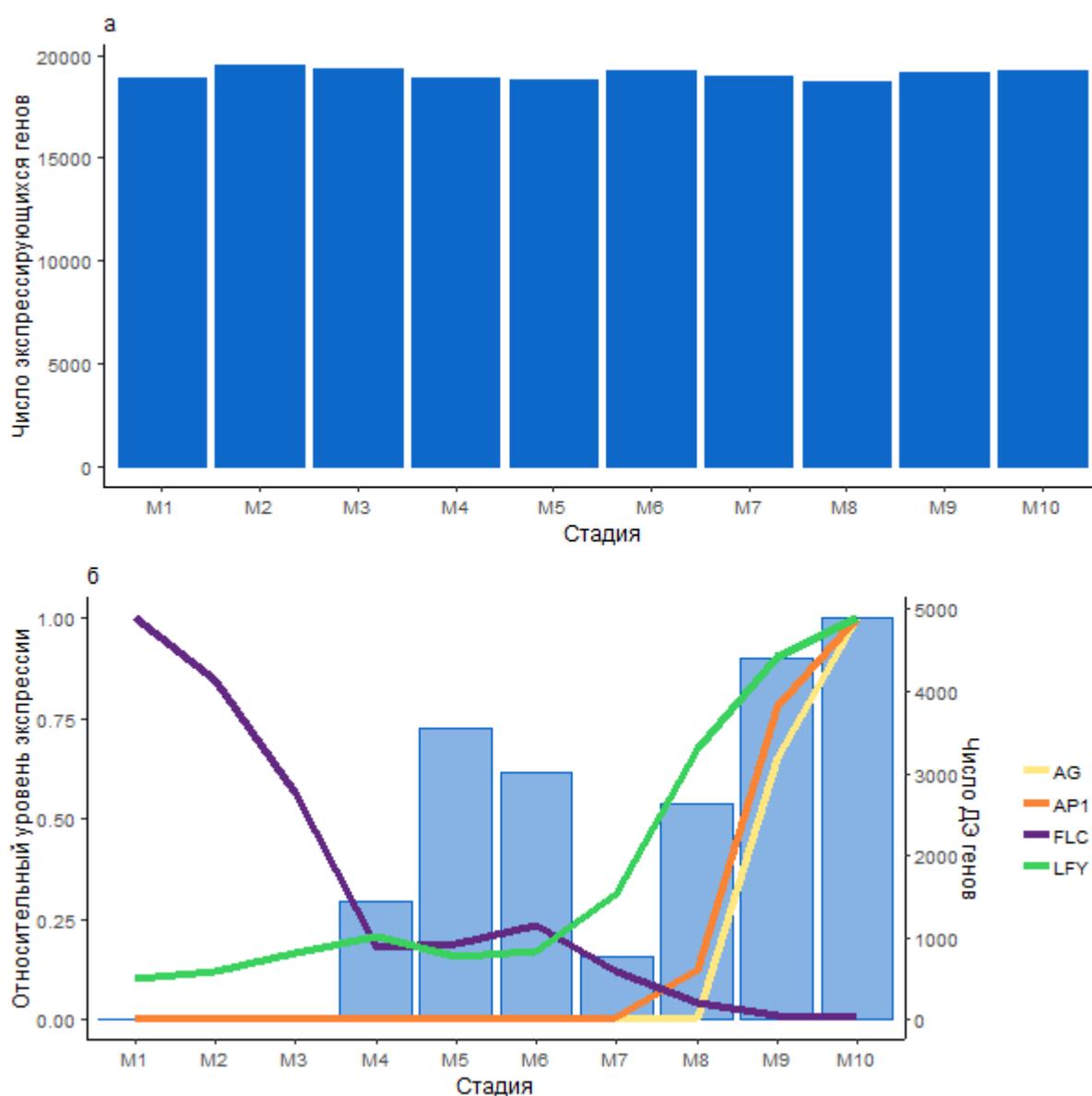
**Рисунок 2.2.** Микрофотографии апикальных меристем побега и фотографии розеток *A. thaliana* на соответствующих стадиях развития. Микрофотографии получены с помощью сканирующего электронного микроскопа; область, которая собиралась с помощью микродиссекции, отмечена на микрофотографиях искусственным зеленым цветом; покрывающие меристему листья удалены.

Анализ числа генов, экспрессирующихся в образцах АМ, показал минимальную разницу (4%) между стадиями: транскрипты наибольшего числа генов (19 480) были найдены на стадии M2, а наименьшего (18 750) – в образце M8 (Рисунок 2.3 а). 64% (21 391) от всех генов,

аннотированных в геноме *A. thaliana* (TAIR10), экспрессировались хотя бы в одном образце, а транскрипты 15 312 генов присутствовали во всех десяти образцах.

### 2.3.2. Динамика ключевых генов – регуляторов цветения

Первым этапом анализа временной серии АМ было описание динамики экспрессии известных генов-регуляторов перехода к цветению. Основным активатором зацветания является ген *LFY*, экспрессия которого возрастает в процессе перехода, а главным репрессором – ген *FLC*, инактивирующийся при зацветании (Michaels and Amasino, 1999). Профили экспрессии генов *LFY* и *FLC* в наших данных соответствуют ожидаемому (Рисунок 2.3 б), что свидетельствует об адекватности собранной нами временной серии.



**Рисунок 2.3.** Общая характеристика временной серии АМ. (а) Число экспрессирующихся генов в каждом образце; (б) Относительный уровень экспрессии основных генов, регулирующих переход к цветению.

Во время вегетативного развития растения периферическая зона АМ закладывает примордии листьев; эта стадия соответствует образцам М1-М4 временной серии (Приложение, Таблица 2.1). В течение этих стадий экспрессия *FLC* уменьшается в пять раз, тогда как *LFY* неизменно экспрессируется на низком уровне. Ген *SVP*, другой негативный регулятор цветения, экспрессируется на высоком уровне как в АМ вплоть до ранних стадий выметывания цветоноса, так и в примордиях цветка (Hartmann et al., 2000). В соответствии с этими данными, в серии наблюдается высокая экспрессия гена *SVP*, растущая от стадии М1 к М4 (Приложение, Рисунок 2.1). Ген *SOC1* является одним из самых ранних маркеров начинающегося перехода к цветению; его экспрессия пятикратно возрастает на стадиях М1-М4 (при этом рост в три раза приходится на стадии М3-М4) (Приложение, Рисунок 2.1). Количество мРНК гена *FD* согласно литературным данным находится на высоком уровне в АМ до перехода к цветению и далее увеличивается (Wigge et al., 2005). В нашей временной серии экспрессия гена *FD* возрастает в 4,5 раза (Приложение, Рисунок 2.1).

На стадиях М4-М6 относительные уровни экспрессии генов *FLC* и *LFY* находятся на низком уровне и равны друг другу (Рисунок 2.3 б), при этом экспрессия *SOC1* уменьшается в два раза, образуя локальный пик на стадии М4 (Приложение, Рисунок 2.1). Известно, что уровень транскрипции генов из семейства *SQUAMOSA PROMOTER BINDING LIKE (SPL)* растет в условиях ДД (Schmid et al., 2003), и два гена из этого семейства, *SPL9* и *SPL15*, участвуют в инициации листовых примордиев (Wang et al., 2009). В соответствии с этим, уровни экспрессии генов *SPL9* и *SPL15* характеризуются локальным пиком на стадии М4, а начиная со стадии М6 – дальнейшим ростом (Приложение, Рисунок 2.1).

На промежутке от М6 до М8 (стадия формирования первых трех примордиев цветков) экспрессия гена *FLC* уменьшается до еще более низкого уровня, а экспрессия *LFY* возрастает в четыре раза по сравнению с М6 (Рисунок 2.3 б). Так же продолжается рост экспрессии гена *SOC1* (в 1,8 раз) (Приложение, Рисунок 2.1). Экспрессия генов *SPL4* и *SPL5* возрастает с нулевого уровня начиная со стадии М7, что соответствует данным *in situ* гибридизации, показывающих отсутствие экспрессии *SPL3*, *SPL4* и *SPL5* в вегетативной меристеме (Wang et al., 2009). Уровень транскрипции генов *SPL9* и *SPL15* продолжает повышаться от стадии М6 к М8, кроме того, начиная со стадии М8, экспрессия гена *API* возрастает с нулевого уровня, а экспрессия *CAL* так же увеличивается по сравнению с фоновыми значениями (Рисунок 2.3 б, Приложение, Рисунок 2.1).

Со стадии М8 начинается рост и развитие цветков. Экспрессия гена *FLC* уменьшается до нуля, а *LFY* – в небольшой степени увеличивается (Рисунок 2.3 б). Начиная с этой стадии активируются гомеозисные гены, определяющие тип органов цветка: экспрессия таких генов,

как *APETALA3*, *PISTILLATA* и *AGAMOUS*, начиная с нулевого уровня, доходит до максимума в М10 (Рисунок 2.3 б).

Таким образом, в соответствии с динамикой экспрессии генов *FLC*, *LFY* и других регуляторов цветения, нашу временную серию меристем можно разделить на четыре основные части: первая, М1-М4, когда экспрессия *FLC* уменьшается, а ген *LFY* экспрессируется на низком уровне (стадия вегетативного развития АМ); вторая, М4-М6, стадия перехода, когда уровни экспрессии *FLC* и *LFY* равны; третья, М6-М8, в которой происходит активация цветения и количество мРНК *LFY* возрастает, а *FLC* убывает; и четвертая, М8-М10, стадия инициации и развития примордиев цветков (Рисунок 2.3 б).

Стадия перехода особенно интересна из-за паритетного уровня транскрипции положительного и отрицательного регуляторов цветения; также следует отметить, что между М5 и М6 не происходит образования новых органов (Приложение, Таблица 2.1), что позволяет сделать предположение о происходящем в этот момент перепрограммировании меристемы из вегетативного состояния в репродуктивное.

### 2.3.3. Общий анализ дифференциально экспрессирующихся генов

Анализ дифференциальной экспрессии между последовательными стадиями (М2 против М1, М3 против М2 и так далее) использовался нами для получения общего представления о процессах, происходящих в АМ. Число ДЭ генов значительно различалось между сравнениями: от 13 между стадиями М2 и М1 до 4 890 между М10 и М9 (Рисунок 2.3 б). Пик числа ДЭ генов при переходе от М8 к М10 связан с образованием примордиев цветков (в соответствии с ростом экспрессии генов, определяющих тип органов цветка, таких, как *AG*, *AP3* и *PI*). Особенно интересно второе локальное увеличение числа ДЭ генов в диапазоне стадий М4-М6: стадию М5 отличают от М4 3 552 ДЭ генов, а М6 от М5 – 3 019 (Рисунок 2.3 б). Этот рост дифференциальной экспрессии совпадает со стадией, определенной нами в разделе 2.3.3 как «стадия перехода», где относительная экспрессия генов *FLC* и *LFY* находится на одном уровне, поэтому гены, экспрессия которых изменяется между стадиями М4-М6, были проанализированы нами в больших подробностях.

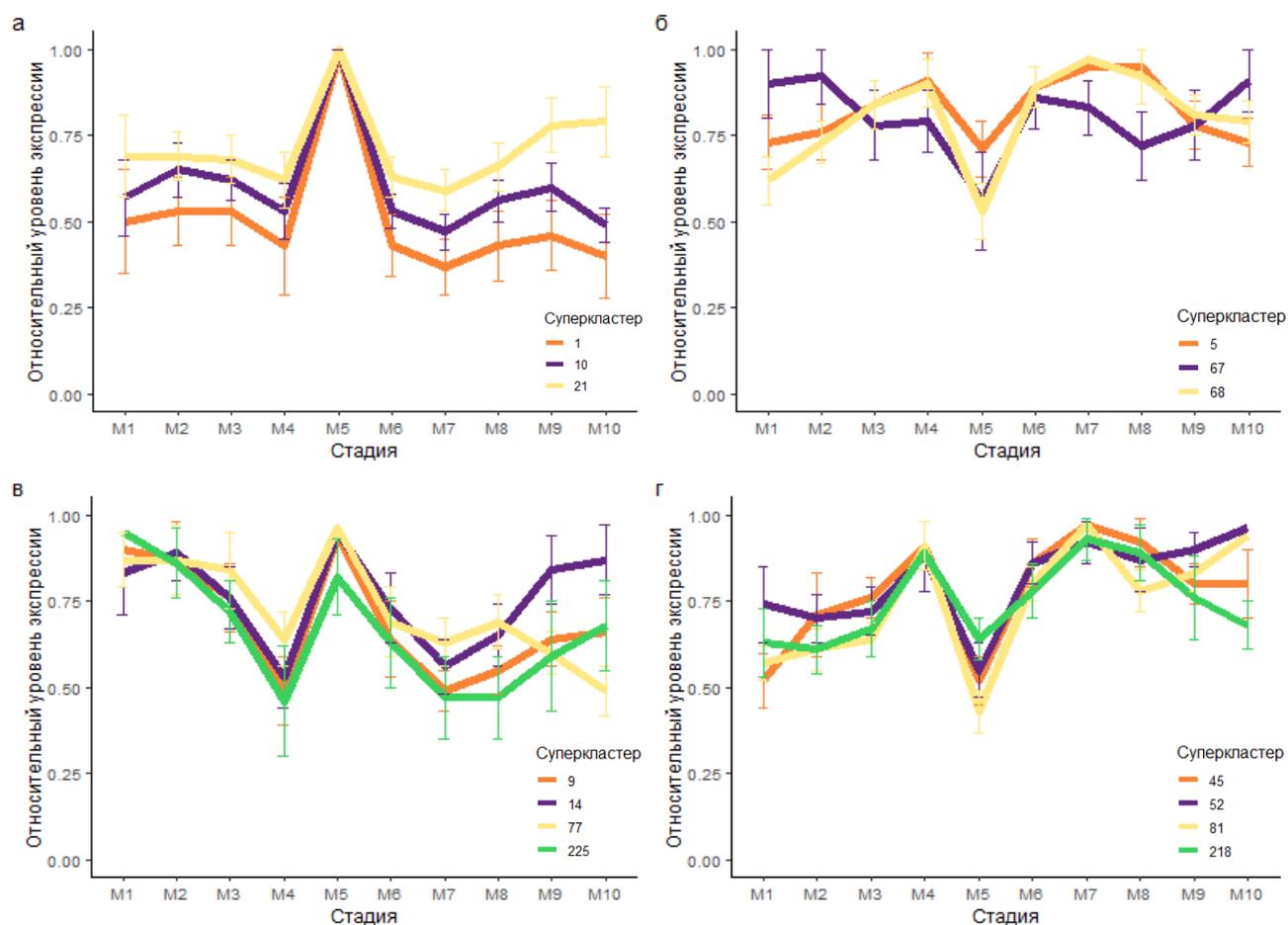
Общая характеристика ДЭ генов с помощью Генной Онтологии (подробнее см. раздел 1.3.3) показала, что гены, экспрессия которых уменьшалась на стадии М5 по сравнению с М4, и гены с растущей экспрессией на стадии М6 по сравнению с М5 были обогащены категориями, связанными организацией цитоскелета и движением (GO:0005874~микротрубочки, GO:0016459~миозиновый комплекс, IPR001752~моторный домен кинезина, GO:0030048~движение, основанное на филаментах актина), модификацией хроматина и репликацией ДНК (GO:0006260~репликация ДНК, хеликазы, GO:0006325~организация хроматина), активностью АТФаз (GO:0005524~связывание АТФ, IPR003593~ядерные АТФазы

AAA+ типа) и киназами (GO:0006468~фосфорилирование аминокислот белка, IPR008271~активный сайт серин/треониновых киназ). Обогащенные категории также совпадали для списков генов, транскрипция которых возрастала в сравнении M5 с M4, и списков генов с уменьшающейся экспрессией на стадии M6 по сравнению с M5. Они относились к сборке нуклеосом (GO:0000785~хроматин, IPR001951~гистон H4, GO:0006323~ упаковка ДНК), тубулинам (GO:0045298~комплекс тубулинов) и биогенезу и структуре рибосом (GO:0005840~рибосома, биосинтез белков). Кроме того, гены, характеризующиеся уменьшением экспрессии в M5 по отношению к M4, были обогащены категориями, описывающими клеточный цикл и репликацию ДНК (GO:0003887~активность ДНК-зависимой ДНК-полимеразы, GO:0022402~процессы клеточного цикла), тогда как гены с возрастанием экспрессии в этом сравнении чаще участвовали в метаболизме нуклеотидов (GO:0006754~биосинтез АТФ, GO:0009141~метаболизм трифосфата нуклеозида).

#### *2.3.4. Кластеризация генов согласно их профилям экспрессии*

Другим способом поиска генов, обладающих характерными профилями экспрессии, является кластерный анализ. Основой для кластеризации служили нормализованные на максимум уровни экспрессии генов (подробнее см. Материалы и Методы); таким образом, в одну группу попадают гены со схожими относительными профилями экспрессии, безотносительно абсолютного числа чтений, приходящихся на ген. Для повышения точности кластеризации нами был применен двухстадийный метод. На первом его этапе гены группируются с помощью метода k-средних. Так как такой тип кластеризации начинается со случайного выбора стартового набора k точек и его результаты различаются между запусками алгоритма, кластеризация методом k-средних была повторена 1 000 раз. На втором этапе на основе числа раз, когда два гена оказывались в одном кластере, была построена матрица расстояний между генами. Эта матрица была использована для иерархической кластеризации (см. Материалы и Методы), в результате которой было получено 2 420 кластеров, содержащих в сумме 18 825 генов (что составляет 88% от числа всех экспрессирующихся генов). Число генов в кластере варьировалось от 51 до 5 с медианой 7, что недостаточно для эффективного анализа обогащения; кроме того, многие кластеры имели схожие профили экспрессии. Поэтому тот же двухстадийный метод был применен для объединения усредненных профилей экспрессии кластеров в более крупные кластеры (далее суперкластеры). В результате было получено 257 суперкластеров, в которые вошли 16 615 генов (77% от числа всех экспрессирующихся в АМ генов).

Экспрессия двенадцати суперкластеров была равномерной на всех стадиях временной серии, кроме M5, где она резко возрастала и достигала максимального уровня (Рисунок 2.4 а).



**Рисунок 2.4.** Усредненные профили экспрессии кластеров с (а) повышением относительной экспрессии на стадии М5; (б) понижением относительной экспрессии на стадии М5; (в) понижением относительной экспрессии на стадии М4 и дальнейшим ростом на стадии М5; (г) повышением относительной экспрессии на стадии М4 и дальнейшим падением на стадии М5.

Гены, входящие в эти суперкластеры, чаще ожидаемого описывались категориями ГО, связанными с биогенезом рибосом, гистонами, нуклеосомами, модификацией хроматина и сплайсингом мРНК (Клерикова et al., 2015), Таблица Приложения 3). Суперкластеры с противоположным профилем экспрессии (то есть, с уменьшением уровня транскрипции на стадии М5 при равномерной экспрессии в остальных образцах, Рисунок 2.4 б) были обогащены категориями, относящимися к АТФазам и связыванию нуклеотидов, хеликазам и пентатрикопептидным повторам (PPR, (Клерикова et al., 2015), Таблица Приложения 3). Еще один характерный профиль транскрипции обладал снижением на стадии М4 и ростом на М5 (Рисунок 2.4 в); гены, входящие в такие суперкластеры, характеризуются терминами, связанными с биогенезом и структурой рибосом и фотосистем, тогда как гены, образующие суперкластеры с обратной динамикой экспрессии (увеличением в М4 и уменьшением в М5, Рисунок 2.4 г), обогащены категориями, описывающими цитоскелет и движение цитоскелета (микротрубочки, кинезины).

### 2.3.5. Анализ профилей экспрессии отдельных классов генов

Анализ генов, дифференциально экспрессирующихся в диапазоне стадий М4-М6, а также генов, входящих в суперкластеры, профиль экспрессии которых меняется на стадиях М4-М6, позволяет предположить, что в ходе инициации цветения происходят изменения в организации хроматина, функционировании цитоскелета, синтезе ДНК и клеточном цикле. Чтобы более глубоко изучить этот вопрос, мы детально рассмотрели динамику экспрессии ряда функциональных групп генов. В их число вошли гены, кодирующие элементы цитоскелета (актины, актин-связывающие белки, факторы полимеризации и деполимеризации актина, тубулины, миозины, кинезины и динеины), гистоны и модификаторы гистонов, ДНК- и РНК-полимеразы, циклины, циклин-зависимые киназы и другие белки, связанные с клеточным циклом.

Экспрессия всех генов, кодирующих актины (кроме специфических для структур цветка), снижается на стадии М4 и возвращается на предыдущий уровень на М5 (Приложение, Рисунок 2.2 а). Одними из важных регуляторов динамической активности цитоскелета являются факторы деполимеризации актина (Actin Depolymerizing Factors, ADF), которые принимают участие в его ремоделировании и контроле деполимеризации (Bamburg, 1999). В геноме *A. thaliana* аннотировано 11 генов, кодирующих ADF (Feng et al., 2006); уровень экспрессии большинства из них уменьшается в образце М4 и возрастает в М5, исключение составляют несколько генов, продукты которых функционируют в пыльце (Ruzicka et al., 2007) (Приложение, Рисунок 2.2 б). Актин-связывающие белки виллины принимают участие как в защите актиновых филаментов от ADF, так и в их разрезании (Khurana et al., 2010). Все пять генов, кодирующих виллины (*VLN1-VLN5*), обладают сниженной транскрипцией на стадии М5 в наших данных (Приложение, Рисунок 2.2 в). Еще один класс белков *A. thaliana*, вовлеченных в стабилизацию актиновых филаментов, – это фимбрины (Wu et al., 2010), которые кодируются генами *FIM1* и *FIM2*; эти гены характеризуются снижением экспрессии на стадии М5 по сравнению с М4 (Приложение, Рисунок 2.2 г).

Миозины класса XI – это белки, обеспечивающие быструю динамику аппарата Гольджи, митохондрий, пероксисом и пластид (Avisar et al., 2008; Peremyslov et al., 2008). Экспрессия генов, кодирующих миозины и миозин-подобные белки, снижена в образце М5. Профили экспрессии генов, кодирующих  $\alpha$ - и  $\beta$ -тубулины имеют локальный пик на М5 (Рисунок 2.5 а), в то время как почти все гены кинезинов имеют сниженную экспрессию на этой стадии (Рисунок 2.5 б).

Гены, экспрессия которых возрастала в образце М5 по сравнению с М4 и снижалась в М6 по отношению к М5, имели обогащение категориями «гистоны», так же, как и суперкластеры с

пиком профиля экспрессии на стадии М5 (Приложение, Таблицы 2.2 и 2.3). Двадцать восемь из тридцати генов *A. thaliana*, кодирующих гистоны, имеют максимальную экспрессию на стадии М5 (Рисунок 2.5 а). Динамическое состояние хроматина поддерживается в том числе с помощью белков-модификаторов гистонов, к которым относятся ацетил-трансферазы, деметилазы и метил-трансферазы (Eichten et al., 2014). Уровень транскрипции большинства генов, кодирующих эти ферменты, уменьшается на стадии М5 (Приложение, Рисунок 2.3).

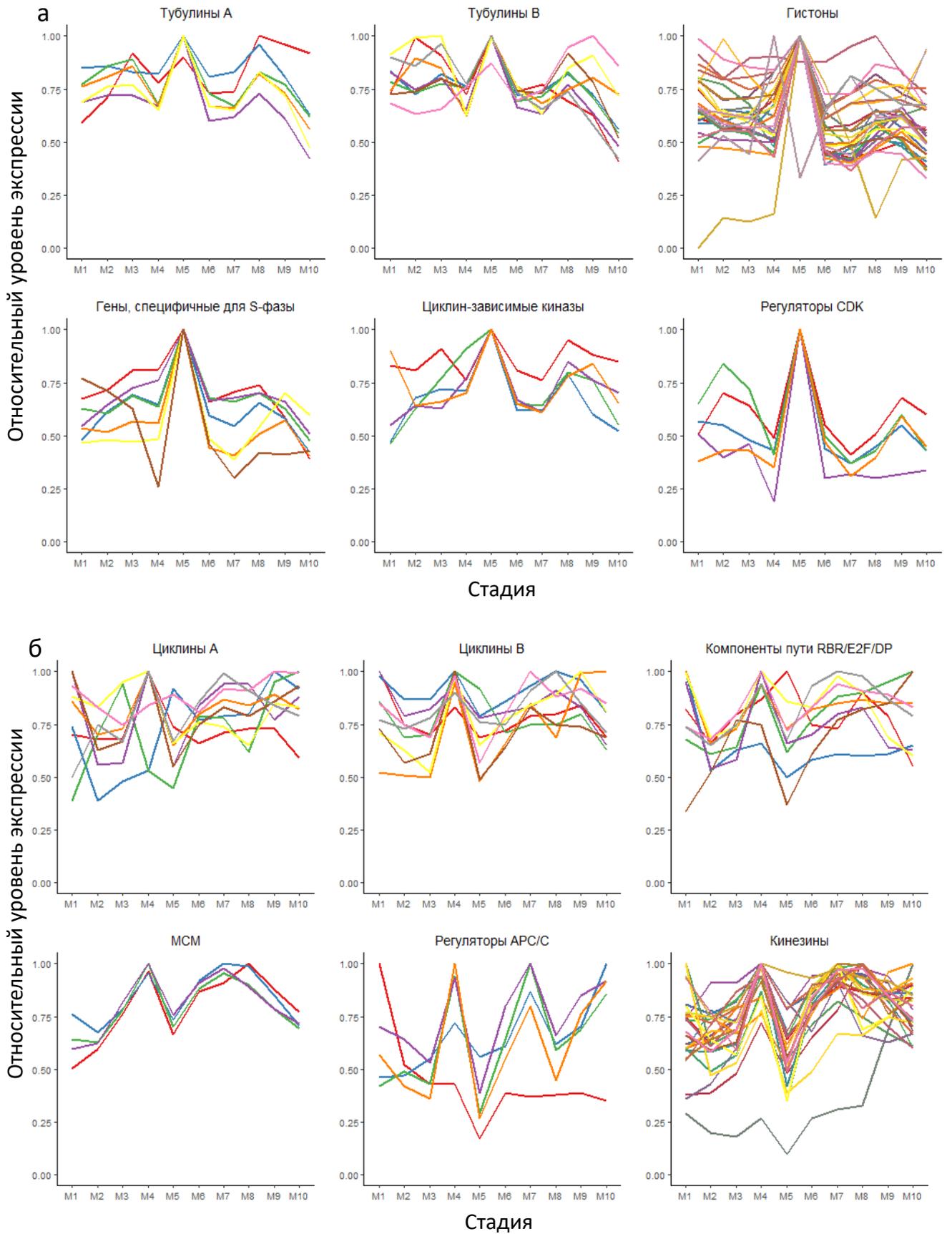
Гены, кодирующие субъединицы ДНК- и РНК-полимераз, характеризуются сниженной экспрессией на стадии М5, при этом некаталитические субъединицы, являющиеся общими для ряда ядерных ДНК-зависимых РНК-полимераз, показывают противоположный профиль (Приложение, Рисунок 2.4).

Что касается регуляторов клеточного цикла, гены циклин-зависимых киназ и их регуляторов имеют максимум экспрессии на стадии М5 (Рисунок 2.5 а), тогда как циклины А, В и Т максимально экспрессируются в М4 (Рисунок 2.5 б).

Таким образом, анализ профилей экспрессии отдельных групп генов позволяет предположить, что на стадии М5 в апикальной меристеме происходят значительные изменения в клеточном цикле.

### *2.3.6. Верификация ключевых стадий временной серии меристем с помощью секвенирования независимо собранных образцов*

В результате проведенного нами анализа временной серии меристем можно сделать вывод, что основные события, связанные с изменением статуса меристемы и переходом к цветению, происходят на стадиях М4-М6. Для подтверждения полученных нами результатов мы повторно собрали временную серию АМ на стадиях М3-М7 в тех же условиях с использованием разработанных нами морфологических маркеров (см. раздел 2.3.1). Биологические повторности этой серии были также в высокой степени схожи между собой ( $r^2$  от 0,93 до 0,99). Списки ДЭ генов для всех последовательных пар образцов совпали со списками из основного эксперимента более чем на 50% в каждом случае, при этом обогащение категориями ГО и других баз данных в большинстве случаев было практически идентичным. Кроме того, медиана ковариации суперкластеров (см. Материалы и Методы) составила 0,95, а для классов генов – 0,9.



**Рисунок 2.5.** Профили экспрессии классов генов, обладающих определенным паттерном: (а) гены, принадлежащие к типу (I) и имеющие пик экспрессии на стадии M5, и (б) гены типа (II), демонстрирующие рост экспрессии на стадиях M4 и M6 вместе с падением на стадии M5.

## 2.4. Обсуждение

### 2.4.1. Морфология растений *A. thaliana* при переходе к цветению

Число розеточных листьев традиционно используется для определения времени перехода к цветению в работах, посвященных изучению мутантов *A. thaliana* по генам регуляции зацветания (Michaels and Amasino, 1999). Для описания растений в полученной нами временной серии, а также для синхронизации стадии развития собираемых растений в пределах одного образца мы использовали общее число листьев розетки совместно с другими количественными характеристиками: число листьев розетки, видимых невооруженным глазом, долю центральной жилки от длины листа, номер наименьшего листа, обладающего трихомами, номер листьев, находящихся на стадии примордиев и, после появления примордиев цветков, – их количество и стадии по (Smyth et al., 1990).

На седьмой день после прорастания растения имели шесть листьев, из которых два было видно невооруженным глазом; первые три листа имели трихомы, а последний, шестой лист был виден как примордий. Вплоть до стадии M5 (11 день после прорастания), число листьев, видимых невооруженным глазом, возрастало на один в день, тогда как общее их число прирастало быстрее и с шести достигло 14 (Приложение, Таблица 2.1). Число листьев, обладающих трихомами, возрастало на два каждый день. Между стадиями M5 и M6 не происходило закладки новых органов и общее число листьев остановилось на 14, при этом небольшие листья увеличились в размерах. Это характеризовало окончание развития розетки, и, на 13-ый день после прорастания (стадия M7), растения закладывали два или три стеблевых листа, в пазухах которых на стадии M10 были видны соцветия паракладиев. После M7 размеры листьев изменялись медленно, а со стадии M8 появлялись первые примордии цветков.

Таким образом, с помощью изучения морфологии растений временную серию AM можно разделить на четыре этапа: M1-M5 - интенсивное вегетативное развитие, быстрый рост розетки; M5-M6 – торможение и отсутствие закладки новых органов; M6-M8 – начало репродуктивного развития, формирование стеблевых листьев; M8-M9 – цветение.

### 2.4.2. Основные паттерны экспрессии генов на стадии перехода

Как анализ дифференциальной экспрессии, так и кластеризация генов и рассмотрение отдельных функциональных групп демонстрируют, что наиболее заметные изменения происходят в профилях экспрессии генов, кодирующих гистоны, тубулины и циклины, транскрипция которых, как хорошо известно, напрямую зависит от стадии клеточного цикла. Паттерны экспрессии этих генов у растений были описаны для искусственно синхронизированных суспензионных клеточных культур (Breune et al., 2002; Menges and Murray, 2002; Menges et al., 2005), однако, насколько нам известно, изменения, происходящие с

клеточным циклом в АМ при переходе к цветению, не были охарактеризованы на молекулярном уровне, хотя меняющаяся скорость пролиферации клеток в момент зацветания подробно документирована в цитологических исследованиях (см. раздел 2.4.8).

Гены, продукты которых участвуют в контроле и прохождении клеточного цикла, обладают двумя основными паттернами экспрессии на стадиях М4-М6: (I) заметный пик на М5 и (II) рост на М4, падение на М5 и возвращение к прежнему уровню на М6 (Рисунок 2.4 а и б).

#### 2.4.3. Паттерны экспрессии генов, связанных с S-фазой

Точная регуляция транскрипции генов гистонов происходит в течении всего клеточного цикла (Таока et al., 1999), при этом количество мРНК всех гистонов синхронно возрастает в клетках, входящих в S-фазу и быстро падает после окончания S-фазы (Menges et al., 2005), что позволяет использовать относительное количество мРНК гистонов и как маркер S-фазы, и как индикатор пролиферативного индекса ткани во всех организмах (Chaubet et al., 1991). Во временной серии АМ почти все гены гистонов и гистон-подобных белков обладают паттерном типа (I) (Рисунок 2.5 а).

Кроме того, гены, кодирующие компоненты комплекса репликации ДНК, демонстрируют паттерн типа (I). В культуре растительных клеток для них показано возрастание в поздней G<sub>1</sub>- и S-фазах (Menges et al., 2002). В нашем эксперименте оба гена *A. thaliana*, кодирующих ядерный антиген пролиферирующих клеток (*PCNA1* и *PCNA2*) имеют максимум экспрессии на стадии М5 (Рисунок 2.5 а). *RPA2/RPA32A* и два *RPA3*-подобных гена (*AT3G52630* и *AT4G18590*), кодирующие субъединицы репликационного белка А, связывающего одноцепочечную ДНК при репликации, имеют такой же профиль, тогда как гены, кодирующие варианты третьей субъединицы ДНК-полимеразы (семейство генов *RPA70*) не показывают дифференциальной экспрессии (Рисунок 2.5 а). Среди других генов, продукты которых участвуют в репликации ДНК, паттерном типа (I) обладает ген *RFC2*, кодирующий фактор репликации С2, и гомолог гена контрольной точки перехода в S-фазу *CDC25* (*AT5G03455*), кодирующий фосфатазу (Рисунок 2.5 а). Последний особенно интересен тем, что в культурах растительных клеток этот ген не показывал изменений в ходе клеточного цикла (Menges et al., 2005), в отличие от его поведения в других организмах; также трансгенная сверхэкспрессия гена дрожжей *CDC25* индуцировало цветение в табаке (Teichmanová et al., 2007).

#### 2.4.4. Паттерны экспрессии генов, связанных с митозом

Важным фактом является то, что паттерн, аналогичный генам S-фазы (пик экспрессии на стадии М5), наблюдается у генов, обычно ассоциируемых с поздней G<sub>2</sub>-фазой и митозом. Гены, кодирующие α- и β-тубулины, имеют локальное повышение экспрессии в образце М5, хотя в этом случае общий профиль отличается увеличенной экспрессией на ранних и поздних стадиях (M1-

M3 и M8-M10) (Рисунок 2.5 а). В время митоза высокий уровень мономеров тубулинов требуется для образования веретена деления, поэтому транскрипция кодирующих их генов увеличена в поздней G<sub>2</sub>/M-фазах клеточного цикла у многих организмов, включая растения (Cho et al., 2001; Breune et al., 2002; Menges and Murray, 2002). Таким образом, на стадии M5 мы наблюдаем одновременное увеличение экспрессии генов, характерных для S-фазы и G<sub>2</sub>/M-фазы.

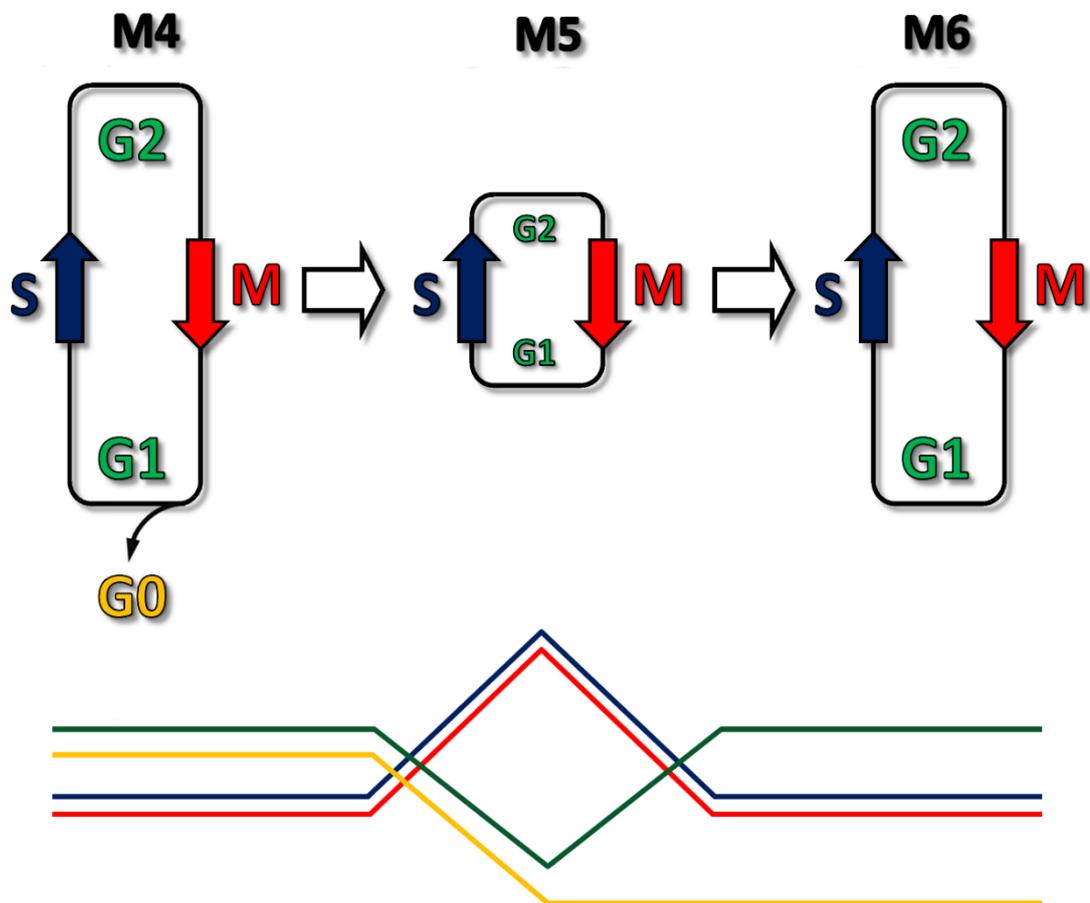
Помимо этого, профиль экспрессии типа (I) на стадии M4-M6 характерен для циклин-зависимых киназ (cyclin dependent kinases, CDK) А- и В-типа (Рисунок 2.5 а). В геноме *A. thaliana* аннотировано несколько классов CDK, и из них один ген класса *CDKA* и четыре *CDKB* непосредственно участвуют в контроле клеточного цикла (Van Leene et al., 2010). В культуре растительных клеток ген *CDKA;1* (ортолог *CDC28* дрожжей и *CDK1* животных, взаимодействующих исключительно с G<sub>1</sub>/S-специфическими циклинами), экспрессируется постоянно и равномерно, тогда как *CDKB1* и *CDKB2* *A. thaliana* связаны с G<sub>2</sub>/M-циклинами (Gegas and Doonan, 2006). Экспрессия двух генов, кодирующих субъединицы циклин-зависимых киназ (*CKS1* и *CKS2*), так же возрастает на стадии M5 (Рисунок 2.5 а) (Menges et al., 2005; Van Leene et al., 2010), как и экспрессия трех генов SIAMESE-подобных ингибиторов клеточного цикла (*SMR2*, *SMR4* и *SMR11*), вовлеченных в контроль митоза (Van Leene et al., 2010).

#### 2.4.5. Предлагаемая модель, описывающая изменение экспрессии генов в точке перехода к цветению

Полученные результаты можно описать моделью, согласно которой в AM на стадии M5 происходит рост числа клеточных делений, сопровождающийся укорочением фаз G<sub>1</sub> и G<sub>2</sub> (Рисунок 2.6). Об этом также свидетельствует обогащение списков генов, меняющих экспрессию на этой стадии, терминами ГО, связанными с биосинтезом нуклеотидов, биогенезом рибосом, трансляцией, функционированием митохондрий и хлоропластными компонентами, то есть категориями, указывающими на пик пролиферации в этой временной точке (Larkin et al., 1989; Lebrun and Freyssinet, 1991).

Однако детальный анализ экспрессионных профилей ряда важных компонентов регуляции клеточного цикла, включая циклины, усложняет картину. Геномы растений включают в себя большие семейства генов, кодирующих циклины и циклин-зависимые киназы (Wang et al., 2004a; Menges et al., 2005). В суспензионных культурах клеток основная часть митотических циклинов А- и В-типа демонстрируют схожий паттерн экспрессии; их продукты накапливаются в поздней G<sub>2</sub> с пиком в раннем митозе (Menges et al., 2005). Исключение составляют гены *CYCA3;1*, *CYCA3;2* и *CYCA3;3*, которые в основном экспрессируются на границе G<sub>1</sub>/S, то есть кодируют G<sub>1</sub>-циклины (Menges et al., 2005; Takahashi et al., 2010). Экспрессия семейств генов *CYCA1*, *CYCB1* и *CYCB2* регулируется общим молекулярным механизмом, включающим

связывание специфичных к М-фазе активационных элементов (MSA-элементов) в промоторах этих генов транскрипционными факторами MYB3R1 и MYB3R4, содержащими домен R1R2R3-Myb (Ito, 2005). В нашей временной серии 13 из этих генов, включая *CYCA3;1*, обладают паттерном типа (II), что коррелирует с профилем экспрессии гена, кодирующего MSA-связывающий ТФ MYB3R1, так же демонстрирующий паттерн типа (II) (Рисунок 2.5 б). Единственный митотический циклин, паттерн экспрессии которого относится к группе (I), – *CYCA1;2*. Это позволяет предположить, что *in vivo* он играет определенную роль в регуляции клеточного цикла, не выявляющуюся при анализе культур клеток. Кроме того, экспрессия генов *CYCA3;2* и *CYCB1;3* относительно равномерна на всем протяжении временной серии, тогда как *CYCA2;1* обладает собственным уникальным паттерном. Таким образом, экспрессия большинства циклинов А- и В-типа соответствует паттерну (II).



**Рисунок 2.6.** Модель, объясняющая изменения в уровнях экспрессии генов, контролирующих клеточный цикл, при переходе к цветению. На стадии М4 клеточный цикл имеет низкую скорость из-за длинных фаз  $G_1$  и  $G_2$ ; кроме того, значительная доля клеток AM находятся вне клеточного цикла в фазе  $G_0$ . В момент перехода (М4-М5) клеточный цикл сокращается за счет уменьшения длительности фаз  $G_1$  и  $G_2$  и многие клетки входят в клеточный цикл из фазы  $G_0$ . Начиная со стадии М6 скорость клеточного цикла вновь снижается. Характерные экспрессионные паттерны генов, регулирующих клеточный цикл, схематически приведены в нижней части рисунка (митотические гены отображены красным, специфические для S-фазы – синим,  $G_1$  и  $G_2$  – зеленым,  $G_0$  – желтым).

Профили экспрессии генов, кодирующих циклины типа D, в наших данных весьма разнообразны. В синхронизированных суспендированных клеточных культурах эти гены имеют пики экспрессии в G<sub>1</sub>-фазе и в G<sub>1</sub>/S-переходе, хотя и с разными профилями: мРНК генов *CYCD5;1* и *CYCD3;3* накапливается в ранней G<sub>1</sub>-фазе и затем убывает по мере движения клеток к S-фазе, тогда как *CYCD4;1*, *CYCD4;2* и *CYCD3;1* увеличивают экспрессию в поздней G<sub>1</sub>-фазе с пиком в G<sub>1</sub>/S (Menges et al., 2002, 2005). Транскрипты девяти циклинов D-типа детектировались в наших данных, из них два (*CYCD3;2* и *CYCD3;3*) обладали паттерном типа (I), три (*CYCD2;1*, *CYCD4;2* и *CYCD6;1*) имели высокий уровень экспрессии только на стадии M1, один (*CYCD1;1*) характеризовался пиком на M9–M10 и остальные не показывали значительного изменения экспрессии в изученных стадиях развития АМ.

Большинство генов, кодирующих компоненты пути E2F/DP/RB, который регулирует G<sub>1</sub>/S-фазу (*DPA*, *DPb*, *E2Fa*, *E2Fb*, *E2Fc*, *RBR*, *WEE1* и *DEL3*, (Menges et al., 2005) имеют паттерн типа (II) (Рисунок 2.5 б), как и все три гена, кодирующие активаторы APC (Anaphase-Promoting Complex) (*CDC20-1*, *CDC20-2* и *CCS52B*), для которых в клеточных культурах показана экспрессия в митозе (Menges et al., 2005).

#### 2.4.6. Расширение предложенной модели

Еще более интересным было наблюдение, что некоторые гены, экспрессирующиеся на одной стадии клеточного цикла в культуре клеток, имели разные паттерны на стадиях M4–M5 временной серии. Например, для тубулинов и кинезинов ранее было показано одновременное увеличение экспрессии в G<sub>2</sub>/M-переходе (Breyne et al., 2002; Menges et al., 2002, 2005). Однако согласно нашим данным об экспрессии в АМ, тубулины имеют профиль типа (I), а кинезины – типа (II) (Рисунок 2.5 а и б). Такая же противонаправленная экспрессия в стадии перехода характерна для семейств специфических для S-фазы генов *PCNA* и *MCM*: *PCNA1* и *PCNA2* экспрессируются согласно профилю типа (I), а несколько генов *MCM* (*MCM3*, *MCM4*, *MCM5*, *MCM6*) – профилю типа (II) (Рисунок 2.5 а и б).

Мы предполагаем, что противоположное направление экспрессии генов, кодирующих регуляторы клеточного цикла, на стадии M4 может объясняться транскрипционной и физиологической реорганизацией в клетках, которые на этой стадии готовятся войти интенсивную пролиферацию на M5. Нами было выдвинуто предположение, что на стадии M4 та часть клеток вегетативной АМ, которая находится в длинном клеточном цикле в протяженной G<sub>2</sub>-фазе, начинает синхронно двигаться к митозу, тогда как клетки, задерживающиеся в G<sub>1</sub>-фазе, готовятся войти в S-фазу. На стадии M5 клетки быстро делятся без долгих G<sub>1</sub>- и G<sub>2</sub>-фаз, а на M6 и дальнейших стадиях пролиферация возвращается на нормальный уровень (Рисунок 2.6). Подготовительные изменения в ходе клеточного цикла в образце M4 могут определять

разнообразные паттерны экспрессии генов-регуляторов на этой стадии. Кроме того, в М5 некоторые неделящиеся клетки (в фазе  $G_0$ ), присутствующие в АМ, могут так же входить в клеточный цикл, что предполагает анализ профилей экспрессии некоторых циклинов D-типа.

#### 2.4.7. Соответствие наблюдаемых изменений в экспрессии генов данным гистологического анализа

Обширные изменения в статусе пролиферации клеток в вегетативной АМ в процессе перехода к цветению были показаны ранее, но не охарактеризованы на молекулярном уровне. Этот феномен был неоднократно описан с помощью цитологических методов для *Silene coeli-rosa* (L.) Godr. и *Sinapis alba* L., а так же для *A. thaliana* и других видов (Gonthier et al., 1987; Francis, 1992; Jacqmar et al., 2003; Bernier, 2013). Следует отметить, что во всех этих исследованиях переход к цветению индуцировался искусственно с помощью изменения фотопериода (переноса растений из условий КД в условия ДД).

В вышеперечисленных работах было показано, что число клеточных делений в АМ резко возрастает в процессе зацветания с уменьшением времени удвоения клеток с 20-80 часов (в зависимости от вида растения и зоны АМ) до нескольких часов (Bodson, 1975; Miller and Lyndon, 1975; Marc and Palmer, 1984; Gonthier et al., 1987; Jacqmar et al., 2003). Уменьшение длительности клеточного цикла происходило в основном за счет ускорения фаз  $G_1$  и  $G_2$  (Marc and Palmer, 1984; Gonthier et al., 1987), хотя так же было известно сокращение S-фазы (Francis and Lyndon, 1979; Gonthier et al., 1987; Jacqmar and Houssa, 1988). Последнее достигалось благодаря активации латентных ориджинов репликации ДНК и наблюдалось сразу перед переходом к репродуктивному развитию (Jacqmar and Houssa, 1988; Durdan et al., 1998). Изменения в ходе  $G_1$ - и  $G_2$ -фаз происходили в несколько этапов.  $G_2$ -фаза являлась самой длинной стадией клеточного цикла в вегетативной АМ, а  $G_1$ -фаза – на более поздних стадиях (Nougarede et al., 1991). В соответствии с детальным изучением происходящих в АМ цитологических процессов, на первой стадии сокращение длительности  $G_2$ -фазы быстро делящихся клеток происходит одновременно с возвращением в цикл неделящихся клеток, находящихся в  $G_2$ -фазе (Gonthier et al., 1987), что создает первую волну митозов в меристеме. Далее происходит укорочение фазы  $G_1$  и переход неделящихся клеток в  $G_0/G_1$ -фазах в S-фазу, давая начало второй волне митозов. Эти наблюдения согласуются с нашей гипотезой о протекающих в АМ процессах.

Кроме того, цитологические изменения в АМ включают в себя частичную синхронизацию клеточных делений (Francis and Lyndon, 1979; Nougarede et al., 1991). В некоторых случаях синхронизация может быть подавлена, что говорит о том, что это может быть скорее сопутствующим активации клеточных делений эффектом (Grose and Lyndon, 1984).

Сложная архитектура меристемы цветковых растений обуславливает гетерогенность клеточной популяции в АМ, включающую зоны быстро делящихся и неделящихся клеток. В процессе перехода к цветению скорость пролиферации клеток меняется разным образом в этих зонах (Bodson, 1975; Marc and Palmer, 1984; Nougarede et al., 1991). Все описанные процессы в совокупности образуют сложную картину, наблюдаемую на молекулярном уровне.

Следует отметить, что так как приведенные выше цитологические наблюдения были сделаны в системе переноса КД/ДД, нельзя точно сказать, какие изменения в делении клеток были вызваны самим фактом индукции, а какие отражают естественный процесс перехода к цветению. Однако по результатам нашего исследования, проведенного в условиях ДД без воздействия на растения, можно сказать, что изменения в клеточном цикле являются частью программы зацветания.

### 3.5. Заключение

Вторая часть нашей работы была посвящена подробному рассмотрению временной серии апикальных меристем, собранной в рамках транскриптомной карты *A. thaliana*. Временная серия АМ покрывала процесс перехода к цветению с разрешением в один день и была собрана в условиях длинного дня, индуктивных для зацветания *A. thaliana*. С помощью анализа дифференциально экспрессирующихся генов мы обнаружили значительные изменения в транскриптоме АМ между десятым и двенадцатым днем после прорастания, когда относительные уровни экспрессии главных репрессора и активатора зацветания, генов *FLC* и *LFY*, были равны друг другу. Кластерный анализ генов позволил выделить группы генов, экспрессия которых согласованно изменяется в этот временной период; большинство из них оказались связанными с регуляцией клеточного цикла, причем на одиннадцатый день после прорастания (стадия М5 временной серии) увеличивается экспрессия как генов, ассоциированных с S-фазой, так и генов, участвующих в митозе.

Для объяснения этой картины мы предлагаем следующую гипотезу: в десятый день после прорастания растений часть клеток меристемы выходит из фазы  $G_0$  в клеточный цикл и/или происходит укорочение фаз  $G_1$  и  $G_2$ , затем, на одиннадцатый день, происходит ускорение клеточного цикла, что может быть связано с реорганизацией меристемы перед активацией гена *LFY*, а на двенадцатый день продолжительность клеточного цикла восстанавливается на прежнем уровне, а состояние АМ меняется из вегетативного в репродуктивное.

## ГЛАВА 3. ОРГАНОСПЕЦИФИЧЕСКИЙ ОТВЕТ НА ХОЛОДОВОЙ СТРЕСС У *ARABIDOPSIS THALIANA*

### 3.1. Введение

Одним из анализов, проведенных нами при изучении транскриптомной карты *A. thaliana* (см. Главу 1), был поиск генов, дифференциально экспрессирующихся между всеми парами образцов, и дальнейший анализ обогащения этих генов категориями ГО. В 6 162 списках ДЭ генов мы обнаружили 1 528 категорий ГО, перепредставленных хотя в одном списке; из них 33 дочерних категорий термина «ответ на стресс» (GO:0006950) присутствовали в среднем в 203 из 6 413 сравнений (от 1 до 742 сравнения, (Klerikova et al., 2016), Таблица S14 Приложения). Активация генов стрессового ответа обнаруживалась не только при сравнении органов, собранных с разных растений, но и при анализе дифференциальной экспрессии между чашелистиками и лепестками одного цветка, что исключает вероятность стрессового воздействия на часть из собираемых образцов.

Таким образом, дифференциальная экспрессия генов стрессового ответа может свидетельствовать об их участии в разных аспектах нормальной жизнедеятельности растения. Как мутация с потерей функцией, так и сверхэкспрессия основных регуляторов реакции на холод, генов *CBF*, приводят к комплексным нарушениям развития растения в нестрессовых условиях (Park et al., 2015; Zhao et al., 2016); кроме того, в качестве рецептора низкой температуры выступает фитохром В, контролирующий зависимый от фотопериода рост (Jung et al., 2016). В транскриптомной карте *A. thaliana* профили экспрессии генов, аннотированных как участвующие в ответе на стресс, чрезвычайно разнообразны, и многие из них отличаются высокой тканевой специфичностью (Приложение, Рисунок 3.1). Это позволяет предположить, что в разных органах и тканях *A. thaliana* развитие стрессовой реакции может отличаться.

Адаптация растений к стрессовым условиям, в том числе к воздействию пониженной температуры, интенсивно изучается разнообразными методами (для обзора см. (Knight and Knight, 2012) и (Guo et al., 2018)). Однако подавляющее большинство исследований было проведено на листьях, проростках или цельных надземных частях растения (Fowler and Thomashow, 2002; Kilian et al., 2007; Robinson and Parkin, 2008; Rasmussen et al., 2013), оставляя в стороне рассмотрение вопроса возможности переноса полученных данных на все растение.

Таким образом, вопрос универсальности и специфичности стрессового ответа в разных органах *A. thaliana* требует дальнейшего изучения и будет рассмотрен нами в Главе 3. Различные стадии холодовой акклиматизации, а также сопряжение стрессового ответа с процессами роста и развития растения описаны в Обзоре литературы; в разделе 3.3 приведены результаты, полученные нами при изучении последствий холодового воздействия на шесть органов *A. thaliana*; полученные результаты обсуждаются в разделе 3.4.

## 3.2. Обзор литературы

### 3.2.1. Приспособление растений к холодовому стрессу

Прикрепленный образ жизни растений определяет их способ справляться с меняющимися условиями окружающей среды. Лишенные возможности избежать неблагоприятных условий, растения развили способность адаптироваться к широкому кругу воздействий и фенотипическую пластичность, позволяющую им приспособлять свою физиологию к внешним изменениям (Bartego-Gil and Salinas, 2018). Один наиболее важных факторов окружающей среды, оказывающий фундаментальное влияние на рост и развитие растения, – это температура. Для каждого вида растений характерен определенный температурный оптимум, зависящий от ареала обитания. При температуре ниже оптимальных значений растение подвергается холодовому стрессу, по механизму повреждений разделяющемуся на две категории: действие пониженных положительных температур и действие температур ниже 0°C (замораживание). Растения, подвергнутые действию низких положительных температур, развивают относительную устойчивость к замораживанию, связанную с способностью подавлять формирование кристаллов льда в клетках, что называется холодной акклиматизацией (Chinnusamy et al., 2010).

Развитие устойчивости к холоду включает в себя многочисленные структурные и метаболические изменения, включая аккумуляцию специфических белков и сахаров, предотвращающих связанную с замерзанием воды в клетке дегидратацию, изменение состава мембран для сохранения их текучести, ферментативный контроль содержания активных форм кислорода в клетке и защиту белков фотосистем и компонентов электрон-транспортной цепи (Ruelland et al., 2009; Theocharis et al., 2012).

Развитие холодной акклиматизации начинается с восприятия и передачи температурного сигнала в клетке. Механизмы, обеспечивающие эти процессы, изучены очень подробно, хотя дальнейшие исследования, безусловно, необходимы (для обзора см. (Guo et al., 2018)). Современная модель восприятия температуры предполагает, что первым этапом реакции на холод служит изменение текучести мембраны и связанная с притоком кальция в клетку реорганизация цитоскелета (Orvar et al., 2000; Pokorna et al., 2004; Hafke et al., 2013).

Следующая стадия включает в себя многочисленные изменения в транскриптоме (Kilian et al., 2007). Сложная система регуляторных белков, основными из которых выступают ТФ из небольшого семейства C-repeat binding factor (CBF), обеспечивает активацию так называемых COR генов (cold-response, отвечающих на холод) (Gilmour et al., 1998; Medina et al., 1999).. Экспрессия генов COR обеспечивает различные аспекты холодной акклиматизации.

При воздействии низкой температуры для лучшей выживаемости растения должен соблюдаться баланс между развитием и ростом и стрессовой реакцией, что достигается благодаря

сложной системе регуляции холодового ответа, учитывающей многочисленные внешние и внутренние сигналы (Hong et al., 2017; Zeng et al., 2017).

### *3.2.2. Механизмы восприятия пониженной температуры*

В первую очередь растение воспринимает внешние стимулы, в том числе пониженную температуру, с помощью рецепторных белков, заякоренных в плазматической мембране. Под действием холода текучесть мембраны снижается, что вызывает изменение мембранного потенциала и конформации трансмембранных белков (Orvar et al., 2000). Эти процессы сопровождаются реорганизацией элементов цитоскелета, необходимой для низкотемпературной индукции экспрессии генов в клетках растения (Pokorna et al., 2004), а также влияющей на поток ионов кальция (Hafke et al., 2013).

Репертуар известных температурных сенсоров растений постоянно пополняется. Среди них можно назвать мембранную протеинкиназу COLD-RESPONSIVE PROTEIN KINASE 1 (CRPK1), функция которой активируется при воздействии холода, что приводит к фосфорилированию белков семейства 14-3-3, переносящих сигнал в ядро (Liu et al., 2017b). Белки 14-3-3 способствуют дестабилизации транскрипционных факторов CBF, таким образом, CRPK1 необходима для предотвращения чрезмерной реакции на холод.

Ca<sup>2+</sup>-проводящие каналы – еще одни важные рецепторы холода, регистрирующие изменение потока ионов кальция через плазматическую мембрану. Большое семейство неселективных катион-проводящих каналов CNGC (Cyclic Nucleotide-Gated Channel), а также семейство GLR (GLutamate Receptor) участвуют в низкотемпературной акклиматизации *A. thaliana* (Meyerhoff et al., 2005; Hong-Bo et al., 2008). Следует отметить, что рецептор, кодируемый геном *AtGLR3.4*, реагирует на механическое воздействие (прикосновение к растению), что говорит об интеграции генетических сетей, контролирующей восприятие растением различных стимулов.

Такого рода интеграция была обнаружена также для холодовой акклиматизации и регуляции развития растений под влиянием света. Связывание рецептора красного света фитохрома В (phyB, подробнее см. раздел 2.2.2) с промоторами целевых генов зависит от температуры: под действием тепла phyB переходит из активной формы (PFR) в неактивную (PR). С помощью этого механизма определение ночного периода (сочетания темноты с понижением температуры) становится более точным (Jung et al., 2016; Legris et al., 2016).

### *3.2.3. Пути передачи холодового сигнала*

После восприятия внешнего стимула плазматической мембраной необходимо передать информацию в ядро, где разворачивается транскрипционный ответ. Наиболее изученными путями передачи холодового сигнала являются кальций-зависимые каскады фосфорилирования

СIPK (CBL-Interacting Protein Kinase, взаимодействующая с CBL протеинкиназа) и MAPK (Mitogen-Activated Protein Kinase — митоген-активируемая протеинкиназа).

Ионы  $\text{Ca}^{2+}$  служат мессенджерами, передающими внешние сигналы в клетку, у большого числа организмов. У растений содержание ионов  $\text{Ca}^{2+}$  в цитозоле возрастает под действием низких температур из-за притока свободного кальция из апопласта и вакуолей через  $\text{Ca}^{2+}$ -каналы (Chinnusamy et al., 2010). Изменение концентрации  $\text{Ca}^{2+}$  регистрируется кальций-связывающими белками, среди которых можно перечислить кальмодулин (CaM), CaM-подобные белки (CML),  $\text{Ca}^{2+}$ -зависимые протеинкиназы (CDPK) и кальциневрин B-подобные белки (CBL) (Kudla et al., 2018). Эти белки обладают мотивом спираль-петля-спираль (EF-hand), взаимодействие которого с ионами кальция запускает каскад сигналинга, приводящего к физиологическому ответу на стресс – изменению экспрессии генов, метаболизма, транспорта ионов и пост-трансляционным модификациям белков (Pareek et al., 2017). Специфичные для растений белки CBL взаимодействуют с серин-треониновыми киназами семейства СIPK, запуская каскад фосфорилирования (Huang et al., 2011).

Каскад фосфорилирования MAPK как правило содержит три протеинкиназы, последовательно фосфорилирующих друг друга. MAP киназа киназа киназа (MAPKKK, MAP3K, MEKK) фосфорилирует MAP киназы киназы (MAP2K, MKK, MEK) по консервативным серин/треониновым аминокислотным остаткам, после чего MAPKK активируются и становятся способны фосфорилировать MAP киназы (MAPK) по треониновым и тирозиновым аминокислотным остаткам. В свою очередь MAPK активируют различные белки-мишени, в том числе регуляторы холодового ответа (Teige et al., 2004). Активность каскада фосфорилирования MAPK регулируется кальций-связывающей киназой CRLK1 (CaM-regulated Receptor-Like Kinase), демонстрируя связность и взаимное регулирование путей ответа на холод на уровне передачи сигнала (Yang et al., 2010).

#### 3.2.4. Транскрипционная регуляция холодового ответа

Восприятие температурного сигнала и передача полученной информации в ядро ведет к активации экспрессии основных транскрипционных регуляторов ответа на холод – генов *CBF* (Gilmour et al., 2004). Три гена *CBF* (*C-REPEAT BINDING FACTOR*; *CBF1*, *CBF2*, *CBF3*) тандемно располагаются в геноме и относятся к семейству транскрипционных факторов APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) (Gilmour et al., 1998; Medina et al., 1999). CBF связываются с последовательностью CCAGC (C-повтор) в промоторах генов, принадлежащих к разным функциональным группам и в целом называемых COR (COLD-RESPONSE, отвечающих на холод), обеспечивая таким образом холодо-индуцибельную активацию их экспрессии (Stockinger et al., 1997). Существование трех генов-регуляторов ответа

на холод обеспечивают дублирование их функции, что подтверждается перекрытием их регулонов, и в следствии этого большую устойчивость системы (Jia et al., 2016; Zhao et al., 2016; Shi et al., 2017).

Интересно отметить, что сверхэкспрессия генов *CBF* и мутация с потерей функции приводят к одинаковому фенотипу растения: снижение скорости роста, карликовость и позднее зацветание (Gilmour et al., 2004; Jia et al., 2016; Zhao et al., 2016). Таким образом, как недостаток, так и избыток продуктов генов *CBF* приводит к уменьшению приспособленности растения, и экспрессия *CBF* должна регулироваться с большой точностью, принимая во внимание не только температурное воздействие, но и целый спектр других внешних и внутренних сигналов.

В условиях нормальной температуры гены *CBF* экспрессируются на низком уровне, но число их транскриптов возрастает уже после нескольких минут обработки холодом (Gilmour et al., 1998; Medina et al., 1999). Быстрый ответ на низкотемпературное воздействие обеспечивает постоянно присутствующий в клетках растения ТФ INDUCER OF CBF EXPRESSION1 (*ICE1*), принадлежащий к семейству белков bHLH (basic helix-loop-helix, «спираль-петля-спираль») (Chinnusamy et al., 2010). *ICE1* и его близкий гомолог *ICE2* связываются с мотивом E-box в промоторах генов *CBF* (Kim et al., 2015).

Холодовая активация белка *ICE1* происходит с помощью нескольких конкурирующих посттрансляционных модификаций. Убиквитин-лигаза HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1 (*HOS1*) убиквитинирует *ICE1*, что ведет к его деградации в 26S протеасоме и ингибированию акклиматизации (Dong et al., 2006), тогда как сумоилирование *ICE1* SUMO (small ubiquitin-related modifier) E3 лигазами *SIZ1* и *SIZ2* снижает уровень его убиквитинирования и разрушения, благодаря чему запускается реакция ответа на холод (Miura et al., 2007). Регуляция активности *ICE1* двумя противодействующими системами посттрансляционных модификаций консервативна как среди двудольных, так и у однодольных (Park et al., 2010).

Еще одна система регуляции *ICE1* с помощью конкуренции посттрансляционных модификаций включает в себя два каскада фосфорилирования MAPK. Протеинкиназы MPK3 и MPK6 фосфорилируют *ICE1*, уменьшая его стабильность и транскрипционную активность, тогда как каскад MEKK1-MKK1/2-MPK4 является их антагонистом, стабилизируя *ICE1* (Zhao et al., 2017; Liu and Zhou, 2018).

В течении суток температура окружающей среды может существенно изменяться, поэтому реакция на холодое воздействие интегрирована с циркадным циклом растения. Гены *CBF* обладают суточной периодичностью экспрессии, регулируемой компонентами циркадного осциллятора: ТФ семейства Myb CIRCADIAN CLOCK-ASSOCIATED1 (*CCA1*) активирует ответ на холод, связываясь с промоторами *CBF* (Dong et al., 2011), тогда как PSEUDO-RESPONSE

REGULATOR5 (PRR5), PRR7, PRR9 и TIMING OF CONSTANS1 (TOC1) группу его репрессируют (Keily et al., 2013).

Гены CBF являются также местом интеграции холодовой акклиматизации и внутренних гормональных сигнальных путей. Так, ТФ brassinостероидного сигнального пути BRASSINAZOLE RESISTANT1 (BZR1) и CESTA (CES) положительно регулируют холодовой ответ, связываясь с промоторами генов CBF (Eremina et al., 2016; Li et al., 2017), тогда как компоненты этиленового сигналинга репрессируют реакцию на пониженную температуру (Jiang et al., 2017). Количество биологически активных гиббереллинов под действием холода снижается, что ведет к накоплению белка DELLA, активирующего холодовой ответ (Richter et al., 2013). Путь абсцизовой кислоты (АБК) контролирует ответ растения на засуху и интегрирован с холодовой реакцией через протеинкиназу OPEN STOMATA1 (OST1), фосфорилирующую и стабилизирующую ICE1 (Ding et al., 2015). С другой стороны, жасмонатный сигналинг негативно влияет на холодовую акклиматизацию благодаря связыванию белков JASMONATE ZIM-DOMAIN (JAZ1 и JAZ4) с ICE1 и ICE2 (Hu et al., 2013).

### 3.2.5. Регулон генов CBF

Гены CBF кодируют транскрипционные факторы, регулирующие экспрессию генов, промоторы которых содержат мотив связывания CBF – С-повтор, а продукты необходимы для защиты от холода (Stockinger et al., 1997). Регулоны CBF изучались с применением различных методов, в том числе создания трансгенных линий *A. thaliana*, сверхэкспрессирующих CBF (Gilmour et al., 2004; Park et al., 2015), получения нокдаунов по этим генам (Novillo et al., 2004) и изучения мутантов *cbf*, в том числе недавно полученного с помощью технологии CRISPR/Cas тройного (Jia et al., 2016).

Среди генов, экспрессия которых изменяется у тройного мутанта *cbf1cbf2cbf3*, 73% обладают С-повтором в промоторной области; таким образом, остальные гены регулируются ТФ CBF опосредованно (Jia et al., 2016). Среди генов регулона CBF были найдены активируемые ими ТФ, такие как RAP2.1 (RELATED TO AP2.1), RAP2.7 и STZ/ZAT10 (SALT TOLERANCE ZINC FINGER), в свою очередь контролируемые субрегулоны CBF (Fowler and Thomashow, 2002), а также регуляторные киназы ERF4, RLP32 и SIPK25 (Jia et al., 2016).

В регулон CBF входят белки-криопротекторы из семейства дигидринов, относящиеся к группе LEA (Late Embryogenesis Abundant). Гидрофильные дигидрины участвуют в стабилизации клеточных мембран и мембранных белков при обезвоживании и воздействии холода (Thomashow, 1999; Breton et al., 2003). Помимо этого, CBF регулируют метаболизм жирных кислот через ADS2 (ACYL-LIPID DESATURASE2, кодирует D9-ацил-липид-десатуразу) (Maruyama et al., 2009), биосинтез пролина и антоцианов (Fowler and Thomashow, 2002; Jia et al.,

2016), концентрацию сахаров в клетке (Gilmour et al., 2004), что также влияет на сохранение текучести мембраны и стабилизацию белков. Денатурированные белки деградируются протеазами, экспрессия которых также растет по действием CBF (Levy et al., 2004).

Гены *CBF* не только сами регулируются разнообразными сигнальными путями, но активируют экспрессию генов, индуцирующихся также засухой, солевым стрессом и фитогормонами. В их число входят *ERD10*, *HVA22D* и *HVA22E*, участвующие в ответе на засуху, *LTI65/RD29B*, индуцирующийся высокой концентрацией солей (Jia et al., 2016).

Гены регулона CBF также находятся под контролем других ТФ, характеризующихся быстрым ответом на понижение температуры, в число которых входят *DEAR1*, *CZF1*, *ZAT10*, *ZAT12*, *ZF* (Park et al., 2015).

### 3.2.6. Баланс между развитием и защитой при холодовой адаптации

В меняющихся условиях окружающей среды растению, ведущему прикрепленный образ жизни, необходимо продолжать рост и развитие даже под стрессовыми воздействиями, затрагивающими эти процессы на всех уровнях, включая деление клеток, их дифференциацию и органогенез. Для правильного протекания роста и развития растения в условиях гипотермии необходимы сложные регуляторные механизмы (Hong et al., 2017; Zeng et al., 2017).

На уровне отдельных клеток важную роль в холодовой реакции имеют активные формы кислорода (АФК). АФК, продуцируемые различными ферментами, служат сигнальными молекулами во многих биотических и абиотических стрессах, включая холодовой. Их действие зависит от концентрации в клетке: небольшое количество АФК служит сигналом для запуска защитного ответа на стресс, тогда как высокие концентрации приводят к разрушению клеточной мембраны и гибели клетки, означая, что защитные механизмы не сработали эффективно. Таким образом, АФК регулируют баланс между ростом растения и защитными механизмами; в случае тяжелых стрессовых повреждений развивающийся в клетках окислительный стресс ограничивает вегетативный рост и репродуктивное развитие растения (Morales and Munné-Bosch, 2016; Mittler, 2017).

Тяжелый холодовой стресс вызывает также повреждения ДНК с последующей клеточной смертью, что было показано на клетках апекса корня *A. thaliana* (Hong et al., 2017).

На уровне органов низкотемпературный стресс приводит к увяданию и обесцвечиванию листьев, высыханию краев листовой пластинки, ускоренному старению и нарушению созревания плодов. Фотосинтетические процессы тесно связаны с производством АФК, поэтому под действием холода развитие хлоропластов риса регулируется специфическими генами, такими как *TCD10* (*THERMO-SENSITIVE CHLOROPHYLL-DEFICIENT MUTANT 10*), *TCD11*, *TSV* (*THERMO-SENSITIVE VIRESCENT MUTANT*) и *TSV3* (Gong et al., 2014; Wu et al., 2016; Sun et al., 2017; Wang

et al., 2017). Развитие боковых корней при холодовом стрессе также требует привлечения дополнительных транскрипционных факторов, таких, как *CYTOKININ RESPONSE FACTOR2* (*CRF2*) и *CRF3* (Jeon et al., 2016; Kim, 2016).

Кроме того, сами гены *CBF* участвуют в росте и развитии растения. Сверхэкспрессия *CBF* приводит к морфологическим отличиям от дикого типа (карликовости) и позднему цветению (Gilmour et al., 2004; Park et al., 2015), а тройной мутант *cbf1/cbf2/cbf3* характеризуется нарушениями в развитии семян, низкой всхожестью и замедленным ростом корней (Zhao et al., 2016).

### 3.2.7 Исследованность ответа на холод

Процесс приспособления растений к неблагоприятным температурным условиям интенсивно изучается на самых разных уровнях: транскриптомном (Gilmour et al., 2004; Park et al., 2015), метаболическом (Kaplan et al., 2004; Song et al., 2016), физиологическом (Hong et al., 2017; Mittler, 2017), популяционном (Varah et al., 2013). Подробная сеть генетической регуляции холодовой акклиматизации, включающая в себя как активацию/репрессию транскрипции генов (Jia et al., 2016), так и многочисленные посттрансляционные модификации белков (Park et al., 2011; Zhao et al., 2017), была создана с помощью изучения воздействия холода на трансгенные линии *A. thaliana*, сверхэкспрессирующие *CBF* (Gilmour et al., 2004; Park et al., 2015), и линии с инактивацией *CBF* (Novillo et al., 2004; Jia et al., 2016).

Большое количество работ посвящено широкомасштабному анализу транскриптома *A. thaliana* при развитии реакции на холод. В их число входит классическая работа Kilian *et al.*, посвященная анализу временной серии холодового воздействия и описывающая экспрессию генов в шести образцах с помощью микрочипов (Kilian et al., 2007), и другие исследования, использующие как технологию микрочипов (Fowler and Thomashow, 2002; Shen et al., 2017), так и секвенирование РНК (Leviatan et al., 2013; Calixto et al., 2018). Примечательно, что во всех перечисленных работах измерение экспрессии проводилось в надземных частях растения (что, как правило, означает целую розетку растения), листьях или семядолях (работы Kilian *et al.* и Shen *et al.* также включали образцы корня).

Проведенный нами анализ публичных экспрессионных данных, содержащихся в базе NCBI, показал, что абсолютное большинство транскриптомных экспериментов по изучению холодового воздействия на *A. thaliana* было проведено с использованием фотосинтезирующих органов. По запросу «Arabidopsis cold rna-seq» в разделе базы NCBI SRA (Sequence Read Archive) с января 2012 г. по апрель 2019 г. было найдено 23 проекта, из которых один сочетал результаты анализа экспрессии в семядолях и гипокотиле проростка (PRJNA320878), в одном проекте изучался холодовой ответ в корнях *A. thaliana* (PRJEB6701), одна работа была посвящена анализу микроРНК и их мишеней и среди других образцов включала в себя цветок (PRJNA190673, (Jeong

et al., 2013) и еще одна работа включена в выдачу поиска по ошибке (PRJNA242915). Аналогичные результаты получаются при поиске по запросу «*Arabidopsis cold microarray*» в разделе GEO базы данных NCBI.

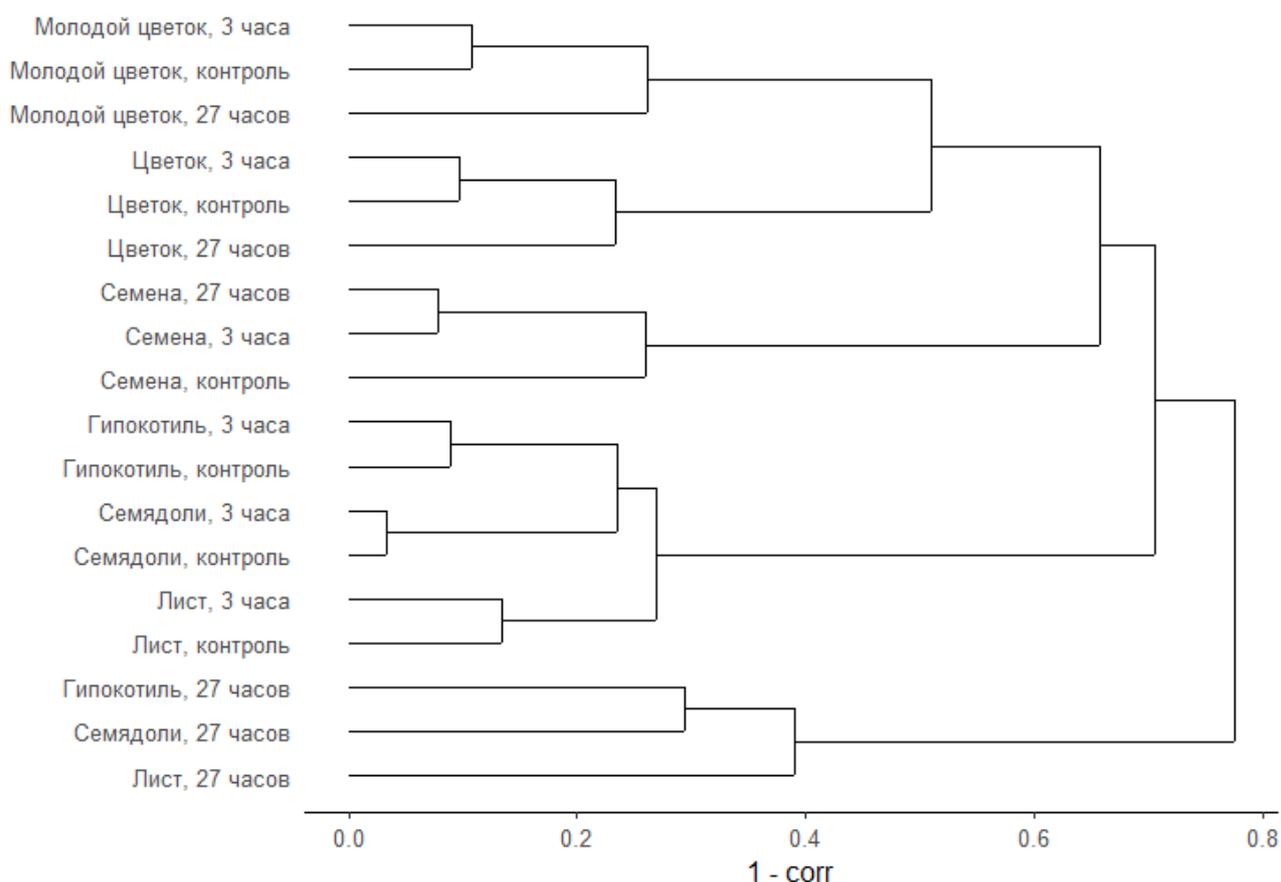
Резюмируя описанные выше факты, можно сказать, что, с одной стороны, изучение транскриптомного ответа на холод у *A. thaliana* ограничено фотосинтезирующими частями растения, но в то же время известна высокая степень интеграции холодовой реакции с процессами роста и развития (Park et al., 2015; Jung et al., 2016; Zhao et al., 2016). Это позволяет предположить различное протекание холодовой акклиматизации в органах и тканях растения. Следует также отметить, что, насколько нам известно, к моменту публикации нашей работы не было проведено ни одного сравнительного исследования ответа на холод в разных органах растения.

### 3.3. Результаты

#### 3.3.1. Выбор образцов и получение первичных данных

Целью нашей работы было изучение разнообразия транскриптомного ответа органов и частей *A. thaliana* на воздействие низких положительных температур. Для этого нами была рассмотрена кластеризация образцов атласа экспрессии генов *A. thaliana* (Рисунок 1.3); среди них мы выбрали наиболее удаленные в отношении транскрипционных профилей образцы, представляющие разные клады дерева кластеризации. Кроме того, на выбор образцов влияла простота их сбора, обеспечивающая легкое воспроизведение стадии развития образца. В результате, нами были взяты в анализ семядоли и гипокотиль однодневного проростка, большой зеленый лист, первый цветок на стадии распускания, цветок на стадии 9 (согласно (Smyth et al., 1990), далее называется «молодой цветок») и зрелые зеленые семена (подробное описание приведено в Приложении, Таблица 3.1).

Для всех выбранных частей растения был собран контрольный образец, выращенный при температуре +22° С, и два образца, подвергшихся воздействию пониженной температуры (+4° С) в течении 3 и 27 часов. Все пробы собирались в двух биологических повторностях.



**Рисунок 3.1.** Кластеризация изучаемых образцов. В качестве меры расстояния между образцами используется вычтенный из единицы  $r^2$ , посчитанный на основе уровней экспрессии генов.

Из собранных образцов была выделена тотальная РНК, использовавшаяся для приготовления библиотек. Библиотеки были секвенированы на платформе Illumina (Illumina, США) с получением 20-25 млн чтений (Klerikova et al., 2019), Таблица S22 Приложения). Пригодность биологических повторностей для дальнейшего анализа была проверена с помощью расчета квадрата коэффициента корреляции Пирсона; средний  $r^2$  составил 0,97 и все значения  $r^2$  лежали в диапазоне 0,9-1 (Приложение, Таблица 3.1). Результаты кластерного анализа на основе  $r^2$  приведены на Рисунке 3.1.

### 3.3.2. Анализ дифференциальной экспрессии

Для выявления генов, отвечающих на холодовое воздействие в различных частях растения, мы проанализировали дифференциальную экспрессию между тремя часами воздействия холода и контролем и между 27 часами и контролем для всех образцов (далее сравнения образцов с контролем будут называться «Название образца-3» и «Название образца-27» для 3 и 27 часов обработки холодом соответственно).

Число ДЭ генов значительно различалось между образцами, от 1 022 в Цветке-3 до 6 867 в Листе-27. Соотношение генов, экспрессия которых уменьшалась и увеличивалась в ответ на холодовое воздействие, также было различным как для образцов в одной временной точке (от 0,57 в Листе-3 до 1,32 в Гипокотиле-3), так и между разным временем стрессового воздействия на один образец (от 0,52 в Цветке-3 до 1,40 в Цветке-27) (Приложение, Таблица 3.2).

В итоге, 15 459 генов являлись ДЭ хотя бы в одном сравнении из 12, то есть их экспрессия активировалась или репрессировалась в ответ на холод; все дальнейшие анализы проводились с использованием только этих генов, полный список которых с указанием изменения уровня экспрессии приведен в Klerikova et al., Таблице S4 Приложения (Klerikova et al., 2019).

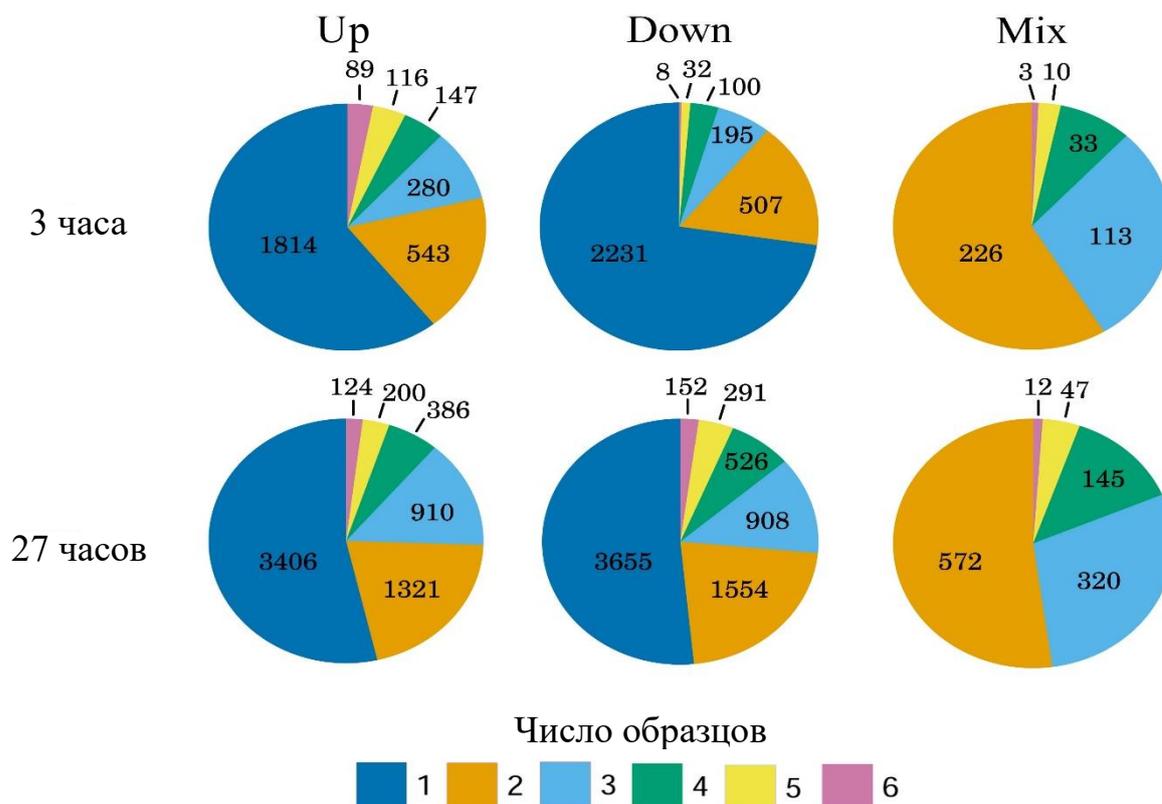
### 3.3.3. Изменение экспрессии генов регулона *CBF* в различных органах

Высокая степень изученности развития реакции растения на снижение температуры позволяет сравнить результаты нашего эксперимента с литературными данными. В работе Park et al. были определены регулоны генов, кодирующих ТФ *CBF1-3*, с помощью секвенирования РНК трансгенных линий со сверхэкспрессией генов *CBF* под конститутивным промотором вируса мозаики цветной капусты (CaMV) 35S (Park et al., 2015). Для образцов Листа, Семядолей и Гипокотиле изменение экспрессии большинства (75% в среднем) ДЭ генов происходило в одну сторону (увеличения или уменьшения) в трансгенных линиях Park et al. и в наших данных (процент совпадения списков генов варьировался от 50% у Гипокотиле-3 до 93% в Листе-27) (Приложение, Таблица 3.3). С другой стороны, для цветков обеих стадий средний процент пересечения списков равнялся 59%, а наименьшее совпадение было у Семян-3 (35%).

### 3.3.4. Определение общих и органоспецифических ДЭ генов

Так как хорошо известно, что ранний и поздний транскрипционные ответы на холодовое воздействие существенно различаются (Kilian et al., 2007), все дальнейшие анализы проводились отдельно для образцов после 3 и 27 часов воздействия пониженной температуры.

Как было сказано в разделе 1.3.5, гены, экспрессия которых увеличивается и уменьшается при действии холода по сравнению с контрольным образцом, могут быть вовлечены в разные биологические процессы, и желательно рассматривать их отдельно. Чтобы оценить изменение профиля экспрессии гена интегрально по всем образцам, 15 459 ДЭ генов были разделены на три категории: гены, экспрессия которых значительно уменьшается хотя бы в одном образце по сравнению с контролем и нигде не увеличивается (категория Down), гены, экспрессия которых возрастает хотя бы в одном образце и нигде не уменьшается (категория Up), и гены, экспрессия которых хотя бы в одном образце снижается и хотя бы в одном увеличивается (Mix). В образцах *A. thaliana* после 3 часов при +4°C число генов в категории Down составило 3 073, в Up – 2 989 и 385 генов относилось к категории Mix. Для 27 часов воздействия соответствующие числа были существенно выше: 7 086 генов – Down, 6 347 – Up и 1 096 – Mix.



**Рисунок 3.2.** Число генов, дифференциально экспрессирующихся в разном числе образцов для трех и 27 часов воздействия холода отдельно.

В каждой из категорий число генов, являющихся ДЭ во всех образцах, было крайне небольшим: Down-3 – 8 генов, Up-3 – 89 и Mix-3 – 3, но несколько возросло для образцов,

собранных после 27 часов гипотермии: 152, 124 и 12 для Down, Up и Mix соответственно (Рисунок 3.2). В категории Up как на 3, так и на 27 часах, наблюдалось значительное обогащение терминами, связанными со стрессовым ответом (например, GO:0009631, «адаптация к холоду») (Klerikova et al., 2019), Таблица Приложения S8).

Для генов, демонстрирующих снижение уровня экспрессии во всех образцах после трех часов холодого воздействия, не было обнаружено обогащения какими-либо категориями ГО даже после присоединения к ним генов, ДЭ в пяти из шести образцов, из-за недостаточного для анализа количества генов (всего 40). Однако гены, относящиеся к группе Down-27 и ДЭ во всех образцах, были существенно обогащены терминами, связанными с фотосинтетическими процессами (Klerikova et al., 2019), Таблица Приложения S8).

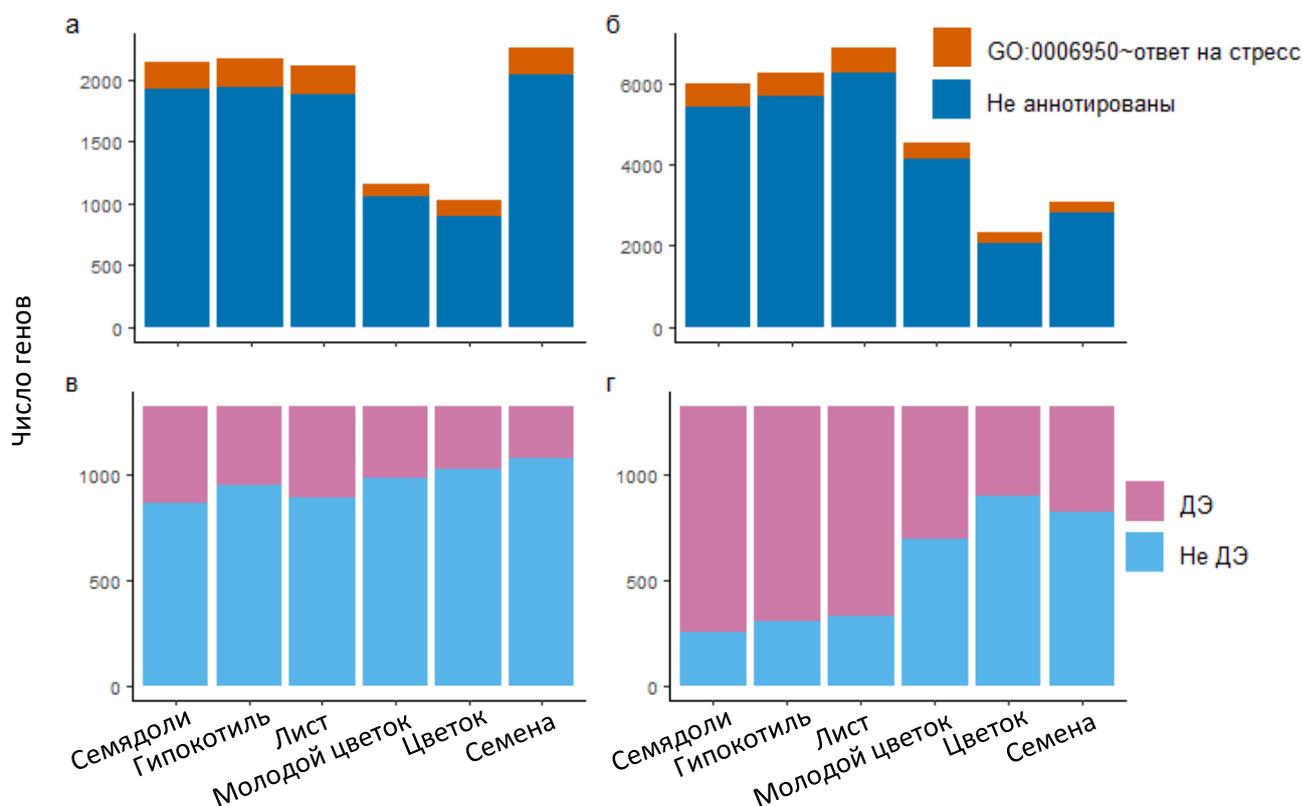
Число генов, изменяющих экспрессию в процессе холодого ответа и являющихся специфическими для конкретных органов и частей растения, различалось в диапазоне от 90 для генов с повышенной экспрессией в Цветке-3 и Молодом Цветке-3 до 1 115 для генов со сниженной экспрессией в Листе-27 (Приложение, Таблица 3.4). Всего в образцах после 3 часов стресса было обнаружено 4 045 органоспецифических генов и 6 780 – после 27 часов.

Уникальные для образцов гены имели чрезвычайно разнообразное обогащение, специфичное для разных органов и частей растения. Гены, обладающие сниженной экспрессией в Семядолях-3, были обогащены терминами, связанными с пентатрикопептидными повторами (PPR) (Klerikova et al., 2019), Таблица S9 Приложения). Эти же категории перепредставлены в генах с увеличенной экспрессией в Семядолях-27 (Klerikova et al., 2019), Таблица S10 Приложения). Среди них ген *SOARI* кодирует PPR-содержащий белок, участвующий в регуляции ответа на холод, недостаток воды и солевой стресс (Jiang et al., 2015). В образце Листа-27 гены со снижением экспрессии были обогащены категориями, относящимися к фотосинтезу (Klerikova et al., 2019), Таблица S10 Приложения). В Семенах-3 и -27 гены со сниженной экспрессией были связаны со сплайсингом РНК, ответом на присутствие грибов и транспортом липидов, а с повышенной – с запасанием липидов (Klerikova et al., 2019), Таблицы S9 и S10 Приложения). Гены с повышенной экспрессией в Молодых Цветках-27 характеризовались перепредставленностью категорий, связанных с катаболическими процессами и развитием пыльцы (Klerikova et al., 2019), Таблица S10 Приложения).

### 3.3.5. Поведение известных генов, участвующих в ответе на холод

Поскольку наряду с общими для стрессового ответа генами существуют также органоспецифические, представляется интересным изучить участие этих генов в развитии устойчивости к холоду. Для этого мы рассмотрели аннотацию ГО этих генов: только небольшая доля ДЭ генов была аннотирована категорией GO:0006950 (ответ на стресс) и дочерними

категориями, варьируясь от 8 до 12% как для трех, так и для 27 часов воздействия холода; при этом абсолютные числа генов различались от 106 в Молодых Цветках-3 до 328 в Листе-27 (Рисунок 3.3 а и б).



**Рисунок 3.3.** Число генов, аннотированных как участвующие в ответе на стресс, среди ДЭ генов после (а) трех и (б) 27 часов холодного воздействия, и число генов COR, ДЭ в изучаемых нами образцах после (в) трех и (г) 27 часов холодного воздействия.

В работе Park et al. был определен список генов, транскрипция которых меняется, во-первых, в трансгенных растениях со сверхэкспрессией одного из генов *CBF*, а во-вторых, при воздействии пониженной температуры; далее в нашей работе именно эти гены будут называться «отвечающими на действие холода» (COR, cold-response) (Park et al., 2015). 60% из 1 322 COR-генов были ДЭ хотя бы в одном образце после трех часов холодного воздействия (из них 38% принадлежали к категории Down, 56% – к Up и только 6% – к Mix) (Klerikova et al., 2019), Таблица S11 Приложения). Если рассматривать образцы в отдельности, процент ДЭ генов среди COR-генов варьировался от 22% в Семенах-3 до 35% в Семядолях-3 (Рисунок 3.3 в). После 27 часов при +4°C картина становится еще более выраженной: 99% COR-генов были ДЭ хотя бы в одном образце, и, в зависимости от образца, процент ДЭ генов составлял от 79-85% в Семядолях-27, Листе-27 и Гипокотиле-27 до 36, 51 и 41% в Цветке-27, Молодом Цветке-27 и Семенах-27 соответственно (Рисунок 3.3 г).

Экспрессия основных генов-регуляторов *CBF1*, *CBF2* и *CBF3* активируется в течении нескольких минут после понижения температуры (Gilmour et al., 1998; Medina et al., 1999), однако

эти гены не уникальны в своем раннем ответе на стресс. В работе Park et al. с использованием временной серии Kilian et al. было идентифицировано еще 27 транскрипционных факторов, имеющих сходный с CBF профиль экспрессии (так называемые гены «первой волны») (Kilian et al., 2007; Park et al., 2015). Первое измерение экспрессии было произведено нами через три часа после начала воздействия низкой температуры, что не является самой ранней из возможных временных точек, однако мы все еще наблюдаем изменение экспрессии всех генов первой волны хотя бы в одном образце. Четыре из тридцати генов были ДЭ во всех образцах после трех часов при +4°C (включая *CBF3*), остальные же имели заметно различающиеся паттерны экспрессионных изменений (Klepikova et al., 2019), Таблицы S13 и S14 Приложения). Экспрессия пяти генов менялась разнонаправленно в разных органах, причем отличия в уровне изменения экспрессии были существенными (например, уровень изменения экспрессии гена *ZAT12* в Гипокотиле-3 – 0,22, а в Листе-3 – 2,96). После 27 часов холодового воздействия разнообразие ответа на стресс стало еще более заметно: в большинстве образцов транскрипция генов первой волны вернулась на уровень контроля или снизилась по сравнению с ним, тогда как в Листе-27 19 генов все еще имели повышенную экспрессию (Klepikova et al., 2019), Таблица S14 Приложения).

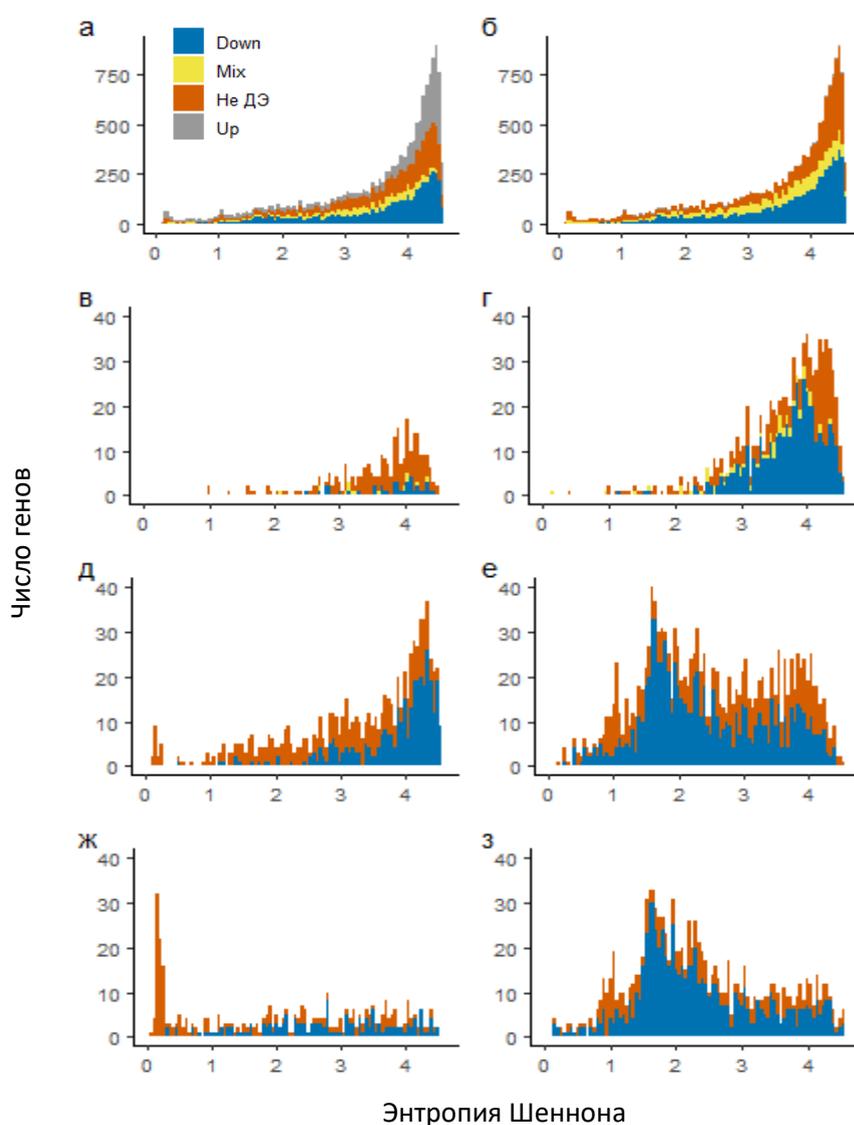
### 3.3.6. Экспрессионные характеристики генов, отвечающих на холодовое воздействие – энтропия Шеннона

Для дальнейшего изучения органоспецифических генов стрессового ответа мы охарактеризовали их с помощью параметров экспрессии, основанных на транскриптомной карте *A. thaliana* (см. Главу 1).

Энтропия Шеннона (H) используется для оценки широты паттерна экспрессии (см. раздел 1.3.7). Распределение энтропии для 15 459 генов, меняющих экспрессию в ответ на стресс, практически совпадает с распределением для всех экспрессирующихся генов (Рисунок 1.8), и, таким образом, сильно смещено в область высоких значений H (соответствующих широкому паттерну экспрессии) и обладает небольшим пиком на очень низкой H, куда входят гены, экспрессирующиеся в органо-, ткане- или стадия-специфичной манере (Рисунок 3.4 а и б). Гены, экспрессия которых изменялась при гипотермии хотя бы в пяти органах, не образовывали пика распределения на низких значениях H (Рисунок 3.4 в и г), а уникальные для образцов генов в большинстве случаев имели сходное с общим распределение энтропии Шеннона (Приложение, Рисунок 3.2).

Исключение составляли несколько образцов, обладающих характерными свойствами. Распределение энтропии генов, имеющих увеличенную экспрессию в образце Лист-3, обладает локальным пиком на низких (0-0,3) значениях энтропии (Рисунок 3.4 д), что говорит об их

специфичной экспрессии в ограниченном числе тканей. Мы рассмотрели паттерн экспрессии этих генов в транскриптомной карте *A. thaliana*. В условиях отсутствия стрессового воздействия экспрессия всех генов за одним исключением была ограничена зрелыми пыльниками и целыми цветками, содержащими пыльники на той же стадии. Хотя эти гены являются уникальными для образца Листа после трех часов холодового воздействия, половина из них (44%) так же были ДЭ в Листе-27 и 33% – в Цветке-27.



**Рисунок 3.4.** Распределение энтропии Шеннона среди всех генов, отвечающих на холодовое воздействие, с указанием типа изменения экспрессии после (а) трех и (б) 27 часов при +4°C; среди генов, ДЭ в пяти или шести образцах при (в) трех и (г) 27 часах действия холода; среди генов, специфически участвующих в холодовой акклиматизации в (д) Листе-3, (е) Семенах-3, (ж) Цветке-27 и (з) Семенах-27.

Функция большинства генов из этой группы заключалась в контроле состояния клеточной стенки и роста пыльцевой трубки. Наиболее выраженными изменениями в уровне транскрипции характеризовались гены, кодирующие пектиновые метилтрансферазы.

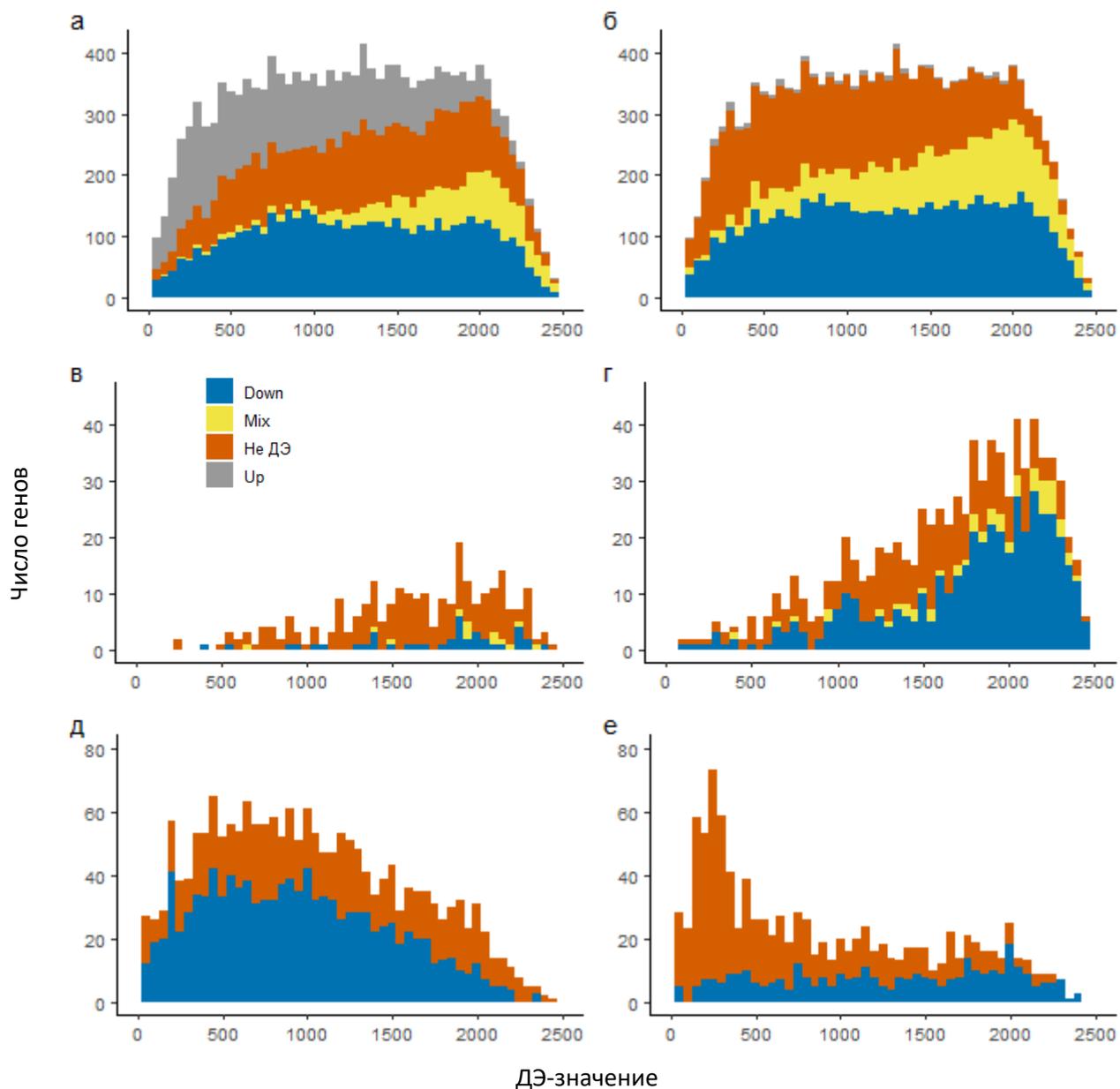
Распределение энтропии  $H$  генов, имеющих повышенную транскрипцию исключительно в Цветке-27, также имеет пик на низких значениях (Рисунок 3.4 ж). Образующие его гены специфичны для пыльцы. Анализ обогащения терминами ГО показывает перепредставленность категорий, связанных с ростом пыльцевой трубки и модификацией клеточной стенки (Klerikova et al., 2019), Таблица Приложения S16).

Еще один образец с определенными особенностями распределения энтропии  $H$  – Семена-3. Как для генов, экспрессия которых повышена после воздействия  $+4^{\circ}\text{C}$  в течении трех часов, так и для генов со сниженной экспрессией распределение энтропии является бимодальным (Рисунок 3.4 е и з). Вместе с этим ДЭ гены образца Семена-27 не имеют пика в области высоких  $H$ , а список генов, энтропия которых имеет средние значения, сильно перекрывается со списком для Семян-3. В обеих временных точках гены, имеющие сниженную экспрессию и среднее значение  $H$ , обладали обогащением терминами, связанными с доменом F-box и с защитой растений (Klerikova et al., 2019), Таблица Приложения S17), тогда как гены, экспрессия которых возрастает, ассоциированы с метаболизмом и запасанием липидов.

### *3.3.7. Экспрессионные характеристики генов, отвечающих на холодовое воздействие – ДЭ-значение*

Еще один параметр экспрессии, использовавшийся в работе над транскриптомной картой *A. thaliana*, – это ДЭ-значение (см. раздел 1.3.5). Мы проанализировали распределение ДЭ-значений для разных групп генов, меняющих свою экспрессию в стрессовых условиях. Распределение ДЭ-значений для всех 15 459 стрессовых генов по сравнению с общим распределением обладает меньшим числом генов с низким ДЭ-значением (ср. Рисунок 1.6 и 3.5 а), а гены, входящие в категории Mix-3 и Mix-27, были смещены в область высоких ДЭ-значений.

ДЭ гены, общие хотя бы для пяти органов, имеют распределение ДЭ-значений, смещенное к правой части (Рисунок 3.5 б), тогда как у органоспецифических ДЭ генов оно немного смещено влево (Приложение, Рисунок 3.3, Рисунок 3.5 в). Большинство образцов не имеет заметных особенностей распределения (Приложение, Рисунок 3.3), за исключением Молодого Цветка-27, где выраженный пик распределения в области ДЭ-значений меньше 500 составляли гены с увеличенной при холодовом воздействии экспрессией и небольшое количество генов с уменьшенной экспрессией (Рисунок 3.5 в). Анализ обогащения категориями ГО связывает эти гены с убиквитин-зависимыми процессами катаболизма белков, а также с регуляцией транскрипции и доменом F-box (Klerikova et al., 2019), Таблица Приложения S18).



**Рисунок 3.5.** Распределение ДЭ-значения среди всех генов, отвечающих на холодовое воздействие, с указанием на тип изменения экспрессии после (а) трех и (б) 27 часов при +4°C; среди генов, ДЭ в пяти или шести образцах при (в) трех и (г) 27 часах действия холода; среди генов, специфически участвующих в холодной акклиматизации в (д) Листе-27, (е) Молодом Цветке-27.

### 3.3.8. Анализ регуляторных элементов в промоторных областях ДЭ генов

Для более глубокого понимания структуры генетических сетей, управляющих холодной акклиматизацией в разных органах и частях растения, мы проанализировали представленность регуляторных элементов различных ТФ в промоторах нескольких групп ДЭ генов (например, ДЭ генов, общих хотя бы для пяти органов, уникальных для образцов ДЭ генов, генов с определенной энтропией). Как и ожидалось, в промоторных регионах ДЭ генов, экспрессия которых возрастает хотя бы в пяти органах (генов «ядра» стрессового ответа) чаще, чем в среднем

по геному, встречались регуляторные элементы транскрипционных факторов CBF (CRT/DRE-регуляторные элементы). В свою очередь, гены «ядра» стрессового ответа, имеющие CRT/DRE-регуляторные элементы в промоторах, обогащены категориями ГО, связанными с ответом на стресс. В их число входят 12 ТФ; регуляторные мотивы этих ТФ также перепредставлены в списке генов, ДЭ хотя бы пяти образцах. Регулируемые ими гены также обогащены стрессовыми терминами. Четыре из этих 12 ТФ принадлежат к семейству ERF/AP2, и их участие в биотическом и абиотическом стрессовом ответе было известно ранее (Tsutsui et al., 2009). В списках генов, ДЭ только в одном образце, перепредставленности регуляторных элементов CBF не наблюдается, что говорит о существовании отличных от них факторов, управляющих органоспецифическим стрессовым ответом. В частности, мы нашли перепредставленность регуляторных элементов 19 ТФ из семейства NAC в промоторах ДЭ генов, экспрессия которых повышается в Молодом Цветке-27.

### 3.4. Обсуждение

#### 3.4.1. Предварительный анализ данных

Рассмотрение единства и уникальности развития устойчивости к низким температурам у растения требует дифференцированного анализа транскриптомного ответа на холод в разных органах. Атлас экспрессии генов *A. thaliana* дает представление о транскриптомной схожести разных органов, тканей и частей растения в нормальных, нестрессовых условиях и позволяет выбрать наиболее далекие образцы. Используя дерево кластеризации (Рисунок 1.3), мы выбрали образцы, коэффициент корреляции  $r^2$  между которыми не превышал 0,6: семядоли и гипокотиль однодневного проростка, большой зеленый лист, первый цветок на стадии распускания, цветок на стадии 9 (согласно (Smyth et al., 1990), далее называется «молодой цветок») и зрелые зеленые семена.

Все образцы собирались в одно и тоже время (6 часов после включения света) при суточном цикле 12 часов света/12 часов темноты, чтобы минимизировать влияние циркадного ритма на результаты, так как хорошо известно, что экспрессия многих генов стрессового ответа подчиняется циркадному ритму (Harmer et al., 2000; Fowler, 2005). Таким образом, растения подвергались воздействию пониженной температуры (+4° C) в течении 3 и 27 часов.

После получения и обработки результатов секвенирования на платформе Illumina, а также проверки на пригодность биологических повторностей к дальнейшим анализам, квадрат коэффициента корреляции Пирсона был использован для построения дерева кластеризации образцов. Результаты кластерного анализа показывают, что семена, цветки и молодые цветки объединялись в отдельные клады по органам, вне зависимости от наличия и продолжительности обработки холодом, тогда как листья, семядоли и гипокотиль были очень похожи друг на друга и частично группировались по времени стрессового воздействия (Рисунок 3.1), что может объясняться фотосинтетической активностью этих органов, определяющей основные процессы, происходящие в процессе адаптации к гипотермии.

Следует, однако, отметить, что несмотря на совместную группировку листьев, семядолей и гипокотилия, расстояние между их кладами составило от 0,26 до 0,29 (Рисунок 3.1). Коэффициенты  $r^2$  между контрольными и подвергшимися трехчасовому стрессу образцами находились в диапазоне от 0,5 до 0,94, а между контролем и 27 часами – от 0,12 до 0,64; таким образом, нам удалось отобрать как в высокой степени разнообразные в отношении транскриптома органы растения, так и выбрать время стрессового воздействия, значительно влияющие на экспрессию генов.

В соответствии с этим, мы обнаружили большое число генов, ДЭ между контрольными образцами контроля и подвергнутыми гипотермии (от 1 022 до 6 867 в зависимости от органа и

времени воздействия). Качественные различия в транскриптомном ответе на стресс в разных органах наблюдаются уже на уровне соотношения числа генов, экспрессия которых активируется и репрессируется в ответ на пониженную температуру (от 0,57 до 1,32; Приложение, Таблица 3.2).

Всего нами было выявлено 15 459 генов, ДЭ хотя бы в одном сравнении из 12, что составляет 46% от всех генов *A. thaliana* (33 323); такая большая доля показывает существенные изменения, происходящие при адаптации к гипотермии во всем растении.

Как обсуждалось в разделе 3.2.7, подавляющее большинство транскриптомных исследований адаптации растений к гипотермии проводилось на листьях или целых проростках, что позволяет с одной стороны проверить, насколько результаты нашего эксперимента совпадают с литературными данными, а с другой – оценить разницу между традиционно изучаемыми органами *A. thaliana* и менее исследованными, такими, как цветки и семена. Для сравнения мы использовали результаты секвенирования РНК листьев трансгенных растений со сверхэкспрессией одного из генов *CBF* (Park et al., 2015). Транскриптомы зеленых частей растения – листа, семядолей и гипокотилия – демонстрировали согласованные изменения в трансгенных линиях и в наших данных: процент генов, экспрессия которых менялась однонаправленно, доходил вплоть до 93% у Листа-27 (Приложение, Таблица 3.3). Однако у Цветка, Молодого цветка и Семян средний процент совпадения был существенно ниже (59% против 75% у зеленых органов) и в минимальном случае составлял 35%. Таким образом, анализ наших данных дает непротиворечивые результаты по сравнению с уже исследованными системами и одновременно с тем позволяет получить новую информацию об органоспецифических паттернах стрессового ответа.

#### 3.4.2. Анализ общих и органоспецифических ДЭ генов

Как было сказано выше (см. раздел 3.4.1), 15 459 генов характеризовались изменениями в экспрессии под действием холода, однако все они имели разные паттерны дифференциальной экспрессии: некоторые были ДЭ во всех образцах, тогда как другие активировались или репрессировались только в отдельных органах. Эта картина также усложнялась различиями в уровнях изменения экспрессии между разными образцами, вплоть до противоположных изменений под действием пониженной температуры.

Для первичного разделения всех ДЭ генов мы использовали направление изменения экспрессии при обработке холодом. Гены были разделены на три категории – Down, Up и Mix. Большинство генов как в случае 3 часов холодового воздействия, так и в случае 27 часов принадлежали к категориям Down (3 073 и 7 086 для 3 и 27 часов соответственно) и Up (2 989 и 6 347), а к категории Mix относилось только небольшое число (385 и 1 096).

Среди разделенных таким образом генов мы провели поиск общих для всех образцов ДЭ генов, которые могут служить универсальной базой стрессового ответа, и уникальных для отдельных органов ДЭ генов, вероятно, обеспечивающих специфичность ответа на холод. Число общих ДЭ генов было крайне небольшим (от 8 генов в Down-3 до 152 в Down-27) (Рисунок 3.2). Тем не менее, категории Up-3 и Up-27 характеризовались перепредставленностью терминов, связанных с развитием реакции на стресс (Klerikova et al., 2019), Таблица Приложения S8). Малое количество общих для всех образцов ДЭ генов и их сильное обогащение категориями ГО, связанными со стрессовым ответом, позволяет предположить, что эти гены являются наиболее базовыми в развитии реакции на стресс и образуют «ядро» ответа на холод. Повышение числа общих генов на 27 часах свидетельствует о значительной специфичности раннего ответа на стресс в органах и частях растения, который становится более универсальным на поздних стадиях стрессовой реакции. Гены группы Down-27, ДЭ в пяти и более образцах, имели обогащение категориями ГО, связанными с фотосинтетическими процессами, что соответствует известному факту ингибирования фотосинтеза при низких температурах (Stitt and Hurrly, 2002) (Klerikova et al., 2019), Таблица Приложения S8). Холодовое воздействие вызывает дисбаланс между продолжающейся активностью фотосистем и зависящей от температуры темновой частью фотосинтеза, из-за чего происходит накопление АФК в клетке. Ингибирование фотосинтеза предотвращает этот процесс, защищая структуры клетки (Khanal et al., 2017).

Другая сторона ответа на холод, которую мы изучали, – это специфический для органов ответ. Здесь необходимо еще раз отметить, что 3 и 27 часов рассматривались отдельно, поэтому ДЭ ген может считаться уникальным для органа после трех часов холодового воздействия, хотя и иметь изменения в экспрессии в других образцах на 27 часах. 62% (4 045) генов из 6 447 были ДЭ только в одном образце после трех часов гипотермии и 47% (6 780) из 14 529 – после 27 часов. Если рассматривать образцы в отдельности, то число органоспецифических генов различалось от 18% до 63% в Цветке-3 и Семенах-3 и от 15% до 35% в Семядолях-27 и Семенах-27 (Приложение, Таблица 3.4). Таким образом, на поздних стадиях развития холодовой акклиматизации транскриптомный ответ на стресс в заметной степени унифицируется.

В то время как общие для всех образцов ДЭ гены были обогащены связанными со стрессом категориями, органоспецифические гены существенно различались по своим функциям между образцами и были связаны с разными процессами. Нам удалось проследить динамику экспрессии генов, кодирующих белки с пентатрикопептидными повторами в семядолях (Klerikova et al., 2019), Таблицы S9 и S10 Приложения): такие гены были перепредставлены среди генов с уменьшением экспрессии после 3 часов и увеличением после 27. Известно, что гены, содержащие этот мотив, играют важную роль в биогенезе органелл, редактировании РНК и общей стрессовой реакции (Lurin et al., 2004; Kotera et al., 2005). Участие гена *SOARI*,

кодирующего PPR-содержащий белок, в АБК-зависимой регуляции ответа на холод, недостаток воды и солевой стресс было известно ранее (Jiang et al., 2015), таким образом в зеленых семядолях развитие холодовой реакции в существенно интегрировано с универсальным ответом на абиотический стресс. Уникальные для Листа-27 гены со сниженной под действием холода экспрессией были обогащены категориями, относящимися к фотосинтезу (Клерикова et al., 2019), Таблица S10 Приложения), так же как и полный список генов, репрессированных в Листе-27, что отражает важность торможения фотосинтеза и уменьшения вреда, наносимого накапливающимися активными формами кислорода при действии абиотического стресса (Morales and Munné-Bosch, 2016; Mittler, 2017). Как в Семенах-3, так и в Семенах-27 ДЭ гены, экспрессия которых понижалась в ответ на стресс, участвовали в регуляции сплайсинга РНК, ответа на заражение грибами и транспорта липидов, а ДЭ гены, число транскриптов которых возрастало, были связаны с запасанием липидов (Клерикова et al., 2019), Таблицы S9 и S10 Приложения). Изменения в составе клеточных жиров, ведущие к стабилизации мембраны, являются известной частью механизма холодовой устойчивости (Li et al., 2015) и могут быть особенно важны в этом случае, так как накопление липидов в развивающемся семени является необходимым для репродукции растения процессом (Penfield et al., 2004). Гены, экспрессия которых возрастала в Молодых Цветках-27, характеризовались категориями ГО, описывающими катаболические процессы и развитие пыльцы (Клерикова et al., 2019), Таблица S10 Приложения).

Большое разнообразие процессов, происходящих при холодовой адаптации в разных органах растения, служит доказательством существования органоспецифического ответа на стресс, расширяющего и дополняющего базовую систему генетической регуляции. Таким образом, выводы, сделанные при изучении реакции на гипотермию одного органа (например, листа), имеют ограниченную применимость к другой части растения (семени), что особенно необходимо учитывать в практическом применении полученных данных, например, при создании устойчивых к холоду трансгенных растений.

### 3.4.3. Неравномерная изученность органов *A. thaliana*

В предыдущем разделе мы описали разделение ДЭ генов на общие для всех образцов и органоспецифические. Как показывает анализ аннотации ГО *A. thaliana*, участие в адаптации к низкой температуре известно для генов, изменяющих экспрессию во всех или в большинстве образцов, тогда как уникальные для образцов гены не демонстрировали обогащения терминами ГО, относящимися к ответу на стресс. Чтобы уточнить долю генов, для которых участие в развитии реакции на холод было известно ранее, мы разделили ДЭ гены каждого образца на аннотированные категорией GO:0006950 (ответ на стресс) и дочерними категориями (8-12%) и неаннотированные (большинство) (Рисунок 3.3 а и б). Таким образом, в аннотации ГО *A. thaliana*

сильно недооценено число генов, отвечающих на воздействие холода, что особенно выражено для органов и частей растения, на которых не были сфокусированы основные исследования стрессового ответа (Рисунок 3.3).

Влияние выбора органа или части растения на результат также хорошо видно при рассмотрении списка генов, входящих в регулон ТФ CBF1-3 и отвечающих на холод, – COR-генов (Park et al., 2015). Почти две трети (60%) из 1 322 COR-генов дифференциально экспрессировались в наших данных после трех часов при +4°C, демонстрируя разнообразное поведение: больше половины относилась к категории Up, еще 38% к Down и совсем немного имели противоположно направленные изменения экспрессии (48 генов, категория Mix). COR-гены были идентифицированы при исследовании листьев *A. thaliana* (Park et al., 2015); при изучении наших данных заметно, что доля ДЭ COR-генов выше для зеленых частей растения (в среднем 31%), нежели для цветков и семян, хотя разница и не очень велика (Рисунок 3.3 в). После 27 часов холодового воздействия это различие становится более заметным: из 1 253 COR-генов, ДЭ хотя в одном образце, в среднем 78% дифференциально экспрессировались в фотосинтезирующих органах по сравнению с 39% в семенах и двух стадиях цветков (Рисунок 3.3 г). Таким образом, холодовой ответ в нефотосинтезирующих органах не только вовлекает дополнительные гены, которые не были ранее ассоциированы со стрессом, но также требует участия иных регуляторов, отличных от генов *CBF*.

Для дальнейшего изучения этого вопроса мы рассмотрели профили экспрессии так называемых генов «первой волны» – 27 ТФ, активирующихся вместе с CBF в течении нескольких минут после понижения температуры (Park et al., 2015). В рассматриваемых нами органах эти транскрипционные регуляторы продемонстрировали разнообразное поведение после трех часов холодового воздействия: четыре гена (*CBF3*, *CIR1*, *CRF3*, *ZF*) имели увеличенную экспрессию во всех образцах, тогда как большинство активировались в разных наборах образцов, причем в Листе-3 возрастала экспрессия 20 ТФ (Klerikova et al., 2019), Таблица S13 Приложения). Особенно интересно отметить разнонаправленные изменения экспрессии некоторых регуляторов (*AXR5*, *ERF11*, *WRKY33*, *ZAT12*), что подчеркивает комплексную картину ответа на стресс во всем растении. После 27 часов пониженной температуры экспрессия большинства ранних ТФ возвращается на уровень контроля, однако в Листе-27 больше половины (19) генов все еще имели повышенную экспрессию (Klerikova et al., 2019), Таблица S14 Приложения).

Таким образом, степень участия основных регуляторов раннего холодового ответа различается в органах растения, что подтверждает необходимость дифференциального исследования акклиматизации к низкой температуре.

#### 3.4.4. Паттерны экспрессии генов, отвечающих на холодовое воздействие

Изучение экспрессионных характеристик генов, вовлеченных в общий и органоспецифичный ответ на гипотермию, помогает понять биологические процессы, изменяющиеся в ответ на понижение температуры в разных органах. Мы использовали данные полученной нами транскриптомной карты *A. thaliana* (Глава 1) – энтропию Шеннона и ДЭ-значение.

Первый параметр, энтропия Шеннона (H), позволяет численно оценить ширину паттерна экспрессии гена, то есть число образцов, в которых детектируются его транскрипты: высокие значения H соотносятся с универсально экспрессируемыми генами, тогда как низкие характерны для органо- или тканеспецифических генов. Все 15 459 генов, экспрессия которых меняется в ответ на стресс в наших данных, имеют распределение, схожее с общим (ср. Рисунок 1.8 и 3.4 а и б), тогда как гены «ядра» стрессового ответа не имели локального пика распределения на низких значениях H, что показывает универсальность экспрессии этих генов в нормальных условиях (Рисунок 3.4 в и г).

Несмотря на небольшое количество уникальных для образцов генов распределение их энтропии Шеннона в большинстве случаев было сходным с распределением для всех генов: так, для Семядолей-3 и -27, Гипокотыля-3, -27, Листа-27 распределение было сильно смещено в область высоких значений H с небольшим количеством тканеспецифических генов (Приложение, Рисунок 3.2).

Некоторые образцы отличались своим распределением. Среди них следует отметить Лист-3, для которого был характерен локальный пик на низких значениях энтропии (Рисунок 3.4 д). Гены, которые его составляют, в нормальных условиях экспрессируются в зрелых пыльниках и участвуют в метаболизме клеточной стенки и контроле роста пыльцевой трубки. Среди них особенно выделялись гены, кодирующие пектиновые метилтрансферазы. Пектин – это один из важных компонентов клеточной стенки, формирующий матрикс, в который погружены остальные полисахариды (целлюлоза, гемицеллюлоза). Пектины синтезируются в аппарате Гольджи в высоко метилэстерированной форме, после чего они могут быть модифицированы пектиновыми метилэстеразами, что приводит к деэстерификации (Micheli, 2001). Отношение содержания эстерифицированных пектинов к деэстерифицированным определяет многие свойства клеточной стенки, такие как жесткость, проницаемость и прочность. Пектиновая метилэстераза PPME1 линеаризует деметилэстерифицированные пектиновые цепи в стенках пыльцевой трубки, при этом уменьшение активности PPME1 у мутанта *A. thaliana ppme1* ведет к уменьшению жесткости клеточной стенки (Jiang et al., 2005). *VGDH1* – еще один ген, кодирующий пектиновую метилэстеразу. *VGDH1* имеет ту же биологическую функцию, что и

его гомолог *VGD1*, обладает линейризующей деметилэстерификационной активностью и модифицирует пектин пыльцевой трубки (Jiang et al., 2005). Как и в случае *PPME1*, утрата функции *VGD1* приводит к уменьшению жесткости клеточной стенки пыльцевой трубки. Экспрессия обоих генов увеличена в два раза в Листе-3 и в 2,5-3 раза в Листе-27. Рост активности пектиновых метилтрансфераз при адаптации к холодному стрессу был показан для других растений (Solecka et al., 2008; Baldwin et al., 2014), и является одним из механизмов холодной акклиматизации, так как увеличение жесткости клеточной стенки дает большую устойчивость к дегидратации и ингибирует рост органа. Таким образом, наши результаты показывают, что реализация этого механизма у *A. thaliana* происходит с помощью рекрутирования специфических для пыльцы генов при развитии реакции на холод в листе.

Гены, составившие пик на низких значениях Н в Листе-3, частично (33%) образуют аналогичный пик в Цветке-27 вместе с другими пыльцеспецифичными генами (Рисунок 3.4 ж). Эти гены вовлечены в регуляцию роста пыльцевой трубки и модификации клеточной стенки (Клерікова et al., 2019), Таблица Приложения S16). Детальное рассмотрение этих генов дает сложную картину. В частности, происходит согласованное повышение экспрессии ингибиторов пектиновых метилэстераз (*PMEI5* и других) одновременно с ростом уровня транскрипции самих пектиновых метилэстераз (*PPME1*, *VGD1*, *VGDH1*, *VGDH2*). Также в Цветке-27 в два раза увеличивается экспрессия гена *CWINV2* – члена семейства генов инвертаз клеточных стенок (*CWIN*), которые кодируют ферменты, расщепляющие сахарозу на глюкозу и фруктозу. Активность белков семейства *CWIN* необходима для развития пыльника и образования жизнеспособной пыльцы (Sharma and Nayyar, 2016). В нормальных условиях экспрессия гена *CWINV2* специфична для пыльцы, тогда как в наших данных он начинает экспрессироваться в листе после трех часов холодного воздействия. В ряде чувствительных к холоду растений холодной стресс снижает стерильность пыльцы (Oliver et al., 2005; Ohnishi et al., 2010). Сравнение сортов риса с разными реакциями на холод показало, что специфический для пыльцы ген *CWIN* имеет сниженную экспрессию в чувствительных к холоду сортах и не меняется в устойчивых (Oliver et al., 2005). Таким образом, возрастание экспрессии гена *CWINV2* предположительно является частью пути холодной акклиматизации, и зачастую происходит одновременно с ростом активности транспортеров моносахаридов, обеспечивая доставку углеводов в пыльники (Oliver et al., 2005). Однако, в наших данных увеличения экспрессии генов семейства *STP*, кодирующих транспортеры моносахаридов у *A. thaliana*, не наблюдается.

Мы предполагаем, что сложное состояние транскриптома, наблюдаемое в образце Цветка-27, является отражением нескольких одновременно происходящих в органе процессов, таких, как продолжение нормального развития и дальнейшее разворачивание реакции холодной акклиматизации. Как упомянуто выше, в ответ на стресс в этом образце менялась экспрессия

генов, вовлеченных в рост пыльцевой трубки. Полярный рост – это необходимое свойство нормального развития пыльцевой трубки и волосков корней, требующее пространственно-временной регуляции расположения фибриллярного F-актина. Члены семейства Rho ГТФаз (ROP) растений выполняют самые разнообразные биологические функции, включая рост кончика пыльцевой трубки и волосков корня, гормональную регуляцию развития и ответ на биотический стресс (Nibau et al., 2006). ГТФазы ROP находятся под контролем положительных регуляторов факторов обмена гуанина (Guanine nucleotide Exchange Factors, GEF) (Li and Liu, 2012). Известно, что семейство ROPGEF (GEF, содержащие домен обмена нуклеотидов ROP, специфический для растений) вовлечено в рост пыльцевой трубки (Gu et al., 2006; Zhang and McCormick, 2007). Белок ROP1 известен тем, что активирует два соперничающих пути: полимеризацию F-актина, контролируруемую продуктом гена *RIC4*, и его разборку, запускаемую белком *RIC3* (Gu et al., 2005). Сборка фибриллярного F-актина ведет к накоплению везикул в клетке, тогда как его деполимеризация требуется для экзоцитоза, а правильный баланс двух процессов делает возможным полярный рост клетки (Lee et al., 2008b). В образце Цветка-27 экспрессия генов *ROPGEF8* и *RIC3* увеличена вдвое, так же, как и экспрессия двух генов профилинов – *PRF4* и *PRF5*. Профилины – регуляторы полимеризации актина в апикальной мембране пыльцевой трубки; мутанты *prf5* и *prf4* характеризуются сниженной скоростью роста пыльцевой трубки. Так как исследованный нами образец Цветка-27 является распустившимся цветком, где опыление уже произошло, увеличение экспрессии генов, ответственных за рост пыльцевой трубки, может говорить об интенсификации скорости фертилизации.

Семена-3 и Семена-27 характеризуются пиком распределения в области средних значений *H*, куда входят как гены со снижающейся, так и растущей экспрессией (Рисунок 3.4 е и з); списки этих генов наполовину перекрываются. Анализ экспрессионных паттернов этих генов в транскриптомной карте *A. thaliana* показал их специфичность для семян. Гены, имеющие сниженную экспрессию, связаны с растительным иммунитетом (термины GO:0031640, GO:0050832) и доменом F-box (Клерикова et al., 2019), Таблица Приложения S17), тогда как гены, экспрессия которых возрастает, обогащены категориями, описывающими метаболизм и запасание липидов. Семейство генов *A. thaliana*, содержащих домен F-box, включает в себя 897 членов; большинство из них плохо изучены (Hua et al., 2011). Среди них в Семенах-3 (и в большинстве случаев в Семенах-27) гены, связанные с реакцией на заражение микроорганизмами, имели пониженный уровень транскрипции. В их число входит ген *SON1*, продукт которого является частью системной защиты от патогенов (Kim et al., 2006), а также гены *SCRL5*, *LCR73*, *RALFL3*, кодирующие и другие белки относятся к небольшим цистеин-богатым противомикробным пептидам, в группу которых также входят дефензины, общие для эукариот

защитные белки (Silverstein et al., 2005). Гены, кодирующие эти белки, в Семенах-3 (и в большинстве случаев в Семенах-27) имеют пониженный уровень транскрипции.

#### *3.4.5. Дифференциальное участие регулируемых холодом генов в нормальном развитии растения*

Второй из рассматриваемых нами параметров, ДЭ-значение, представляет собой число парных сравнений образцов в транскриптомной карте, в которых данный ген был ДЭ, и отражает как разнообразие профиля экспрессии гена, так и ширину его паттерна (раздел 1.3.5). Распределение ДЭ-значений для 15 459 генов, экспрессия которых меняется под действием холода, не является случайной подвыборкой распределения всех генов и имеет меньше генов с низким ДЭ-значением (Рисунки 1.6 и 3.5 а и б), причем особенно смещены в область высоких ДЭ-значений гены категорий Mix-3 и Mix-27. Высоким ДЭ-значением обладают гены, транскрипты которых обнаруживаются во многих тканях в значимо различающемся количестве. Таким образом, гены, имеющие противоположные изменения экспрессии в холодовом ответе (категория Mix), вероятно вовлечены в разные процессы в разных тканях.

ДЭ гены, общие хотя бы для пяти органов, имеют распределение ДЭ-значений, смещенное к правой части (Рисунок 3.5 в и г). Как и энтропия  $H$ , это говорит о широком паттерне экспрессии базовых генов холодового ответа и поддерживает идею дифференциального использования этих генов в разных процессах в разных тканях.

У органоспецифических ДЭ генов оно немного смещено влево (Рисунок 3.5 д и е, Приложение, Рисунок 3.3). Из всех образцов только гены, характерные для Молодого Цветка-27, имели выраженные особенности распределения: пик в области ДЭ-значений меньше 500, образованный в основном генами, активирующимися в ответ на стресс (Рисунок 3.5 е). Гены, имеющие низкие ДЭ-значения могут равномерно экспрессироваться в большинстве образцов, либо быть тканеспецифическим (см. раздел 1.4.4). Большинство генов с низким ДЭ-значением, дифференциально экспрессирующиеся в Молодом Цветке-27, в транскриптомной карте, охватывающей нормальное развитие растения, были специфичны для молодых пыльников (принадлежащих цветкам стадии 9 в соответствии со Smyth et al.) и для молодых цветков на той же стадии (Smyth et al., 1990). Анализ обогащения этих генов категориями ГО показал значимую перепредставленность терминов, связанных с убиквитин-зависимыми процессами катаболизма белков, а также с регуляцией транскрипции и доменом F-box (Клерикова et al., 2019), Таблица Приложения S18). Убиквитинирование – это один из ключевых регуляторных механизмов посттрансляционной модификации; присоединение к белкам убиквитинов и дальнейшее их расщепление в 26S-протеасоме служит для модулирования уровня белка и удаления дефектных белков. Участие в ответе на абиотические стрессы было продемонстрировано для некоторых

генов, кодирующих компоненты системы убиквитиновой 26S-протеасомы (Lyzenga and Stone, 2012).

Многие белки, содержащие домен F-box, являются субъединицами комплекса SCF (SKP1, CULLIN, FBX и RBX1), представляющего собой убиквитиновую лигазу типа E3 (Callis, 2014). Среди генов, экспрессия которых возрастает в образце Молодого Цветка-27, 63 аннотированны как содержащие домен F-box. Кроме того, известно, что некоторые SKP-подобные белки *A. thaliana* (ARABIDOPSIS SKP-LIKE, ASK), так же являющиеся частью комплекса SCF, напрямую взаимодействуют с F-box-содержащими белками (Kuroda et al., 2012). В образце Молодого Цветка-27 восемь генов из семейства ASK обладают значимо повышенной экспрессией.

В работе Gladman et al. два белка из семейства NO APICAL MERISTEM/ARABIDOPSIS TRANSCRIPTION ACTIVATION FACTOR-1/CUP-SHAPED COTYLEDONS-2 (NAC), NAC78 и NAC56, были определены как положительные регуляторы генов системы убиквитиновой 26S-протеасомы (Gladman et al., 2016). Хотя мы не видим изменения экспрессии этих генов в Молодом цветке-27, список ДЭ генов включает в себя несколько других генов NAC, таких как NAC023 и NAC063, которые также могут указывать на участие системы убиквитирования белков в ответе на холододовый стресс в развивающемся цветке.

#### 3.4.6. Привлечение разных транскрипционных факторов в органах *A. thaliana*

Изучение представленности регуляторных элементов различных ТФ в промоторах генов позволяет находить ключевые регуляторы общей и органоспецифической холододовой акклиматизации.

Базовые гены стрессового ответа (ДЭ в пяти или шести органах) главным образом регулируются ТФ семейства CBF. В число активируемых ими генов выходят транскрипционные факторы из семейств ERF/AP2, TCP, NAC, WRKY, Myb, а также содержащие домен «цинковые пальцы B-box-типа» и гомеобокс. Эти ТФ вместе с CBF образуют основную сеть генетической регуляции, универсально действующую во всем растении.

Регуляция тканеспецифических генов стрессового ответа требует определенного набора ТФ для каждого из органов. Так, для развития холододовой акклиматизации в семенах и цветке необходимо сочетание ТФ семейств bZIP и NAC, тогда как в листе в основном функционируют регуляторы из семейства ERF/AP2. В генах, экспрессия которых изменяется в Молодом Цветке-3, чаще встречаются регуляторные элементы ТФ ERF/AP2, bZIP и Myb, а после 27 часов холододового воздействия основную регуляцию выполняют транскрипционные факторы семейства NAC.

### 3.5. Заключение

В Главе 3 мы рассматривали вопрос универсальности развития реакции на холодовой стресс в разных органах *A. thaliana*. Предпосылки к предположению о дифференциальной регуляции низкотемпературной акклиматизации находятся как в известных фактах сопряжения ответа на холод и нормального развития растения (Park et al., 2015; Jung et al., 2016; Zhao et al., 2016), так и в большом разнообразии экспрессионных профилей известных генов стрессовой реакции.

Для решения этой задачи мы проанализировали транскриптомы шести органов *A. thaliana*, подвергнутых холодовому воздействию в течении трех и 27 часов. Значительное число генов было дифференциально экспрессируемым хотя бы в одном образце, при этом хорошо изученные органы (лист, семядоли и гипокотиль) продемонстрировали изменения, схожие с результатами других исследований, тогда как семена, и две стадии развития цветка заметно отличались.

Основу ответа на холод составляет небольшое число генов, ДЭ во всех образцах. Эти гены были хорошо известны и ранее, о чем говорит существенное их обогащение терминами ГО, связанными с ответом на стресс. «Ядерные» гены холодовой реакции входят в регулон транскрипционных факторов СВФ.

Число уникальных для образцов генов различалось от сотен до тысяч. В каждом органе эти гены были вовлечены в разные процессы – от регуляции фотосинтеза в листьях до роста и развития пыльцевой трубки в цветке; для органов также была характерна регуляция стрессового ответа разными семействами ТФ. На примере листа мы показали, что для развития холодовой акклиматизации могут привлекаться продукты генов, в нормальных условиях исключительно экспрессирующихся в других частях растения (пыльце).

Таким образом, на основании полученных нами данных можно заключить, что в основе ответа на холодовой стресс лежит единый для всего растения механизм, вовлекающий регуляторы семейства СВФ, который модифицируется в каждом органе в зависимости от преимущественно происходящих в них процессов.

## ГЛАВА 4. ТРАНСКРИПТОМНАЯ КАРТА *CAPSELLA BURSA-PASTORIS*

### 4.1. Введение

Транскриптомная карта *A. thaliana*, описанная нами в Главе 1, была создана с большим разрешением; избыточное число образцов необходимо для решения разнообразных задач – определения точного паттерна гена, поиска стабильно экспрессирующихся генов, сравнения экспрессии ортологичных генов *A. thaliana* и вида, обладающего транскриптомной картой меньшего размера (Penin et al., 2019). В том числе высокоразрешенный атлас экспрессии может быть использован для планирования транскриптомной карты другого вида, что было сделано нами для растения *Capsella bursa-pastoris*, также относящемуся к семейству Крестоцветных. Морфология *C. bursa-pastoris* и *A. thaliana* довольно схожа, кроме того, в геномах этих видов сохранилась большая доля ортологичных генов.

*C. bursa-pastoris* – распространенный по всему миру недавний аллотетраплоид, служащий удобной моделью ранних последствий полиплоидизации. Отсутствие секвенированного генома долгое время препятствовало проведению массового анализа транскрипции, однако использование генетического картирования и анализ сегрегирующей популяции в работе Kasianov et al. позволил решить эту проблему (Kasianov et al., 2017).

Изучение разнообразия транскриптома *C. bursa-pastoris* с помощью атласа экспрессии генов представляет большой интерес. До сих пор существует ограниченное число транскриптомных карт недавних полиплоидов, сохранивших большое число гомеологичных генов; как пример можно привести атласы *Camelina sativa* (Kagale et al., 2016) и *Arachis hypogaea* (Clevenger et al., 2016). Сравнение экспрессии гомеологичных генов в наборе хорошо различающихся в транскрипционном отношении образцов может позволить изучить процессы функциональной дивергенции гомеологов, их суб- и неофункционализацию, либо утрату функции.

## 4.2. Обзор литературы

### 4.2.1. Полиплоидизация – основной путь эволюции растительных геномов

Полиплоидизация или кратное увеличение числа хромосом – один из основных механизмов эволюции растений. Полногеномные дубликации (ПГД) многократно повторялись на протяжении всего времени существования растений: начиная с гексаплоидного общего предка настоящих двудольных (Vekemans et al., 2012) и ПГД в ранней эволюции однодольных (Tang et al., 2010), полиплоидизация происходила во многих таксонах (Paterson et al., 2004, 2012; The French–Italian Public Consortium for Grapevine Genome Characterization, 2007; Murat et al., 2015).

После ПГД основную проблему как на клеточном, так и на организменном уровне составляет отсутствие налаженного взаимодействия новообразованных субгеномов, что приводит к существенным нарушениям регуляции экспрессии генов, белок-белковых взаимодействий и сайленсинга мобильных генетических элементов (Ågren et al., 2016; Kryvokhyzha et al., 2019). Для разрешения подобных коллизий после полиплоидизации начинается процесс перестройки генома и транскриптома, в ходе которого один из гомеологов (ортологичных генов из разных субгеномов) может утратить функциональность и псевдогенезироваться, приобрести новую функцию (неофункционализация), либо может начаться расхождение паттернов экспрессии гомеологичных генов и их субфункционализация; в совокупности эти процессы приводят к диплоидизации генома (Otto and Whitton, 2000; Douglas et al., 2015).

Субфункционализация гомеологов или утрата одного из них служат основными путями разрешения межгеномного конфликта у многих видов, включая *A. thaliana* (Jeffrey Chen et al., 2004; Groszmann et al., 2011), *Tragopogon mirus* Ownbey (козлобородник, (Buggs et al., 2010), сою (Roulin et al., 2013) и хлопок (Adams and Wendel, 2005; Chaudhary et al., 2009). Сопровождающие этот процесс структурные перестройки (как на хромосомном уровне, так и в масштабах отдельных генов) происходили при древней полногеномной дубликации у *A. thaliana* (Thomas et al., 2006) и у недавнего естественного полиплоида *T. mirus* (Koh et al., 2010); у кукурузы и пшеницы наблюдаются древние и современные потери одного из гомеологов (Feldman et al., 1997; Shaked et al., 2001; Mochida et al., 2003; Schnable et al., 2011), у аллополиплоидов из рода *Brassica* процесс диплоидизации сопровождается многочисленными транслокациями и транспозициями участков генома (Song et al., 1995).

У многих, хотя и не у всех видов растений преимущественно сохраняются гены из одного из субгеномов, которые, кроме того, экспрессируются на более высоком уровне, чем их гомеологи (Garsmeur et al., 2014; Woodhouse et al., 2014); это явление называется геномным доминированием и продемонстрировано для разнообразных растений, включая кукурузу из

Злаков и представителей семейства Brassicaceae – *A. suecica*, *Brassica rapa* и *Camelina sativa* (Chang et al., 2010; Schnable et al., 2011; Woodhouse et al., 2014; Kagale et al., 2016; Pont and Salse, 2017).

Следует отметить, что полиплодия чрезвычайно широко распространена среди сельскохозяйственных видов, таких, как пшеница, кукуруза, капуста и хлопок, поэтому ее последствия, в том числе геномное доминирование, хорошо изучены (Yang et al., 2016). Однако применимость этих наблюдений к некультивируемым растениям ограничена из-за прямого влияния на эволюцию сельскохозяйственных видов искусственного отбора и селекции.

#### 4.2.2. Последствия полиплоидизации для транскриптома

Последствия полиплоидизации для транскриптома растения изучались для большого разнообразия видов. Между гомеологичными генами субгеномов полиплоида нередко наблюдается дифференциальная экспрессии; источником этого различия может быть как унаследованная от предковых видов разница в уровне транскрипции, так и возникшая в процессе коэволюции двух субгеномов.

Немедленные последствия полиплоидизации преимущественно изучаются с помощью лабораторной реконструкции аллополиплоидов (Jackson and Chen, 2010). Сравнение транскриптомов таких растений с их предковыми видами позволяет понять степень сохранения уровней экспрессии после полиплоидизации. Такие работы были сделаны на реконструированных аллополиплоидах родов *Arabidopsis* и *Senecio*, а также хлопка и пшеницы. Доля генов, сохраняющих предковый уровень экспрессии, составляла 65-95% в аллотетраплоидах *Arabidopsis* (Wang et al., 2006), 30-70 % у хлопка (Flagel et al., 2008), 40-73% в межвидовых гибридах *Senecio* (Hegarty et al., 2006), и вплоть до 80% в синтетической пшенице (Pumphrey et al., 2009). Таким образом, это значение варьировалось в широких пределах даже в рамках одного эксперимента, показывая в большой мере стохастическую природу установления уровней экспрессии генов в новообразованном полиплоиде. Это подтверждается исследованием транскриптомов искусственных полиплоидов рода *Arabidopsis*, в котором было показано быстрое возникновение дифференциальной эпигенетической регуляции, случайным образом снижающей уровень экспрессии одного из гомеологов у синтетического гибрида *A. thaliana* и *A. arenosa* (Wang et al., 2004b).

Предположительно, сохранение предкового уровня экспрессии генов позволяет сбалансировать дозу гена при появлении функциональной избыточности после полиплоидизации (Veitia et al., 2008). С другой стороны, дивергенция уровней экспрессии ТФ и других регуляторов транскрипции может приводить к репрограммированию сетей генетической регуляции и объяснять новые, неаддитивные, фенотипы аллополиплоидов.

Следует отметить, несмотря на известный факт непрямой зависимости количества белка от числа мРНК кодирующего его гена, разница в накоплении белков, кодируемых двумя гомеологами, малоизучена. У искусственно ресинтезированного аллотетраплоида *V. parvis* более 25% из 1600 исследованных полипептидов двух субгеномов накапливаются неравномерно, однако глобальных различий в экспрессии генов определенных функциональных категорий или функционировании метаболических путей из двух субгеномов обнаружено не было (Albertin et al., 2007).

Помимо немедленных последствий полиплоидизации хорошо изучены различия в транскриптомах давно возникших полиплоидов. Паттерны экспрессии 43% гомеологов у аллополиплоидного хлопка, возникшего около 1-2 миллионов л.н., различались (Udall and Wendel, 2006; Yoo et al., 2013); схожие результаты были получены для сои, претерпевшей два раунда полиплоидизации 13 и 59 миллионов л.н. (Roulin et al., 2013). Однако взаимодействие гомеологов у недавних полиплоидов остается слабоизученным, и большая часть работ была выполнена на пшенице, *Triticum aestivum*, возникшей около 10 тысяч л.н. и все это время находившейся под интенсивным искусственным отбором (Matsuoka, 2011). Около 55% генов пшеницы характеризовались тканеспецифической экспрессией гомеологов, причем разные органы демонстрировали отличия в степени смещения уровней экспрессии гомеологов: наиболее сильное расхождение было свойственно тычинкам (Mutti et al., 2017), а в листьях преимущественно обнаруживались транскрипты гомеологов из субгенома В (Nomura et al., 2005). Недавняя публикация полного генома *T. aestivum* позволила существенно углубить понимание взаимодействия между субгеномами и суб- и неофункционализации гомеологов (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018).

Таким образом, большинство исследований полиплоидизации затрагивают либо немедленные последствия межвидовых скрещиваний, либо уже завершающийся процесс диплоидизации древних полиплоидов, тогда как недавние полиплоидные виды, уже прошедшие некоторый эволюционный путь, изучены слабо.

#### 4.2.3. *Capsella bursa-pastoris* – модельный объект для изучения ранней эволюции

Принадлежащая к семейству Крестоцветных *Capsella bursa-pastoris* (L.) Medik (пастушья сумка) – распространенное по всему миру сорное растение, использующееся в качестве модели для изучения ранних последствий полиплоидизации (Han et al., 2015) и других аспектов эволюции растений (Hintz et al., 2006; Nutt et al., 2006). Происхождение *C. bursa-pastoris* долгое время оставалось предметом дискуссии (Slotte et al., 2006; St Onge et al., 2012), но последние исследования установили, что *C. bursa-pastoris* возникла в результате скрещивания *C. rubella/grandiflora* и *C. orientalis* около 100 000–300 000 л.н. (Douglas et al., 2015). Из-за крайней

близости геномов родительских видов точная сборка полного генома *C. bursa-pastoris* была сопряжена с большими трудностями, однако применение анализа сегрегирующей популяции и генетического картирования позволило получить полногеномную сборку *C. bursa-pastoris* с разделением скаффолдов на два субгенома (Kasianov et al., 2017).

Поиск геномного доминирования и несимметричной экспрессии гомеологичных генов *C. bursa-pastoris* проводился неоднократно, однако в разных исследованиях были получены противоречивые результаты: при изучении экспрессии в 31 линии пастушьей сумки со всего мира не было найдено преимущественной экспрессии генов одного из субгеномов (Douglas et al., 2015), что подтвердилось для 25 линий из Китая (Huang et al., 2018), тогда как более высокая экспрессия субгенома, происходящего от *C. rubella/grandiflora* была показана для растений из Европы и Ближнего Востока (Крувокхызха et al., 2019). Следует отметить, что эти работы были проведены на одном органе растения (листьях). Таким образом, интерес представляет как общее изучение разнообразия транскриптома *C. bursa-pastoris* в разных органах и тканях, так и выяснения степени асимметрии экспрессии гомеологов из двух субгеномов.

### 4.3. Результаты

#### 4.3.1. Сбор образцов, секвенирование и первичная обработка данных

Высокая степень детализации атласа экспрессии генов *A. thaliana* (Глава 1) позволяет подробно изучить транскрипционную схожесть различных органов и частей растения и использовать эту информацию при планировании транскриптомной карты другого объекта. В разделе 1.4.1 мы рассматривали основные клады образцов атласа *A. thaliana* (Рисунок 1.3); так как образцы внутри каждой группы схожи между собой, для транскриптомной карты меньшего разрешения можно выбрать только один репрезентативный образец. Морфологическое сходство *A. thaliana* и *C. bursa-pastoris* позволяет без затруднений воспроизводить стадию развития органа или части растения, благодаря чему в основу атласа экспрессии генов *C. bursa-pastoris* легли десять наиболее транскрипционно различающихся образцов: корень, пластинка и жилка молодого листа, апикальная меристема, органы цветка, семена на двух стадиях (далее «основная транскриптомная карта», для описания образцов см. Приложение, Таблица 4.1). Также нами был собран дополнительный набор образцов семядолей однодневных проростков, подвергнутых холодному воздействию (+4°C) в течении 3, 15 и 27 часов, а также соответствующий контроль. Все образцы были собраны в двух биологических повторностях.

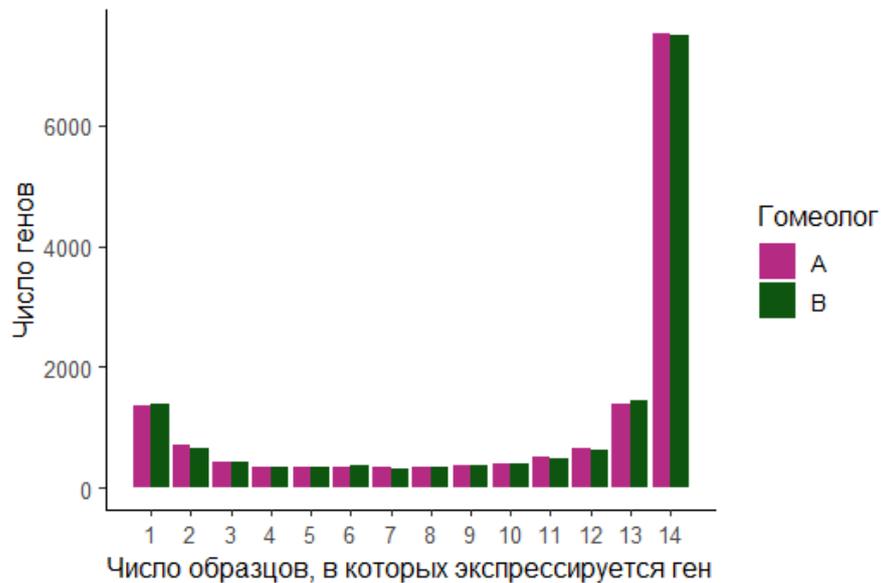
Выделенная из образцов РНК была использована для приготовления библиотек кДНК, секвенированных на платформе Illumina. Для каждой пробы было получено 30-45 млн одноконцевых чтений длиной 108 п.н. Проверка схожести биологических повторностей с помощью квадрата коэффициента корреляции Пирсона показала их пригодность для дальнейшего анализа ( $r^2$  варьировался от 0,91 до 0,99 со средним 0,98, Приложение Таблица 4.1).

#### 4.3.2. Общие экспрессионные характеристики транскриптома *C. bursa-pastoris*

Транскриптомная карта позволяет подробно описать экспрессионные особенности вида (см. Главу 1). *C. bursa-pastoris* – модельный объект ранней эволюции полиплоидов (Han et al., 2015), поэтому основное внимание в нашей работе уделено сравнительному изучению экспрессии гомеологичных генов, принадлежащих к двум субгеномам – А (происходящем от родительского вида *C. rubella/grandiflora*) и В (*C. orientalis*) (Douglas et al., 2015). Скаффолды секвенированного в нашей лаборатории генома *C. bursa-pastoris* были отнесены к одному из двух субгеномов, что позволило обнаружить 16 798 пар гомеологичных генов (Kasianov et al., 2017).

Для общей оценки разнообразия транскриптома *C. bursa-pastoris* мы определили число экспрессирующихся генов без разделения на субгеномы. Всего 15 036 (45%) генов экспрессировались во всех образцах и 30 041 (89%) – хотя бы в одном образце. Наименьшее число экспрессирующихся генов (18 772, 56%) насчитывалось в апикальной меристеме побега, а наибольшее (24 380, 73%) – в корне. Во всех образцах число транскрибирующихся генов из

субгеномов А и В было практически одинаковым (9 386 гомеологов А и 9 386 гомеологов В экспрессировалось в образце АМ, и 12 187 и 12 193 – в образце корня). Транскрипты обоих гомеологичных генов детектировались в каждом из образцов более чем в 95% случаев (Приложение, Таблица 4.2). Если рассматривать субгеномы отдельно, 15 014 гомеологов А экспрессировались хотя бы в одном образце, 7 529 – во всех образцах, а транскрипты 1 731 генов, принадлежащих к субгеному А, обнаружены не были. Соответствующие значения для субгенома В составили 15 027 генов (хотя бы в одном образце), 7 507 (во всех образцах) и 1 718 (не экспрессировались).



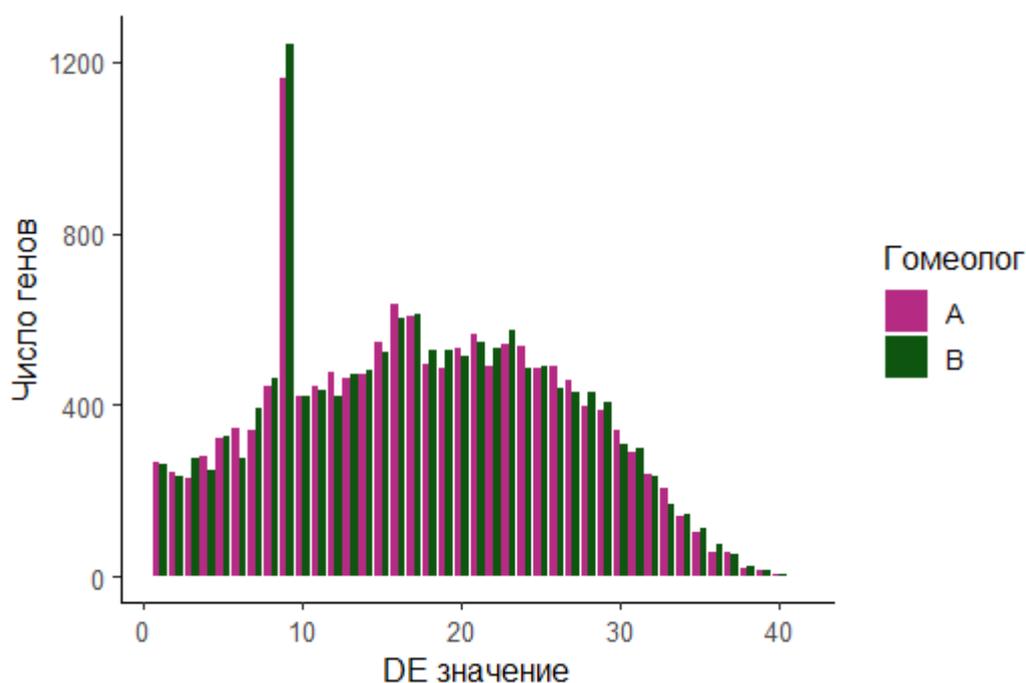
**Рисунок 4.1.** Распределение генов по числу образцов, в которых ген экспрессируется. Гомеологи из субгеномов А и В приведены отдельно.

Для описания транскрипционного паттерна генов мы посчитали число образцов, в которых экспрессируется ген. Большинство генов экспрессировались во всех или почти во всех (больше 11) образцах (Рисунок 4.1). Если рассматривать только основную часть транскриптомной карты (10 образцов), то распределение сохраняется (Приложение, Рисунок 4.1). Второй пик образуют гены с узким паттерном экспрессии, транскрипты которых были найдены только в одном или двух образцах (Рисунок 4.1). Гомеологи А и В одинаково распределялись по числу образцов, в которых они экспрессировались (Рисунок 4.1).

#### 4.3.3. Анализ дифференциальной экспрессии

Чтобы оценить транскриптомное разнообразие образцов, мы проанализировали дифференциальную экспрессию между всеми парами образцов основной транскриптомной карты. Число ДЭ генов варьировалось от 2 311 между апикальной меристемой и черешком листа до 18 272 между пыльником и корнем (Приложение, Таблица 4.3).

Для описания профиля экспрессии генов мы посчитали ДЭ-значение, отражающее количество парных сравнений, в которых ген был ДЭ (см. Главу 1). У 30 028 генов (89% от всех генов) ДЭ-значение было больше нуля; максимальное ДЭ-значение составило 41 из возможных, в соответствии с числом сравнений, 45. Распределение ДЭ-значений не различалось между субгеномами А и В (Рисунок 4.2), однако число пар, в которых оба гомеолога были либо одновременно ДЭ, либо не ДЭ, было низким (679 пар). Основная часть (10 165 пар) показала согласованную экспрессию (т.е. оба гомеолога ДЭ или не ДЭ) в одних сравнениях, но несогласованную (т.е. только гомеолог А ДЭ или только гомеолог В ДЭ) в других (Приложение, Таблица 4.4). Число пар, которые показали только несогласованную экспрессию (т.е. во всех сравнениях только гомеолог А или только В были ДЭ), так же было очень низким (541 для субгенома А и 551 - для В). Квадрат корреляции Пирсона для ДЭ-значений гомеологов А и В составил 0,65.



**Рисунок 4.2.** Распределение ДЭ-значений у генов, дифференциально экспрессирующихся хотя бы между одной парой образцов. Гомеологи из субгеномов А и В приведены отдельно.

В дополнительном наборе данных образцы, подвергнутые воздействию холода в течении 3, 15 и 27 часов при +4°C, сравнивались между собой и с контрольным образцом. Число ДЭ генов лежало в диапазоне от 3 270 между 3 часами и контролем до 8 790 между 27 часами и контролем. Общее число генов, экспрессия которых изменялась в ответ на низкотемпературное воздействие хотя бы в одном из сравнений, составило 14 596 (7 285 из них принадлежали к субгеному А и 7 311 – к субгеному В). У *C. bursa-pastoris* и *A. thaliana* индуцированные стрессом изменения продемонстрировали большое сходство. Реакция на понижение температуры хорошо изучена у

*A. thaliana* (см. Главу 3); с помощью изучения развития холодной акклиматизации у трансгенных растений был идентифицирован основной набор генов, активирующихся в ответ на стресс (гены COR, (Park et al., 2015). 792 из 1 322 генов COR *A. thaliana* имели пару ортологов в геноме *C. bursa-pastoris*, и экспрессия большинства из них (652 гомеологов А, 660 гомеологов В) изменялась в ответ на холодное воздействие.

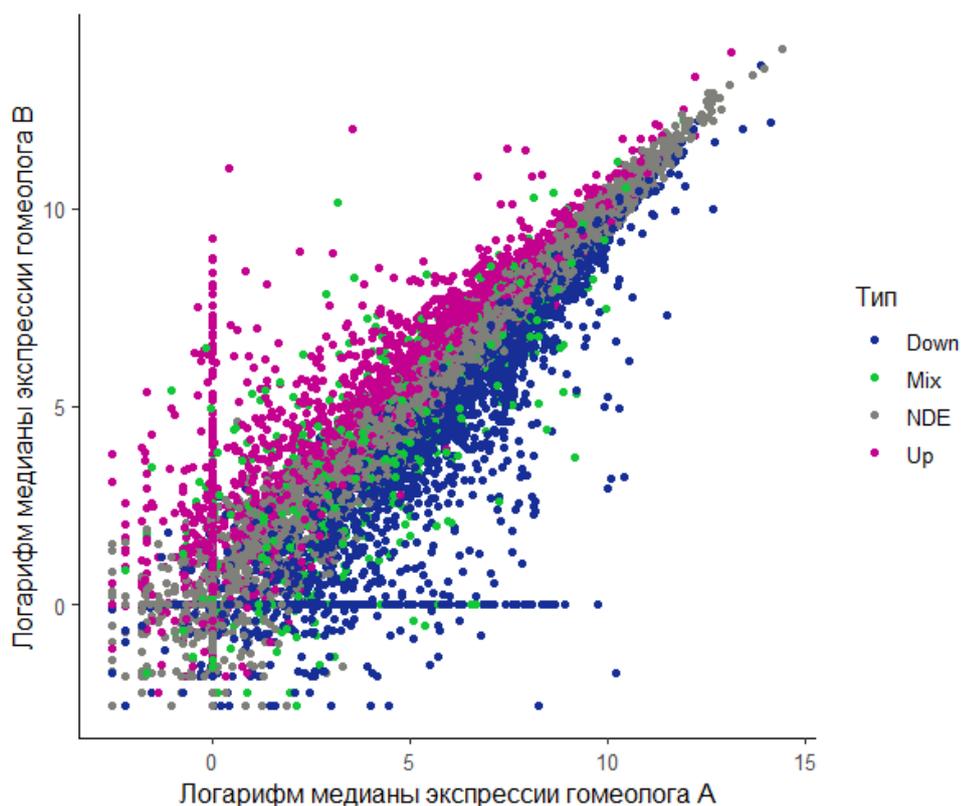
#### 4.3.4. Анализ дифференциальной экспрессии гомеологичных генов

В целом, профили экспрессии гомеологичных генов были схожи между собой: в разных образцах  $r^2$  между уровнями экспрессии гомеологов А и В варьировался от 0,8 в апикальной меристеме до 0,9 в пластинке листа (Приложение, Таблица 4.5). Для поиска различий в уровнях транскрипции гомеологов мы проанализировали дифференциальную экспрессию внутри пар гомеологов: в каждом образце экспрессия гомеолога В сравнивалась с гомеологом А и отбирались пары, экспрессия в которых значимо различалась (далее ДЭАВ пары).

**Таблица 4.1.** Число ДЭАВ пар в каждом образце.

Образец	Уровень экспрессии гомеолога В		Всего ДЭАВ пар
	снижен	повышен	
Меристема	206	212	418
Черешок листа	924	831	1755
Пластинка листа	810	748	1558
Семена	226	193	419
Чашелистики	1477	1403	2880
Лепестки	1383	1285	2668
Пыльники	1418	1405	2823
Семяпочки	1062	1019	2081
Сухие семена	680	620	1300
Корень	745	715	1460

Число ДЭАВ пар заметно различалось между образцами: от 418 в зрелых семенах (SD) до 2 880 в чашелистиках (SP, Таблица 4.1). Аналогично ДЭ-значению (раздел 4.3.3), мы вычислили ДЭАВ-значение для каждой пары гомеологов, В основной транскриптомной карте 7 376 пар гомеологов (44% от общего числа пар) была ДЭ хотя бы в одном образце (то есть имели ДЭАВ-значение один и больше). Экспрессия большинства гомеологов изменялась однонаправленно: в 3 274 парах гомеолог В был понижен по сравнению с А хотя бы в одном образце, в 3 235 – повышен. Только 867 пар гомеологов показали разное направление изменения экспрессии в двух и более образцах; такие пары были обогащены категориями «цитохром P450», «связывание кислорода» и «гем» (Рисунок 4.3, Приложение, Таблица 4.6). Распределение ДЭАВ-значений не различалось для пар, в которых экспрессия гомеолога В была понижена или повышена (Приложение, Рисунок 4.2).



**Рисунок 4.3.** Корреляция экспрессии гомеологичных генов. Цветом отмечено направление изменения экспрессии при сравнении гомеолога В и гомеолога А: синий, Down – гомеолог В имеет сниженную экспрессию по сравнению с гомеологом А хотя бы в одном образце и нигде не имеет повышенной; маджента, Up – гомеолог В имеет повышенную экспрессию по сравнению с гомеологом А хотя бы в одном образце и нигде не имеет пониженной; зеленый, Mix – гомеолог В имеет повышенную экспрессию по сравнению с гомеологом А хотя бы в одном образце и пониженную хотя бы в одном образце; серый, NDE – гомеологи В и А не дифференциально экспрессируются ни в одном образце.

Среди генов, дифференциально экспрессирующихся в ответ на холодовое воздействие (14 596 генов, см. раздел 4.3.3), 3 883 гомеологов А и 3 855 гомеологов В входили в пары, ДЭАВ хотя бы в одном из образцов. Распределение ДЭАВ-значений для таких генов не отличалось от общего для всех пар.

#### 4.3.5. Изучение промоторных областей гомеологичных генов

Для выяснения причины возникновения дифференциальной экспрессии между гомеологичными генами совместно с Кулаковским И.В. мы провели анализ промоторных областей гомеологов. В каждой паре гомеологов промоторные участки (определенные как 500 п.н. до старта транскрипции гена) могут одновременно иметь (или одновременно не иметь) сайт посадки определенного транскрипционного фактора. Такая ситуация далее называлась нами «симметрия». В противном случае сайтом посадки может обладать только один из гомеологичных генов, что приводит к «асимметрии» по этому сайту. Следует предполагать, что

к появлению дифференциальной экспрессии между гомеологами приводит ситуация асимметрии по сайту связывания.

Благодаря наличию секвенированных геномов родительских видов – *C. rubella* и *C. orientalis*, мы рассматривали промоторные области пар гомеологов и их ортологов в родительских видах. Как симметрия, так и асимметрия в сайтах связывания ТФ могут возникать как в результате возникновения сайта связывания в субгеноме А или В *C. bursa-pastoris* по сравнению с предковым видом, так и в результате исчезновения. Для каждого сайта связывания ТФ (отдельно для случаев возникновения и утраты) с помощью теста Фишера была проанализирована четырехпольная таблица: число ДЭАВ гомеологичных пар с асимметрией по этому сайту, число ДЭАВ гомеологичных пар с симметрией, число не-ДЭАВ гомеологичных пар с асимметрией по этому сайту, число не-ДЭАВ гомеологичных пар с симметрией. По результатам теста Фишера были найдены сайты связывания ТФ, значимо чаще встречающиеся в ДЭАВ парах в случае асимметрии: в транскриптомной карте – O2, TGA1A, TCP16(1), PIF3(1), PIF3(2), TGA1, VZIP910(2), TGA2(2); в наборе образцов, подвергнутых стрессовому воздействию – CBF1-3, AT1G12630, MYB30, MYB73, AT1G72740. Следует отметить, что наибольший вклад в разницу между ДЭАВ и не-ДЭАВ гомеологичными генами вносили сайты связывания, возникшие в промоторе одного из гомеологов по сравнению с родительскими видами.

## 4.4. Обсуждение

### 4.4.1. Сбор образцов, секвенирование и первичная обработка данных

Одна из важных задач эволюционной биологии – изучение сохранения и изменения функций ортологичных генов у разных видов. У растений многочисленные полногеномные дубликации создают большие группы ортологичных генов; в этом случае информации о консервативности нуклеотидных последовательностей недостаточно для исследования эволюции функции (Penin et al., 2019). Например, при соответствии одному гену первого вида нескольких генов во втором, функция может сохраняться только у одного из ортологов, тогда как остальные могут утратить ее или претерпеть неофункционализацию.

Косвенным указанием на сохранение биологической функции может служить совпадение паттерна экспрессии ортологичных генов у двух видов. Наиболее полным отображением транскрипционного профиля гена в настоящее время служат транскриптомные карты (Stelpflug et al., 2016). Сравнение подробных атласов экспрессии помогает получить информацию о функциональной дивергенции, даже если сравнения нуклеотидных последовательностей не хватает для разрешения этого вопроса (Penin et al., 2019).

Аллополиплоидные виды особенно интересны тем, что в них совместно функционируют два генома, ранее независимо существовавших в двух видах. Гомеологичные гены из разных субгеномов у родительских видов выполняли одну и ту же функцию, и после полиплоидизации возникает функциональная избыточность, со временем ведущая к суб- и неофункционализации, либо утрате одного из гомеологов. У молодых аллополиплоидов, произошедших от близких видов, таких как *C. bursa-pastoris*, с помощью транскриптомных данных можно оценить степень функциональной дивергенции гомеологов.

Морфологическое сходство и сохранение большого числа (13 123) ортологичных генов (Kasianov et al., 2017) *C. bursa-pastoris* и *A. thaliana* позволяет собрать образцы *C. bursa-pastoris*, ориентируясь на транскриптомную карту *A. thaliana*, благодаря чему нам удалось получить наиболее разнообразные транскрипционные профили: максимальный  $r^2$  между образцами был 0,67 (Приложение, Таблица 4.1).

### 4.4.2. Отсутствие отличий между субгеномами по общим экспрессионным характеристикам

Общие транскриптомные характеристики *C. bursa-pastoris* не отличались от других видов, для которых были созданы атласы экспрессии генов (подробнее см. раздел 1.2.6). Во всех образцах экспрессировалось 45% генов как в целом, так и отдельно по субгеномам, что полностью соответствует значениям у других видов (42-50%). Транскрипты 89% генов в целом и гомеологов в отдельности были найдены хотя бы в одном образце, что ненамного превышает типичный для растений диапазон в 75-85%. Таким образом, нами не была обнаружена разница

между субгеномами А и В ни по общим экспрессионным характеристикам, ни по числу гомеологов двух субгеномов, экспрессирующихся в каждом из образцов (Приложение, Таблица 4.2).

Паттерн экспрессии гена был описан нами наиболее простым способом через число образцов, в которых экспрессируется ген (Рисунок 4.1). Как и для транскриптомной карты *A. thaliana*, содержащей существенно больше проб, распределение экспрессирующихся генов по числу образцов обладало двумя пиками - наибольший охватывал 64% генов, обладающих широким паттерном экспрессии, а второй пик (14% генов) отражал тканеспецифическую экспрессию. Распределение гомеологов из субгеномов А и В по числу образцов, в которых они экспрессировались, полностью совпадало, соответственно, массового изменения паттерна экспрессии гомеологичных генов найдено не было.

#### 4.4.3. Различия во вкладе в дифференциальную экспрессию двух субгеномов

Общее разнообразие транскриптома можно оценить с помощью анализа дифференциальной экспрессии между всеми парами образцов (раздел 1.4.4). Мы обнаружили 30 028 генов, ДЭ хотя бы в одном из 45 сравнений, что подтверждает значительное разнообразие собранных нами образцов. Между субгеномами А и В не было найдено разницы по распределению ДЭ-значения (Рисунок 4.2), то есть степень вклада субгеномов в разнообразие транскриптома одинакова. Однако если рассматривать отдельные гомеологичные пары, то во многих из них экспрессия гомеологов А и В вела себя по-разному в разных сравнениях, иногда изменяясь одновременно, иногда только у одного гомеолога, и доля пар, показавших полностью согласованную экспрессию, была очень низка (Приложение, Таблица 4.4).

Как общее рассмотрение разнообразия экспрессии, так и отдельный анализ генов, участвующих в развитии холодовой реакции, показал отсутствие сайленсинга одного из субгеномов или других нарушений паттернов экспрессии (так называемого «транскриптомного шока»). Это наблюдение согласуется с предыдущими работами, показавшими, что транскриптомный шок возникает немедленно после гибридизации и быстро стабилизируется в последующих поколениях (Hegarty et al., 2006; Zhang et al., 2016a, 2016b). Несмотря на высокую (0,8-0,9) корреляцию уровней экспрессии гомеологов, мы обнаружили около тысячи гомеологичных пар, обладающих разными паттернами экспрессии. Это несоответствие может объясняться двумя способами. Во-первых, так как *C. bursa-pastoris* является гибридом между разными, хотя и близкородственными, видами, уровень и паттерн экспрессии некоторых гомеологичных генов может быть наследием диплоидных предков (Buggs et al., 2014). В другом случае наблюдаемая разница в экспрессии гомеологов может быть результатом расхождения в

их регуляции. В обоих случаях, эти гены особенно интересны для исследования функционального расхождения между субгеномами А и В.

#### 4.4.4. Различия в паттернах экспрессии гомеологичных генов

Для дальнейшего изучения различий в экспрессии мы рассмотрели дифференциальную экспрессию внутри пар гомеологов. Несмотря на схожее число экспрессирующихся генов в разных образцах, органы *C. bursa-pastoris* характеризовались отличающимся более чем в шесть раз числом ДЭАВ пар (Таблица 4.1). Количество гомеологов В, экспрессия которых возрастала или снижалась по сравнению с гомеологом А, было одинаковым как в каждом образце, так и в целом (Рисунок 4.3), и распределение их ДЭАВ-значений также не отличалось (Приложение, Рисунок 4.2). Таким образом, мы показали отсутствие асимметричного сайленсинга одного из субгеномов. Кроме того, отсутствие обогащения ДЭАВ пар категориями ГО указывает на неспецифическое снижение экспрессии генов обоих субгеномов, не связанное с целыми регуляторными или метаболическими путями.

Гены, участвующие в регуляции холодной акклиматизации, по распределению ДЭАВ-значений не отличались от всех генов, показывая, отсутствие более быстрого расхождения паттернов экспрессии. Этот результат отличается от выводов предыдущих исследований, в которых дифференциальная экспрессия между гомеологами чаще была свойственна стрессовым генам (Zhang et al., 2016a, 2016b); по-видимому, преимущественное расхождение транскрипционных паттернов генов, регулирующих ответ на стресс, не универсально среди полиплоидных видов.

Среди всех ДЭАВ пар преобладали ситуации, когда только один гомеолог экспрессировался или имел значительно более высокую экспрессию. Так как нами были собраны чрезвычайно разнообразные органы и части растения, отсутствие детектируемой экспрессии одного из гомеологов является сильным указанием на потерю функции. Например, уровень экспрессии гомеолога А гена *FLC* (одного из ключевых регуляторов перехода к цветению у *A. thaliana*, см. Главу 2) значимо снижен во всех образцах по сравнению с гомеологом В. В более редких случаях транскрипты обоих гомеологов обнаруживались в разном количестве в разных органах, что может говорить о субфункционализации или неофункционализации.

Совместно с Кулаковским И.В. нами было показано, что асимметрия экспрессии объясняется разницей в регуляторных областях ДЭАВ пар, причем эти различия не унаследованы от предковых видов, а возникли в ходе совместной эволюции двух субгеномов *C. bursa-pastoris*. В промоторах гомеологов, дифференциально отвечающих на холодный стресс, были найдены изменения в сайтах связывания ТФ семейства CBF, что отражает включение только одного из

гомеологов в регуляторную сеть ответа на стресс; кроме того, наблюдались существенные изменения в регуляции зависимого от света развития растения.

#### 4.4.5. Интеграция транскриптомных данных в базу TraVA

Организация экспрессионных данных в доступном для пользователей формате чрезвычайно важна для их повторного использования. Типы транскриптомных баз данных обсуждаются в разделе 1.2.11, а описание созданной нами базы Transcriptome Variation and Analysis (TraVA), представляющей результаты анализа транскриптомной карты *A. thaliana* приведено в разделе 1.4.8. К моменту публикации нашей работы число широкомасштабных экспрессионных экспериментов для *C. bursa-pastoris* было крайне ограниченным (Slotte et al., 2007; Douglas et al., 2015; Kryvokhyzha et al., 2016), и эти данные не были организованы в удобном для повторного анализа формате.

Мы интегрировали результаты анализа экспрессии *C. bursa-pastoris* в базу данных TraVA, которая доступна по адресу <http://travadb.org/browse/Species=Cbp>. Этот раздел базы ориентирован в первую очередь на анализ дифференциальной экспрессии между гомеологичными генами в пределах каждого из образцов и демонстрирует уровни транскрипции пар гомеологов в 14 образцах, а также уровень изменения их экспрессии, если эта пара дифференциально экспрессируется.

#### 4.5. Заключение

Эволюция цветковых растений тесно связана с многочисленными полногеномными дупликациями. Анализ ныне существующих полиплоидов может пролить свет на процессы, происходившие на протяжении последних 140 миллионов лет. Для изучения ранних последствий полиплоидизации хорошей моделью служит естественный аллотетраплоид *C. bursa-pastoris*, обладающая рядом достоинств, таких, как небольшой геном, который был секвенирован с разделением на субгеномы, сохранение большого числа пар гомеологичных генов и близость к наиболее исследованному модельному растению *A. thaliana*.

В нашей работе мы изучали асимметрию экспрессии генов, принадлежащих к двум субгеномам А и В *C. bursa-pastoris*. С помощью создания и анализа транскриптомной карты, включавшей десять образцов различных органов и частей растения и четыре образца, подвергнутых холодовому стрессу, мы обнаружили отсутствие глобальной асимметрии экспрессии двух субгеномов и специфического сайленсинга одного из них. В тоже время было обнаружено значительное число гомеологичных пар, гены которых дифференциально экспрессируются хотя бы в одном образце, при этом число пар, в которых экспрессия гомеолога В была снижена или повышена по сравнению с гомеологом А, было одинаковым. Таким образом, мы наблюдаем субфункционализацию и утрату функции гомеологов, происходящие равным образом в обоих субгеномах.

Результаты транскриптомного анализа были добавлены в созданную нами базу данных Transcriptome Variation and Analysis (TraVA), позволяющей визуализировать уровень транскрипции гомеологичных генов и их дифференциальную экспрессию.

## МАТЕРИАЛЫ И МЕТОДЫ

### 1. Получение гомозиготных растений *A. thaliana*

Одно растение *A. thaliana* линии Col-0 (CS70000) выращивались в условиях, предотвращающих переопыление. После самоопыления одно семя было высажено, и новое растение было выращено в таких же условиях, после чего эта процедура была повторена три раза для увеличения гомозиготности линии.

### 2. Выращивание растений *A. thaliana* и *C. bursa-pastoris*

Семена были высажены на смесь вермикулита и почвы в соотношении 1:1 и для синхронизации всхожести пять дней выдерживались при +4°C. Растения, в дальнейшем подвергаемые стрессовым воздействиям (Глава 1, 3 и 4), не обрабатывались пониженной температурой. Далее растения были перенесены в климатическую камеру POL-EKO Aparatura (Poland) и выращивались в ней при световом цикле 16 часов света и 8 часов темноты, +22 °C и 50% относительной влажности. В качестве источника освещения использовались лампы Philips Master TL5 HO 54 W/840; расстояние от розеток растений до ламп составляло 27 см.

### 3. Сбор образцов *A. thaliana* и *C. bursa-pastoris*

Описание образцов приведено в соответствующих таблицах каждой из глав; все образцы собирались с помощью ручной микродиссекции иглами от медицинских шприцов. Все образцы были собраны в двух биологических повторностях; каждая повторность содержала материал с 15 растений. Собранные образцы немедленно фиксировались в ингибирующем РНКазы буфере RNAlater (Qiagen, Venlo, Netherlands). Сбор всех образцов (если не указано иное) производился с 10 до 11 часов после включения света в климатической камере.

### 4. Сбор образцов, подвергнутых холодового воздействию

Для всех образцов *A. thaliana* в Главе 3 контрольный образец собирался с 8 до 9 часов после включения света. На следующий день в 5 часов после включения света температура в климатической камере устанавливалась на + 4 °C. Образцы подвергались холодовому воздействию в течении 3 и 27 часов и собирались с 8 до 9 часов после включения света. Все образцы были собраны в двух биологических повторностях; каждая повторность содержала материал с 15 растений.

Для образцов *C. bursa-pastoris* в Главе 4 семядоли без черешков были собраны с однодневных проростков с 10 до 11 часов после включения света. Независимо выращенные однодневные проростки были подвергнуты холодовому воздействию (+ 4 °C) в течении 3 и 27

часов, либо в течении 15 часов. Все образцы собирались с 10 до 11 часов после включения света в двух биологических повторностях; каждая повторность содержала материал с 15 растений.

## 5. Выделение РНК и секвенирование

Эта часть работы не выполнялась диссертантом и приведена здесь для полноты описания методов.

Тотальная РНК была выделена с помощью набора RNeasy Plant Kit (Qiagen) в соответствии с указаниями фирмы-разработчика. Библиотеки кДНК для секвенирования на платформе Illumina были приготовлены с помощью набора TruSeq RNA Sample Prep Kits v2 (Illumina, San Diego, CA, USA) в соответствии с указаниями фирмы-разработчика.

Секвенирование библиотек кДНК было проведено на платформе Illumina HiSeq2000 с получением одиночных чтений длиной 50 п.н. (Главы 1, 2, 3) или 108 п.н. (Глава 4).

## 6. Триммирование, картирование и подсчет числа чтений, приходящихся на ген

Эта часть работы не выполнялась диссертантом и приведена здесь для полноты описания методов.

Чтения были триммированы с использованием программы CLC Genomics Workbench 6.5.1 (CLC bio, Denmark) со следующими параметрами: «качество – 0,05; обрезать неоднозначные нуклеотиды – 2; удалять 5' терминальный нуклеотид – 1; удалять 3' терминальный нуклеотид» – 1; удалять чтения длиной менее 25 п.н.». Триммированные чтения были картированы с помощью CLC Genomics Workbench на референсный геном *A. thaliana* (TAIR10) с разрешением только уникально картированных чтений (параметры length fraction = 1, similarity fraction = 0.95).

В Главе 4 картирование чтений проводилось с помощью созданного в лаборатории референсного транскриптома *C. bursa-pastoris*, содержащего регионы гомеологичных генов, выравнивающиеся без инделей.

В Главе 1, для проверки влияния неуникального картирования на экспрессию генов данные также были картированы той же программой, но с разрешением множественно картируемых чтений (до 10 совпадений).

Для каждого гена было посчитано общее число чтений, приходящихся на ген, с помощью суммирования всех чтений, картирующихся на этот ген.

Для нормализации на размер библиотеки использовался метод медианы, предложенный в (Anders and Huber, 2010).

## 7. Определение экспрессирующихся генов

Гены, имеющие не менее 16 (Глава 1) или 5 (Глава 2) чтений в каждой из биологических повторностей данного образца, считались экспрессирующимися в этом образце. Экспрессирующимся назывался ген, который экспрессируется хотя бы в одном образце.

## 8. Определение параметра Z

Для каждого гена в каждом образце параметр Z был посчитан следующим образом:

$$Z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i}$$

где  $i$  – номер гена,  $j$  – номер образца,  $x_{ij}$  – экспрессия гена  $i$  в образце  $j$ ,  $\mu_i$  – средняя экспрессия гена  $i$  по всем образцам,  $\sigma_i$  – стандартное отклонение гена  $i$  по всем образцам.

## 9. Анализ дифференциальной экспрессии

Дифференциально экспрессирующиеся гены были определены с помощью пакета статистической среды R «DESeq» (Главы 1, 2, (Anders and Huber, 2010) или «DESeq2» (Главы 3, 4 (Love et al., 2014)). В качестве пороговых значений использовалось значение FDR (false discovery rate) меньше 0,05 и уровень изменения экспрессии больше 2.

## 10. Определение ДЭ-значения

В Главе 1 для одновременного анализа 3 081 парных сравнений между образцами вводилась дополнительная поправка на множественное тестирование. Все значения FDR для 33 323 генов в 3 081 сравнении были взяты вместе и функция R «p.adjust» из пакета «stats» была применена для подсчета FDR на основе этих значений. Далее гены были отобраны, как ДЭ, с применением порогов вторичный FDR < 0,05 и уровень изменения экспрессии больше 2.

После этого ДЭ-значение гена было посчитано как число парных сравнений образцов, в которых данный ген был ДЭ.

## 11. Анализ обогащения списков генов категориями ГО

Анализ обогащения списков генов категориями ГО и других баз данных проводился с помощью веб-приложения «DAVID gene functional annotation tool» (Huang et al., 2009b, 2009a) с пороговыми значениями FDR < 0,05 и уровень измененности представленности категории больше 2.

## 12. Определение стабильных генов

Для каждого гена был посчитан коэффициент вариации (CV) как стандартное отклонение по всем образцам, деленное на среднее по всем образцам. В Главе 1 CV рассчитывался как для транскриптомной карты в отдельности, так и для объединения транскриптомной карты и набора образцов, подвергнутых стрессовым воздействиям.

### 13. Определение энтропии Шеннона

Энтропия Шеннона (H) рассчитывалась как описано в (Schug et al., 2005). Так как некоторые, но не все, образцы были собраны в виде временных серий, образцы были сгруппированы, используя иерархическую кластеризацию: образцы, дистанция между которыми была меньше 0,3 объединялись в группы и экспрессия генов в них усреднялась. Группировка образцов описана в (Klerikova et al., 2016), Таблица S15 Приложения.

### 14. Кластеризация

Дерево иерархической кластеризации (Главы 1, 3) было получено с помощью функции «hclust» из пакета «stats» статистической среды R.

В Главе 2 кластеризация профилей экспрессии генов проводилась следующим образом. Нормализованные с помощью метода медианы значения общего числа чтений на ген были поделены для каждого гена на максимальное значение. Далее, гены, экспрессирующиеся хотя бы в одном образце, были кластеризованы методом k-средних с  $k=1\ 500$ . Кластеризация проводилась функцией «kmeans» пакета R «stats» в 1 000 независимых запусков. Далее на основании результатов этой кластеризации была построена матрица расстояний. Для каждой пары генов считалось N – число кластеризаций, в которых эти два гена были в одном кластере;  $1\ 000-N$  – мера расстояния между этой парой генов. Полученная матрица расстояний использовалась для иерархической кластеризацией функцией «hclust» из пакета R «fastcluster» (Müllner, 2013). Получившееся дерево было разрезано с помощью функции «cutreeDynamic» из пакета R «dynamicTreeCut» при минимальном размере кластера 5 (Langfelder et al., 2008).

Для получения суперкластеров уровни экспрессии генов в каждом кластере были усреднены, и кластеризация средних профилей была проведена так, как описано выше (при кластеризации k-средними  $k=100$ ).

### 15. Сканирующая электронная микроскопия (СЭМ)

Для получения фотографий под СЭМ образцы фиксировались в 70% этаноле, после чего последовательно переносились в 80% этанол на 15 минут, в 96% этанол на 15 минут, в смесь этанол:ацетон (1:1) на один час и затем трижды в чистый свежий ацетон на 30 минут. Затем образцы были высушены в аппарате для сушки в критической точке, расположены на железном столике и покрыты платиной и палладием 10-20 мм толщиной. Фотографирование проводилось с помощью электронного микроскопа JSM-6380 (JEOL, Tokyo, Japan) с вольтажом 15–20 kV. Фотографии под СЭМ были обработаны, используя Adobe Photoshop.

## 16. Количественная ПЦР в реальном времени

Эта часть работы не выполнялась диссертантом и приведена здесь для полноты описания методов.

Для верификации результатов, полученных с помощью секвенирования РНК, мы провели ОТ-ПЦР в реальном времени генов *LFY*, *FLC*, *API*, *AG*, маркирующих переход к цветению. Гены *AT4G33380* и *AT4G34270*, стабильно экспрессирующиеся в широком спектре тканей (Czechowski, 2005), были использованы в качестве референсов. РНК была выделена тем же способом (раздел 4), для синтеза кДНК использовалась обратная транскриптаза SuperScript II (Invitrogen). ПЦР была проведена с использованием набора 2x KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems, South Africa) на приборе Eco Real-Time PCR (Illumina, USA) при следующей программе: 95 °C – 5 мин. (1 цикл), 95 °C – 10 с, 60 °C – 30 с (45 циклов). Уровень экспрессии генов был рассчитан с использованием метода ddCt (Livak and Schmittgen, 2001).

## 17. Определение перепредставленности мотивов ТФ в регуляторных областях генов

Эта часть работы не выполнялась диссертантом и приведена здесь для полноты описания методов.

Для определения ключевых факторов, регулирующих дифференциально экспрессирующиеся гены, была использована аннотация мишеней транскрипционных факторов, полученная с помощью метода ampDAP-seq (O'Malley et al., 2016). Для каждого набора ДЭ генов рассматривались все ТФ. Относительное обогащение мишеней ТФ в списке ДЭ генов было посчитано как  $\log_2$  (числа целей-ДЭ) / (числа целей-не ДЭ). Статистическая значимость обогащения рассчитывалась с помощью точного теста Фишера с поправкой на множественное тестирование FDR.

## 18. Размещение полученных данных в базе NCBI

Все полученные экспрессионные данные были размещены в базе данных NCBI Sequence Read Archive. Номера соответствующих записей: PRJNA314076 – Глава 1, транскриптомная карта *A. thaliana*, PRJNA324514 – Глава 1, образцы *A. thaliana*, подвергнутые стрессовой обработке, PRJNA268115 – Глава 2, временная серия меристем *A. thaliana*, PRJNA411947 – Глава 3, разные органы *A. thaliana*, подвергнутые воздействию холода, PRJNA347542 – Глава 4, транскриптомная карта *C. bursa-pastoris*.

## ВЫВОДЫ

1. С использованием методов высокопроизводительного секвенирования создана транскриптомная карта модельного объекта биологии растений *Arabidopsis thaliana*, включающая 79 органов и тканей на разных стадиях развития. Анализ транскриптомной карты:

- показал, что в наборе изучаемых образцов детектируются транскрипты более 99% генов, потенциально способных к экспрессии вне зависимости от внешних условий;
- подтвердил, что только небольшая часть (1,8%) генов обладает узким паттерном экспрессии, указывая на их вероятное участие в определении тканевой специфичности, тогда как большая часть генов экспрессируются во всех или большинстве тканей растения;
- продемонстрировал, что гены, обладающие равномерной экспрессией во всех образцах, участвуют в разных биологических процессах, но чаще связаны с функционированием мембран;
- показал, что различные семейства транскрипционных факторов обладают разным уровнем тканевой специфичности: для семейств MADS, LOB, LIM и MYB характерны более узкие паттерны экспрессии, тогда как SWI/SNF, SWI3, SNF2, CAMTA, DDT и FAR, как правило, присутствуют во всех или в большинстве органах и тканях *Arabidopsis thaliana*.

2. Проведенный с помощью атласа экспрессии генов *Arabidopsis thaliana* анализ биологических процессов показал, что:

- в момент перехода к цветению в апикальной меристеме побега *Arabidopsis thaliana* происходит критическое изменение динамики клеточных делений, которое, вероятно, связано с сокращением длительности клеточного цикла за счет фаз G<sub>1</sub> и G<sub>2</sub>;
- развитие холодового стресса у *Arabidopsis thaliana* сочетает универсальные для всего растения процессы с тканеспецифичными; в том числе показано, что в некоторых случаях тканеспецифичный ответ на стресс происходит с помощью рекрутирования генов, в нормальных условиях отвечающие за развитие других частей растения.

3. Создана транскриптомная карта *Capsella-bursa-pastoris* – аллотетраплоидного вида, эволюционно и морфологически близкого к *Arabidopsis thaliana*. Показано отсутствие геномного доминирования, то есть преимущественной экспрессии генов, принадлежащих одному из субгеномов, и случайный характер инактивации одного гена из пары гомеологов на ранней стадии эволюции полиплоидов. Сравнение профилей экспрессии гомеологов с известными данными о функциональной эволюции генов показало, что изменение паттерна и/или уровня экспрессии является свидетельством изменения функции гена.

## СПИСОК ЛИТЕРАТУРЫ

1. Чайлахян, М.Х. Гормональная теория развития растений / М.Х. Чайлахян. – Москва: Изд-во Академии наук СССР, 1937. – 197 с.
2. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium / Z. Su [и др.] // *Nature Biotechnology*. – 2014. – Т. 32. – № 9. – С. 903-914.
3. A Developmental Transcriptome Map for Allotetraploid *Arachis hypogaea* / J. Clevenger [и др.] // *Frontiers in Plant Science*. – 2016. – Т. 7.
4. A gene expression map of *Arabidopsis thaliana* development / M. Schmid [и др.] // *Nature Genetics*. – 2005. – Т. 37. – № 5. – С. 501-506.
5. A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies / Y. Jiao [и др.] // *Nature Genetics*. – 2009. – Т. 41. – № 2. – С. 258-263.
6. An expression atlas of rice mRNAs and small RNAs / K. Nobuta [и др.] // *Nature Biotechnology*. – 2007. – Т. 25. – № 4. – С. 473-477.
7. Analysis of *Arabidopsis* floral transcriptome: detection of new florally expressed genes and expansion of Brassicaceae-specific gene families / L. Zhang [и др.] // *Frontiers in Plant Science*. – 2015. – Т. 5. – Analysis of *Arabidopsis* floral transcriptome.
8. Architecture and evolution of a minute plant genome / E. Ibarra-Laclette [и др.] // *Nature*. – 2013. – Т. 498. – № 7452. – С. 94-98.
9. Barrero-Gil J. Gene Regulatory Networks Mediating Cold Acclimation: The CBF Pathway / J. Barrero-Gil, J. Salinas // *Survival Strategies in Extreme Cold and Desiccation* / ред. М. Iwaya-Inoue, М. Sakurai, М. Uemura. – Singapore: Springer Singapore, 2018. – Т. 1081. – Gene Regulatory Networks Mediating Cold Acclimation. – С. 3-22.
10. Briggs W.R. Photoreceptors in Plant Photomorphogenesis to Date. Five Phytochromes, Two Cryptochromes, One Phototropin, and One Superchrome / W.R. Briggs // *PLANT PHYSIOLOGY*. – 2001. – Т. 125. – № 1. – С. 85-88.
11. Durdan S.F. Activation of latent origins of DNA replication in florally determined shoot meristems of long-day and short-day plants: *Silene coeli-rosa* and *Pharbitis nil* / S.F. Durdan, R.J. Herbert, D. Francis // *Planta*. – 1998. – Т. 207. – Activation of latent origins of DNA replication in florally determined shoot meristems of long-day and short-day plants. – № 2. – С. 235-240.
12. Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.) / Y. Zhang [и др.] // *Scientific Reports*. – 2016a. – Т. 6. – С. 21476.
13. Genome sequence and analysis of the tuber crop potato / X. Xu [и др.] // *Nature*. – 2011. – Т. 475. – № 7355. – С. 189-195.
14. Genome sequence of the palaeopolyploid soybean / J. Schmutz [и др.] // *Nature*. – 2010. – Т. 463. – № 7278. – С. 178-183.
15. Marc J. Variation in Cell-Cycle Time and Nuclear DNA Content in the Apical Meristem of *Helianthus annuus* L. During the Transition to Flowering / J. Marc, J.H. Palmer // *American Journal of Botany*. – 1984. – Т. 71. – № 4. – С. 588.
16. Ng M. Function and evolution of the plant MADS-box gene family / M. Ng, M.F. Yanofsky // *Nature Reviews Genetics*. – 2001. – Т. 2. – № 3. – С. 186-195.
17. Rossel J.B. Global Changes in Gene Expression in Response to High Light in *Arabidopsis* / J.B. Rossel // *PLANT PHYSIOLOGY*. – 2002. – Т. 130. – № 3. – С. 1109-1120.
18. Shannon S. Genetic Interactions That Regulate Inflorescence Development in *Arabidopsis* / S. Shannon // *THE PLANT CELL ONLINE*. – 1993. – Т. 5. – № 6. – С. 639-655.

19. Sharma K.D. Regulatory Networks in Pollen Development under Cold Stress / K.D. Sharma, H. Nayyar // *Frontiers in Plant Science*. – 2016. – Т. 7.
20. The tomato genome sequence provides insights into fleshy fruit evolution / S. Sato [и др.] // *Nature*. – 2012. – Т. 485. – № 7400. – С. 635-641.
21. Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat / H. Zhang [и др.] // *Scientific Reports*. – 2016b. – Т. 6. – С. 26363.
22. A comprehensive analysis of interaction and localization of Arabidopsis SKP1-like (ASK) and F-box (FBX) proteins / H. Kuroda [и др.] // *PLoS One*. – 2012. – Т. 7. – № 11. – С. e50009.
23. A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis / S. Cao [и др.] // *The Plant Cell*. – 2014. – Т. 26. – № 3. – С. 1009-1017.
24. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) / J. Yu [и др.] // *Science* (New York, N.Y.). – 2002. – Т. 296. – № 5565. – С. 79-92.
25. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*) / S.A. Goff [и др.] // *Science* (New York, N.Y.). – 2002. – Т. 296. – № 5565. – С. 92-100.
26. A dynamic gene expression atlas covering the entire life cycle of rice / L. Wang [et al.] // *The Plant Journal*. – 2010. – Vol. 61. – № 5. – P. 752-766.
27. A gene atlas of the mouse and human protein-encoding transcriptomes / A.I. Su [et al.] // *Proceedings of the National Academy of Sciences*. – 2004. – Vol. 101. – № 16. – P. 6062-6067.
28. A gene expression atlas of the model legume *Medicago truncatula* / V.A. Benedito [et al.] // *The Plant Journal*. – 2008. – Vol. 55. – № 3. – P. 504-513.
29. A gene expression map of shoot domains reveals regulatory mechanisms / C. Tian [et al.] // *Nature Communications*. – 2019. – Vol. 10. – № 1.
30. A High Resolution Map of the Arabidopsis thaliana Developmental Transcriptome Based on RNA-seq Profiling / A.V. Klepikova [et al.] // *The Plant Journal*. – 2016.
31. A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche / R.K. Yadav [et al.] // *Development*. – 2014. – Vol. 141. – № 13. – P. 2735-2744.
32. A MADS domain gene involved in the transition to flowering in Arabidopsis / R. Borner [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2000. – Т. 24. – № 5. – С. 591-599.
33. A pair of related genes with antagonistic roles in mediating flowering signals / Y. Kobayashi [и др.] // *Science* (New York, N.Y.). – 1999. – Т. 286. – № 5446. – С. 1960-1962.
34. A Rho family GTPase controls actin dynamics and tip growth via two counteracting downstream pathways in pollen tubes / Y. Gu [и др.] // *The Journal of Cell Biology*. – 2005. – Т. 169. – № 1. – С. 127-138.
35. A Sacrifice-for-Survival Mechanism Protects Root Stem Cell Niche from Chilling Stress / J.H. Hong [et al.] // *Cell*. – 2017. – Vol. 170. – № 1. – P. 102-113.e14.
36. A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain / K. Davie [et al.] // *Cell*. – 2018. – Vol. 174. – № 4. – P. 982-998.e20.
37. A Sorghum bicolor expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums / N. Shakoor [et al.] // *BMC Plant Biology*. – 2014. – Vol. 14. – № 1. – P. 35.
38. A survey of best practices for RNA-seq data analysis / A. Conesa [et al.] // *Genome Biology*. – 2016. – Vol. 17. – № 1.

39. A Transcriptome Atlas of *Physcomitrella patens* Provides Insights into the Evolution and Development of Land Plants / C. Ortiz-Ramírez [et al.] // *Molecular Plant*. – 2016. – Vol. 9. – № 2. – P. 205-220.
40. A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing / S.-Y. Chen [et al.] // *Scientific Reports*. – 2017. – Vol. 7. – № 1.
41. Activation tagging of the floral inducer FT / I. Kardailsky [и др.] // *Science (New York, N.Y.)*. – 1999. – Т. 286. – № 5446. – С. 1962-1965.
42. Adams K.L. Polyploidy and genome evolution in plants / K.L. Adams, J.F. Wendel // *Current Opinion in Plant Biology*. – 2005. – Т. 8. – № 2. – С. 135-141.
43. Advances and current challenges in calcium signaling / J. Kudla [и др.] // *The New Phytologist*. – 2018. – Т. 218. – № 2. – С. 414-431.
44. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears / M.F. Rai [и др.] // *Journal of Orthopaedic Research: Official Publication of the Orthopaedic Research Society*. – 2018. – Т. 36. – № 1. – С. 484-497.
45. Ågren J.A. Transposable element evolution in the allotetraploid *Capsella bursa-pastoris* / J.A. Ågren, H.-R. Huang, S.I. Wright // *American Journal of Botany*. – 2016. – Vol. 103. – № 7. – P. 1197-1202.
46. An advanced reference genome of *Trifolium subterraneum* L. reveals genes related to agronomic performance / P. Kaur [et al.] // *Plant Biotechnology Journal*. – 2017. – Vol. 15. – № 8. – P. 1034-1046.
47. An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors / J. Jin [et al.] // *Molecular Biology and Evolution*. – 2015. – Vol. 32. – № 7. – P. 1767-1773.
48. An atlas and analysis of bovine skeletal muscle long noncoding RNAs / X.F. Liu [и др.] // *Animal Genetics*. – 2017. – Т. 48. – № 3. – С. 278-286.
49. An atlas of gene expression from seed to seed through barley development / A. Druka [et al.] // *Functional & Integrative Genomics*. – 2006. – Vol. 6. – № 3. – P. 202-211.
50. An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets / D. Winter [et al.] // *PLoS ONE*. – 2007. – Vol. 2. – № 8. – P. e718.
51. An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development / S.C. Stelpflug [et al.] // *The Plant Genome*. – 2016. – Vol. 9. – № 1. – P. 0.
52. An H. CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis / H. An // *Development*. – 2004. – Vol. 131. – № 15. – P. 3615-3626.
53. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants: Soybean transcriptome atlas / M. Libault [et al.] // *The Plant Journal*. – 2010. – An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. – P. no-no.
54. An Overview of Signaling Regulons During Cold Stress Tolerance in Plants / A. Pareek [и др.] // *Current Genomics*. – 2017. – Т. 18. – № 6. – С. 498-511.
55. An update to database TraVA: organ-specific cold stress response in *Arabidopsis thaliana* / A.V. Klepikova [et al.] // *BMC Plant Biology*. – 2019. – Vol. 19. – An update to database TraVA. – № S1.
56. Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering / S. Torti [и др.] // *The Plant Cell*. – 2012. – Т. 24. – № 2. – С. 444-462.
57. Anders S. Differential expression analysis for sequence count data / S. Anders, W. Huber // *Genome Biology*. – 2010. – Т. 11. – № 10. – C. R106.

58. Andrés F. The genetic basis of flowering responses to seasonal cues / F. Andrés, G. Coupland // *Nature Reviews Genetics*. – 2012. – Vol. 13. – № 9. – P. 627-639.
59. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage / H. Tang [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2010. – Т. 107. – № 1. – С. 472-477.
60. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization / T. Csorba [et al.] // *Proceedings of the National Academy of Sciences*. – 2014. – Vol. 111. – № 45. – P. 16160-16165.
61. Arabidopsis Biological Resource Center. // The Ohio State University. – 2015. Адрес сайта <https://abrc.osu.edu/seed-handling> (дата обращения 31 марта 2019 г.).
62. Arabidopsis FIMBRIN5, an actin bundling factor, is required for pollen germination and pollen tube growth / Y. Wu [и др.] // *The Plant Cell*. – 2010. – Т. 22. – № 11. – С. 3745-3763.
63. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* / Arabidopsis Genome Initiative // *Nature*. – 2000. – Т. 408. – № 6814. – С. 796-815.
64. Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability / S. Laubinger [и др.] // *Development (Cambridge, England)*. – 2006. – Т. 133. – № 16. – С. 3213-3222.
65. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27 / F. Turck [и др.] // *PLoS genetics*. – 2007. – Т. 3. – № 6. – С. e86.
66. Arabidopsis thaliana: a model plant for genome analysis / D.W. Meinke [и др.] // *Science (New York, N.Y.)*. – 1998. – Т. 282. – Arabidopsis thaliana. – № 5389. – С. 662, 679-682.
67. Arabidopsis VILLIN1 and VILLIN3 have overlapping and distinct activities in actin bundle formation and turnover / P. Khurana [и др.] // *The Plant Cell*. – 2010. – Т. 22. – № 8. – С. 2727-2748.
68. Assis R. Neofunctionalization of young duplicate genes in *Drosophila* / R. Assis, D. Bachtrog // *Proceedings of the National Academy of Sciences*. – 2013. – Vol. 110. – № 43. – P. 17409-17414.
69. ASYMMETRIC LEAVES2-LIKE1 gene, a member of the AS2/LOB family, controls proximal-distal patterning in Arabidopsis petals / A. Chalfun-Junior [et al.] // *Plant Molecular Biology*. – 2005. – Vol. 57. – ASYMMETRIC LEAVES2-LIKE1 gene, a member of the AS2/LOB family, controls proximal? – № 4. – P. 559-575.
70. AtGLR3.4, a glutamate receptor channel-like gene is sensitive to touch and cold / O. Meyerhoff [и др.] // *Planta*. – 2005. – Т. 222. – № 3. – С. 418-427.
71. Bamberg J.R. Proteins of the ADF/cofilin family: essential regulators of actin dynamics / J.R. Bamberg // *Annual Review of Cell and Developmental Biology*. – 1999. – Т. 15. – Proteins of the ADF/cofilin family. – С. 185-230.
72. Bargmann B.O.R. Fluorescence Activated Cell Sorting of Plant Protoplasts / B.O.R. Bargmann, K.D. Birnbaum // *Journal of Visualized Experiments*. – 2010. – № 1.
73. Bar-Joseph Z. Studying and modelling dynamic biological processes using time-series gene expression data / Z. Bar-Joseph, A. Gitter, I. Simon // *Nature Reviews. Genetics*. – 2012. – Т. 13. – № 8. – С. 552-564.
74. Bäurle I. Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of Arabidopsis targets / I. Bäurle, C. Dean // *PloS One*. – 2008. – Т. 3. – № 7. – С. e2733.

75. Becker J.D. Transcriptional Profiling of Arabidopsis Tissues Reveals the Unique Characteristics of the Pollen Transcriptome / J.D. Becker // *PLANT PHYSIOLOGY*. – 2003. – Vol. 133. – № 2. – P. 713-725.
76. Bernier G. My favourite flowering image: the role of cytokinin as a flowering signal / G. Bernier // *Journal of Experimental Botany*. – 2013. – Т. 64. – My favourite flowering image. – № 18. – С. 5795-5799.
77. Bioengineering plant resistance to abiotic stresses by the global calcium signal system / S. Hong-Bo [и др.] // *Biotechnology Advances*. – 2008. – Т. 26. – № 6. – С. 503-510.
78. Birnbaum K. A Gene Expression Map of the Arabidopsis Root / K. Birnbaum // *Science*. – 2003. – Vol. 302. – № 5652. – P. 1956-1960.
79. Blake J.A. Ten Quick Tips for Using the Gene Ontology / J.A. Blake // *PLoS Computational Biology*. – 2013. – Vol. 9. – № 11. – P. e1003343.
80. Blue light-dependent interaction of CRY2 with SPA1 regulates COP1 activity and floral initiation in Arabidopsis / Z. Zuo [и др.] // *Current biology: CB*. – 2011. – Т. 21. – № 10. – С. 841-847.
81. Bodson M. Variation in the Rate of Cell Division in the Apical Meristem of *Sinapis alba* During Transition to Flowering / M. Bodson // *Annals of Botany*. – 1975. – Vol. 39. – № 3. – P. 547-554.
82. Brassinosteroids participate in the control of basal and acquired freezing tolerance of plants / M. Eremina [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2016. – Т. 113. – № 40. – С. E5982-E5991.
83. Briggs W.R. Phototropins 1 and 2: versatile plant blue-light receptors / W.R. Briggs, J.M. Christie // *Trends in Plant Science*. – 2002. – Т. 7. – Phototropins 1 and 2. – № 5. – С. 204-210.
84. BZR1 Positively Regulates Freezing Tolerance via CBF-Dependent and CBF-Independent Pathways in Arabidopsis / H. Li [и др.] // *Molecular Plant*. – 2017. – Т. 10. – № 4. – С. 545-559.
85. Calcium/calmodulin-regulated receptor-like kinase CRLK1 interacts with MEKK1 in plants / T. Yang [и др.] // *Plant Signaling & Behavior*. – 2010. – Т. 5. – № 8. – С. 991-994.
86. Callis J. The ubiquitination machinery of the ubiquitin system / J. Callis // *The Arabidopsis Book*. – 2014. – Т. 12. – С. e0174.
87. Capovilla G. Control of flowering by ambient temperature / G. Capovilla, M. Schmid, D. Pose // *Journal of Experimental Botany*. – 2015. – Vol. 66. – № 1. – P. 59-69.
88. Capsella as a model system to study the evolutionary relevance of floral homeotic mutants / P. Nutt [et al.] // *Plant Systematics and Evolution*. – 2006. – Vol. 259. – № 2-4. – P. 217-235.
89. Castillejo C. The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering / C. Castillejo, S. Pelaz // *Current biology: CB*. – 2008. – Т. 18. – № 17. – С. 1338-1343.
90. Catching a “hopeful monster”: shepherd’s purse (*Capsella bursa-pastoris*) as a model system to study the evolution of flower development / M. Hintz [et al.] // *Journal of Experimental Botany*. – 2006. – Vol. 57. – Catching a “hopeful monster”. – № 13. – P. 3531-3542.
91. CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in Arabidopsis / F. Novillo [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2004. – Т. 101. – № 11. – С. 3985-3990.
92. Cell cycle-regulated gene expression in Arabidopsis / M. Menges [и др.] // *The Journal of Biological Chemistry*. – 2002. – Т. 277. – № 44. – С. 41987-42002.
93. Chapter 2 Cold Signalling and Cold Acclimation in Plants / E. Ruelland [et al.] // *Advances in Botanical Research*. – Elsevier, 2009. – Vol. 49. – P. 35-150.

94. Characterization of genes in the ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB) family in *Arabidopsis thaliana*, and functional and molecular comparisons between AS2 and other family members / Y. Matsumura [et al.] // *The Plant Journal*. – 2009. – Vol. 58. – № 3. – P. 525-537.
95. Chinnusamy V. Gene regulation during cold stress acclimation in plants / V. Chinnusamy, J.-K. Zhu, R. Sunkar // *Methods in Molecular Biology (Clifton, N.J.)*. – 2010. – Т. 639. – С. 39-55.
96. Chromatin-mediated feed-forward auxin biosynthesis in floral meristem determinacy / N. Yamaguchi [и др.] // *Nature Communications*. – 2018. – Т. 9. – № 1. – С. 5290.
97. CIPK7 is involved in cold response by interacting with CBL1 in *Arabidopsis thaliana* / C. Huang [и др.] // *Plant Science: An International Journal of Experimental Plant Biology*. – 2011. – Т. 181. – № 1. – С. 57-64.
98. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape / R.C. O'Malley [и др.] // *Cell*. – 2016. – Т. 165. – № 5. – С. 1280-1292.
99. Clustering of high throughput gene expression data / H. Pirim [et al.] // *Computers & Operations Research*. – 2012. – Vol. 39. – № 12. – P. 3046-3061.
100. Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd's Purse (*Capsella bursa-pastoris*) / K.R. St Onge [и др.] // *Molecular Biology and Evolution*. – 2012. – Т. 29. – № 7. – С. 1721-1733.
101. Cold-induced repression of the rice anther-specific cell wall invertase gene OSINV4 is correlated with sucrose accumulation and pollen sterility / S.N. Oliver [et al.] // *Plant, Cell and Environment*. – 2005. – Vol. 28. – № 12. – P. 1534-1551.
102. Comparative Analysis of Developmental Transcriptome Maps of *Arabidopsis thaliana* and *Solanum lycopersicum* / A.A. Penin [et al.] // *Genes*. – 2019. – Vol. 10. – № 1. – P. 50.
103. Comparative analysis of the transcriptomes of *Populus trichocarpa* and *Arabidopsis thaliana* suggests extensive evolution of gene expression regulation in angiosperms / T. Quesada [et al.] // *New Phytologist*. – 2008. – Vol. 180. – № 2. – P. 408-420.
104. Comparative transcriptomics in the Triticeae / A.W. Schreiber [et al.] // *BMC Genomics*. – 2009. – Vol. 10. – № 1. – P. 285.
105. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells / S. Zhao [et al.] // *PLoS ONE*. – 2014. – Vol. 9. – № 1. – P. e78644.
106. Comparison of the transcriptional landscapes between human and mouse tissues / S. Lin [et al.] // *Proceedings of the National Academy of Sciences*. – 2014. – Vol. 111. – № 48. – P. 17224-17229.
107. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection / M. Libault [и др.] // *Plant Physiology*. – 2010. – Т. 152. – № 2. – С. 541-552.
108. Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa* / C. Tong [et al.] // *BMC Genomics*. – 2013. – Vol. 14. – № 1. – P. 689.
109. Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage / D.-H. Jeong [и др.] // *Plant Physiology*. – 2013. – Т. 162. – № 3. – С. 1225-1245.
110. Conversion of leaves into petals in *Arabidopsis* / S. Pelaz [и др.] // *Current biology: CB*. – 2001. – Т. 11. – № 3. – С. 182-184.
111. COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in *Arabidopsis* / L.-J. Liu [и др.] // *The Plant Cell*. – 2008a. – Т. 20. – № 2. – С. 292-306.

112. Crevillén P. Regulation of the floral repressor gene FLC: the complexity of transcription in a chromatin context / P. Crevillén, C. Dean // *Current Opinion in Plant Biology*. – 2011. – Т. 14. – № 1. – С. 38-44.
113. Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses / S.-C. Jiang [и др.] // *Plant Molecular Biology*. – 2015. – Т. 88. – № 4-5. – С. 369-385.
114. CYTOKININ RESPONSE FACTOR2 (CRF2) and CRF3 Regulate Lateral Root Development in Response to Cold Stress in Arabidopsis / J. Jeon [и др.] // *The Plant Cell*. – 2016. – Т. 28. – № 8. – С. 1828-1843.
115. Czechowski T. Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis / T. Czechowski // *PLANT PHYSIOLOGY*. – 2005. – Vol. 139. – № 1. – P. 5-17.
116. DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in Arabidopsis / T. Tsutsui [и др.] // *Journal of Plant Research*. – 2009. – Т. 122. – № 6. – С. 633-643.
117. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome / G. Zhang [et al.] // *Genome Research*. – 2010. – Vol. 20. – № 5. – P. 646-654.
118. DELLA proteins physically interact with CONSTANS to regulate flowering under long days in Arabidopsis / F. Xu [и др.] // *FEBS letters*. – 2016. – Т. 590. – № 4. – С. 541-549.
119. Demarsy E. Higher plants use LOV to perceive blue light / E. Demarsy, C. Fankhauser // *Current Opinion in Plant Biology*. – 2009. – Т. 12. – № 1. – С. 69-74.
120. Differential expression of genes important for adaptation in *Capsella bursa-pastoris* (Brassicaceae) / T. Slotte [и др.] // *Plant Physiology*. – 2007. – Т. 145. – № 1. – С. 160-173.
121. Differential Mechanisms of Photosynthetic Acclimation to Light and Low Temperature in Arabidopsis and the Extremophile *Eutrema salsugineum* / N. Khanal [и др.] // *Plants* (Basel, Switzerland). – 2017. – Т. 6. – № 3.
122. Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization / W. Albertin [и др.] // *BMC genomics*. – 2007. – Т. 8. – С. 56.
123. Dissection of floral induction pathways using global expression analysis / M. Schmid [и др.] // *Development* (Cambridge, England). – 2003. – Т. 130. – № 24. – С. 6001-6012.
124. Distinct Patterns of Genetic Variation Alter Flowering Responses of Arabidopsis Accessions to Different Daylengths / A. Giakountis [et al.] // *PLANT PHYSIOLOGY*. – 2010. – Vol. 152. – № 1. – P. 177-191.
125. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis / A. Samach [и др.] // *Science* (New York, N.Y.). – 2000. – Т. 288. – № 5471. – С. 1613-1616.
126. Dong M.A. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in Arabidopsis / M.A. Dong, E.M. Farré, M.F. Thomashow // *Proceedings of the National Academy of Sciences of the United States of America*. – 2011. – Т. 108. – № 17. – С. 7241-7246.
127. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution / L. Flagel [и др.] // *BMC biology*. – 2008. – Т. 6. – С. 16.
128. Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity / B.L. Orvar [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2000. – Т. 23. – № 6. – С. 785-794.

129. Eichten S.R. Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation / S.R. Eichten, R.J. Schmitz, N.M. Springer // *Plant Physiology*. – 2014. – Т. 165. – Epigenetics. – № 3. – С. 933-947.
130. Eisenberg E. Human housekeeping genes, revisited / E. Eisenberg, E.Y. Levanon // *Trends in Genetics*. – 2013. – Vol. 29. – № 10. – P. 569-574.
131. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology / J. Waese [et al.] // *The Plant Cell*. – 2017. – Vol. 29. – ePlant. – № 8. – P. 1806-1821.
132. Establishment of the Lotus japonicus Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation / J. Verdier [et al.] // *The Plant Journal*. – 2013. – Vol. 74. – № 2. – P. 351-362.
133. Evaluation of Reference Genes for Accurate Normalization of Gene Expression for Real Time-Quantitative PCR in *Pyrus pyrifolia* Using Different Tissue Samples and Seasonal Conditions / T. Imai [et al.] // *PLoS ONE*. – 2014. – Vol. 9. – № 1. – P. e86492.
134. Exploring the Temperature-Stress Metabolome of Arabidopsis / F. Kaplan [et al.] // *PLANT PHYSIOLOGY*. – 2004. – Vol. 136. – № 4. – P. 4159-4168.
135. Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon* / R. Sibout [и др.] // *The New Phytologist*. – 2017. – Т. 215. – № 3. – С. 1009-1025.
136. Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses / W. Chen [и др.] // *The Plant Cell*. – 2002. – Т. 14. – № 3. – С. 559-574.
137. Expression profiling and bioinformatic analyses of a novel stress-regulated multispanning transmembrane protein family from cereals and Arabidopsis / G. Breton [и др.] // *Plant Physiology*. – 2003. – Т. 132. – № 1. – С. 64-74.
138. F-box proteins FKF1 and LKP2 act in concert with ZEITLUPE to control Arabidopsis clock progression / A. Baudry [и др.] // *The Plant Cell*. – 2010. – Т. 22. – № 3. – С. 606-622.
139. Feng Y. Comparative study of rice and Arabidopsis actin-depolymerizing factors gene families / Y. Feng, Q. Liu, Q. Xue // *Journal of Plant Physiology*. – 2006. – Т. 163. – № 1. – С. 69-79.
140. Finding the active genes in deep RNA-seq gene expression studies / T. Hart [et al.] // *BMC Genomics*. – 2013. – Vol. 14. – № 1. – P. 778.
141. FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis / M. Sawa [и др.] // *Science (New York, N.Y.)*. – 2007. – Т. 318. – № 5848. – С. 261-265.
142. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering / Y.H. Song [и др.] // *Science (New York, N.Y.)*. – 2012. – Т. 336. – № 6084. – С. 1045-1049.
143. FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis / T. Imaizumi [и др.] // *Nature*. – 2003. – Т. 426. – № 6964. – С. 302-306.
144. Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana / S. Melzer [и др.] // *Nature Genetics*. – 2008. – Т. 40. – № 12. – С. 1489-1492.
145. Food and Agricultural Organization of the United Nations. Food and Agricultural Organization of the United Nations Cereal Supply and Demand Brief. Адрес сайта: <http://www.fao.org/worldfoodsituation/csdb/en/> (дата обращения 1 августа 2019 г.).
146. Fowler S. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway / S. Fowler, M.F. Thomashow // *The Plant Cell*. – 2002. – Т. 14. – № 8. – С. 1675-1690.
147. Fowler S.G. Low Temperature Induction of Arabidopsis CBF1, 2, and 3 Is Gated by the Circadian Clock / S.G. Fowler // *PLANT PHYSIOLOGY*. – 2005. – Vol. 137. – № 3. – P. 961-968.

148. Francis D. Synchronisation of cell division in the shoot apex of *Silene* in relation to flower initiation / D. Francis, R.F. Lyndon // *Planta*. – 1979. – Т. 145. – № 2. – С. 151-157.
149. Francis D. The cell cycle in plant development / D. Francis // *New Phytologist*. – 1992. – Vol. 122. – № 1. – P. 1-20.
150. Frequent Introgressions from Diploid Species Contribute to the Adaptation of the Tetraploid Shepherd's Purse (*Capsella bursa-pastoris*) / T.-S. Han [et al.] // *Molecular Plant*. – 2015. – Vol. 8. – № 3. – P. 427-438.
151. FTIP1 Is an Essential Regulator Required for Florigen Transport / L. Liu [et al.] // *PLoS Biology*. – 2012. – Vol. 10. – № 4. – P. e1001313.
152. Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species / S. Alves-Carvalho [et al.] // *The Plant Journal*. – 2015. – Vol. 84. – № 1. – P. 1-19.
153. Functional characterization of the SIZ/PIAS-type SUMO E3 ligases, OsSIZ1 and OsSIZ2 in rice / H.C. Park [и др.] // *Plant, Cell & Environment*. – 2010. – Т. 33. – № 11. – С. 1923-1934.
154. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification / D. Vekemans [и др.] // *Molecular Biology and Evolution*. – 2012. – Т. 29. – № 12. – С. 3793-3806.
155. Gegas V.C. Expression of cell cycle genes in shoot apical meristems / V.C. Gegas, J.H. Doonan // *Plant Molecular Biology*. – 2006. – Т. 60. – № 6. – С. 947-961.
156. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays / S. Brenner [et al.] // *Nature Biotechnology*. – 2000. – Vol. 18. – № 6. – P. 630-634.
157. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium / M. Ashburner [и др.] // *Nature Genetics*. – 2000. – Т. 25. – Gene ontology. – № 1. – С. 25-29.
158. Gene-Sharing Networks Reveal Organizing Principles of Transcriptomes in Arabidopsis and Other Multicellular Organisms / S. Li [et al.] // *The Plant Cell*. – 2012. – Vol. 24. – № 4. – P. 1362-1378.
159. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes / A. Vlasova [et al.] // *Genome Biology*. – 2016. – Vol. 17. – № 1.
160. Genome organization of more than 300 defensin-like genes in Arabidopsis / K.A.T. Silverstein [и др.] // *Plant Physiology*. – 2005. – Т. 138. – № 2. – С. 600-610.
161. Genome-scale cold stress response regulatory networks in ten Arabidopsis thaliana ecotypes / P. Barah [и др.] // *BMC genomics*. – 2013. – Т. 14. – С. 722.
162. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis / C. Lurin [и др.] // *The Plant Cell*. – 2004. – Т. 16. – № 8. – С. 2089-2103.
163. Genome-Wide Analysis of Spatial Gene Expression in Arabidopsis Flowers / F. Wellmer [et al.] // *The Plant Cell*. – 2004. – Vol. 16. – № 5. – P. 1314-1326.
164. Genome-wide analysis of the cyclin family in Arabidopsis and comparative phylogenetic analysis of plant cyclin-like proteins / G. Wang [и др.] // *Plant Physiology*. – 2004. – Т. 135. – № 2. – С. 1084-1099.
165. Genome-wide atlas of transcription during maize development: Maize gene atlas / R.S. Sekhon [et al.] // *The Plant Journal*. – 2011. – Vol. 66. – Genome-wide atlas of transcription during maize development. – № 4. – P. 553-563.
166. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids / J. Wang [и др.] // *Genetics*. – 2006. – Т. 172. – № 1. – С. 507-517.

167. Genome-Wide Survey of Cold Stress Regulated Alternative Splicing in *Arabidopsis thaliana* with Tiling Microarray / N. Leviatan [et al.] // PLoS ONE. – 2013. – Vol. 8. – № 6. – P. e66511.
168. Gibberellic acid signaling is required for ambient temperature-mediated induction of flowering in *Arabidopsis thaliana* / V.C. Galvão [et al.] // The Plant Journal. – 2015. – Vol. 84. – № 5. – P. 949-962.
169. Gibberellin Regulates the *Arabidopsis* Floral Transition through miR156-Targeted SQUAMOSA PROMOTER BINDING-LIKE Transcription Factors / S. Yu [et al.] // The Plant Cell. – 2012. – Vol. 24. – № 8. – P. 3320-3332.
170. GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains / S. Fowler [и др.] // The EMBO journal. – 1999. – Т. 18. – GIGANTEA. – № 17. – C. 4679-4688.
171. Gilmour S.J. *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities / S.J. Gilmour, S.G. Fowler, M.F. Thomashow // Plant Molecular Biology. – 2004. – Т. 54. – № 5. – C. 767-781.
172. Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control / M. Menges [и др.] // The Plant Journal: For Cell and Molecular Biology. – 2005. – Т. 41. – № 4. – C. 546-566.
173. Global Prediction of Tissue-Specific Gene Expression and Context-Dependent Gene Networks in *Caenorhabditis elegans* / M.D. Chikina [et al.] // PLoS Computational Biology. – 2009. – Vol. 5. – № 6. – P. e1000417.
174. Gonthier R. Changes in cell-cycle duration and growth fraction in the shoot meristem of *Sinapis* during floral transition / R. Gonthier, A. Jacqumard, G. Bernier // Planta. – 1987. – Т. 170. – № 1. – C. 55-59.
175. Goremykin V.V. Analysis of the *Amborella trichopoda* Chloroplast Genome Sequence Suggests That *Amborella* Is Not a Basal Angiosperm / V.V. Goremykin // Molecular Biology and Evolution. – 2003. – Vol. 20. – № 9. – P. 1499-1505.
176. Grose S. Inhibition of growth and synchronised cell division in the shoot apex in relation to flowering in *Silene* / S. Grose, R.F. Lyndon // Planta. – 1984. – Т. 161. – № 4. – C. 289-294.
177. Guo X. Cold signaling in plants: Insights into mechanisms and regulation: Cold stress signaling / X. Guo, D. Liu, K. Chong // Journal of Integrative Plant Biology. – 2018. – Vol. 60. – Cold signaling in plants. – № 9. – P. 745-756.
178. Hanano S. *Arabidopsis* TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression / S. Hanano, K. Goto // The Plant Cell. – 2011. – Т. 23. – № 9. – C. 3172-3184.
179. Hansey C.N. Cell Wall Composition and Ruminant Digestibility of Various Maize Tissues Across Development / C.N. Hansey, A.J. Lorenz, N. de Leon // BioEnergy Research. – 2010. – Vol. 3. – № 1. – P. 28-37.
180. Hanzawa Y. A single amino acid converts a repressor to an activator of flowering / Y. Hanzawa, T. Money, D. Bradley // Proceedings of the National Academy of Sciences. – 2005. – Vol. 102. – № 21. – P. 7748-7753.
181. Heo J.B. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA / J.B. Heo, S. Sung // Science (New York, N.Y.). – 2011. – Т. 331. – № 6013. – C. 76-79.

182. High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution / A.S. Kasianov [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2017. – Т. 91. – № 2. – С. 278-291.
183. Ho W.W.H. Structural features determining flower-promoting activity of *Arabidopsis* FLOWERING LOCUS T / W.W.H. Ho, D. Weigel // *The Plant Cell*. – 2014. – Т. 26. – № 2. – С. 552-564.
184. Homeologue-specific expression divergence in the recently formed tetraploid *Capsella bursa-pastoris* (Brassicaceae) / H.-R. Huang [et al.] // *New Phytologist*. – 2018. – Vol. 220. – № 2. – P. 624-635.
185. Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners / P.L. Chang [и др.] // *Genome Biology*. – 2010. – Т. 11. – № 12. – С. R125.
186. Honma T. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs / T. Honma, K. Goto // *Nature*. – 2001. – Т. 409. – № 6819. – С. 525-529.
187. Housekeeping genes as internal standards: use and limits / O. Thellin [и др.] // *Journal of Biotechnology*. – 1999. – Т. 75. – Housekeeping genes as internal standards. – № 2-3. – С. 291-295.
188. How many genes are there in plants (... and why are they there)? / L. Sterck [et al.] // *Current Opinion in Plant Biology*. – 2007. – Vol. 10. – № 2. – P. 199-203.
189. Huang D.W. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists / D.W. Huang, B.T. Sherman, R.A. Lempicki // *Nucleic Acids Research*. – 2009a. – Т. 37. – Bioinformatics enrichment tools. – № 1. – С. 1-13.
190. Huang D.W. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources / D.W. Huang, B.T. Sherman, R.A. Lempicki // *Nature Protocols*. – 2009b. – Т. 4. – № 1. – С. 44-57.
191. Huminiecki L. Congruence of tissue expression profiles from Gene Expression Atlas, SAGEmap and TissueInfo databases / L. Huminiecki, A.T. Lloyd, K.H. Wolfe // *BMC genomics*. – 2003. – Т. 4. – № 1. – С. 31.
192. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris* / G.M. Douglas [et al.] // *Proceedings of the National Academy of Sciences*. – 2015. – Vol. 112. – № 9. – P. 2806-2811.
193. Identification of three kinds of mutually related composite elements conferring S phase-specific transcriptional activation / K. Taoka [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 1999. – Т. 18. – № 6. – С. 611-623.
194. Integration of omic networks in a developmental atlas of maize / J.W. Walley [et al.] // *Science*. – 2016. – Vol. 353. – № 6301. – P. 814-818.
195. Integration of spatial and temporal information during floral induction in *Arabidopsis* / P.A. Wigge [и др.] // *Science (New York, N.Y.)*. – 2005. – Т. 309. – № 5737. – С. 1056-1059.
196. Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate / S.J. Liljegrén [и др.] // *The Plant Cell*. – 1999. – Т. 11. – № 6. – С. 1007-1018.
197. International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon* / International Brachypodium Initiative // *Nature*. – 2010. – Т. 463. – № 7282. – С. 763-768.
198. InterPro in 2019: improving coverage, classification and access to protein sequence annotations / A.L. Mitchell [et al.] // *Nucleic Acids Research*. – 2019. – Vol. 47. – InterPro in 2019. – № D1. – P. D351-D360.

199. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences / T. Slotte [и др.] // *American Journal of Botany*. – 2006. – Т. 93. – № 11. – С. 1714-1724.
200. Involvement of the sieve element cytoskeleton in electrical responses to cold shocks / J.B. Hafke [и др.] // *Plant Physiology*. – 2013. – Т. 162. – № 2. – С. 707-719.
201. Ito M. Conservation and diversification of three-repeat Myb transcription factors in plants / M. Ito // *Journal of Plant Research*. – 2005. – Т. 118. – № 1. – С. 61-69.
202. Jackson S. Genomic and expression plasticity of polyploidy / S. Jackson, Z.J. Chen // *Current Opinion in Plant Biology*. – 2010. – Vol. 13. – № 2. – P. 153-159.
203. Jacqumard A. Cell division and morphological changes in the shoot apex of *Arabidopsis thaliana* during floral transition / A. Jacqumard, I. Gadisseur, G. Bernier // *Annals of Botany*. – 2003. – Т. 91. – № 5. – С. 571-576.
204. Jacqumard A. DNA fiber replication during a morphogenetic switch in the shoot meristematic cells of a higher plant / A. Jacqumard, C. Houssa // *Experimental Cell Research*. – 1988. – Т. 179. – № 2. – С. 454-461.
205. Jasmonate regulates the inducer of *cbf* expression-C-repeat binding factor/DRE binding factor1 cascade and freezing tolerance in *Arabidopsis* / Y. Hu [и др.] // *The Plant Cell*. – 2013. – Т. 25. – № 8. – С. 2907-2924.
206. Jerzmanowski A. SWI/SNF chromatin remodeling and linker histones in plants / A. Jerzmanowski // *Biochimica Et Biophysica Acta*. – 2007. – Т. 1769. – № 5-6. – С. 330-345.
207. Kim J. CYTOKININ RESPONSE FACTORS Gating Environmental Signals and Hormones / J. Kim // *Trends in Plant Science*. – 2016. – Т. 21. – № 12. – С. 993-996.
208. Knight M.R. Low-temperature perception leading to gene expression and cold tolerance in higher plants / M.R. Knight, H. Knight // *The New Phytologist*. – 2012. – Т. 195. – № 4. – С. 737-751.
209. Koh J. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (Asteraceae) / J. Koh, P.S. Soltis, D.E. Soltis // *BMC genomics*. – 2010. – Т. 11. – С. 97.
210. Koornneef M. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana* / M. Koornneef, C.J. Hanhart, J.H. van der Veen // *Molecular & general genetics: MGG*. – 1991. – Т. 229. – № 1. – С. 57-66.
211. Kotera E. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts / E. Kotera, M. Tasaka, T. Shikanai // *Nature*. – 2005. – Т. 433. – № 7023. – С. 326-330.
212. Kozera B. Reference genes in real-time PCR / B. Kozera, M. Rapacz // *Journal of Applied Genetics*. – 2013. – Vol. 54. – № 4. – P. 391-406.
213. Langfelder P. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R / P. Langfelder, B. Zhang, S. Horvath // *Bioinformatics (Oxford, England)*. – 2008. – Т. 24. – Defining clusters from a hierarchical cluster tree. – № 5. – С. 719-720.
214. Lebrun M. Nucleotide sequence and characterization of a maize cytoplasmic ribosomal protein S11 cDNA / M. Lebrun, G. Freyssinet // *Plant Molecular Biology*. – 1991. – Т. 17. – № 2. – С. 265-268.
215. Levy M. A single recessive mutation in the proteolytic machinery of *Arabidopsis* chloroplasts impairs photoprotection and photosynthesis upon cold stress / M. Levy, A. Bachmair, Z. Adam // *Planta*. – 2004. – Т. 218. – № 3. – С. 396-405.
216. Li Z. ROPGEF1 and ROPGEF4 are functional regulators of ROP11 GTPase in ABA-mediated stomatal closure in *Arabidopsis* / Z. Li, D. Liu // *FEBS letters*. – 2012. – Т. 586. – № 9. – С. 1253-1258.

217. Lin C. Blue light receptors and signal transduction / C. Lin // *The Plant Cell*. – 2002. – T. 14 Suppl. – C. S207-225.
218. Lin W. -c. The Arabidopsis LATERAL ORGAN BOUNDARIES-Domain Gene ASYMMETRIC LEAVES2 Functions in the Repression of KNOX Gene Expression and in Adaxial-Abaxial Patterning / W. -c. Lin // *THE PLANT CELL ONLINE*. – 2003. – Vol. 15. – № 10. – P. 2241-2252.
219. Liu Y. MAPping Kinase Regulation of ICE1 in Freezing Tolerance / Y. Liu, J. Zhou // *Trends in Plant Science*. – 2018. – T. 23. – № 2. – C. 91-93.
220. Livak K.J. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method / K.J. Livak, T.D. Schmittgen // *Methods (San Diego, Calif.)*. – 2001. – T. 25. – № 4. – C. 402-408.
221. Love M.I. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 / M.I. Love, W. Huber, S. Anders // *Genome Biology*. – 2014. – Vol. 15. – № 12. – P. 550.
222. Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression / S.J. Gilmour [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 1998. – T. 16. – № 4. – C. 433-442.
223. Lyzenga W.J. Abiotic stress tolerance mediated by protein ubiquitination / W.J. Lyzenga, S.L. Stone // *Journal of Experimental Botany*. – 2012. – T. 63. – № 2. – C. 599-616.
224. Maize Gene Atlas Developed by RNA Sequencing and Comparative Evaluation of Transcriptomes Based on RNA Sequencing and Microarrays / R.S. Sekhon [et al.] // *PLoS ONE*. – 2013. – Vol. 8. – № 4. – P. e61005.
225. MAP Kinase Cascades Regulate the Cold Response by Modulating ICE1 Protein Stability / C. Zhao [и др.] // *Developmental Cell*. – 2017. – T. 43. – № 5. – C. 618-629.e5.
226. Mara C.D. The Arabidopsis floral homeotic proteins APETALA3 and PISTILLATA negatively regulate the BANQUO genes implicated in light signaling / C.D. Mara, T. Huang, V.F. Irish // *The Plant Cell*. – 2010. – T. 22. – № 3. – C. 690-702.
227. Matsuoka Y. Evolution of Polyploid Triticum Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in their Diversification / Y. Matsuoka // *Plant and Cell Physiology*. – 2011. – Vol. 52. – Evolution of Polyploid Triticum Wheats under Cultivation. – № 5. – P. 750-764.
228. Members of a novel class of Arabidopsis Rho guanine nucleotide exchange factors control Rho GTPase-dependent polar growth / Y. Gu [и др.] // *The Plant Cell*. – 2006. – T. 18. – № 2. – C. 366-381.
229. Menges M. Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity / M. Menges, J.A.H. Murray // *The Plant Journal: For Cell and Molecular Biology*. – 2002. – T. 30. – № 2. – C. 203-212.
230. Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A / K. Maruyama [и др.] // *Plant Physiology*. – 2009. – T. 150. – № 4. – C. 1972-1980.
231. Metabolite Profiling of adh1 Mutant Response to Cold Stress in Arabidopsis / Y. Song [и др.] // *Frontiers in Plant Science*. – 2016. – T. 7. – C. 2072.
232. Meyer E. Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research: Panicum hallii transcriptome and gene atlas / E. Meyer, T.L. Logan, T.E. Juenger // *The Plant Journal*. – 2012. – Vol. 70. – Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research. – № 5. – P. 879-890.

233. Michaels S.D. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering / S.D. Michaels, R.M. Amasino // *The Plant Cell*. – 1999. – Т. 11. – № 5. – С. 949-956.
234. Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology / F. Micheli // *Trends in Plant Science*. – 2001. – Т. 6. – Pectin methylesterases. – № 9. – С. 414-419.
235. Miller M.B. The cell cycle in vegetative and floral shoot meristems measured by a double labelling technique / M.B. Miller, R.F. Lyndon // *Planta*. – 1975. – Vol. 126. – № 1. – P. 37-43.
236. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data / A. Brazma [и др.] // *Nature Genetics*. – 2001. – Т. 29. – № 4. – С. 365-371.
237. Mittler R. ROS Are Good / R. Mittler // *Trends in Plant Science*. – 2017. – Т. 22. – № 1. – С. 11-19.
238. Mochida K. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags / K. Mochida, Y. Yamazaki, Y. Ogihara // *Molecular genetics and genomics: MGG*. – 2003. – Т. 270. – № 5. – С. 371-377.
239. Model selection reveals control of cold signalling by evening-phased components of the plant circadian clock / J. Keily [и др.] // *The Plant Journal*. – 2013. – Т. 76. – № 2. – С. 247-257.
240. Modulation of floral development by a gibberellin-regulated microRNA / P. Achard [et al.] // *Development*. – 2004. – Vol. 131. – № 14. – P. 3357-3365.
241. Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis / U. Hartmann [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2000. – Т. 21. – Molecular cloning of SVP. – № 4. – С. 351-360.
242. Molecular control of seasonal flowering in rice, arabidopsis and temperate cereals / R. Shrestha [и др.] // *Annals of Botany*. – 2014. – Т. 114. – № 7. – С. 1445-1458.
243. Montaigu A. de. Plant development goes like clockwork / A. de Montaigu, R. Tóth, G. Coupland // *Trends in genetics: TIG*. – 2010. – Т. 26. – № 7. – С. 296-306.
244. Morales M. Oxidative Stress: A Master Regulator of Plant Trade-Offs? / M. Morales, S. Munné-Bosch // *Trends in Plant Science*. – 2016. – Т. 21. – Oxidative Stress. – № 12. – С. 996-999.
245. Müllner D. fastcluster : Fast Hierarchical, Agglomerative Clustering Routines for R and Python / D. Müllner // *Journal of Statistical Software*. – 2013. – Vol. 53. – fastcluster. – № 9.
246. Mutational Evidence for the Critical Role of CBF Transcription Factors in Cold Acclimation in Arabidopsis / C. Zhao [и др.] // *Plant Physiology*. – 2016. – Т. 171. – № 4. – С. 2744-2759.
247. Mutti J.S. Evolution of Gene Expression Balance Among Homeologs of Natural Polyploids / J.S. Mutti, R.K. Bhullar, K.S. Gill // *G3: Genes|Genomes|Genetics*. – 2017. – Vol. 7. – № 4. – P. 1225-1237.
248. Myosin XI-K Is required for rapid trafficking of Golgi stacks, peroxisomes, and mitochondria in leaf cells of *Nicotiana benthamiana* / D. Avisar [и др.] // *Plant Physiology*. – 2008. – Т. 146. – № 3. – С. 1098-1108.
249. Nibau C. RAC/ROP GTPases: «hubs» for signal integration and diversification in plants / C. Nibau, H. Wu, A.Y. Cheung // *Trends in Plant Science*. – 2006. – Т. 11. – RAC/ROP GTPases. – № 6. – С. 309-315.
250. Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat / M. Pumphrey [и др.] // *Genetics*. – 2009. – Т. 181. – № 3. – С. 1147-1157.
251. Nougarede A. Location of cell cycle changes in relation to morphological changes in the shoot apex of *Silene coeli-rosa* immediately before sepal initiation / A. Nougarede, D. Francis, P. Rondet // *Protoplasma*. – 1991. – Vol. 165. – № 1-3. – P. 1-10.

252. Novel *as1* and *as2* defects in leaf adaxial-abaxial polarity reveal the requirement for *ASYMMETRIC LEAVES1* and *2* and *ERECTA* functions in specifying leaf adaxial identity / L. Xu [и др.] // *Development* (Cambridge, England). – 2003. – Т. 130. – № 17. – С. 4097-4107.
253. Ohnishi S. Low temperature stress at different flower developmental stages affects pollen development, pollination, and pod set in soybean / S. Ohnishi, T. Miyoshi, S. Shirai // *Environmental and Experimental Botany*. – 2010. – Vol. 69. – № 1. – P. 56-62.
254. Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock / S.L. Harmer [и др.] // *Science* (New York, N.Y.). – 2000. – Т. 290. – № 5499. – С. 2110-2113.
255. Organ-specific expression of different histone H3 and H4 gene subfamilies in developing and adult maize / N. Chaubet [и др.] // *Plant Molecular Biology*. – 1991. – Т. 17. – № 4. – С. 935-940.
256. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids / M.R. Woodhouse [et al.] // *Proceedings of the National Academy of Sciences*. – 2014. – Vol. 111. – № 14. – P. 5283-5288.
257. *OST1* Kinase Modulates Freezing Tolerance by Enhancing *ICE1* Stability in *Arabidopsis* / Y. Ding [et al.] // *Developmental Cell*. – 2015. – Vol. 32. – № 3. – P. 278-289.
258. Otto S.P. Polyploid incidence and evolution / S.P. Otto, J. Whitton // *Annual Review of Genetics*. – 2000. – Т. 34. – С. 401-437.
259. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae) / D. Kryvokhyzha [et al.] // *PLOS Genetics*. – 2019. – Vol. 15. – № 2. – P. e1007949.
260. Paterson A.H. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics / A.H. Paterson, J.E. Bowers, B.A. Chapman // *Proceedings of the National Academy of Sciences*. – 2004. – Vol. 101. – № 26. – P. 9903-9908.
261. Perception of UV-B by the *Arabidopsis* *UVR8* protein / L. Rizzini [и др.] // *Science* (New York, N.Y.). – 2011. – Т. 332. – № 6025. – С. 103-106.
262. Photoexcited *CRY2* interacts with *CIB1* to regulate transcription and floral initiation in *Arabidopsis* / H. Liu [и др.] // *Science* (New York, N.Y.). – 2008b. – Т. 322. – № 5907. – С. 1535-1539.
263. Photoperiodic and thermosensory pathways interact through *CONSTANS* to promote flowering at high temperature under short days / V. Fernández [et al.] // *The Plant Journal*. – 2016. – Vol. 86. – № 5. – P. 426-440.
264. Phylogenetic comparison of F-Box (FBX) gene superfamily within the plant kingdom reveals divergent evolutionary histories indicative of genomic drift / Z. Hua [и др.] // *PloS One*. – 2011. – Т. 6. – № 1. – С. e16219.
265. Phytochrome B integrates light and temperature signals in *Arabidopsis* / M. Legris [и др.] // *Science* (New York, N.Y.). – 2016. – Т. 354. – № 6314. – С. 897-900.
266. *PHYTOCHROME-DEPENDENT LATE-FLOWERING* accelerates flowering through physical interactions with phytochrome B and *CONSTANS* / M. Endo [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2013. – Т. 110. – № 44. – С. 18017-18022.
267. Phytochromes function as thermosensors in *Arabidopsis* / J.-H. Jung [et al.] // *Science*. – 2016. – Vol. 354. – № 6314. – P. 886-889.
268. *PIF3* is a negative regulator of the *CBF* pathway and freezing tolerance in *Arabidopsis* / B. Jiang [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2017. – Т. 114. – № 32. – С. E6695-E6702.

269. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors / J. Jin [и др.] // *Nucleic Acids Research*. – 2014. – Т. 42. – PlantTFDB 3.0. – № Database issue. – С. D1182-1187.
270. Plasma Membrane CRPK1-Mediated Phosphorylation of 14-3-3 Proteins Induces Their Nuclear Import to Fine-Tune CBF Signaling during Cold Response / Z. Liu [et al.] // *Molecular Cell*. – 2017. – Vol. 66. – № 1. – P. 117-128.e5.
271. Ponger L. Determinants of CpG islands: expression in early embryo and isochore structure / L. Ponger, L. Duret, D. Mouchiroud // *Genome Research*. – 2001. – Т. 11. – № 11. – С. 1854-1860.
272. Pont C. Wheat paleohistory created asymmetrical genomic evolution / C. Pont, J. Salse // *Current Opinion in Plant Biology*. – 2017. – Т. 36. – С. 29-37.
273. Promoter features related to tissue specificity as measured by Shannon entropy / J. Schug [и др.] // *Genome Biology*. – 2005. – Т. 6. – № 4. – С. R33.
274. Quail P.H. Photosensory perception and signalling in plant cells: new paradigms? / P.H. Quail // *Current Opinion in Cell Biology*. – 2002. – Т. 14. – Photosensory perception and signalling in plant cells. – № 2. – С. 180-188.
275. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray / M. Schena [et al.] // *Science*. – 1995. – Vol. 270. – № 5235. – P. 467-470.
276. Quesada V. Regulated RNA processing in the control of Arabidopsis flowering / V. Quesada, C. Dean, G.G. Simpson // *The International Journal of Developmental Biology*. – 2005. – Т. 49. – № 5-6. – С. 773-780.
277. Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome / C.P.G. Calixto [et al.] // *The Plant Cell*. – 2018. – P. tpc.00177.2018.
278. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes / M. Feldman [и др.] // *Genetics*. – 1997. – Т. 147. – № 3. – С. 1381-1387.
279. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution / K. Song [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 1995. – Т. 92. – № 17. – С. 7719-7723.
280. Rataj K. Message ends: RNA 3' processing and flowering time control / K. Rataj, G.G. Simpson // *Journal of Experimental Botany*. – 2014. – Т. 65. – Message ends. – № 2. – С. 353-363.
281. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*) / B. Chaudhary [и др.] // *Genetics*. – 2009. – Т. 182. – № 2. – С. 503-517.
282. Redox regulation of plant stem cell fate / J. Zeng [и др.] // *The EMBO journal*. – 2017. – Т. 36. – № 19. – С. 2844-2855.
283. Re-Evaluation of Florigen Transport Kinetics with Separation of Functions by Mutations That Uncouple Flowering Initiation and Long-Distance Transport / M. Endo [et al.] // *Plant and Cell Physiology*. – 2018. – Vol. 59. – № 8. – P. 1621-1629.
284. Regulation of the Arabidopsis CBF regulon by a complex low-temperature regulatory network / S. Park [et al.] // *The Plant Journal*. – 2015. – Vol. 82. – № 2. – P. 193-207.
285. Removal of redundant contigs from de novo RNA-Seq assemblies via homology search improves accurate detection of differentially expressed genes / H. Ono [et al.] // *BMC Genomics*. – 2015. – Vol. 16. – № 1.
286. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres / A.H. Paterson [и др.] // *Nature*. – 2012. – Т. 492. – № 7429. – С. 423-427.

287. Repression of flowering by the miR172 target SMZ / J. Mathieu [и др.] // *PLoS biology*. – 2009. – Т. 7. – № 7. – С. e1000148.
288. Research progress on the autonomous flowering time pathway in Arabidopsis / J.-Z. Cheng [и др.] // *Physiology and Molecular Biology of Plants: An International Journal of Functional Plant Biology*. – 2017. – Т. 23. – № 3. – С. 477-485.
289. Reserve mobilization in the Arabidopsis endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1 / S. Penfield [и др.] // *The Plant Cell*. – 2004. – Т. 16. – № 10. – С. 2705-2718.
290. Rho-GTPase-dependent filamentous actin dynamics coordinate vesicle targeting and exocytosis during tip growth / Y.J. Lee [и др.] // *The Journal of Cell Biology*. – 2008a. – Т. 181. – № 7. – С. 1155-1168.
291. Richter R. Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis / R. Richter, E. Bastakis, C. Schwechheimer // *Plant Physiology*. – 2013. – Т. 162. – № 4. – С. 1992-2004.
292. RNA-seq analysis of an apical meristem time series reveals a critical point in Arabidopsis thaliana flower initiation / A.V. Klepikova [et al.] // *BMC Genomics*. – 2015. – Vol. 16. – № 1.
293. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome / A.J. Severin [et al.] // *BMC Plant Biology*. – 2010. – Vol. 10. – RNA-Seq Atlas of Glycine max. – № 1. – P. 160.
294. Robinson S.J. Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature / S.J. Robinson, I.A.P. Parkin // *BMC genomics*. – 2008. – Т. 9. – С. 434.
295. Role of SVP in the control of flowering time by ambient temperature in Arabidopsis / J.H. Lee [и др.] // *Genes & Development*. – 2007. – Т. 21. – № 4. – С. 397-402.
296. Saha P. Assessing Reference Genes for Accurate Transcript Normalization Using Quantitative Real-Time PCR in Pearl Millet [*Pennisetum glaucum* (L.) R. Br.] / P. Saha, E. Blumwald // *PLoS ONE*. – 2014. – Vol. 9. – № 8. – P. e106308.
297. Sawa M. GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana / M. Sawa, S.A. Kay // *Proceedings of the National Academy of Sciences*. – 2011. – Vol. 108. – № 28. – P. 11698-11703.
298. Schnable J.C. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss / J.C. Schnable, N.M. Springer, M. Freeling // *Proceedings of the National Academy of Sciences of the United States of America*. – 2011. – Т. 108. – № 10. – С. 4069-4074.
299. Selection of Reference Genes for Gene Expression Studies in Siberian Apricot (*Prunus sibirica* L.) Germplasm Using Quantitative Real-Time PCR / J. Niu [et al.] // *PLoS ONE*. – 2014. – Vol. 9. – № 8. – P. e103900.
300. Selection of Reference Genes for Quantitative Real-Time PCR Normalization in *Panax ginseng* at Different Stages of Growth and in Different Organs / J. Liu [et al.] // *PLoS ONE*. – 2014. – Vol. 9. – № 11. – P. e112177.
301. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat / H. Shaked [и др.] // *The Plant Cell*. – 2001. – Т. 13. – № 8. – С. 1749-1759.
302. Shen P.-C. Microarray meta-analysis to explore abiotic stress-specific gene expression patterns in Arabidopsis / P.-C. Shen, A.-L. Hour, L. Liu // *Botanical Studies*. – 2017. – Т. 58. – № 1. – С. 22.

303. Shifting the limits in wheat research and breeding using a fully annotated reference genome / The International Wheat Genome Sequencing Consortium (IWGSC) [et al.] // *Science*. – 2018. – Vol. 361. – № 6403. – P. eaar7191.
304. Single-Cell Genomic Analysis in Plants / Y. Yuan [и др.] // *Genes*. – 2018. – Т. 9. – № 1.
305. Sites of actin filament initiation and reorganization in cold-treated tobacco cells / J. Pokorna [et al.] // *Plant, Cell and Environment*. – 2004. – Vol. 27. – № 5. – P. 641-653.
306. SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in *Arabidopsis* / K. Miura [и др.] // *The Plant Cell*. – 2007. – Т. 19. – № 4. – C. 1403-1414.
307. Smyth D.R. Early flower development in *Arabidopsis* / D.R. Smyth, J.L. Bowman, E.M. Meyerowitz // *The Plant Cell*. – 1990. – Т. 2. – № 8. – C. 755-767.
308. SOC1 translocated to the nucleus by interaction with AGL24 directly regulates leafy / J. Lee [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2008b. – Т. 55. – № 5. – C. 832-843.
309. Solecka D. Are Pectins Involved in Cold Acclimation and De-acclimation of Winter Oil-seed Rape Plants? / D. Solecka, J. Żebrowski, A. Kacperska // *Annals of Botany*. – 2008. – Vol. 101. – № 4. – P. 521-530.
310. Soneson C. A comparison of methods for differential expression analysis of RNA-seq data / C. Soneson, M. Delorenzi // *BMC Bioinformatics*. – 2013. – Vol. 14. – № 1. – P. 91.
311. Song J. Remembering the prolonged cold of winter / J. Song, J. Irwin, C. Dean // *Current biology: CB*. – 2013. – Т. 23. – № 17. – C. R807-811.
312. Song Y.H. Flowering time regulation: photoperiod- and temperature-sensing in leaves / Y.H. Song, S. Ito, T. Imaizumi // *Trends in Plant Science*. – 2013. – Т. 18. – Flowering time regulation. – № 10. – C. 575-583.
313. Spatiotemporal secretion of PEROXIDASE36 is required for seed coat mucilage extrusion in *Arabidopsis* / T. Kunieda [и др.] // *The Plant Cell*. – 2013. – Т. 25. – № 4. – C. 1355-1367.
314. SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for *Arabidopsis* gynoecium and fruit development / M. Groszmann [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2011. – Т. 68. – № 5. – C. 816-829.
315. Stefanović S. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? / S. Stefanović, D.W. Rice, J.D. Palmer // *BMC evolutionary biology*. – 2004. – Т. 4. – Long branch attraction, taxon sampling, and the earliest angiosperms. – C. 35.
316. Stitt M. A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in *Arabidopsis* / M. Stitt, V. Hurry // *Current Opinion in Plant Biology*. – 2002. – Т. 5. – A plant for all seasons. – № 3. – C. 199-206.
317. Stochastic and Epigenetic Changes of Gene Expression in *Arabidopsis* Polyploids / J. Wang [et al.] // *Genetics*. – 2004. – Vol. 167. – № 4. – P. 1961-1973.
318. Stockinger E.J. *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit / E.J. Stockinger, S.J. Gilmour, M.F. Thomashow // *Proceedings of the National Academy of Sciences of the United States of America*. – 1997. – Т. 94. – № 3. – C. 1035-1040.
319. Structural alteration of cell wall pectins accompanies pea development in response to cold / L. Baldwin [и др.] // *Phytochemistry*. – 2014. – Т. 104. – C. 37-47.
320. Talbert P.B. Environmental responses mediated by histone variants / P.B. Talbert, S. Henikoff // *Trends in Cell Biology*. – 2014. – Vol. 24. – № 11. – P. 642-650.

321. Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana* / J. Van Leene [и др.] // *Molecular Systems Biology*. – 2010. – Т. 6. – С. 397.
322. Temperature Sensing Is Distributed throughout the Regulatory Network that Controls FLC Epigenetic Silencing in Vernalization / R.L. Antoniou-Kourounioli [et al.] // *Cell Systems*. – 2018. – Vol. 7. – № 6. – P. 643-655.e9.
323. The ancient subclasses of *Arabidopsis* Actin Depolymerizing Factor genes exhibit novel and differential expression / D.R. Ruzicka [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2007. – Т. 52. – № 3. – С. 460-472.
324. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV / The Angiosperm Phylogeny Group // *Botanical Journal of the Linnean Society*. – 2016. – Vol. 181. – № 1. – P. 1-20.
325. The *Arabidopsis* CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration / J. Medina [и др.] // *Plant Physiology*. – 1999. – Т. 119. – № 2. – С. 463-470.
326. The *Arabidopsis* E3 ubiquitin ligase HOS1 negatively regulates CONSTANS abundance in the photoperiodic control of flowering / A. Lazaro [и др.] // *The Plant Cell*. – 2012. – Т. 24. – № 3. – С. 982-999.
327. The *Arabidopsis* FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex / C.A. Helliwell [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2006. – Т. 46. – № 2. – С. 183-192.
328. The *Arabidopsis* H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor / S. Zheng [et al.] // *Nature Communications*. – 2019. – Vol. 10. – № 1.
329. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools / P. Lamesch [et al.] // *Nucleic Acids Research*. – 2012. – Vol. 40. – The *Arabidopsis* Information Resource (TAIR). – № D1. – P. D1202-D1210.
330. The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC / F. Liu [и др.] // *Molecular Cell*. – 2007. – Т. 28. – № 3. – С. 398-407.
331. The ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves / E. Semiarti [и др.] // *Development (Cambridge, England)*. – 2001. – Т. 128. – № 10. – С. 1771-1783.
332. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses / J. Kilian [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2007. – Т. 50. – The AtGenExpress global stress expression data set. – № 2. – С. 347-363.
333. The B73 maize genome: complexity, diversity, and dynamics / P.S. Schnable [и др.] // *Science (New York, N.Y.)*. – 2009. – Т. 326. – The B73 maize genome. – № 5956. – С. 1112-1115.
334. The cbfs triple mutants reveal the essential functions of CBFs in cold acclimation and allow the definition of CBF regulons in *Arabidopsis* / Y. Jia [и др.] // *The New Phytologist*. – 2016. – Т. 212. – № 2. – С. 345-353.
335. The chromatin-remodelling factor PICKLE interacts with CONSTANS to promote flowering in *Arabidopsis* / Y. Jing [et al.] // *Plant, Cell & Environment*. – 2019.
336. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors / J. Putterill [и др.] // *Cell*. – 1995. – Т. 80. – № 6. – С. 847-857.

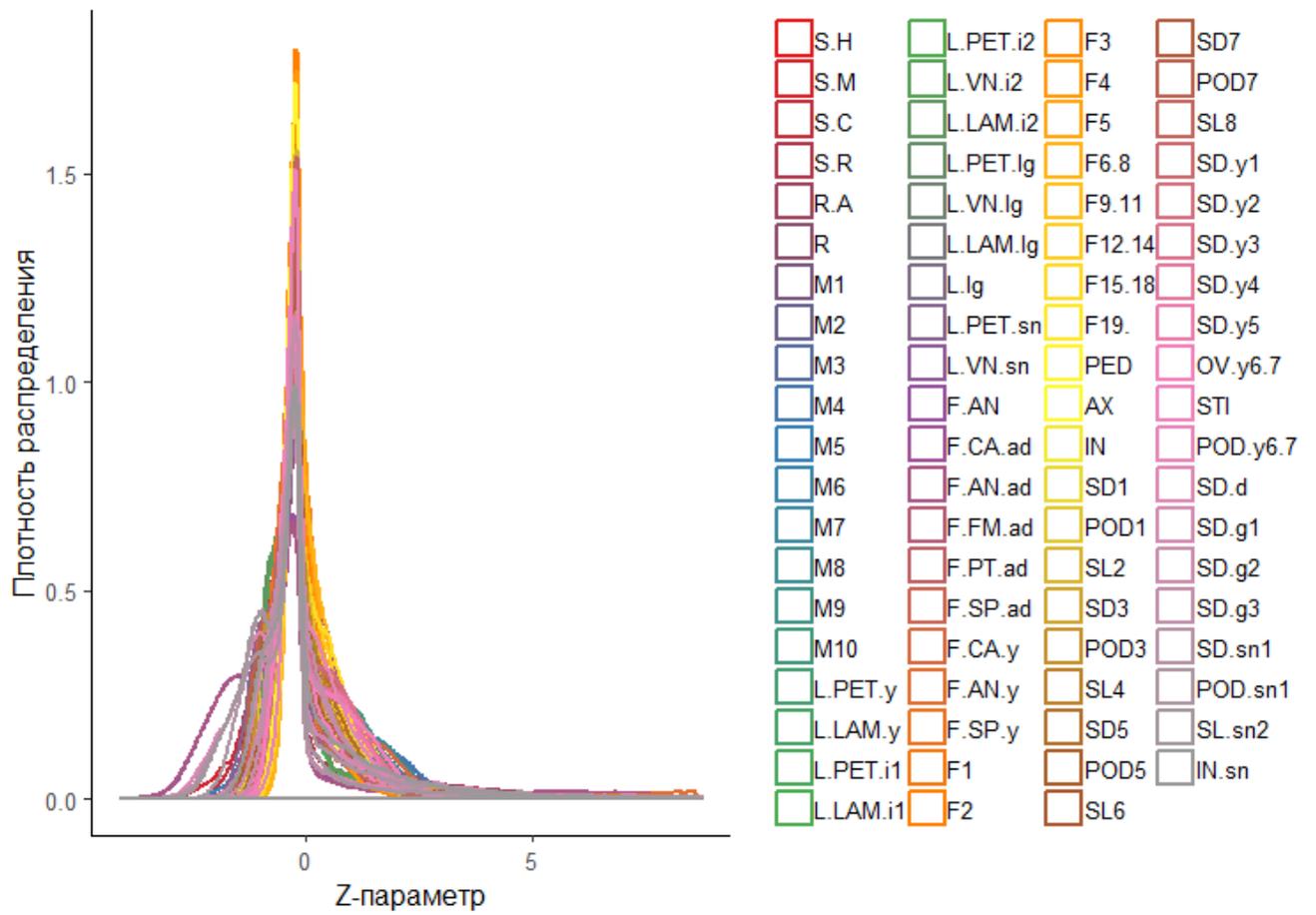
337. The development of an Arabidopsis model system for genome-wide analysis of polyploidy effects / Z. Jeffrey Chen [и др.] // *Biological Journal of the Linnean Society*. Linnean Society of London. – 2004. – Т. 82. – № 4. – С. 689-700.
338. The developmental transcriptome atlas of the biofuel crop *Camelina sativa* / S. Kagale [et al.] // *The Plant Journal*. – 2016. – Vol. 88. – № 5. – P. 879-894.
339. The fate of duplicated genes in a polyploid plant genome / A. Roulin [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2013. – Т. 73. – № 1. – С. 143-153.
340. The F-box protein FKF1 inhibits dimerization of COP1 in the control of photoperiodic flowering / B.-D. Lee [et al.] // *Nature Communications*. – 2017. – Vol. 8. – № 1.
341. The fission yeast mitotic activator *cdc25* and sucrose induce early flowering synergistically in the day-neutral *Nicotiana tabacum* cv. Samsun / M. Teichmanová [и др.] // *The New Phytologist*. – 2007. – Т. 176. – № 4. – С. 804-812.
342. The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation / C.C. Sheldon [и др.] // *The Plant Cell*. – 1999. – Т. 11. – The FLF MADS box gene. – № 3. – С. 445-458.
343. The flowering time regulator *CONSTANS* is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element / S.B. Tiwari [и др.] // *The New Phytologist*. – 2010. – Т. 187. – № 1. – С. 57-66.
344. The French–Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla / The French–Italian Public Consortium for Grapevine Genome Characterization // *Nature*. – 2007. – Vol. 449. – № 7161. – P. 463-467.
345. The gene expression landscape of pine seedling tissues / R.A. Cañas [et al.] // *The Plant Journal*. – 2017. – Vol. 91. – № 6. – P. 1064-1087.
346. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection / J. Yang [et al.] // *Nature Genetics*. – 2016. – Vol. 48. – № 10. – P. 1225-1232.
347. The influence of population structure on gene expression and flowering time variation in the ubiquitous weed *Capsella bursa-pastoris* (Brassicaceae) / D. Kryvokhyzha [et al.] // *Molecular Ecology*. – 2016. – Vol. 25. – № 5. – P. 1106-1121.
348. The legacy of diploid progenitors in allopolyploid gene expression patterns / R.J.A. Buggs [и др.] // *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. – 2014. – Т. 369. – № 1648.
349. The *Medicago* genome provides insight into the evolution of rhizobial symbioses / N.D. Young [и др.] // *Nature*. – 2011. – Т. 480. – № 7378. – С. 520-524.
350. The *Medicago sativa* gene index 1.2: a web-accessible gene expression atlas for investigating expression differences between *Medicago sativa* subspecies / J.A. O'Rourke [et al.] // *BMC Genomics*. – 2015. – Vol. 16. – The *Medicago sativa* gene index 1.2. – № 1.
351. The *Medicago truncatula* gene expression atlas web server / J. He [et al.] // *BMC Bioinformatics*. – 2009. – Vol. 10. – № 1. – P. 441.
352. The microRNA-regulated SBP-Box transcription factor *SPL3* is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1* / A. Yamaguchi [и др.] // *Developmental Cell*. – 2009. – Т. 17. – № 2. – С. 268-278.

353. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments / S.A. Bustin [et al.] // *Clinical Chemistry*. – 2009. – Vol. 55. – The MIQE Guidelines. – № 4. – P. 611-622.
354. The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis* / M. Teige [и др.] // *Molecular Cell*. – 2004. – Т. 15. – № 1. – С. 141-152.
355. The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1 / C.-H. Dong [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2006. – Т. 103. – № 21. – С. 8281-8286.
356. The organization and expression of a maize ribosomal protein gene family / J.C. Larkin [и др.] // *Genes & Development*. – 1989. – Т. 3. – № 4. – С. 500-509.
357. The *Physcomitrella patens* gene atlas project: large-scale RNA-seq based expression data / P.-F. Perroud [et al.] // *The Plant Journal*. – 2018. – Vol. 95. – The *Physcomitrella patens* gene atlas project. – № 1. – P. 168-182.
358. The Plant List. The Plant List Version 1.1. Адрес сайта: <http://www.theplantlist.org/> (дата обращения 1 августа 2019 г.).
359. The precise regulation of different COR genes by individual CBF transcription factors in *Arabidopsis thaliana* / Y. Shi [и др.] // *Journal of Integrative Plant Biology*. – 2017. – Т. 59. – № 2. – С. 118-133.
360. The Proteasome Stress Regulon Is Controlled by a Pair of NAC Transcription Factors in *Arabidopsis* / N.P. Gladman [и др.] // *The Plant Cell*. – 2016. – Т. 28. – № 6. – С. 1279-1296.
361. The rice OsV4 encoding a novel pentatricopeptide repeat protein is required for chloroplast development during the early leaf stage under cold stress / X. Gong [и др.] // *Journal of Integrative Plant Biology*. – 2014. – Т. 56. – № 4. – С. 400-410.
362. The Rice Pentatricopeptide Repeat Gene TCD10 is Needed for Chloroplast Development under Cold Stress / L. Wu [и др.] // *Rice (New York, N.Y.)*. – 2016. – Т. 9. – № 1. – С. 67.
363. The rice TCD11 encoding plastid ribosomal protein S6 is essential for chloroplast development at low temperature / W.-J. Wang [и др.] // *Plant Science*. – 2017. – Т. 259. – С. 1-11.
364. The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in *Arabidopsis* / J.-H. Jung [и др.] // *The Plant Journal: For Cell and Molecular Biology*. – 2012. – Т. 69. – № 4. – С. 577-588.
365. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization / R.F. McCormick [et al.] // *The Plant Journal*. – 2018. – Vol. 93. – The *Sorghum bicolor* reference genome. – № 2. – P. 338-354.
366. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis* / I. Searle [и др.] // *Genes & Development*. – 2006. – Т. 20. – № 7. – С. 898-912.
367. The Transcriptome of the Reference Potato Genome *Solanum tuberosum* Group Phureja Clone DM1-3 516R44 / A.N. Massa [et al.] // *PLoS ONE*. – 2011. – Vol. 6. – № 10. – P. e26801.
368. The unified ICE-CBF pathway provides a transcriptional feedback control of freezing tolerance during cold acclimation in *Arabidopsis* / Y.S. Kim [и др.] // *Plant Molecular Biology*. – 2015. – Т. 89. – № 1-2. – С. 187-201.
369. The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from de novo assembly and quantification of RNA-seq data provides insights into seed maturation mechanisms / S. Yao [et al.] // *The Plant Journal*. – 2016. – Vol. 88. – № 2. – P. 318-327.

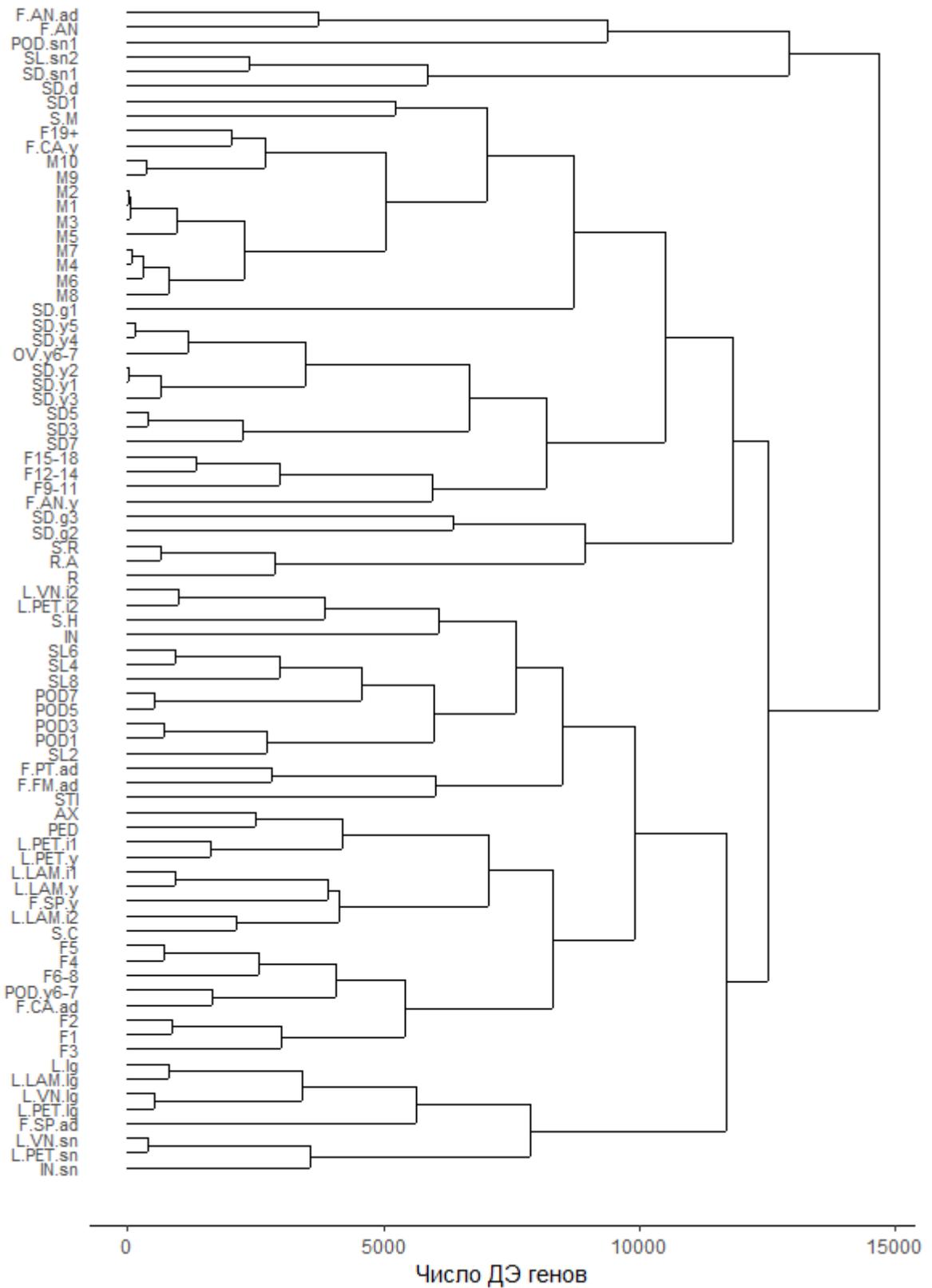
370. The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis* / T. Laux [и др.] // *Development* (Cambridge, England). – 1996. – Т. 122. – № 1. – С. 87-96.
371. Theocharis A. Physiological and molecular changes in plants grown at low temperatures / A. Theocharis, C. Clément, E.A. Barka // *Planta*. – 2012. – Т. 235. – № 6. – С. 1091-1105.
372. Thomas B.C. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes / B.C. Thomas, B. Pedersen, M. Freeling // *Genome Research*. – 2006. – Т. 16. – № 7. – С. 934-946.
373. Thomashow M.F. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms / M.F. Thomashow // *Annual Review of Plant Physiology and Plant Molecular Biology*. – 1999. – Т. 50. – PLANT COLD ACCLIMATION. – С. 571-599.
374. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat / T. Nomura [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2005. – Т. 102. – № 45. – С. 16490-16495.
375. Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment / J.K. Kim [et al.] // *Journal of Experimental Botany*. – 2006. – Vol. 58. – № 3. – P. 415-424.
376. TiSGeD: a database for tissue-specific genes / S.-J. Xiao [и др.] // *Bioinformatics* (Oxford, England). – 2010. – Т. 26. – TiSGeD. – № 9. – С. 1273-1275.
377. Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus* / R. Buggs [и др.] // *The New Phytologist*. – 2010. – Т. 186. – № 1. – С. 175-183.
378. TobEA: an atlas of tobacco gene expression from seed to senescence / K.D. Edwards [et al.] // *BMC Genomics*. – 2010. – Vol. 11. – TobEA. – № 1. – P. 142.
379. TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks / M. Zouine [et al.] // *The Plant Journal*. – 2017. – Vol. 92. – № 4. – P. 727-735.
380. Transcription factor PIF4 controls the thermosensory activation of flowering / S.V. Kumar [и др.] // *Nature*. – 2012. – Т. 484. – № 7393. – С. 242-245.
381. Transcriptional regulation and function during the human cell cycle / R.J. Cho [и др.] // *Nature Genetics*. – 2001. – Т. 27. – № 1. – С. 48-54.
382. Transcriptome analysis during cell division in plants / P. Breyne [и др.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2002. – Т. 99. – № 23. – С. 14825-14830.
383. Transcriptome Analysis of Leaves, Flowers and Fruits Perisperm of *Coffea arabica* L. Reveals the Differential Expression of Genes Involved in Raffinose Biosynthesis / S.T. Ivamoto [et al.] // *PLOS ONE*. – 2017. – Vol. 12. – № 1. – P. e0169595.
384. Transcriptome database resource and gene expression atlas for the rose / A. Dubois [et al.] // *BMC Genomics*. – 2012. – Vol. 13. – № 1. – P. 638.
385. Transcriptome responses to combinations of stresses in *Arabidopsis* / S. Rasmussen [и др.] // *Plant Physiology*. – 2013. – Т. 161. – № 4. – С. 1783-1794.
386. Transcriptome Shock after Interspecific Hybridization in *Senecio* Is Ameliorated by Genome Duplication / M.J. Hegarty [et al.] // *Current Biology*. – 2006. – Vol. 16. – № 16. – P. 1652-1659.
387. Transposase-derived transcription factors regulate light signaling in *Arabidopsis* / R. Lin [и др.] // *Science* (New York, N.Y.). – 2007. – Т. 318. – № 5854. – С. 1302-1305.
388. TSV, a putative plastidic oxidoreductase, protects rice chloroplasts from cold stress during development by interacting with plastidic thioredoxin Z / J. Sun [и др.] // *The New Phytologist*. – 2017. – Т. 215. – № 1. – С. 240-255.

389. Turck F. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage / F. Turck, F. Fornara, G. Coupland // *Annual Review of Plant Biology*. – 2008. – Т. 59. – Regulation and identity of florigen. – С. 573-594.
390. Two Arabidopsis cyclin A3s possess G1 cyclin-like features / I. Takahashi [и др.] // *Plant Cell Reports*. – 2010. – Т. 29. – № 4. – С. 307-315.
391. Two class XI myosins function in organelle trafficking and root hair development in Arabidopsis / V.V. Peremyslov [и др.] // *Plant Physiology*. – 2008. – Т. 146. – № 3. – С. 1109-1116.
392. Two evolutionarily distinct classes of paleopolyploidy / O. Garsmeur [и др.] // *Molecular Biology and Evolution*. – 2014. – Т. 31. – № 2. – С. 448-454.
393. Ubiquitin and Ubiquitin-like Modifiers in Plants / H.J. Park [et al.] // *Journal of Plant Biology*. – 2011. – Vol. 54. – № 5. – P. 275-285.
394. Udall J.A. Polyploidy and Crop Improvement / J.A. Udall, J.F. Wendel // *Crop Science*. – 2006. – Vol. 46. – № Supplement\_1. – P. S-3.
395. Understanding Brassicaceae evolution through ancestral genome reconstruction / F. Murat [et al.] // *Genome Biology*. – 2015. – Vol. 16. – № 1.
396. Understanding the biochemical basis of temperature-induced lipid pathway adjustments in plants / Q. Li [и др.] // *The Plant Cell*. – 2015. – Т. 27. – № 1. – С. 86-103.
397. Utility of RNA Sequencing for Analysis of Maize Reproductive Transcriptomes / R.M. Davidson [et al.] // *The Plant Genome Journal*. – 2011. – Vol. 4. – № 3. – P. 191.
398. Validation of reference genes for quantitative real-time PCR during latex regeneration in rubber tree / X. Long [et al.] // *Gene*. – 2015. – Vol. 563. – № 2. – P. 190-195.
399. VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract / L. Jiang [и др.] // *The Plant Cell*. – 2005. – Т. 17. – № 2. – С. 584-596.
400. Veitia R.A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects / R.A. Veitia, S. Bottani, J.A. Birchler // *Trends in genetics: TIG*. – 2008. – Т. 24. – Cellular reactions to gene dosage imbalance. – № 8. – С. 390-397.
401. Vinogradov A.E. Isochores and tissue-specificity / A.E. Vinogradov // *Nucleic Acids Research*. – 2003. – Т. 31. – № 17. – С. 5212-5220.
402. Wang J.-W. miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana / J.-W. Wang, B. Czech, D. Weigel // *Cell*. – 2009. – Т. 138. – № 4. – С. 738-749.
403. Willyard C. New human gene tally reignites debate / C. Willyard // *Nature*. – 2018. – Vol. 558. – № 7710. – P. 354-355.
404. Wollenberg A.C. Natural variation in the temperature range permissive for vernalization in accessions of Arabidopsis thaliana / A.C. Wollenberg, R.M. Amasino // *Plant, Cell & Environment*. – 2012. – Т. 35. – № 12. – С. 2181-2191.
405. Yano R. Melonnet-DB, a Grand RNA-Seq Gene Expression Atlas in Melon (Cucumis melo L.) / R. Yano, S. Nonaka, H. Ezura // *Plant and Cell Physiology*. – 2018. – Vol. 59. – № 1. – P. e4-e4.
406. Yoo M.-J. Homoeolog expression bias and expression level dominance in allopolyploid cotton / M.-J. Yoo, E. Szadkowski, J.F. Wendel // *Heredity*. – 2013. – Т. 110. – № 2. – С. 171-180.
407. Zhang Y. A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in Arabidopsis thaliana / Y. Zhang, S. McCormick // *Proceedings of the National Academy of Sciences of the United States of America*. – 2007. – Т. 104. – № 47. – С. 18830-18835.

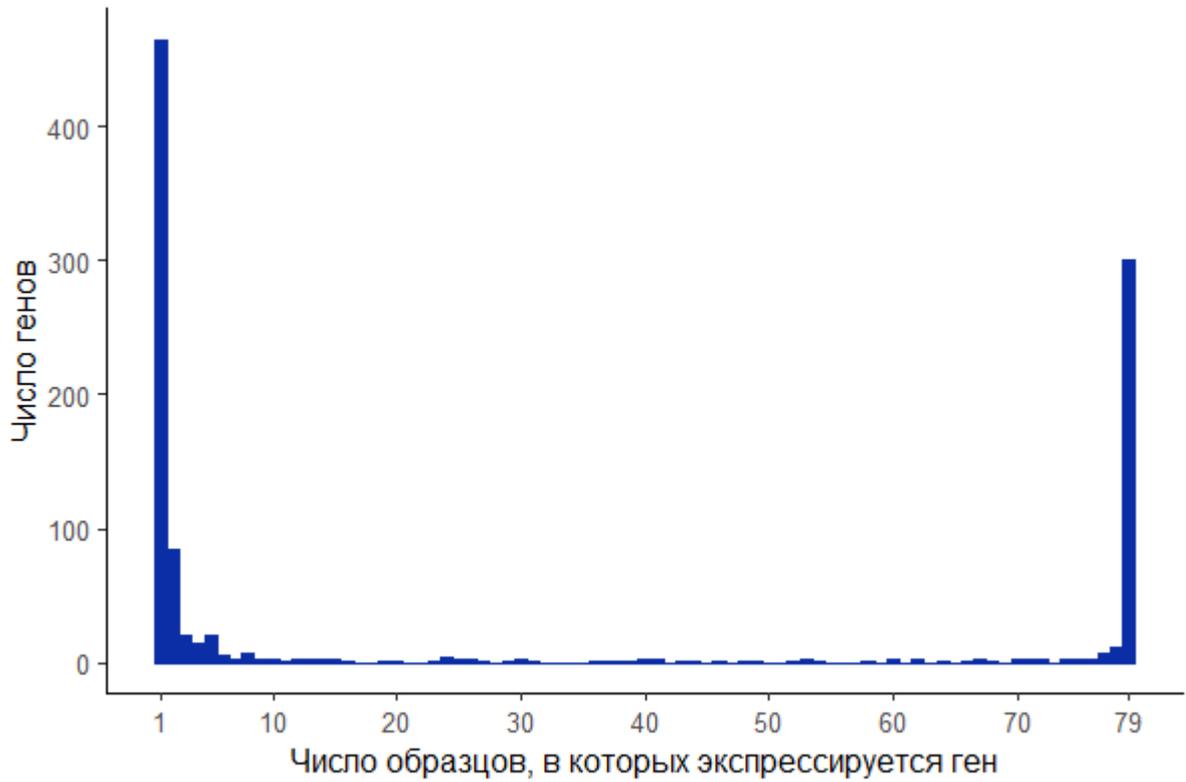




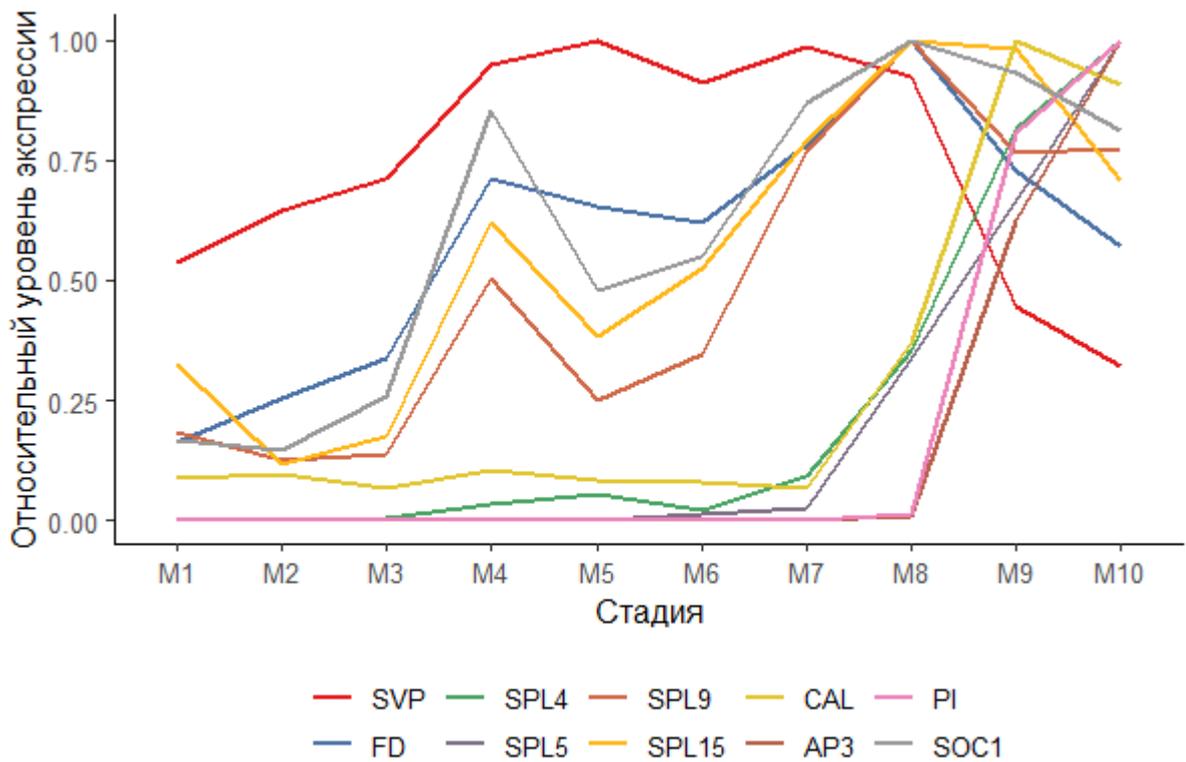
**Рисунок 1.3.** Распределение параметра  $Z$  для каждого из образцов транскриптомной карты.



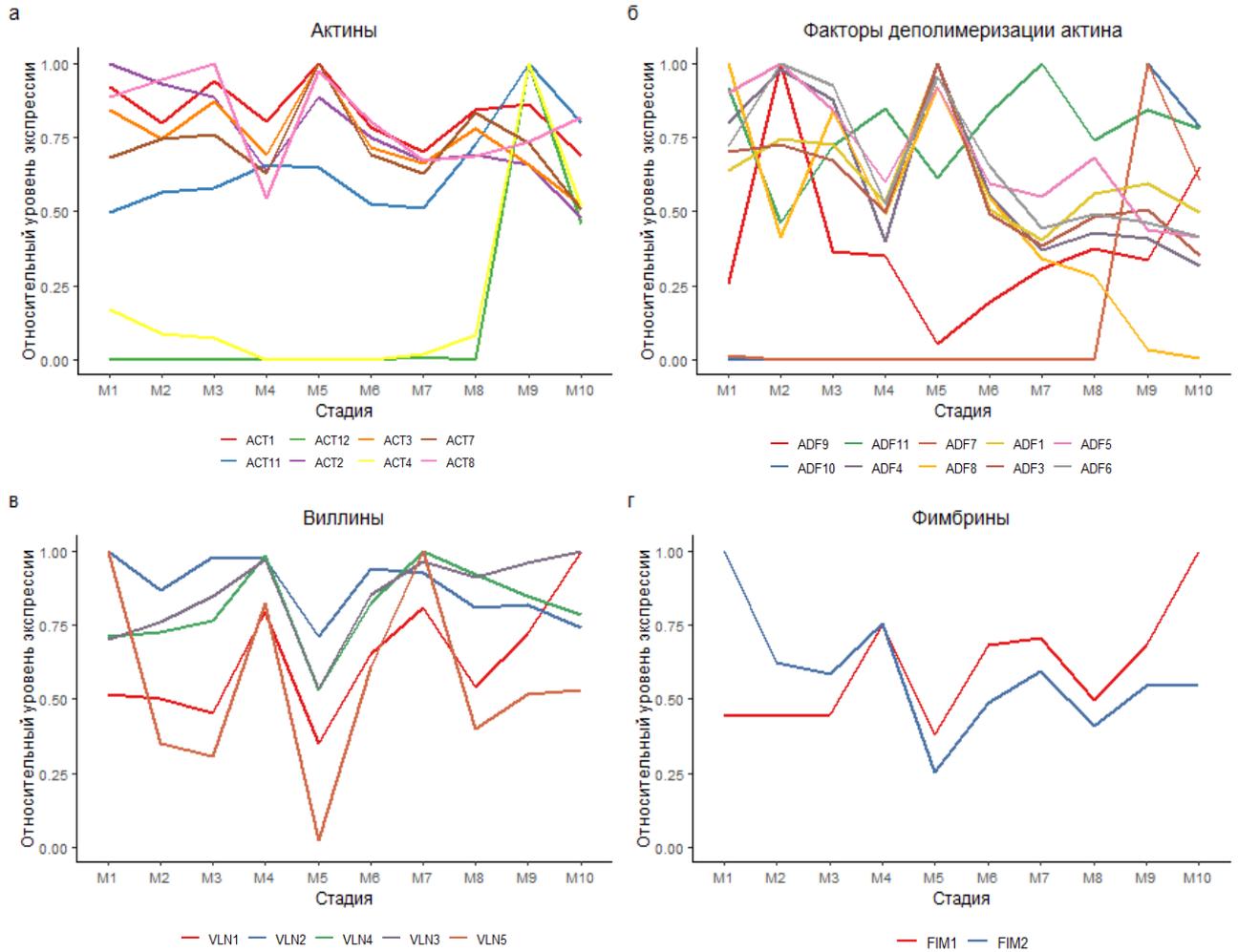
**Рисунок 1.4.** Иерархическая кластеризация образцов. В качестве метрики расстояния используется число генов, ДЭ между каждой парой образцов.



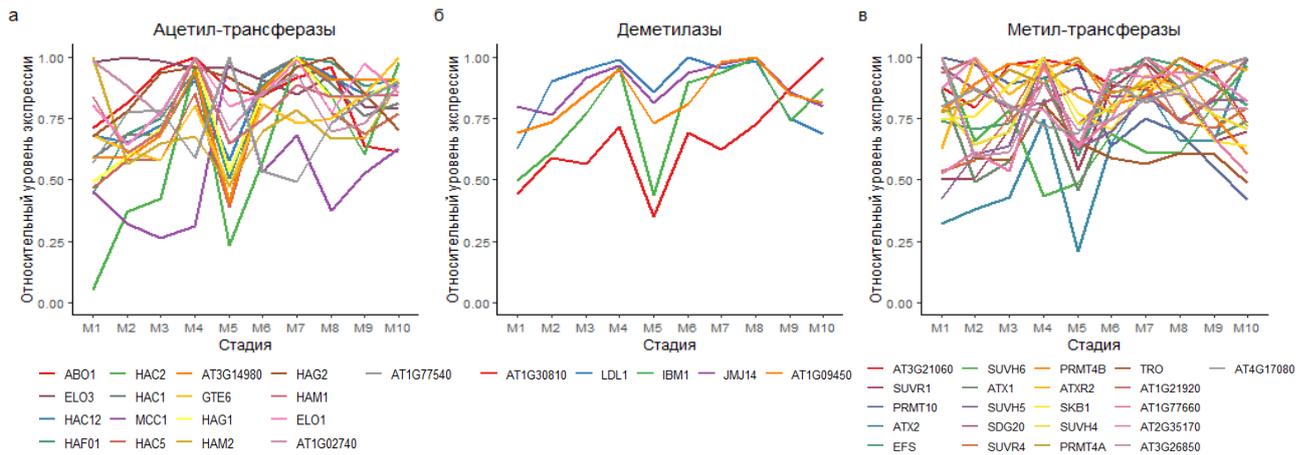
**Рисунок 1.5.** Распределение числа образцов, в которых экспрессируются гены с ДЭ значением менее ста.



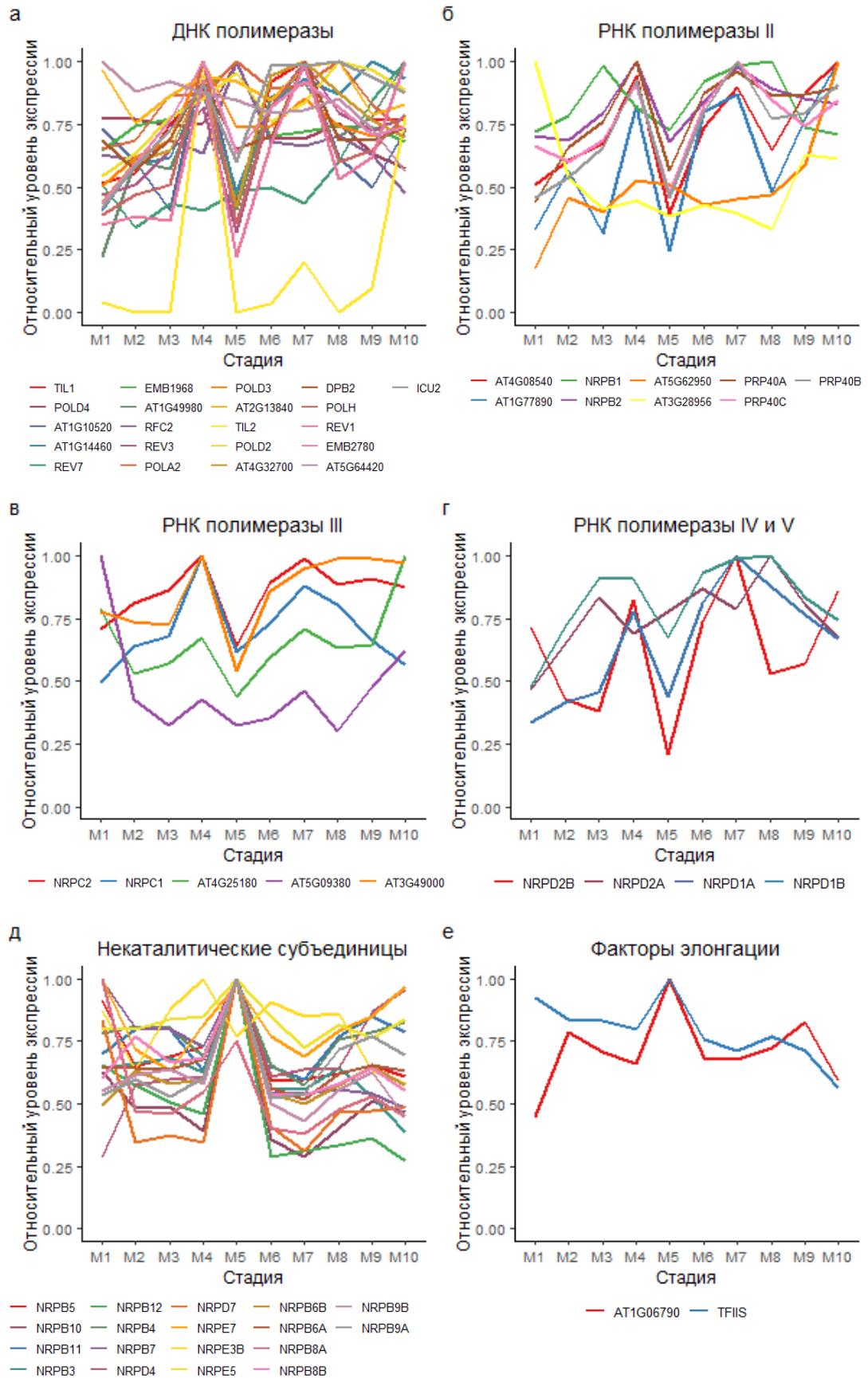
**Рисунок 2.1.** Профили экспрессии генов, регулирующих процесс перехода к цветению и участвующих в развитии цветков.



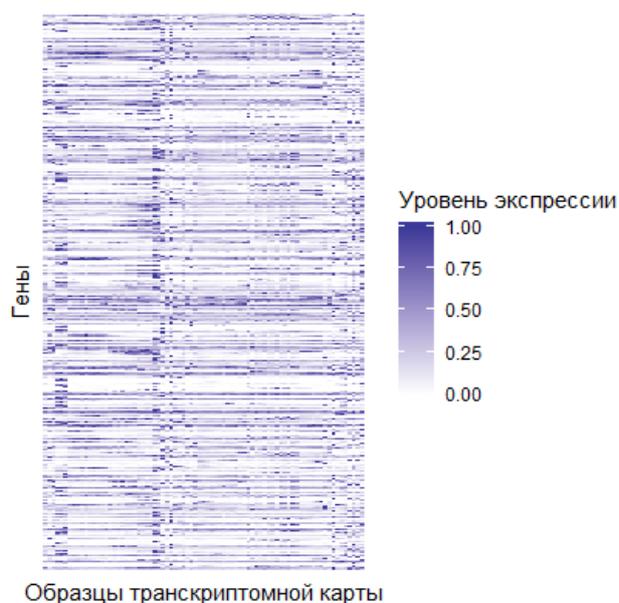
**Рисунок 2.2.** Профили экспрессии генов, кодирующих (а) актины, и регуляторов полимеризации актинов: (б) факторов деполимеризации актина, (в) виллинов и (г) фимбринов.



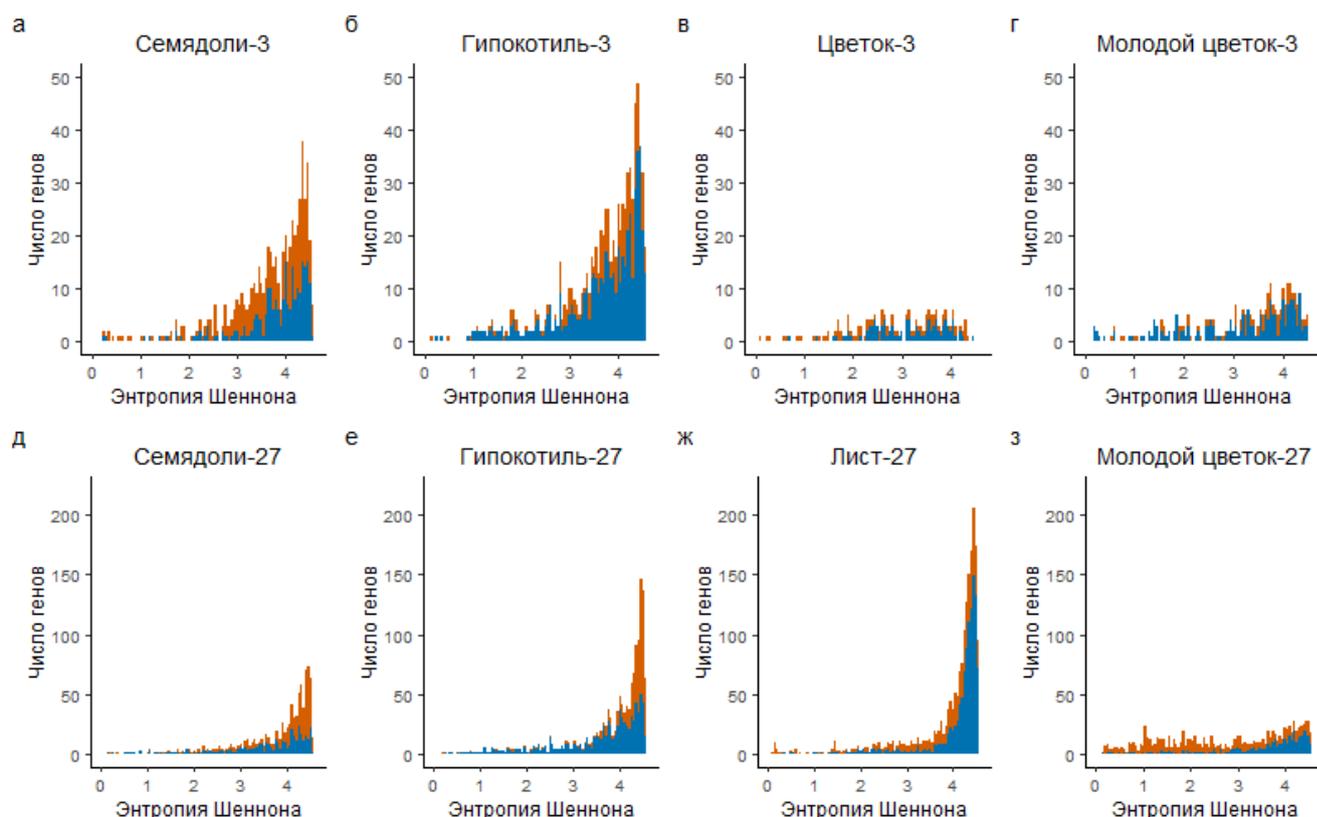
**Рисунок 2.3.** Профили экспрессии генов, кодирующих модификаторы гистонов: (а) ацетил-трансферазы, (б) деметилазы и (в) метил-трансферазы.



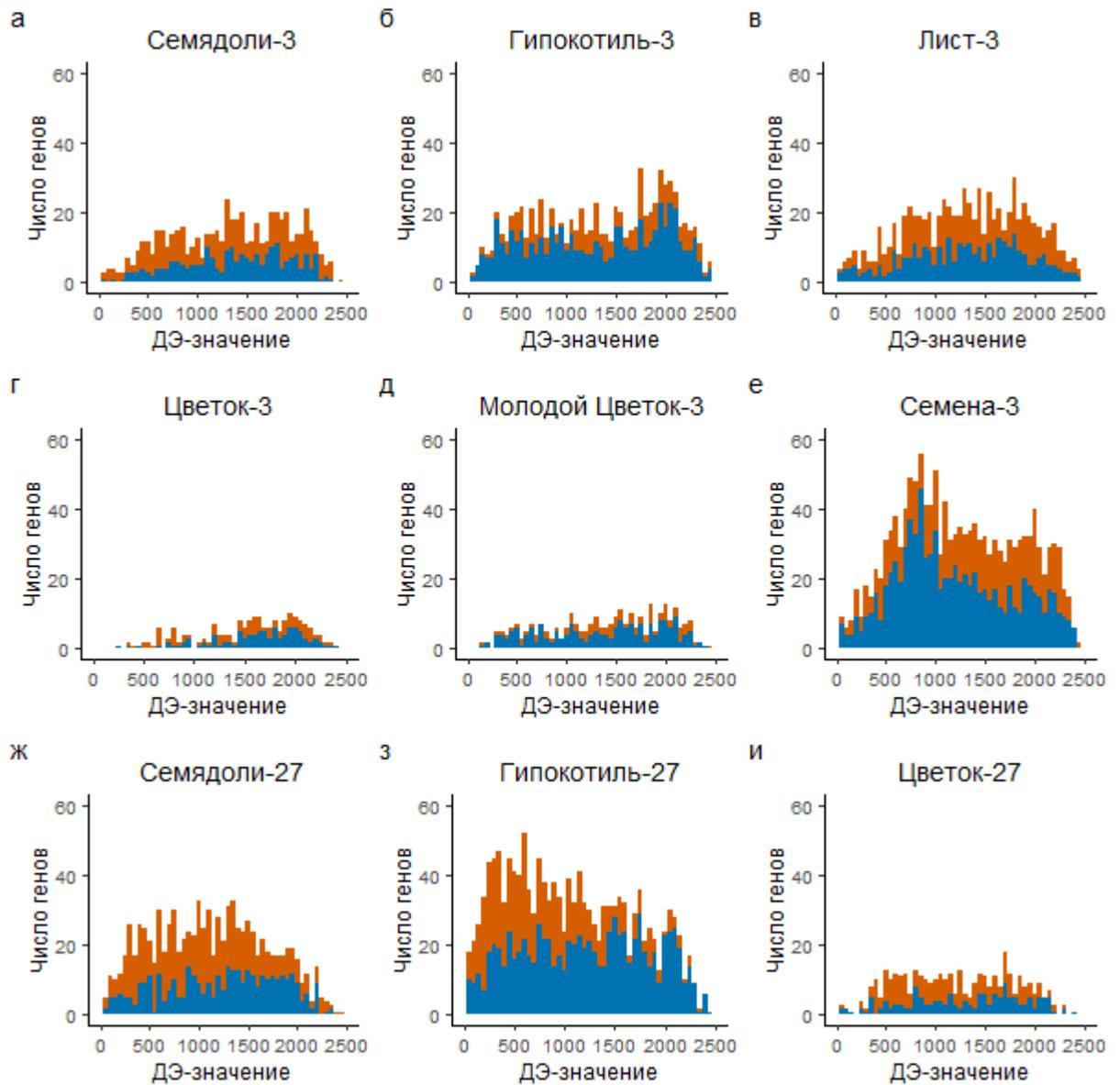
**Рисунок 2.4.** Профили экспрессии генов, кодирующих субъединицы ДНК- и РНК-полимераз: (а) ДНК-полимераз, (б) РНК-полимераз класса II, (в) РНК-полимераз класса III, (г) РНК-полимераз класса IV и V, (д) некодирующих субъединиц и (е) факторов элонгации.



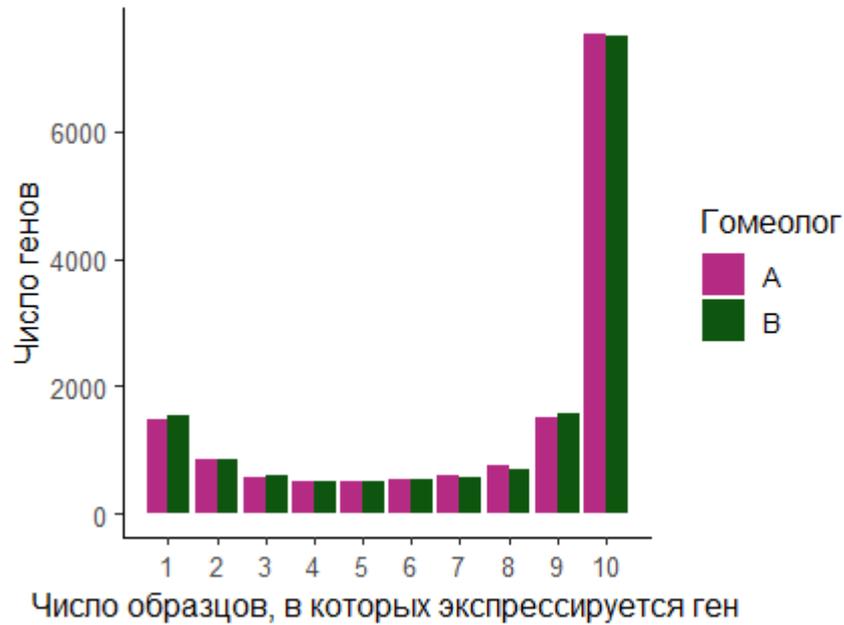
**Рисунок 3.1.** Тепловая карта, отражающая разнообразие уровней экспрессии генов, участвующих в ответе на стрессовые воздействия. Каждая строка тепловой карты представляет собой ген, аннотированный категорией «ответ на стресс» в Генной Онтологии; каждая колонка – образец транскриптомной карты. Приведены относительные уровни экспрессии, нормализованные для каждого гена на максимальное значение.



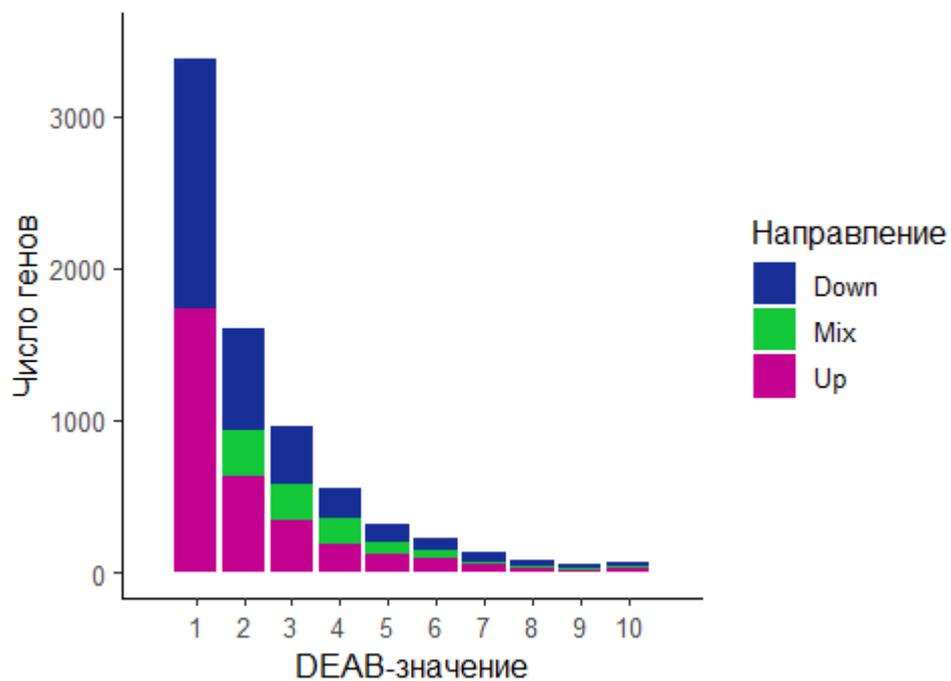
**Рисунок 3.2.** Распределение энтропии Шеннона генов, специфически отвечающих на холодовое воздействие в (а) Семядолях-3, (б) Гипокотиле-3, (в) Цветке-3, (г) Молодом цветке-3, (д) Семядолях-27, (е) Гипокотиле-27, (ж) Листе-27, (з) Молодом цветке-27. Синим отмечены гены, экспрессия которых снижается под действием холода, красным – гены, уровень транскрипции которых возрастает.



**Рисунок 3.3.** Распределение ДЭ-значений генов, специфически отвечающих на холодное воздействие в (а) Семядолях-3, (б) Гипокотиле-3, (в) Листе-3, (г) Цветке-3, (д) Молодом цветке-3, (е) Семенах-3, (ж) Семядолях-27, (з) Гипокотиле-27, (и) Цветке-27. Синим отмечены гены, экспрессия которых снижается под действием холода, красным – гены, уровень транскрипции которых возрастает.



**Рисунок 4.1.** Распределение генов по числу образцов основной транскриптомной карты *C. bursa-pastoris*, в которых ген экспрессируется. Гомеологи из субгеномов А и В приведены отдельно.



**Рисунок 4.2.** Распределение ДЭАВ значений дифференциально экспрессирующихся пар гомеологов. Цветами обозначено направление изменения экспрессии: синий (Down) – экспрессия гомеолога В снижена по сравнению с гомеологом А хотя бы в одном образце и нигде не повышена, маджента (Up) – экспрессия гомеолога В повышена по сравнению с гомеологом А хотя бы в одном образце и нигде не снижена, зеленый (Mix) – экспрессия гомеолога В снижена по сравнению с гомеологом А хотя бы в одном образце и повышена по сравнению с гомеологом А хотя бы в одном образце.

**Таблица 1.1.** Список транскриптомных карт, рассмотренных в обзоре литературы Главы 1.

Публикация	Объект	Число образцов	Метод изучения экспрессии
Schmid et al., 2005	<i>Arabidopsis thaliana</i>	79	Микрочип
Druka et al., 2006	<i>Hordeum vulgare</i>	15	Микрочип
Nobuta et al., 2007	<i>Oryza sativa</i>	18	MPSS
Benedito et al., 2008	<i>Medicago truncatula</i>	18	Микрочип
Quesada et al., 2008	<i>Populus trichocarpa</i>	5	Микрочип
Jiao et al., 2009	<i>Oryza sativa</i>	40	Микрочип
Schreiber et al., 2009	<i>Triticum aestivum</i>	13	Микрочип
Wang et al., 2010	<i>Oryza sativa</i>	39	Микрочип
Edwardset al., 2010	<i>Nicotiana tabacum</i>	19	Микрочип
Zhang et al., 2010	<i>Oryza sativa</i>	8	Секвенирование РНК
Libault et al., 2010	<i>Glycine max</i>	9	Секвенирование РНК
Severin et al., 2010	<i>Glycine max</i>	14	Секвенирование РНК
Sekhon et al., 2011	<i>Zea mays</i>	60	Микрочип
Massa et al., 2011	<i>Solanum tuberosum</i>	32	Секвенирование РНК
Fasoli et al., 2012	<i>Vitis vinifera</i>	54	Микрочип
Meyer et al., 2012	<i>Panicum hallii</i>	8	Секвенирование РНК
Dubois et al., 2012	<i>Rosa chinensis</i>	13	Секвенирование РНК
Verdier et al., 2013	<i>Lotus japonicus</i>	15	Микрочип
Sekhon et al., 2013	<i>Zea mays</i>	18	Секвенирование РНК
Tong et al., 2013	<i>Brassica rapa</i>	6	Секвенирование РНК
Shakoor et al., 2014	<i>Sorghum bicolor</i>	78	Микрочип
Ibarra-Laclette et al., 2015	<i>Persea americana</i>	9	Секвенирование РНК
O'Rourke et al., 2015	<i>Medicago sativa</i>	12	Секвенирование РНК
Alves-Carvalho et al., 2015	<i>Pisum sativum</i>	20	Секвенирование РНК
Ortiz-Ramirez et al., 2016	<i>Physcomitrella patens</i>	10	Микрочип
Kagale et al., 2016	<i>Camelina sativa</i>	12	Секвенирование РНК
Clevenger et al., 2016	<i>Arachis hypogaea</i>	22	Секвенирование РНК
Stelpflug et al., 2016	<i>Zea mays</i>	79	Секвенирование РНК
Walley et al., 2016	<i>Zea mays</i>	23	Секвенирование РНК
Vlasova et al., 2016	<i>Phaseolus vulgaris</i>	27	Секвенирование РНК
Yao et al., 2016	<i>Vigna unguiculata</i>	27	Секвенирование РНК
Sibout et al., 2017	<i>Brachypodium distachyon</i>	9	Микрочип
Ivamoto et al., 2017	<i>Coffea arabica</i>	7	Секвенирование РНК
Canas et al., 2017	<i>Pinus pinaster</i>	14	Секвенирование РНК
Kaur et al., 2017	<i>Trifolium subterraneum</i>	5	Секвенирование РНК
McCormick et al., 2018	<i>Sorghum bicolor</i>	47	Секвенирование РНК
Yano et al., 2018	<i>Cucumis melo</i>	30	Секвенирование РНК

**Таблица 1.2.** Описание образцов транскриптомной карты и квадраты коэффициентов корреляции Пирсона биологических повторностей каждого образца.

Код	Описание образца	$r^2$
S.H	Гипокотиль 1-дневного проростка	0,98
S.M	Меристема, развивающиеся первые два листа и прилежащие ткани однодневного проростка	0,87
S.C	Семядоли с черешками 1-дневного проростка	0,99
S.R	Корень 1-дневного проростка	0,97
R.A	Верхушка корня 7-дневного растения	0,94
R	Остальная часть корня 7-дневного растения	0,99
M1	Апикальная меристема 7-дневного растения	0,97
M2	Апикальная меристема 8-дневного растения	0,98
M3	Апикальная меристема 9-дневного растения	0,98
M4	Апикальная меристема 10-дневного растения	0,97
M5	Апикальная меристема 11-дневного растения	0,96
M6	Апикальная меристема 12-дневного растения	0,97
M7	Апикальная меристема 13-дневного растения	0,99
M8	Апикальная меристема 14-дневного растения	0,96
M9	Апикальная меристема 15-дневного растения	1
M10	Апикальная меристема 16-дневного растения	0,99
L.PET.y	Черешок третьего листа меньше 0.3 см	0,99
L.LAM.y	Пластинка с жилкой третьего листа меньше 0.3 см	0,98
L.PET.i1	Черешок третьего листа 0.5 см	0,99
L.LAM.i1	Пластинка с жилкой третьего листа 0.5 см	1
L.PET.i2	Черешок третьего листа 1 см	0,98
L.VN.i2	Жилка третьего листа 1 см	0,99
L.LAM.i2	Пластинка без жилок третьего листа 1 см	0,98
L.PET.lg	Черешок третьего листа на стадии распускания первого цветка	0,98
L.VN.lg	Жилка третьего листа на стадии распускания первого цветка	0,97
L.LAM.lg	Пластинка третьего листа на стадии распускания первого цветка	0,96
L.lg	Целый лист третьего листа на стадии распускания первого цветка	0,97
L.PET.sn	Черешок желтеющего третьего листа (на стадии зрелых семян нижних стручков)	0,87
L.VN.sn	Жилка желтеющего третьего листа (на стадии зрелых семян нижних стручков)	0,83
F.AN	Зрелые раскрывшиеся пыльники с пыльцой	1
F.CA.ad	Пестик взрослого цветка с нераскрывшимися пыльниками (третий цветок на стадии распускания первого цветка)	0,97
F.AN.ad	Пыльники взрослого цветка с нераскрывшимися пыльниками (третий цветок на стадии распускания первого цветка)	0,98
F.FM.ad	Тычиночные нити взрослого цветка с нераскрывшимися пыльниками (третий цветок на стадии распускания первого цветка)	0,96
F.PT.ad	Лепестки взрослого цветка с нераскрывшимися пыльниками (третий цветок на стадии распускания первого цветка)	0,98
F.SP.ad	Чашелистики взрослого цветка с нераскрывшимися пыльниками (третий цветок на стадии распускания первого цветка)	0,98
F.CA.y	Пестик цветка на стадии 9	1

F.AN.y	Пыльники цветка на стадии 10	0,93
F.SP.y	Чашелистики цветка на стадии 11	0,99
F1	Первый цветок на стадии распускания первого цветка	0,99
F2	Второй цветок на стадии распускания первого цветка	0,97
F3	Третий цветок на стадии распускания первого цветка	1
F4	Четвертый цветок на стадии распускания первого цветка	0,96
F5	Пятый цветок на стадии распускания первого цветка	0,98
F6-8	Шестой-восьмой цветки на стадии распускания первого цветка	0,99
F9-11	Девятый-одиннадцатый цветки на стадии распускания первого цветка	1
F12-14	Двенадцатый-четырнадцатый цветки на стадии распускания первого цветка	0,99
F15-18	Пятнадцатый-восемнадцатый цветки на стадии распускания первого цветка	0,92
F19+	Начиная с девятнадцатого цветка все соцветие на стадии распускания первого цветка	0,99
PED	Цветоножки цветков на стадии распускания первого цветка	1
AX	Ось соцветия на стадии распускания первого цветка	0,99
IN	Первое междоузлие растения на стадии распускания первого цветка	0,99
SD1	Семена из зрелого зеленого первого стручка (1 см длина)	0,99
POD1	Стручок без семян из зрелого зеленого первого стручка (1 см длина)	0,98
SL2	Стручок второй целиком при зрелом зеленом первом стручке (1 см длина)	1
SD3	Семена третьего стручка при зрелом зеленом первом стручке (1 см длина)	0,9
POD3	Стручок третий без семян при зрелом зеленом первом стручке (1 см длина)	0,99
SL4	Стручок четвертый целиком при зрелом зеленом первом стручке (1 см длина)	0,96
SD5	Семена пятого стручка при зрелом зеленом первом стручке (1 см длина)	1
POD5	Стручок пятый без семян при зрелом зеленом первом стручке (1 см длина)	0,99
SL6	Стручок шестой целиком при зрелом зеленом первом стручке (1 см длина)	0,98
SD7	Семена седьмого стручка при зрелом зеленом первом стручке (1 см длина)	0,99
POD7	Стручок седьмой без семян при зрелом зеленом первом стручке (1 см длина)	0,98
SL8	Стручок восьмой целиком при зрелом зеленом первом стручке (1 см длина)	0,98
SD.y1	Семена из первого стручка сразу после опадения околоцветника	0,97
SD.y2	Семена из второго стручка сразу после опадения околоцветника первого стручка	0,98
SD.y3	Семена из третьего стручка сразу после опадения околоцветника первого стручка	0,96
SD.y4	Семена из четвертого стручка сразу после опадения околоцветника первого стручка	0,97
SD.y5	Семена из пятого стручка сразу после опадения околоцветника первого стручка	0,96
OV.y6-7	Семяпочки из пестиков шестого и седьмого цветков сразу после опадения околоцветника первого стручка (до оплодотворения)	0,99
STI	Рыльцевая ткань пестиков шестого и седьмого цветков сразу после опадения околоцветника первого стручка (до оплодотворения)	0,99

POD.y6-7	Ткань плодолистиков шестого и седьмого цветков без семяпочек и рыльцевой ткани сразу после опадения околоцветника первого стручка (до оплодотворения)	0,99
SD.d	Сухие семена	0,98
SD.g1	Первый день после замачивания на фильтровальной бумаге	0,98
SD.g2	Второй день после замачивания на фильтровальной бумаге	1
SD.g3	Проростки на фильтровальной бумаге	0,99
SD.sn1	Семена первого желтеющего стручка (легко раскрывается)	0,98
POD.sn1	Первый желтеющий стручок без семян	0,99
SL.sn2	Второй стручок целиком при первом желтеющем цветке	0,99
IN.sn	Первое междоузлие желтеющее	0,98

**Таблица 1.3.** Обогащение терминами ГО и других баз данных списка генов, имеющих ДЭ значение больше 2 300.

База данных	Термин ГО или других баз данных	Уровень обогащения	FDR
COG_ONTOLOGY	Secondary metabolites biosynthesis, transport, and catabolism	4.04	0.00
GOTERM_BP_FAT	GO:0016139~glycoside catabolic process	29.78	0.00
GOTERM_BP_FAT	GO:0016145~S-glycoside catabolic process	27.49	0.00
GOTERM_BP_FAT	GO:0019759~glycosinolate catabolic process	27.49	0.00
GOTERM_BP_FAT	GO:0019762~glucosinolate catabolic process	27.49	0.00
GOTERM_BP_FAT	GO:0019253~reductive pentose-phosphate cycle	25.53	0.00
GOTERM_BP_FAT	GO:0019685~photosynthesis, dark reaction	23.83	0.00
GOTERM_BP_FAT	GO:0044273~sulfur compound catabolic process	23.16	0.00
GOTERM_BP_FAT	GO:0019757~glycosinolate metabolic process	18.95	0.00
GOTERM_BP_FAT	GO:0016143~S-glycoside metabolic process	18.95	0.00
GOTERM_BP_FAT	GO:0019760~glucosinolate metabolic process	18.95	0.00
GOTERM_BP_FAT	GO:0015977~carbon utilization by fixation of carbon dioxide	17.87	0.02
GOTERM_BP_FAT	GO:0019758~glycosinolate biosynthetic process	16.43	0.00
GOTERM_BP_FAT	GO:0019761~glucosinolate biosynthetic process	16.43	0.00
GOTERM_BP_FAT	GO:0016144~S-glycoside biosynthetic process	16.43	0.00
GOTERM_BP_FAT	GO:0016137~glycoside metabolic process	11.35	0.00
GOTERM_BP_FAT	GO:0009611~response to wounding	6.48	0.00
GOTERM_BP_FAT	GO:0015979~photosynthesis	6.47	0.00
GOTERM_BP_FAT	GO:0006790~sulfur metabolic process	5.92	0.00
GOTERM_BP_FAT	GO:0009414~response to water deprivation	5.75	0.00
GOTERM_BP_FAT	GO:0009753~response to jasmonic acid stimulus	5.73	0.00
GOTERM_BP_FAT	GO:0009415~response to water	5.47	0.00
GOTERM_BP_FAT	GO:0044275~cellular carbohydrate catabolic process	5.34	0.01
GOTERM_BP_FAT	GO:0042545~cell wall modification	5.29	0.02
GOTERM_BP_FAT	GO:0009409~response to cold	4.35	0.00
GOTERM_BP_FAT	GO:0016051~carbohydrate biosynthetic process	3.99	0.01
GOTERM_BP_FAT	GO:0019748~secondary metabolic process	3.52	0.00
GOTERM_BP_FAT	GO:0009416~response to light stimulus	3.45	0.00
GOTERM_BP_FAT	GO:0009314~response to radiation	3.34	0.00

GOTERM_BP_FAT	GO:0009266~response to temperature stimulus	3.20	0.04
GOTERM_BP_FAT	GO:0009628~response to abiotic stimulus	2.64	0.00
GOTERM_BP_FAT	GO:0009719~response to endogenous stimulus	2.20	0.02
GOTERM_BP_FAT	GO:0010033~response to organic substance	2.18	0.00
GOTERM_BP_FAT	GO:0055114~oxidation reduction	2.16	0.00
GOTERM_CC_FAT	GO:0010598~NAD(P)H dehydrogenase complex (plastoquinone)	30.35	0.02
GOTERM_CC_FAT	GO:0048046~apoplast	5.08	0.00
GOTERM_CC_FAT	GO:0009535~chloroplast thylakoid membrane	4.40	0.00
GOTERM_CC_FAT	GO:0055035~plastid thylakoid membrane	4.40	0.00
GOTERM_CC_FAT	GO:0042651~thylakoid membrane	4.38	0.00
GOTERM_CC_FAT	GO:0034357~photosynthetic membrane	4.05	0.00
GOTERM_CC_FAT	GO:0044436~thylakoid part	3.77	0.00
GOTERM_CC_FAT	GO:0031976~plastid thylakoid	3.67	0.00
GOTERM_CC_FAT	GO:0009534~chloroplast thylakoid	3.67	0.00
GOTERM_CC_FAT	GO:0031984~organelle subcompartment	3.65	0.00
GOTERM_CC_FAT	GO:0009579~thylakoid	3.46	0.00
GOTERM_CC_FAT	GO:0030312~external encapsulating structure	3.04	0.00
GOTERM_CC_FAT	GO:0009941~chloroplast envelope	3.04	0.01
GOTERM_CC_FAT	GO:0009570~chloroplast stroma	3.04	0.02
GOTERM_CC_FAT	GO:0009526~plastid envelope	3.03	0.01
GOTERM_CC_FAT	GO:0009532~plastid stroma	3.02	0.01
GOTERM_CC_FAT	GO:0005618~cell wall	2.98	0.00
GOTERM_CC_FAT	GO:0044434~chloroplast part	2.74	0.00
GOTERM_CC_FAT	GO:0044435~plastid part	2.65	0.00
GOTERM_CC_FAT	GO:0005576~extracellular region	2.53	0.00
INTERPRO	IPR000528:Plant lipid transfer protein/Par allergen	15.95	0.04
INTERPRO	IPR013770:Plant lipid transfer protein and hydrophobic protein, helical	9.44	0.00
INTERPRO	IPR003612:Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor	6.69	0.01
KEGG_PATHWAY	ath00966:Glucosinolate biosynthesis	16.87	0.01
PIR_SUPERFAMILY	PIRSF001654:phospholipid transfer protein	18.33	0.01
SMART	SM00499:AAI	11.03	0.00
SP_PIR_KEYWORDS	calvin cycle	34.13	0.01
SP_PIR_KEYWORDS	photorespiration	31.50	0.02
SP_PIR_KEYWORDS	Monooxygenase	5.18	0.00
SP_PIR_KEYWORDS	glycosidase	4.85	0.00
SP_PIR_KEYWORDS	cell wall biogenesis/degradation	4.25	0.04
SP_PIR_KEYWORDS	signal	3.13	0.00
SP_PIR_KEYWORDS	disulfide bond	3.12	0.00
SP_PIR_KEYWORDS	oxidoreductase	3.08	0.00
SP_PIR_KEYWORDS	glycoprotein	3.04	0.00
SP_PIR_KEYWORDS	chloroplast	2.94	0.00
SP_PIR_KEYWORDS	iron	2.86	0.04
SP_PIR_KEYWORDS	alternative splicing	2.50	0.00
SP_PIR_KEYWORDS	Secreted	2.45	0.04

**Таблица 1.4.** Обогащение терминами ГО и других баз данных списка генов, имеющих ДЭ значение меньше 100.

База данных	Термин ГО или других баз данных	Уровень обогащения	FDR
GOTERM_BP_FAT	GO:0006396~RNA processing	3.63	0.00
GOTERM_BP_FAT	GO:0000154~rRNA modification	8.11	0.00
GOTERM_BP_FAT	GO:0046907~intracellular transport	3.33	0.00
GOTERM_BP_FAT	GO:0006886~intracellular protein transport	3.86	0.00
GOTERM_BP_FAT	GO:0034660~ncRNA metabolic process	3.91	0.00
GOTERM_BP_FAT	GO:0034613~cellular protein localization	3.71	0.00
GOTERM_BP_FAT	GO:0016192~vesicle-mediated transport	3.64	0.00
GOTERM_BP_FAT	GO:0034470~ncRNA processing	4.47	0.00
GOTERM_BP_FAT	GO:0070727~cellular macromolecule localization	3.49	0.00
GOTERM_BP_FAT	GO:0006364~rRNA processing	5.37	0.00
GOTERM_BP_FAT	GO:0016072~rRNA metabolic process	5.37	0.00
GOTERM_BP_FAT	GO:0009451~RNA modification	4.96	0.00
GOTERM_BP_FAT	GO:0006397~mRNA processing	4.89	0.00
GOTERM_BP_FAT	GO:0016071~mRNA metabolic process	4.54	0.00
GOTERM_CC_FAT	GO:0005732~small nucleolar ribonucleoprotein complex	5.79	0.00
GOTERM_CC_FAT	GO:0048475~coated membrane	6.27	0.03
GOTERM_CC_FAT	GO:0030117~membrane coat	6.27	0.03
INTERPRO	IPR017451:F-box associated type 1	3.28	0.00
UP_SEQ_FEATURE	domain:F-box	2.26	0.01
UP_SEQ_FEATURE	metal ion-binding site:Manganese	6.79	0.02
UP_SEQ_FEATURE	repeat:PPR 5	2.54	0.04
UP_SEQ_FEATURE	repeat:PPR 4	2.51	0.05

**Таблица 2.1.** Морфологические маркеры, использованные для синхронизации стадии развития растений, находящихся на близких стадиях развития. Номера стадий развития цветка приведены по Smyth et al., 1990.

Стадия	День после прорастания	Число листьев, видимых невооруженным глазом	Доля жилки и номер листа	Номер первого листа с трихомами	Номера листьев на стадии примордиев	Число цветков (стадии)
M1	7	2		3	6	-
M2	8	3	1,2 (50)	4	8	-
M3	9	4	1,2 (90)	6	9	-
M4	10	5	1,2 (90)	8	13	-
M5	11	6	3 (50)	9	14	-
M6	12	6	3 (90)	9	14	-
M7	13	7		10	16/17	-
M8	14	8		10		3 (1-2)
M9	15	9		10		4 (1-4)
M10	16	10				5 (2-5)

**Таблица 3.1.** Описание изучаемых образцов; для каждого органа приведен квадрат корреляции Пирсона для биологических повторностей контроля, образца, подвергнутого воздействию холода в течении 3 и 27 часов.

Образец	Описание образца	Стадия развития растения	$r^2$ , контроль	$r^2$ , 3 часа холода	$r^2$ , 27 часов холода
Семядоли	Семядоли без черешков	Однодневный проросток	0.99	0.99	0.99
Гипокотиль	Гипокотиль от места прикрепления семядолей до корня	Однодневный проросток	0.99	0.97	1
Лист	Целый третий лист	Распускание первого цветка	0.97	0.99	0.97
Цветок	Целый первый цветок без цветоножки	Распускание первого цветка	0.99	0.98	0.98
Молодой цветок	Целый 12-ый цветок без цветоножки (стадия 9 по to Smyth et al., 1999)	Распускание первого цветка	0.98	0.95	0.9
Семена	Семена из пятого стручка	Зеленый первый стручок, 1,5 см	0.94	0.99	0.9

**Таблица 3.2.** Число дифференциально экспрессирующихся генов.

Образец	Число ДЭ генов			Соотношение числа ДЭ генов со сниженной и повышенной экспрессией
	со сниженной экспрессией	с повышенной экспрессией	всего	
Семядоли-3	855	1286	2141	0.66
Семядоли-27	2885	3092	5977	0.93
Гипокотиль-3	1238	937	2175	1.32
Гипокотиль-27	3546	2699	6245	1.31
Лист-3	765	1351	2116	0.57
Лист-27	3580	3287	6867	1.09
Цветок-3	352	670	1022	0.53
Цветок-27	1350	961	2311	1.4
Молодой цветок-3	603	560	1163	1.08
Молодой цветок-27	2388	2134	4522	1.12
Семена-3	1112	1142	2254	0.97
Семена-27	1766	1327	3093	1.33

**Таблица 3.3.** Направление изменения экспрессии генов, входящих в регулон ТФ СВФ, под действием холода в органах *A. thaliana*.

Гены, экспрессия которых возрастает под 35S:CBF				
Образец	Изменение экспрессии в наших данных			% совпадения
	Снижение	Рост	Не ДЭ	
Семядоли-3	1	90	45	66
Гипокотиль-3	2	68	66	50
Лист-3	0	112	24	82
Цветок-3	0	89	47	65
Молодой цветок-3	0	82	54	60
Семена-3	2	47	87	35
Семядоли-27	0	107	29	79
Гипокотиль-27	3	108	25	79
Лист-27	0	127	9	93
Цветок-27	1	76	59	56
Молодой цветок-27	9	76	51	56
Семена-27	1	66	69	49

Гены, экспрессия которых снижается под 35S:CBF				
Образец	Изменение экспрессии в наших данных			% совпадения
	Снижение	Рост	Не ДЭ	
Семядоли-3	8	4	28	20
Гипокотиль-3	5	4	31	13
Лист-3	12	5	23	30
Цветок-3	12	0	28	30
Молодой цветок-3	5	0	35	13
Семена-3	4	3	33	10
Семядоли-27	28	0	12	70
Гипокотиль-27	28	0	12	70
Лист-27	26	2	12	65
Цветок-27	15	0	25	38
Молодой цветок-27	16	3	21	40
Семена-27	8	3	29	20

**Таблица 3.4.** Число органоспецифических ДЭ генов.

Образец	Число ДЭ генов после трех часов		Число ДЭ генов после 27 часов	
	со сниженной экспрессией	с повышенной экспрессией	со сниженной экспрессией	с повышенной экспрессией
Семядоли	236	346	372	551
Гипокотиль	543	274	852	580
Лист	333	425	1115	769
Цветок	99	90	158	217
Молодой цветок	196	90	354	729
Семена	824	589	724	359

**Таблица 4.1.** Описание образцов транскриптомной карты *C. bursa-pastoris*. Для каждого образца приведен квадрат коэффициента корреляции Пирсона между биологическими повторностями.

ID образца	Образец	$r^2$
M	Апикальная меристема	0.907
L.P	Черешок листа	0.978
L.L	Пластинка листа	0.997
SD	Зрелые зеленые семена	0.96
SP	Чашелистики	0.991
PT	Лепестки	0.989
AN	Пыльники	0.971
OV	Семяпочки до оплодотворения	0.993
SD.d	Сухие семена	0.991
R	Корень	0.994
C.C	Семядоли, контроль	0.982
C.3	Семядоли, 3 часа холода	0.99
C.15	Семядоли, 15 часов холода	0.986
C.27	Семядоли, 27 часов холода	0.999

**Таблица 4.2.** Число экспрессирующихся гомеологов в каждом из образцов.

Образец	Число экспрессирующихся генов	Число экспрессирующихся гомеологов A	Число экспрессирующихся гомеологов B	Число пар, в которых оба гомеолога экспрессируются
M	18772	9386	9386	8399
L.P	22857	11440	11417	10942
L.L	22901	11428	11473	10986
SD	23193	11625	11568	10942
SP	22977	11516	11461	10881
PT	22422	11249	11173	10606
AN	23332	11690	11642	10768
OV	22788	11384	11404	10857
SD.d	21366	10672	10694	9829
R	24380	12187	12193	11666
C.C	23153	11583	11570	11069
C3	22955	11519	11436	10922
C15	22793	11398	11395	10809
C27	22580	11320	11260	10700

**Таблица 4.3.** Число дифференциально экспрессирующихся генов. Выше диагонали представлено число генов, экспрессия которых возрастает в образце из колонки по сравнению с образцом из строки; ниже диагонали – число генов, экспрессия которых снижается.

Образец	M	L.P	L.L	SD	SP	PT	AN	OV	SD.d	R
M	-	1164	2244	3502	6268	5128	7965	1858	4320	4746
L.P	1147	-	2060	5032	7104	6146	9381	3594	6613	6034
L.L	1627	1447	-	4759	6933	6626	9421	3503	6631	6136
SD	1421	2542	2655	-	5923	5282	7608	2277	3947	4863
SP	4951	7269	7060	6946	-	3123	7645	7576	6402	8306
PT	4591	7258	7566	6850	4543	-	7608	7129	7109	8265
AN	5528	8824	8723	7288	6176	4846	-	8715	6690	9181
OV	2060	4608	5180	5034	7627	6247	9138	-	6465	7020
SD.d	3971	7136	7309	5432	7601	7865	8107	7328	-	8577
R	3114	5030	5473	5639	7344	5951	9091	5119	6450	-

**Таблица 4.4.** Число пар гомеологов, в которых только один гомеолог, либо оба являются ДЭ в сравнениях разных образцов.

Пара гомеологов, в которой	Число пар
оба гомеолога ДЭ, только А ДЭ и только В ДЭ	10165
оба гомеолога ДЭ и только А ДЭ	1671
оба гомеолога ДЭ и только В ДЭ	1638
оба гомеолога ДЭ	679
только А или В - ДЭ, но не оба одновременно	313
только А - ДЭ	551
только В - ДЭ	543

**Таблица 4.5.** Квадрат коэффициента корреляции Пирсона между гомеологичными генами в каждом из образцов.

Образец	$r^2$	Образец	$r^2$
M	0.80	OV	0.87
L.P	0.89	SD.d	0.83
L.L	0.90	R	0.88
SD	0.87	C.C	0.88
SP	0.85	C3	0.88
PT	0.85	C15	0.87
AN	0.81	C27	0.87

**Таблица 4.6.** Обогащение терминами ГО и других баз данных списка генов, имеющих разно направление изменения экспрессии в двух и более образцах.

База данных	Термин ГО или других баз данных	Уровень обогащения	FDR
GOTERM_MF_FAT	GO:0015238~drug transporter activity	4.65	0.05
INTERPRO	IPR002213:UDP-glucuronosyl/UDP-glucosyltransferase	4.62	0.01
INTERPRO	IPR017972:Cytochrome P450, conserved site	3.74	0.00
INTERPRO	IPR001128:Cytochrome P450	3.69	0.00
INTERPRO	IPR002401:Cytochrome P450, E-class, group I	3.67	0.00
GOTERM_MF_FAT	GO:0019825~oxygen binding	3.50	0.00
INTERPRO	IPR017973:Cytochrome P450, C-terminal region	3.46	0.01
SP_PIR_KEYWORDS	Monoxygenase	3.29	0.01
SP_PIR_KEYWORDS	heme	3.18	0.00
GOTERM_MF_FAT	GO:0046906~tetrapyrrole binding	3.07	0.00
GOTERM_MF_FAT	GO:0020037~heme binding	2.94	0.00
SP_PIR_KEYWORDS	iron	2.93	0.00
GOTERM_MF_FAT	GO:0005506~iron ion binding	2.54	0.00
COG_ONTOLOGY	Secondary metabolites biosynthesis, transport, and catabolism	2.51	0.00
SP_PIR_KEYWORDS	oxidoreductase	2.25	0.00