

На правах рукописи

Клепикова Анна Владимировна

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

03.01.09 – математическая биология, биоинформатика

03.02.07 – генетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва – 2020

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук.

Научный руководитель:

Пенин Алексей Александрович

кандидат биологических наук

зав. лабораторией

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации имени А.А. Харкевича Российской академии наук

Официальные оппоненты:

Самсонова Мария Георгиевна

доктор биологических наук

зав. лабораторией, профессор

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

Голденкова-Павлова Ирина Васильевна

доктор биологических наук

ведущий научный сотрудник, доцент

Федеральное государственное бюджетное учреждение науки Институт физиологии растений им. К.А. Тимирязева Российской академии наук

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»

Защита диссертации состоится «__» _____ 2020 года в _____ на заседании диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации имени А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке ИППИ РАН, а также на сайте ИППИ РАН по адресу: <http://iitp.ru/ru/dissertation/>

Автореферат разослан «__» _____ 2020 г.

Ученый секретарь диссертационного совета
доктор биологических наук, профессор

Рожкова Г.И.

Общая характеристика работы

Актуальность темы исследования. Изучение функций генов и сетей генетической регуляции процессов развития и жизнедеятельности организма входит число фундаментальных задач биологии. Классические методы исследования функции гена времязатратны и требуют значительных усилий для создания мутантных и трансгенных организмов и последующего их анализа, поэтому эта область долгое время ограничивалась группой модельных видов, таких как *Arabidopsis thaliana* или *Drosophila melanogaster*, а на немодельные проводилась экстраполяция полученных результатов.

Косвенным указанием на функцию гена может служить паттерн его экспрессии в разных клетках, тканях и органах. В последние несколько десятилетий развитие технологий привело к созданию новой области, посвященной массовому анализу экспрессии генов, – транскриптомики. Большим ее разделом является создание транскриптомных карт, или атласов экспрессии генов, отражающих уровень экспрессии всех генов организма во всех его органах и тканях на разных стадиях развития. С помощью транскриптомной карты возможно очень точно описать паттерн экспрессии гена и таким образом приблизиться к пониманию его функции.

Создание высокоразрешенного транскриптомного атласа остается сложной задачей как в отношении сбора образцов, так и последующего получения экспрессионных данных. Однако на фоне прогресса в транскриптомных технологиях новые транскриптомные карты самых разных организмов, в том числе растений, продолжают создаваться. Самым исследованным видом среди растений остается *A. thaliana*, использующийся в качестве основного источника информации о функциях и паттернах экспрессии генов. Первая транскриптомная карта растения была создана для *A. thaliana*, и, хотя в ее состав входит большое число образцов, использование микрочипов для измерения уровня экспрессии генов сокращает ее применимость как вследствие известных ограничений технологии, так и из-за неполного охвата генов *A. thaliana*.

Таким образом, для изучения функций генов путем анализа паттерна их экспрессии необходима подробная транскриптомная карта модельного объекта, созданная современными методами секвенирования РНК. Также важно подчеркнуть необходимость организации полученных экспрессионных данных в удобные для пользователей базы, предоставляющие возможности визуализации паттернов экспрессии генов и проведения разнообразных анализов, благодаря чему существенно повышается повторная используемость данных.

Цели и задачи. Целью настоящего исследования было создание транскриптомных карт *A. thaliana* и *Capsella bursa-pastoris* и оценка возможности использования экспрессионных данных для анализа биологических процессов. В рамках поставленной цели решались следующие задачи:

- 1) создание и анализ детализированных транскриптомных карт *A. thaliana* и *C. bursa-pastoris*;
- 2) изучение динамики экспрессии генов в меристеме *A. thaliana* при переходе к цветению;
- 3) изучение тканевой специфичности ответа на холодовой стресс у *A. thaliana*.

Научная новизна и практическая значимость. Наша работа была посвящена созданию и анализу транскриптомной карты *A. thaliana*, а также рассмотрению отдельных ее аспектов: временной серии апикальных меристем при переходе к цветению и органоспецифического ответа на холодовой стресс. Кроме того, высокая детализация атласа экспрессии *A. thaliana* была использована для планирования транскриптомной карты *C. bursa-pastoris*.

В первой главе описывается создание высокоразрешенной транскриптомной карты *A. thaliana* методом секвенирования РНК. Транскриптомная карта включала в себя ряд образцов, ранее не изучавшихся подробно с точки зрения транскриптомики. Анализ

транскриптомной карты позволил обнаружить экспрессию практически всех белок-кодирующих генов и определить основные экспрессионные параметры как транскриптома в целом, так и отдельных образцов. Мы провели поиск генов, стабильно экспрессирующихся в разнообразных органах, тканях и условиях выращивания, что было первым широкомасштабным анализом стабильности экспрессии с использованием результатов секвенирования РНК. Также нами были рассмотрены паттерны экспрессии ранее неизученных транскрипционных факторов и установлена их тканевая специфичность. Полученные данные были организованы в созданную нами базу данных Transcriptome Variation Analysis (TraVA), позволяющую визуализировать экспрессионные профили генов и проводить анализ дифференциальной экспрессии.

Во второй главе мы провели анализ временной серии апикальных меристем, что явилось первым экспрессионным исследованием процесса перехода к цветению в высоком разрешении. Анализ образцов, собираемых ежедневно в течении десяти дней, позволил обнаружить стадию в переходе к цветению, в которой массово изменяется экспрессия генов, ассоциированных с регуляцией клеточного цикла. Такое поведение было объяснено нами укорочением клеточного цикла в момент перестройки апикальной меристемы за счет фаз G₁ и G₂.

Третья глава была посвящена анализу генов, экспрессия которых изменяется в ответ на холодовое воздействие. Нами впервые было показано, что помимо базовой реакции на холод, универсальной для всего растения, в каждом органе функционируют дополнительные регуляторные сети, модифицирующие общий ответ в зависимости от особенностей органа.

В четвертой главе работы транскриптомная карта *A. thaliana* была использована как основа для создания атласа экспрессии генов аллотетраплоида *C. bursa-pastoris*. Это исследование представляет собой как первый транскриптомный атлас этого вида, так и наиболее масштабный анализ экспрессии генов *C. bursa-pastoris*. Анализ транскриптомной карты позволил установить отсутствие асимметричного сайленсинга гомеологов одного из субгеномов на ранних стадиях эволюции аллополиплоидов.

Полученные результаты имеют как теоретическую, так и практическую значимость. Анализ транскриптомной карты *A. thaliana* позволил выявить основные закономерности экспрессии как генов, так и тканей и органов, определить паттерны экспрессии генов, в том числе ранее не изучавшихся транскрипционных факторов, и определить стабильно экспрессирующихся в разных органах и тканях гены. Изучение паттернов экспрессии генов является отправной точкой для понимания их функции, а правильно выбранные стабильные гены необходимы для корректного проведения экспрессионных анализов с помощью метода ОТ-ПЦР в реальном времени.

Созданная на основе атласа экспрессии генов *A. thaliana* база данных Transcriptome Variation Analysis (TraVA) начиная с 2016 года набирает 4000 уникальных пользователей в год. Данные транскриптомной карты используются в основной базе данных *A. thaliana* TAIR (arabidopsis.org), где они приводятся по умолчанию в качестве экспрессионного паттерна гена. С использованием базы TraVA проводится учебный курс «Plant Bioinformatics» на платформе Coursera (<https://ru.coursera.org/learn/plant-bioinformatics>).

Анализ временной серии меристем позволил охарактеризовать транскриптомные изменения, происходящие в апикальной меристеме при переходе к цветению, и, кроме того, получить высокоразрешенную временную серию, повторный анализ которой проводится другими научными группами.

С помощью транскриптомной карты *C. bursa-pastoris* были установлены отсутствие геномного доминирования на уровне экспрессии и случайный характер инактивации одного гена из пары гомеологов на ранней стадии эволюции полиплоидов. Данные транскриптомной карты *C. bursa-pastoris* и результаты анализа экспрессии в различных органах под низкотемпературным воздействием были интегрированы в базу данных TraVA и также используются как публичный ресурс.

Положения, выносимые на защиту.

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 *A. thaliana*.

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

- в момент перехода к цветению в апикальной меристеме побега *A. thaliana* происходит критическое изменение динамики клеточных делений, которое, вероятно, связано с сокращением длительности клеточного цикла за счет фаз G₁ и G₂;
- развитие холодового стресса у *A. 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 *A. 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.

Личный вклад автора в исследование. Все результаты, включенные в диссертацию, получены лично соискателем или при его непосредственном участии, за исключением разделов 1.3.2, 2.3.6 и 3.3.8, выполненных соавторами публикации. Обсуждение и интерпретация результатов осуществлялись автором совместно с научным руководителем и соавторами публикаций.

Степень достоверности и апробация результатов. По материалам диссертации опубликовано пять статей в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях и семинарах: The 26th International Conference on Arabidopsis Research (ICAR) 2015, Париж, Франция; The 29th International Conference on Arabidopsis Research (ICAR) 2018, Турку, Финляндия; Life of Genomes 2018, Казань, Россия; The 44th the Federation of European Biochemical Societies Congress (FEBS) 2018, Прага, Чехия; The 11th International Conference on Bioinformatics of Genome Regulation and Structure\ Systems Biology 2018, Новосибирск Россия; The International Conference on Polyploidy 2019, Гент, Бельгия; VII съезд Вавиловского общества генетиков и селекционеров (ВОГиС), 2019, Санкт-Петербург, Россия.

Структура и объем работы. Диссертация изложена на 179 страницах 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.

Содержание работы

Глава 1 диссертации посвящена созданию и анализу транскриптомной карты *A. thaliana*. Анализ паттерна экспрессии гена служит удобным непрямым методом изучения его биологической функции, что особенно важно для немодельных объектов. Наиболее полное описание паттерна экспрессии представляет собой транскриптомную карту, или атлас экспрессии генов, – оценку глобальной экспрессии генов во всех органах и тканях в ходе всего онтогенеза организма от эмбриогенеза до старения. Обзор литературы Главы 1 освещает современное состояние области транскриптомики, посвященной созданию атласов экспрессии генов. Основой для обзора послужили 37 транскриптомных карт растений, созданные в период с 2005 по 2018 год с помощью технологии микрочипов и секвенирования РНК. Хотя с каждым годом число публикуемых атласов экспрессии возрастает все быстрее, представленность таксонов остается неравномерной. Так, 34 транскриптомных карты созданы для покрытосеменных растений, а однодольные представлены исключительно Злаками. Разрешение атласов существенно отличается как в отношении числа изученных образцов (от 5 до 79), так и объема полученных данных (от 9 Мб до 4,5 Gb для секвенирования РНК), однако общие характеристики транскриптома (число экспрессирующихся генов, профили экспрессии образцов, кластеризация органов и тканей) были в большой степени схожи у всех растений. Модельный объект биологии растений *A. thaliana* по-прежнему остается самым изученным растением; информация о функциях его генов зачастую переносится на другие виды, что делает важным создание подробной транскриптомной карты *A. thaliana* с помощью современных методов анализа экспрессии генов.

В разделе результатов Главы 1 описана структура собранной нами транскриптомной карты *A. thaliana* и приведены результаты ее анализа. Всего было собрано 79 образцов различных органов, тканей и частей растения, подробно описывающих его морфологию; отдельное внимание уделялось малоизученным в отношении транскриптома образцам (таким, как апикальная меристема побега) и временным сериям (листьев, семян, апикальных меристем) (Рисунок 1). Все образцы (транскриптомной карты, а также все, описанные в последующих главах диссертации) собирались в двух биологических повторностях, каждая из которых представляла собой пул из 15 растений. В качестве дополнительного набора данных мы использовали ранее собранные в лаборатории временные серии листовых проб, подвергнутых воздействию пониженной температуры (+4°C), повышенной температуры (+42°C) и механическим повреждениям (протыкание листьев иглой). РНК, выделенная из образцов, была секвенирована на платформе HiSeq2000 (Illumina, США) с получением суммарно 3,6 млрд чтений.

В качестве общей характеристики транскриптома *A. thaliana* мы использовали число экспрессирующихся генов как в отдельных образцах, так и по всей транскриптомной карте в целом. Мы использовали «строгий» порог экспрессируемости: на ген должно приходиться 16 или более нормализованных на размер библиотеки чтений в каждой из двух биологических повторностей образца. 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 генов, экспрессирующихся в каждом из образцов, было очень схожим (от 15 525 до 19 613). При добавлении к данным транскриптомной карты набора образцов, подвергнутых воздействию абиотических стрессов, обнаружили транскрипты дополнительных 26 генов. 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). Анализ обогащения этих генов терминами Генной Онтологии (ГО) показал 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).

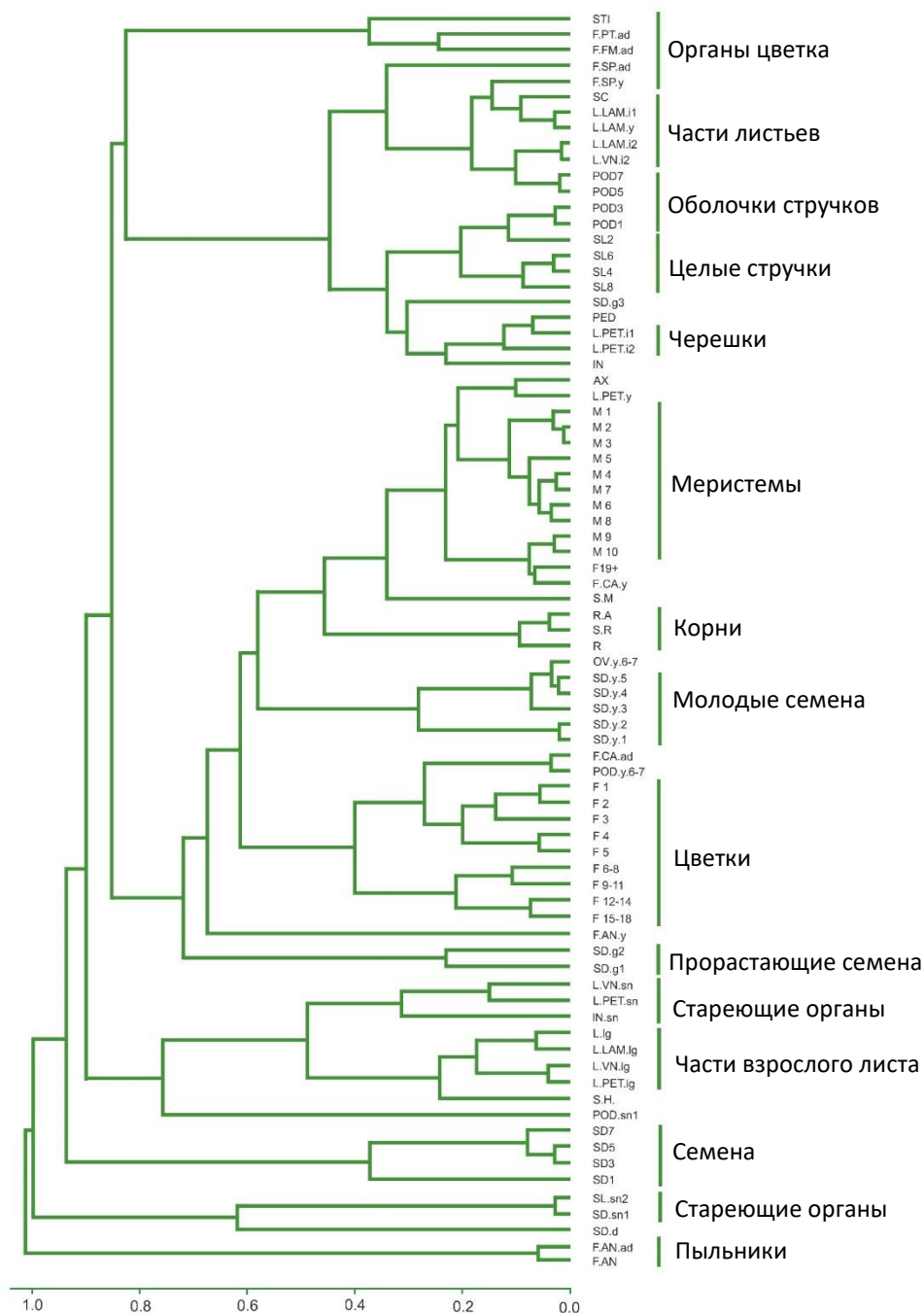


Рисунок 1. Дерево кластеризации образцов транскриптомной карты. В качестве меры расстояния между парой образцов (ось X) использовался вычтенный из единицы квадрат корреляции Пирсона.

В качестве первичного описания паттерна экспрессии гена мы использовали число образцов, в которых этот ген транскрибируется (Рисунок 2а); распределение этого значения 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. Гены, обладающие более широким паттерном экспрессии, имели более высокий уровень экспрессии (Рисунок 2б).

Для сравнения профилей экспрессии отдельных образцов 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. 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).

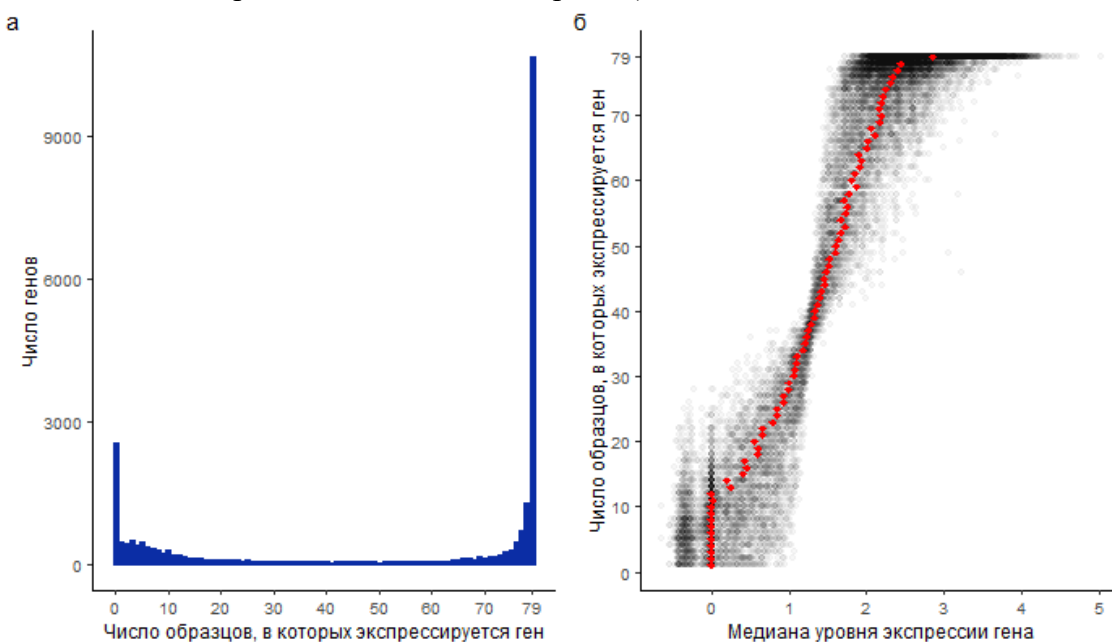


Рисунок 2. Экспрессионные характеристики генов. (а) Распределение генов по числу образцов, в которых ген экспрессируется (порог экспрессируемости составляет 16 и более чтений в каждой из биологических повторностей); (б) Распределение числа образцов, в которых экспрессируется ген, в зависимости от медианного уровня экспрессии гена.

Общая вариативность транскриптома *A. thaliana* была оценена с помощью анализа дифференциальной экспрессии генов между всеми возможными парами образцов (3 081 попарное сравнение). Число дифференциально экспрессирующихся (ДЭ) генов между разными парами образцов варьировалось от 14 до 15 149 ДЭ генов. Мы проанализировали перепредставленность категорий ГО в списках генов, ДЭ между каждой парой образцов; гены, экспрессия которых возрастает или снижается в одной из проб, рассматривались 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 списках ДЭ генов; к ним относятся такие категории, как «фотосинтетические мембраны» и «темновая стадия фотосинтеза». Кроме того, часто оказывались дифференциально экспрессирующимися гены, вовлеченные в регуляцию хроматина и деления клеток, такие как «коровые гистоны», «хроматин», «фазы клеточного цикла», «циклин, С-конец». 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 3. Гены, обладающие наибольшим ДЭ-значением, были обогащены терминами, относящимися к фотосинтезу и связанному с ним метаболизму, а гены с наименьшим – категориями, связанными с процессингом РНК, локализацией и транспортом.

Для поиска стабильно экспрессирующихся генов в качестве меры равномерности экспрессии мы использовали коэффициент вариации (CV), 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 генов, имеющих наименьший CV, терминами ГО показал перепредставленность категорий, связанных с транспортом нуклеиновых кислот и белков, мембранами и процессингом РНК. Чтобы проверить, сохраняет ли экспрессия этих генов стабильность в других условиях выращивания, мы добавили результаты секвенирования РНК растений, подвергнутых воздействию абиотических стрессов, 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. 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*.

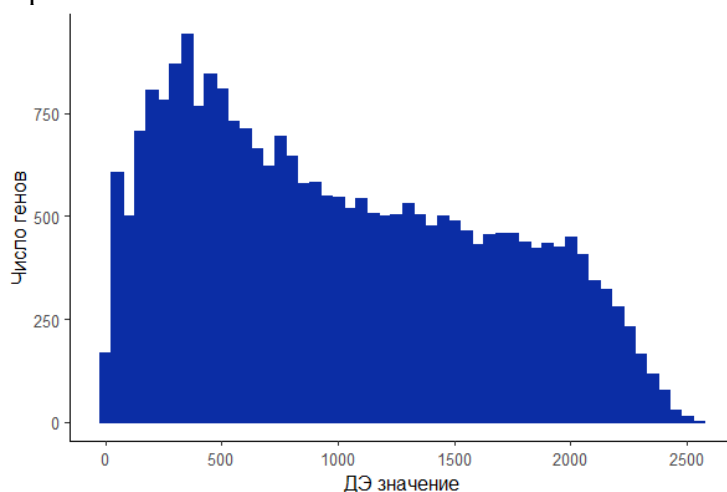


Рисунок 3. Распределение ДЭ-значений генов, дифференциально экспрессирующихся хотя бы в одном сравнении образцов.

Для решения обратной задачи – изучения тканевой специфичности экспрессии генов – мы использовали такую меру ширины паттерна экспрессии, как энтропия Шеннона (H). Распределение энтропии H смещено в область высоких значений, что отражает широкий паттерн экспрессии большинства генов (Рисунок 4). Второй локальный максимум, значительно меньших размеров, находится в области значений H 0,1-0,25 и образуется генами, специфично экспрессирующимися в определенных образцах.

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, и такими клеточными компонентами, как мембраны. Гены, энтропия которых была ниже 0,15, характеризовались категориями «межклеточный сигналинг», «пектинэстеразная активность», «клеточная стенка».

Полнотранскриптомный анализ транскрипционных факторов полезен для понимания функционирования сетей генетической регуляции, контролирующей разные аспекты развития и жизнедеятельности организма. Транскрипты 2 148 из 2 757 ТФ *A. thaliana* были обнаружены хотя бы в одном образце нашей транскриптомной карты. Для изучения паттерна экспрессии регуляторных генов мы проанализировали распределения значений энтропии Шеннона для всех семейств ТФ. В большинстве семейств встречались как тканеспецифичные ТФ, так и экспрессирующиеся во всех образцах гены. Наиболее узкий

паттерн экспрессии имели такие классы ТФ, как MADS, LOB, LIM и MYB, а семейства SWI/SNF-SWI3, SNF2, CAMTA, DDT и FAR были самыми широко экспрессирующимися. Внутри классов ТФ корреляция экспрессии отсутствовала.

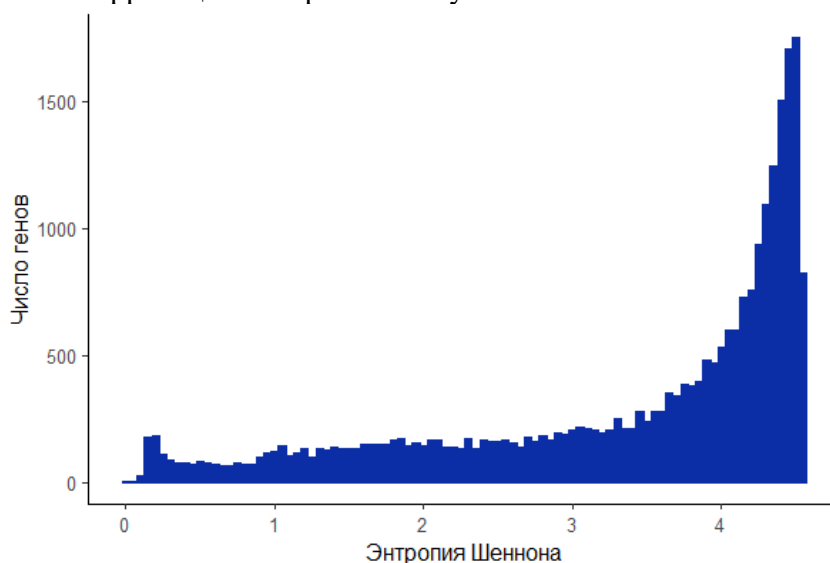


Рисунок 4. Распределение энтропии Шеннона экспрессирующихся в транскриптомной карте генов.

Результаты, полученные в Главе 1, обсуждаются в разделе 1.4. Анализ основанной на технологии микрочипов транскриптомной карты *A. thaliana* позволил выбрать наиболее различающиеся образцы и дополнить их малоизученными органами и частями растения, в том числе организованными во временные серии, что позволяет изучать динамику транскриптома и конструировать сети генетической регуляции. Кластеризация по профилям экспрессии образцов транскриптомной карты (Рисунок 1) отражает анатомию растения, и образцы преимущественно группируются по морфологическому и физиологическому сходству.

Общая характеристика транскриптома *A. thaliana* была проведена с помощью подсчета числа экспрессирующихся генов. Основные параметры, полученные нами, совпадали с другими растениями – 77% генов экспрессировались хотя бы в одном образце, 40% – во всех, образцы отличались друг от друга не более чем на 20%. Транскрипты большинства генов были обнаружены практически во всех образцах, то есть являлось необходимым для базовой жизнедеятельности клеток и тканей, вторую группу составляли тканеспецифичные гены (Рисунок 2а). Гены, экспрессирующиеся в большем числе образцов, характеризуются более высокой экспрессией (Рисунок 2б), тогда как тканеспецифические гены имеют меньшую медиану уровня экспрессии, что также свидетельствует о структурной роли первых и регуляторной – вторых. Включение в анализ дополнительного набора образцов, подвергнутых действию абиотических стрессов, практически не увеличило число экспрессирующихся генов, показывая полноту описания транскриптома *A. thaliana* нашими образцами. Среди неэкспрессирующихся генов преобладали связанные с защитой от биотического стресса, что может объясняться отсутствием среди использованных нами данных образцов, подвергнутых воздействию биотического стресса.

Изучение профилей экспрессии образцов с помощью параметра Z показало, что большинство образцов имеют похожие профили экспрессии с исключениями, объясняющимися низким уровнем экспрессии большинства генов в образце (желтеющие семена) или присутствием высокоэкспрессируемых органоспецифических генов (пыльники).

Для поиска различий между образцами мы провели анализ дифференциальной экспрессии между всеми парами образцов (3 081 сравнений). Многие из ДЭ генов связаны с фотосинтетическими процессами, что отражает разницу профилей экспрессии зеленых и

незеленых частей растения, а также с регуляцией хроматина и делением клеток, что вызвано отличиями между быстро развивающимися и зрелыми органами. Также мы оценили разнообразие экспрессии генов в транскриптомной карте с помощью меры «ДЭ-значение гена» (Рисунок 3). Гены, имеющие минимальное ДЭ-значение (ниже 100), были как тканеспецифическими, так и стабильно экспрессирующимися во всех образцах, и участвовали в базовых биологических процессах, таких, как процессинг РНК, транспорт и локализация белков. Функции многих из генов с ДЭ-значением больше 2 300 изучались ранее; среди их описания категориями ГО перепредставлены термины, связанные с фотосинтетическими процессами.

Изучение генов, обладающих равномерным уровнем экспрессии в разных органах и тканях растения, имеет как практическую значимость (получение референсов для сравнения экспрессии с помощью ОТ-ПЦР в реальном времени), так и для понимания, какие клеточные процессы требуют стабильной экспрессии участвующих в них генов во всем организме. Наиболее стабильно экспрессирующиеся гены (с $CV < 0,3$) участвуют в таких процессах, как «процессинг мРНК», «внутриклеточный транспорт» и «транспорт с помощью везикул». Используя дополнительный набор стрессовых данных совместно с транскриптомной картой, мы отобрали 27 наиболее стабильных генов (с $CV \leq 0,2$). Эти гены 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.

Изучение тканевой специфичности экспрессии генов является одной из важных задач транскриптомики, решение которой ведет к пониманию онтогенеза тканей и органов, а найденные гены могут использоваться в таких прикладных областях, как клеточный сортинг или создание узконаправленных трансгенных конструкций. Анализ паттернов экспрессии генов с помощью энтропии Шеннона (Рисунок 4) позволил обнаружить гены, обеспечивающих основы жизнедеятельности всех типов клеток и участвующих в таких процессах, как транспорт нуклеиновых кислот, метаболизм липопротеинов и мембран. Распределение энтропии Шеннона в нашей транскриптомной карте повторяет форму распределения *N* кукурузы, демонстрируя межвидовую универсальность соотношения широко экспрессирующихся и тканеспецифичных генов. Наиболее тканеспецифические гены 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.

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. 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. 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. Рассмотрение паттернов экспрессии ТФ из разных, даже хорошо изученных семейств (таких, как MADS), позволило определить точный паттерн экспрессии ряда генов, что может быть использовано для дальнейшего изучения их функции. Хотя многие ТФ одного семейства формируют белковые комплексы (белки субъединиц рибосом, «флоральные квартеты») распределение коэффициента корреляции Спирмена внутри классов генов не показывает наличия координированной экспрессии.

На основе полученной нами транскриптомная карта *A. thaliana* и результатов ее анализа была создана публично доступная база данных Transcriptome Variation Analysis

(TraVA, travadb.org). База данных предоставляет возможность визуализации уровня экспрессии как одного, так и многих генов во всех образцах атласа, а также анализа дифференциальной экспрессии. К марту 2019-го года посещаемость базы данных TraVA достигла 4000 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.

В Главе 2 мы рассматриваем временную серию апикальной меристемы (АМ) побега *A. thaliana*, входящую в состав собранной нами транскриптомной карты. Переход к цветению, то есть изменение типа развития АМ с вегетативного на репродуктивное, является крайне важным этапом в жизни растений, а для однолетников, таких, как *A. thaliana*, в большой степени определяет репродуктивный успех. Своевременность зацветания контролируется сложной сетью генетической регуляции, сопрягающей ряд внешних и внутренних стимулов. Основа этой сети подробно охарактеризована с помощью изучения мутантов и трансгенных линий с нарушением зацветания, однако к настоящему времени существует ограниченное число работ, посвященных полнотранскриптомной динамике в ходе переключения развития растения с вегетативного на репродуктивное. В обзоре литературы Главы 2 мы рассматриваем генетический контроль зацветания *A. thaliana*; основное внимание уделяется фотопериодическому (зависящему от долготы светового дня) пути инициации цветения. В листьях рецепторные белки воспринимают свет разных длин волн, далее сигнал передается в АМ с помощью транспортируемого белка FLOWERING LOCUS T (FT), где он активирует экспрессию основных регуляторов зацветания, генов *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*, *LEAFY (LFY)* и *APETALA1 (API)*. Кроме того, в обзоре кратко описаны другие пути инициации цветения – вернализационный, автономный, температурный и гормональный.

В составе транскриптомной карты пулы из 15 АМ собирались ежедневно с 7 по 16 день после прорастания растений (далее – стадии М1-М10). Для снижения биологической вариабельности растений и синхронизации стадии их развития мы разработали систему морфологических маркеров, основанную на количественных характеристиках розеточных листьев и примордиев. 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 (FLOWERING LOCUS C)*, инактивирующийся при зацветании; их профили экспрессии соответствуют ожидаемому (Рисунок 5), что свидетельствует об адекватности собранной нами временной серии.

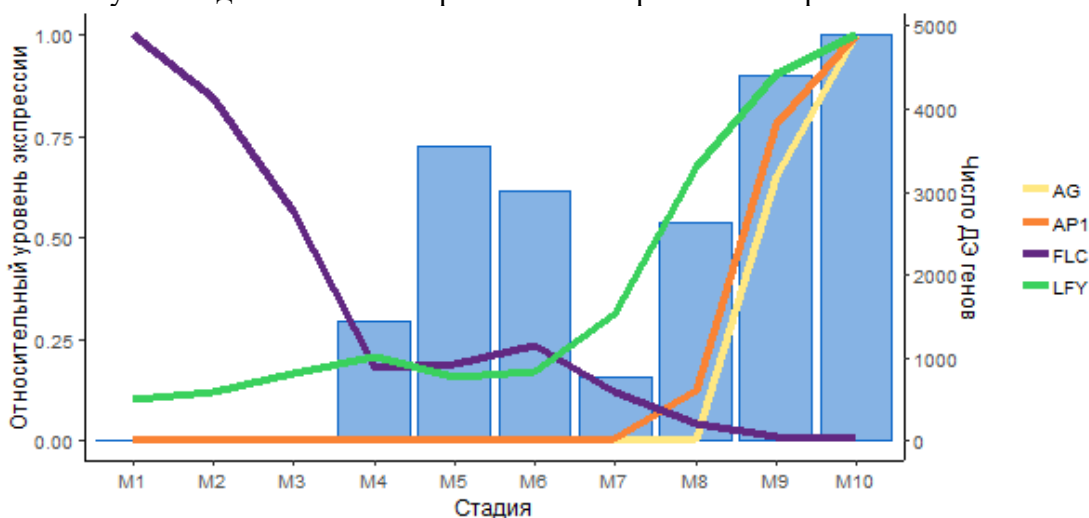


Рисунок 5. Относительный уровень экспрессии основных генов, регулирующих переход к цветению во временной серии АМ.

Описание морфологии растений и поведение ключевых генов, контролирующих переход к цветению, позволило разделить временную серию на четыре этапа. Стадии М1-М4 представляют собой вегетативное развитие, когда периферическая зона АМ закладывает примордии листьев, экспрессия *FLC* уменьшается в пять раз, тогда как *LFY* неизменно экспрессируется на низком уровне. Количество транскриптов гена *SOC1*, одного из самых ранних маркеров начинающегося перехода к цветению пятикратно возрастает на стадиях М1-М4 (при этом рост в три раза приходится на стадии М3-М4). На стадии перехода - М4-М6 - относительные уровни экспрессии генов *FLC* и *LFY* находятся на низком уровне и равны друг другу (Рисунок 5), при этом экспрессия *SOC1* уменьшается в два раза, образуя локальный пик на стадии М4. Между стадиями М5 и М6 не происходит формирования новых органов, что позволяет сделать предположение о происходящем в этот момент перепрограммировании меристемы из вегетативного состояния в репродуктивное. На промежутке от М6 до М8 - активация цветения (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 5). Так же продолжается рост экспрессии гена *SOC1* (в 1,8 раз). Начиная со стадии М8, экспрессия гена *API* возрастает с нулевого уровня (Рисунок 5). Со стадии М8 - инициации и развития примордиев цветков - начинается рост и развитие цветков. Экспрессия гена *FLC* уменьшается до нуля, а *LFY* – в небольшой степени увеличивается, активируется экспрессия гомеозисных генов, определяющих тип органов цветка (*APETALA3*, *PISTILLATA* и *AGAMOUS*) (Рисунок 5).

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 5). 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 5). Этот рост дифференциальной экспрессии совпадает с определенной выше стадией перехода. 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 его движением, модификацией хроматина и репликацией ДНК. Обогащенные категории также совпадали для списков генов, транскрипция которых возрастала в сравнении М5 с М4, и списков генов с уменьшающейся экспрессией на стадии М6 по сравнению с М5; эти категории относились к сборке нуклеосом, тубулинам и биогенезу и структуре рибосом. Кроме того, гены, характеризующиеся ростом экспрессии в М5 по отношению к М4, чаще участвовали в метаболизме нуклеотидов.

Другим способом поиска генов, обладающих характерными профилями экспрессии, является кластерный анализ. Для повышения точности кластеризации нами был применен двухстадийный метод, с кластеризацией профилей экспрессии генов с помощью метода *k*-средних в тысяче повторностей, построении матрицы расстояния между генами на основании числа раз, когда два гена входили в один кластер, и использования этой матрицы для иерархической кластеризации. Затем двухстадийный метод был применен второй раз для кластеризации усредненных профилей экспрессии кластеров и получения суперкластеров. В итоге, 16 615 генов (77% от числа экспрессирующихся в АМ генов) вошли в состав 257 суперкластеров.

Экспрессия двенадцати суперкластеров была равномерной на всех стадиях временной серии, кроме М5, где она резко возрастала и достигала максимального уровня (Рисунок 6а). Гены, входящие в эти суперкластеры, чаще ожидаемого описывались категориями ГО, связанными с биогенезом рибосом, гистонами, нуклеосомами, модификацией хроматина и

сплайсингом мРНК. Суперкластеры с противоположным профилем экспрессии (то есть, с уменьшением уровня транскрипции на стадии М5 при равномерной экспрессии в остальных образцах, Рисунок 6б) были обогащены категориями, относящимися к АТФазам и связыванию нуклеотидов, хеликазам и пентатрикопептидным повторам. Еще один характерный профиль транскрипции обладал снижением на стадии М4 и ростом на М5 (Рисунок 6в); 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 6г), обогащены категориями, описывающими цитоскелет и движение цитоскелета (микротрубочки, кинезины).

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 глубже изучить этот вопрос, мы детально рассмотрели динамику экспрессии ряда функциональных групп генов. В их число вошли гены, кодирующие элементы цитоскелета, 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, схожее поведение демонстрируют гены, кодирующие факторы деполимеризации актина (ADF) и регулирующие динамическую активность цитоскелета, а также миозины класса XI (белки, обеспечивающие быструю динамику аппарата Гольджи, митохондрий, пероксисом и пластид). Профили экспрессии генов, кодирующих α - и β -тубулины имеют локальный пик на М5, в то время как почти все гены кинезинов имеют сниженную экспрессию на этой стадии. Двадцать восемь из тридцати генов *A. thaliana*, кодирующих гистоны, имеют максимальную экспрессию на стадии М5. Динамическое состояние хроматина поддерживается в том числе с помощью белков-модификаторов гистонов, к которым относятся ацетил-трансферазы, деметилазы и метил-трансферазы. Уровень транскрипции большинства генов, кодирующих эти ферменты, уменьшается на стадии М5.

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

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

Точная регуляция транскрипции генов гистонов происходит в течении всего клеточного цикла, при этом количество мРНК всех гистонов синхронно возрастает в клетках, входящих в S-фазу и быстро падает после окончания S-фазы, что позволяет использовать относительное количество мРНК гистонов и как маркер S-фазы, и как индикатор пролиферативного индекса ткани во всех организмах. Во временной серии АМ почти все гены гистонов и гистон-подобных белков обладают паттерном типа (I). Кроме того, гены, в культуре клеток возрастающих в поздней G₁- и S-фазах и кодирующие компоненты комплекса репликации ДНК, демонстрируют паттерн типа (I): гены кодирующих ядерный антиген пролиферирующих клеток (*PCNA1* и *PCNA2*), *RPA2/RPA32A* и два *RPA3*-подобных гена, кодирующие субъединицы репликационного белка А,

связывающего одноцепочечную ДНК при репликации, ген *RFC2*, кодирующий фактор репликации С2, и гомолог гена контрольной точки перехода в S-фазу *CDC25*, кодирующий фосфатазу.

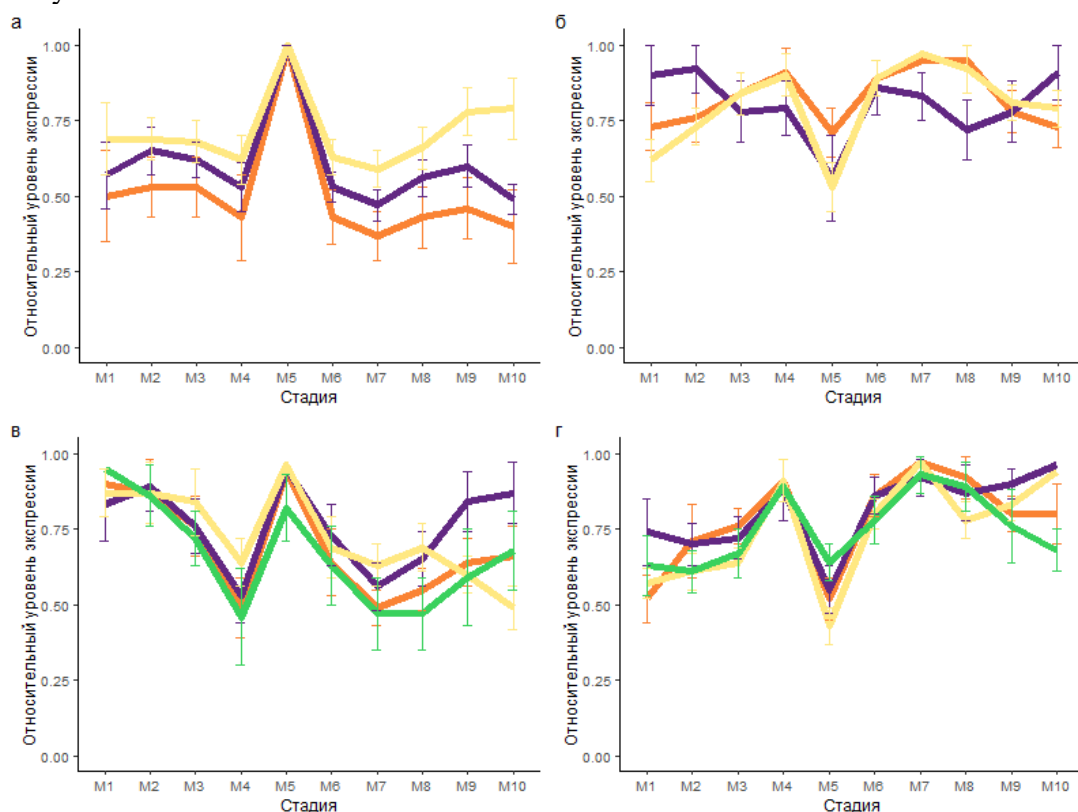


Рисунок 6. Усредненные профили экспрессии кластеров с (а) повышением относительной экспрессии на стадии М5; (б) понижением относительной экспрессии на стадии М5; (в) понижением относительной экспрессии на стадии М4 и дальнейшим ростом на стадии М5; (г) повышением относительной экспрессии на стадии М4 и дальнейшим падением на стадии М5.

Важным фактом является то, что паттерн, аналогичный генам S-фазы (пик экспрессии на стадии М5), наблюдается у генов, обычно ассоциируемых с поздней G₂-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 (М1-М3 и М8-М10). Помимо этого, профиль экспрессии типа (I) на стадии М4-М6 характерен для циклин-зависимых киназ (CDK) А- и В-типа.

Полученные результаты можно описать моделью, согласно которой в АМ на стадии М5 происходит рост числа клеточных делений, сопровождающийся укорочением фаз G₁ и G₂ (Рисунок 7). Об этом также свидетельствует обогащение ДЭ генов на стадии М5, 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.

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. В суспензионных культурах клеток основная часть митотических циклинов А- и В-типа накапливается в поздней G₂ с пиком в раннем митозе. Исключение составляют гены *CYCA3;1*, *CYCA3;2* и *CYCA3;3*, которые в основном экспрессируются на границе G₁/S, то есть кодируют G₁-циклины. Экспрессия семейств генов *CYCA1*, *CYCB1* и *CYCB2* регулируется общим молекулярным механизмом, включающим связывание MSA-элементов в промоторах этих генов транскрипционными факторами MYB3R1 и MYB3R4. 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. 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. Таким образом, экспрессия большинства циклинов А- и В-типа соответствует паттерну (II).

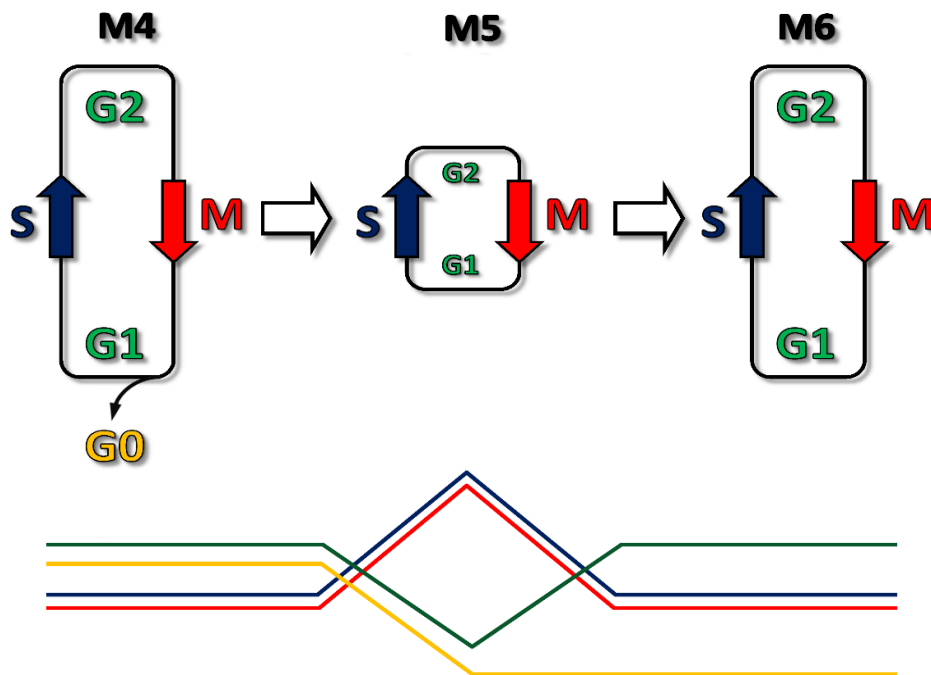


Рисунок 7. Модель, объясняющая изменения в уровнях экспрессии генов, контролирующих клеточный цикл, при переходе к цветению. На стадии М4 клеточный цикл имеет низкую скорость из-за длинных фаз G₁ и G₂; кроме того, значительная доля клеток АМ находятся вне клеточного цикла в фазе G₀. В момент перехода (М4-М5) клеточный цикл сокращается за счет уменьшения длительности фаз G₁ и G₂ и многие клетки входят в клеточный цикл из фазы G₀. 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, G₁ и G₂ – зеленым, G₀ – желтым).

Большинство генов, кодирующих компоненты пути E2F/DP/RB, который регулирует G₁/S-фазу имеют паттерн типа (II), 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. Более того, некоторые гены, экспрессирующиеся на одной стадии клеточного цикла в культуре клеток, имели разные паттерны на стадиях М4-М5 временной серии. Например, для тубулинов и кинезинов ранее было показано одновременное увеличение экспрессии в G₂/M-переходе. Однако согласно нашим данным об экспрессии в АМ, тубулины имеют профиль типа (I), а кинезины - типа (II). 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* – профилю типа (II).

Мы предполагаем, что противоположное направление экспрессии генов, кодирующих регуляторы клеточного цикла, на стадии М4 может объясняться транскрипционной и физиологической реорганизацией в клетках, которые на этой стадии готовятся войти интенсивную пролиферацию на М5. На стадии М4 часть клеток вегетативной АМ, которая находится в длинном клеточном цикле в протяженной G₂-фазе, начинает синхронно двигаться к митозу, тогда как клетки, задерживающиеся в G₁-фазе, готовятся войти в S-фазу. На стадии М5 клетки быстро делятся без долгих G₁- и G₂-фаз, а на М6 и дальнейших стадиях пролиферация возвращается на нормальный уровень. Подготовительные изменения в ходе клеточного цикла в образце М4 могут определять разнообразные паттерны экспрессии генов-регуляторов на этой стадии. Кроме того, в М5 некоторые неделящиеся клетки (в фазе G₀), присутствующие в АМ, могут также входить в

клеточный цикл, что предполагает анализ профилей экспрессии некоторых циклинов D-типа.

При анализе транскриптомной карты *A. thaliana* мы обнаружили, что во многих попарных сравнениях перепредставлены гены, аннотированные термином «ответ на стресс» (GO:0006950). Экспрессия генов стрессовой реакции активировалась 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. В транскриптомной карте *A. thaliana* профили экспрессии генов, аннотированных в Генной Онтологии как участвующие в ответе на стресс, чрезвычайно разнообразны, и многие из них отличаются высокой тканевой специфичностью. Это позволяет предположить, что в разных органах и тканях *A. thaliana* развитие стрессовой реакции может отличаться. В **Главе 3** мы рассматриваем транскриптомные изменения, происходящие в разных органах *A. thaliana* под воздействием пониженной температуры. Обзор литературы посвящен рассмотрению регуляции ответа на холод: механизмов восприятия пониженной температуры рецепторными белками, передачи полученного сигнала в ядро, транскрипционной регуляции холодового ответа (с подробным описанием ТФ семейства *CBF* (*C-REPEAT BINDING FACTOR*; *CBF1*, *CBF2*, *CBF3*), физиологической реакции растения на холод и ее интеграции с процессами роста и развития. Подробная сеть генетической регуляции холодовой акклиматизации, включающая в себя как активацию/репрессию транскрипции генов, так и многочисленные посттрансляционные модификации белков, была создана с помощью изучения воздействия холода на трансгенные линии *A. thaliana*, сверхэкспрессирующие *CBF*, и линии с инактивацией *CBF*. Хотя транскриптомный ответ *A. thaliana* на пониженную температуру неоднократно исследовался, все работы были проведены на фотосинтезирующих частях растения (розетке, листьях, семядолях), и, таким образом универсальность холодовой реакции во всем растении остается неизученной.

Мы использовали транскриптомную карту *A. thaliana*, чтобы выбрать наиболее транскрипционно разнообразные образцы: семядоли и гипокотиль однодневного проростка, большой зеленый лист, первый цветок на стадии распускания, цветок на стадии 9 (далее называется «молодой цветок») и зрелые зеленые семена. 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.

Для выявления генов, отвечающих на холодое 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 различалось от 1 022 в Цветке-3 до 6 867 в Листе-27; всего 15 459 генов были ДЭ хотя бы в одном сравнении из 12. Большинство (75%) генов, входящих в регулон ТФ *CBF1-3*, были в ДЭ в образцах Листа, Семядолей и Гипокотилия, тогда как в цветках обеих стадий и семенах доля ДЭ генов из регулона различалась от 35 до 59%.

В дальнейших анализах образцы после 3 и 27 часов воздействия пониженной температуры рассматривались отдельно. Чтобы оценить изменение профиля экспрессии гена интегрально по всем образцам, 15 459 ДЭ генов были разделены на три категории: гены, экспрессия которых значимо 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 каждой из категорий число генов, являющихся ДЭ во всех образцах, было крайне невелико, но несколько возросло для образцов, собранных после 27 часов гипотермии

(Рисунок 8). В категории Up как на 3, так и на 27 часах, наблюдалось значительное обогащение терминами, связанными со стрессовым ответом. Гены, относящиеся к группе Down-27 и ДЭ во всех образцах, были существенно обогащены терминами, связанными с фотосинтетическими процессами.

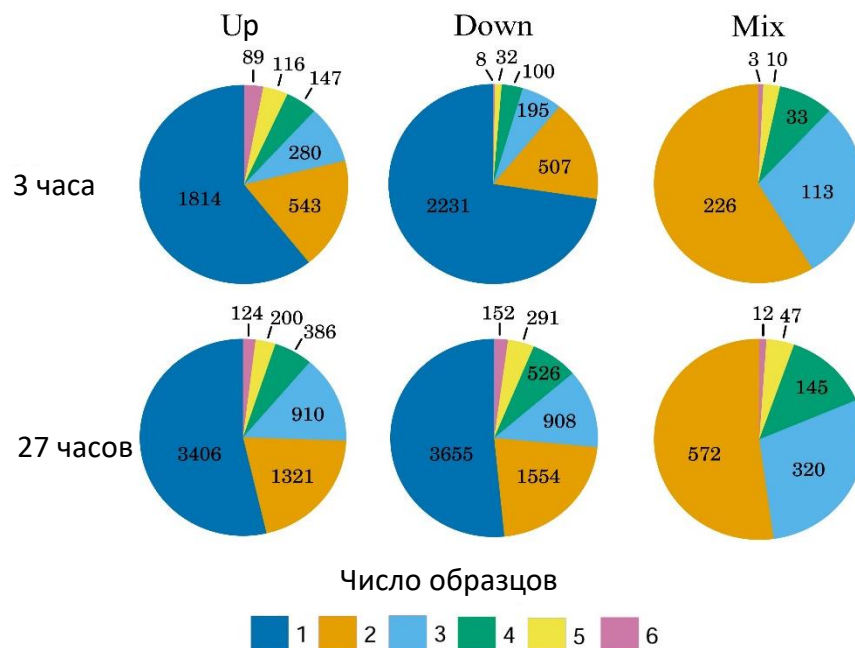


Рисунок 8. Число генов, дифференциально экспрессирующихся в разном числе образцов для трех и 27 часов воздействия холода отдельно.

Число генов, изменяющих экспрессию в процессе холодового 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. 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 и увеличенной в Семядолях-27, были обогащены терминами, связанными с пентатрикопептидными повторами. В образце Листа-27 гены со снижением экспрессии были обогащены категориями, относящимися к фотосинтезу. 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. Гены с повышенной экспрессией в Молодых Цветках-27 характеризовались перепредставленностью категорий, связанных с катаболическими процессами и развитием пыльцы.

Чтобы проверить, известно ли участие найденных нами органоспецифических генов в ответе на холод, мы рассмотрели аннотацию ГО этих генов: только небольшая доля (от 8 до 12%) ДЭ генов была аннотирована категорией GO:0006950 (ответ на стресс) и дочерними категориями. Кроме того, мы рассмотрели поведение так называемых генов COR («отвечающих на действие холода», cold-response), транскрипция которых меняется, во-первых, в трансгенных растениях со сверхэкспрессией одного из генов *CBF*, а во-вторых, при воздействии пониженной температуры. Процент ДЭ генов среди COR-генов был существенно выше для фотосинтезирующих органов (79-85%), чем для цветков и семян (36-51%).

Помимо генов *CBF1-3* 27 ТФ активируются в течении нескольких минут после понижения температуры и служат ранними регуляторами стрессового ответа. Четыре из тридцати ТФ были ДЭ во всех образцах после трех часов при +4°C (включая *CBF3*), остальные же имели заметно различающиеся паттерны экспрессионных изменений.

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.

Для дальнейшего изучения органоспецифических генов стрессового ответа мы охарактеризовали их с помощью параметров экспрессии, основанных на транскриптомной карте *A. thaliana*. Первым из параметров, который мы использовали, была энтропия Шеннона, отражающая ширину паттерна экспрессии. Распределение энтропии для 15 459 генов, меняющих экспрессию в ответ на стресс, практически совпадает с распределением для всех экспрессирующихся генов (Рисунок 9 а и б). Гены, экспрессия которых изменялась при гипотермии хотя бы в пяти органах, не образовывали пика распределения на низких значениях H (Рисунок 9 в и г), а уникальные для образцов генов в большинстве случаев имели сходное с общим распределение энтропии Шеннона, за исключением нескольких образцов. Распределение энтропии генов, имеющих увеличенную экспрессию в образцах Лист-3 и Цветок-27, обладает локальным пиком на низких (0-0,3) значениях энтропии (Рисунок 9 д и ж); в транскриптомной карте *A. thaliana* эти гены экспрессируются исключительно в зрелых пыльниках и целых цветках, содержащими пыльники на той же стадии. В Семенах-3 и -27 распределение энтропии Шеннона обладает пиком в области средних значений (Рисунок 9 е и з), образующие его гены со сниженной экспрессией ассоциированы с терминами, связанными с доменом F-box и с защитой растений, тогда как гены, экспрессия которых возрастает, ассоциированы с метаболизмом и запасанием липидов.

Еще один параметр экспрессии, использовавшийся в описании транскриптомной картой *A. thaliana*, – ДЭ-значение. Распределение ДЭ-значений для всех 15 459 стрессовых генов по сравнению с общим распределением обладает меньшим числом генов с низким ДЭ-значением. ДЭ гены, общие хотя бы для пяти органов, имеют распределение ДЭ-значений, смещенное к правой части, тогда как у органоспецифических ДЭ генов оно немного смещено влево. Большинство образцов не имеет заметных особенностей распределения, 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. Анализ обогащения категориями ГО связывает эти гены с убиквитин-зависимыми процессами катаболизма белков, а также с регуляцией транскрипции и доменом F-box.

Для более глубокого понимания структуры генетических сетей, управляющих холодовой акклиматизацией в разных органах и частях растения, мы проанализировали представленность регуляторных элементов различных ТФ в промоторах разных групп ДЭ генов. Как и ожидалось, в промоторных регионах ДЭ генов, экспрессия которых возрастает хотя бы в пяти органах чаще, чем в среднем по геному, встречались регуляторные элементы ТФ СВФ. В списках генов, ДЭ только в одном образце, перепредставленности регуляторных элементов СВФ 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.

В Обсуждении Главы 3 рассматриваются основные выводы, сделанные на основании полученных результатов. Рассмотрение единства и уникальности развития устойчивости к низким температурам у растения требует дифференцированного анализа транскриптомного ответа на холод в разных органах. 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). Всего нами было выявлено 15 459 генов, ДЭ хотя бы в одном сравнении из 12, что составляет 46% от всех генов *A. thaliana*; такая большая доля показывает существенные изменения, происходящие при адаптации к гипотермии во всем растении.

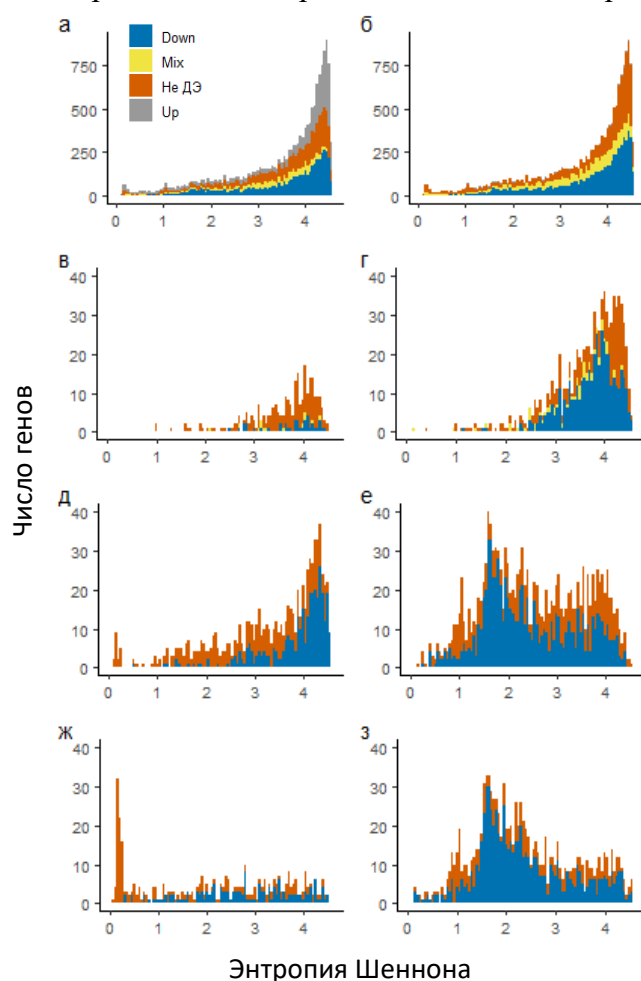


Рисунок 9. Распределение энтропии Шеннона среди всех генов, отвечающих на холодное воздействие, с указанием типа изменения экспрессии после (а) трех и (б) 27 часов при +4°C; среди генов, ДЭ в пяти или шести образцах при (в) трех и (г) 27 часах действия холода; среди генов, специфически участвующих в холодной акклиматизации в (д) Листе-3, (е) Семенах-3, (ж) Цветке-27 и (з) Семенах-27.

Для проверки, насколько результаты нашего эксперимента совпадают с литературными данными, мы использовали результаты секвенирования РНК листьев трансгенных растений со сверхэкспрессией одного из генов *CBF*. Транскриптомы зеленых частей растения – листа, семядолей и гипокотыля – демонстрировали согласованные изменения в трансгенных линиях и в наших данных, 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.

Мы разделили гены на общие (ДЭ во всех или хотя бы в пяти образцах) и тканеспецифичные, экспрессия которых изменялась только в одном из образцов. 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. Другая сторона ответа на холод, которую мы изучали, – это специфический для органов ответ. 62% генов были ДЭ только в одном образце после трех часов гипотермии и 47% – после 27 часов. Таким образом, на поздних стадиях развития холодовой акклиматизации транскриптомный ответ на стресс в заметной степени унифицируется. Органоспецифические гены существенно различались по своим функциям между образцами и были связаны с разными процессами. В зеленых семядолях развитие холодовой реакции в существенно интегрировано с универсальным ответом на абиотический стресс; специфика транскриптомных изменений в Листе-27 отражает важность торможения фотосинтеза и уменьшения вреда, наносимого накапливающимися активными формами кислорода при действии абиотического стресса. Как в Семенах-3, так и в Семенах-27 ДЭ гены, число транскриптов которых возросло, были связаны с запасанием липидов. Изменения в составе клеточных жиров, ведущие к стабилизации мембраны, являются известной частью механизма холодовой устойчивости и могут быть особенно важны в этом случае, так как накопление липидов в развивающемся семени является необходимым для репродукции растения процессом. 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.

Сравнение полученных нами данных с литературными хорошо демонстрирует неравномерную изученность холодового ответа в разных органах *A. thaliana*. Хорошо изученные гены, отвечающие на холод и входящие в регулон ТФ СВF1-3, дифференциально экспрессировались в основном в зеленых частях растения (78% ДЭ генов по сравнению с 39%). 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 СВF. Для дальнейшего изучения этого вопроса мы рассмотрели профили экспрессии так называемых генов «первой волны» – 27 ТФ, активирующихся вместе с СВF в течении нескольких минут после понижения температуры. В рассматриваемых нами органах эти транскрипционные регуляторы продемонстрировали разнообразное поведение после трех часов холодового воздействия. Таким образом, степень участия основных регуляторов раннего холодового ответа различается в органах растения, что подтверждает необходимость дифференциального исследования акклиматизации к низкой температуре.

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. Энтропия Шеннона (H) позволяет численно оценить ширину паттерна экспрессии гена; гены «ядра» стрессового ответа не имели локального пика распределения на низких значениях H, что показывает универсальность экспрессии этих генов в нормальных условиях (Рисунок 9 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, однако некоторые образцы имели уникальные особенности. Среди них следует отметить Лист-3, для которого был характерен локальный пик на низких значениях энтропии (Рисунок 9 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, влияющие на жесткость клеточной стенки и устойчивость к дегидратации. Таким образом, реализация специфической для листа программы защиты от

холода требует рекрутирования специфических для пыльцы генов. Аналогичный пик на низких значениях H в Цветке-27 также образовывали пыльцеспецифичные гены (Рисунок 9 ж), вовлеченные в регуляцию роста пыльцевой трубки и модификации клеточной стенки.

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. Из всех образцов только гены, характерные для Молодого Цветка-27, образовывали выраженный пик в области ДЭ-значений меньше 500. Большинство генов с низким ДЭ-значением, дифференциально экспрессирующиеся в Молодом Цветке-27, в транскриптомной карте, охватывающей нормальное развитие растения, были специфичны для молодых пыльников и для молодых цветков на той же стадии. Анализ обогащения этих генов категориями ГО показал значимую перепредставленность терминов, связанных с убиквитин-зависимыми процессами катаболизма белков; участие в ответе на абиотические стрессы было продемонстрировано для некоторых генов, кодирующих компоненты системы убиквитиновой 26S-протеасомы.

Изучение представленности регуляторных элементов различных ТФ в промоторах генов позволяет находить ключевые регуляторы общей и органоспецифической холодовой акклиматизации. Базовые гены стрессового ответа (ДЭ в пяти или шести органах) главным образом регулируются ТФ семейства CBF. В число активируемых ими генов входят транскрипционные факторы из семейств ERF/AP2, TCP, NAC, WRKY, Myb, а также содержащие домен «цинковые пальцы В-бок-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 холодовой 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 холодового воздействия основную регуляцию выполняют транскрипционные факторы семейства NAC.

В Главе 4 мы использовали подробную транскриптомную карту *A. thaliana* для планирования атласа экспрессии генов *Capsella bursa-pastoris*, также относящемуся к семейству Крестоцветных. *C. bursa-pastoris* – распространенный по всему миру недавний аллотетраплоид, служащий удобной моделью ранних последствий полиплоидизации. Сравнение экспрессии гомеологичных генов в транскриптомной карте может позволить изучить процессы функциональной дивергенции гомеологов. В обзоре литературы Главы 4 мы рассматриваем последствия полиплоидизации - основного пути эволюции растительных геномов: процессы перестройки генома и транскриптома, в результате которых гены могут потерять 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. 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; это явление называется геномным доминированием и продемонстрировано для разнообразных растений, включая Злаки и представителей семейства Brassicaceae. Поиск геномного доминирования и несимметричной экспрессии гомеологичных генов *C. bursa-pastoris* проводился неоднократно, однако в разных исследованиях были получены противоречивые результаты. 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.

Созданный нами атлас экспрессии генов *C. bursa-pastoris* включает в себя десять наиболее транскрипционно различающихся образцов: корень, пластинка и жилка молодого листа, апикальная меристема, органы цветка, семена на двух стадиях, а также дополнительный набор данных: однодневных проростков, подвергнутых холодовому воздействию (+4°C) в течении 3, 15 и 27 часов, а также соответствующий контроль.

C. bursa-pastoris – модельный объект ранней эволюции полиплоидов, 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*). 45% генов экспрессировались во всех образцах и 89% – хотя бы в одном образце, а разница между образцами составляла 56-73% экспрессирующихся генов. Во всех образцах число транскрибирующихся генов из субгеномов А и В было практически одинаковым, и более чем в 95% случаев транскрипты обоих гомеологичных генов детектировались в каждом из образцов. Большинство генов экспрессировались во всех или почти во всех (больше 11) образцах (Рисунок 10). Второй пик образуют гены с узким паттерном экспрессии, транскрипты которых были найдены только в одном или двух образцах (Рисунок 10). Гомеологи А и В одинаково распределялись по числу образцов, в которых они экспрессировались (Рисунок 10).

Чтобы оценить транскриптомное разнообразие образцов, мы проанализировали дифференциальную экспрессию между всеми парами образцов основной транскриптомной карты. Число ДЭ генов варьировалось от 2 311 до 18 272. У 89% 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, однако число пар, в которых оба гомеолога были либо одновременно ДЭ, либо не ДЭ, было низким (679 пар). 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), причем ортологи генов, отвечающих на холод у *A. thaliana* (гены COR) в 60% случаев имели измененную экспрессию при холодовом воздействии у *C. bursa-pastoris*.

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. 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). Число ДЭАВ пар заметно различалось между образцами: от 418 в зрелых семенах (SD) до 2 880 в чашелистиках (SP). Аналогично ДЭ-значению, 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. 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 11).

Среди генов *C. bursa-pastoris*, дифференциально экспрессирующихся в ответ на холодовое воздействие, 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.

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 возникшая после полиплоидизации функциональная избыточность со временем ведут к суб- и неофункционализации, либо утрате одного из гомеологов. У

молодых аллополиплоидов, таких как *C. bursa-pastoris*, с помощью транскриптомных данных можно оценить степень функциональной дивергенции гомеологов.

Общие транскриптомные характеристики *C. bursa-pastoris* (по всем генам и отдельно по субгеномам) находились в пределах значений, полученных для других видов растений. 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. Между субгеномами А и В не было найдено разницы по распределению ДЭ-значения, то есть степень вклада субгеномов в разнообразие транскриптома одинакова. Однако если рассматривать отдельные гомеологичные пары, то во многих из них гомеологи А и В являлись ДЭ в разном наборе образцов. 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. 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.

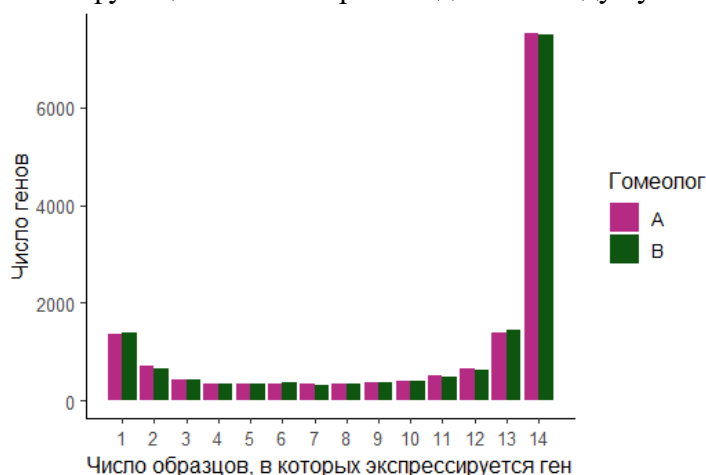


Рисунок 10. Распределение генов по числу образцов, в которых ген экспрессируется. Гомеологи из субгеномов А и В приведены отдельно.

Для дальнейшего изучения различий в экспрессии мы рассмотрели дифференциальную экспрессию внутри пар гомеологов. Количество гомеологов В, экспрессия которых возрастала или снижалась по сравнению с гомеологом А, было одинаковым как в каждом образце, так и в целом (Рисунок 11), и распределение их ДЭАВ-значений также не отличалось. 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 холодовой 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; 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. В более редких случаях транскрипты обоих гомеологов обнаруживались в разном количестве в разных органах, что может говорить о субфункционализации или неофункционализации.

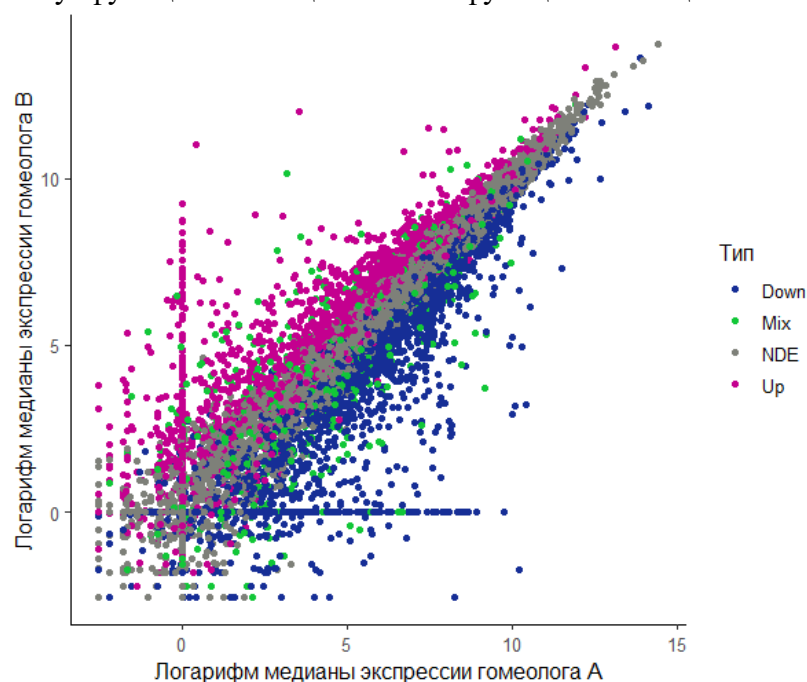


Рисунок 11. Корреляция экспрессии гомеологичных генов. Цветом отмечено направление изменения экспрессии при сравнении гомеолога В и гомеолога А: синий, Down – гомеолог В имеет сниженную экспрессию по сравнению с гомеологом А хотя бы в одном образце и нигде не имеет повышенной; маджента, Up – гомеолог 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.

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.

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 G₁ и G₂;
- 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* – аллотерераплоидного 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 публикаций по теме диссертации

По теме диссертации опубликовано пять статей в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (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.