

Клепикова Анна Владимировна

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

03.02.07 – генетика

03.01.09 – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Новосибирск – 2020

Работа выполнена в лаборатории геномики растений, ФГБУН Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, г. Москва

Научный руководитель:

Пенин Алексей Александрович
кандидат биологических наук, зав. лабораторией, ФГБУН Институт проблем передачи информации имени А.А. Харкевича Российской академии наук, г. Москва

Официальные оппоненты:

Голденкова-Павлова Ирина Васильевна

доктор биологических наук, ведущий научный сотрудник, доцент, ФГБУН «Институт физиологии растений им. К.А. Тимирязева Российской академии наук», г. Москва

Землянская Елена Васильевна

кандидат биологических наук, старший научный сотрудник, ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН», г. Новосибирск

Ведущая организация:

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

Защита диссертации состоится « 2 » декабря 2020 г. на утреннем заседании диссертационного совета Д 003.011.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН» в конференц-зале Института по адресу: 630090, г. Новосибирск, пр. ак. Лаврентьева 10. Тел. +7 (383) 3634906, факс +7(383) 3331278. e-mail: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института: www.bionet.nsc.ru.

Автореферат разослан « ___ » _____ 2020 г.

Ученый секретарь
диссертационного совета
доктор биологических наук



Т.М. Хлебодарова

Общая характеристика работы

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 – GO (Gene Ontology). 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.

Цели и задачи. 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*.

Научная новизна и практическая значимость. 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.

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-provart/>) 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) 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.

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 G₁ и G₂.
3. Развитие холодового стресса 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.

Личный вклад автора в исследование. Все результаты, включенные в диссертацию, получены лично соискателем или при его непосредственном участии. Создание и секвенирование библиотек и первичная обработка чтений (картирование на референсный геном) было выполнено соавторами публикации в процессе совместной работы. Автором были подготовлены данные для размещения в базе TraVA. 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; 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, созданных в период с 2005 по 2018 год. Представленность среди них таксонов растений неравномерна: 34 атласа созданы для покрытосеменных растений, а однодольные представлены исключительно Злаками. Разрешение атласов отличается в отношении как числа изученных образцов, так и объема полученных данных, однако общие характеристики транскриптома (число экспрессирующихся генов, профили экспрессии образцов, кластеризация органов и тканей) были схожи у всех растений.

В разделе результатов Главы 1 описана структура созданной 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).

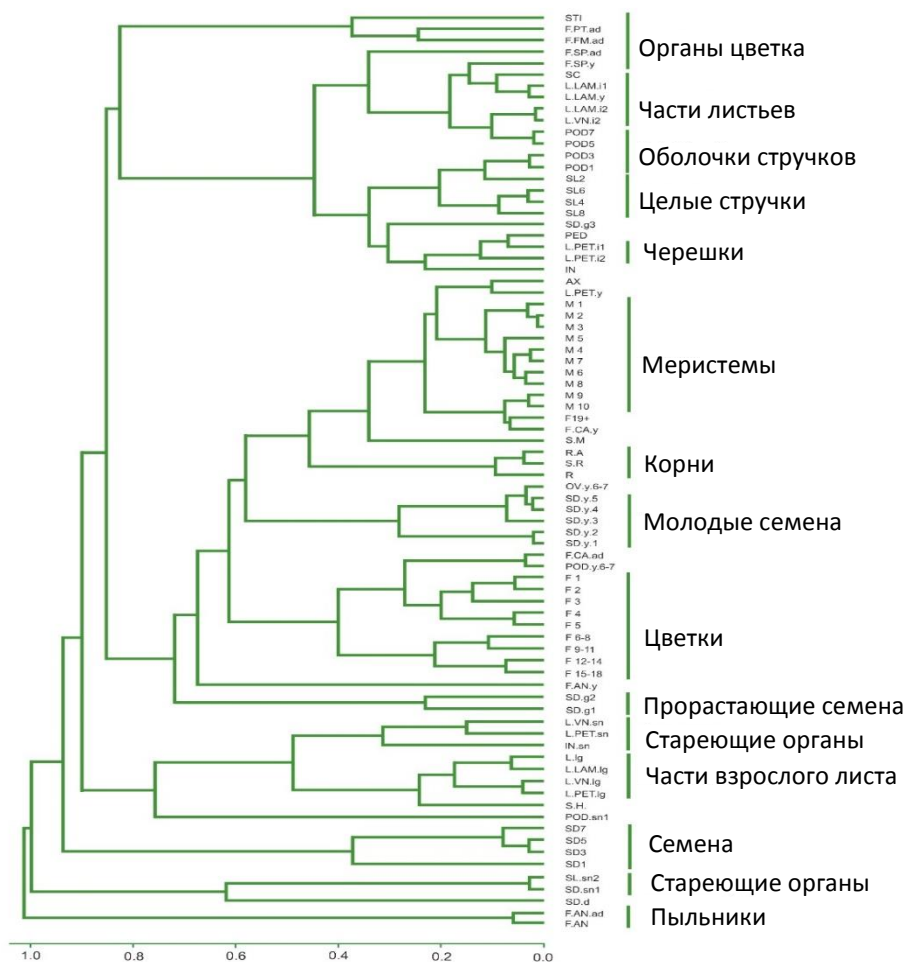


Рисунок 1. Дерево кластеризации образцов транскриптомной карты. В качестве меры расстояния между парой образцов (ось X) использовался вычтенный из единицы квадрат корреляции Пирсона.

Секвенировав собранные образцы, мы обнаружили экспрессию 24 621 из 27 201 белок-кодирующих генов *A. thaliana*; во всех 79 образцах экспрессировалось 40% генов, а число генов, экспрессирующихся в каждом из образцов, было схожим. Также были схожи профили экспрессии образцов, описанные с помощью параметра Z (Z-score). При добавлении набора образцов, подвергнутых воздействию абиотических стрессов, обнаружили транскрипты еще 26 генов. 2 580 генов не экспрессировалось ни в транскриптомной карте, ни в стрессовых образцах.

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. Наиболее часто были ДЭ гены, участвующие в фотосинтетических процессах и регуляции хроматина и деления клеток. Для описания разнообразия экспрессии генов мы ввели интегральную меру «ДЭ-значение гена» – число парных сравнений, в которых данный ген являлся ДЭ. ДЭ-значение отражает, во-первых, ширину паттерна экспрессии и, во-вторых, статистически достоверную разницу в уровнях экспрессии в этих образцах. ДЭ-значение позволяет определить гены, которые вовлечены в орган-специфические либо базовые процессы организма. Распределение ДЭ-значений представлено на Рисунке 2 а. Гены, обладающие наибольшим ДЭ-значением, участвовали в фотосинтезе и связанном с ним метаболизме, а гены с наименьшим – в процессинге, локализации и транспорте РНК.

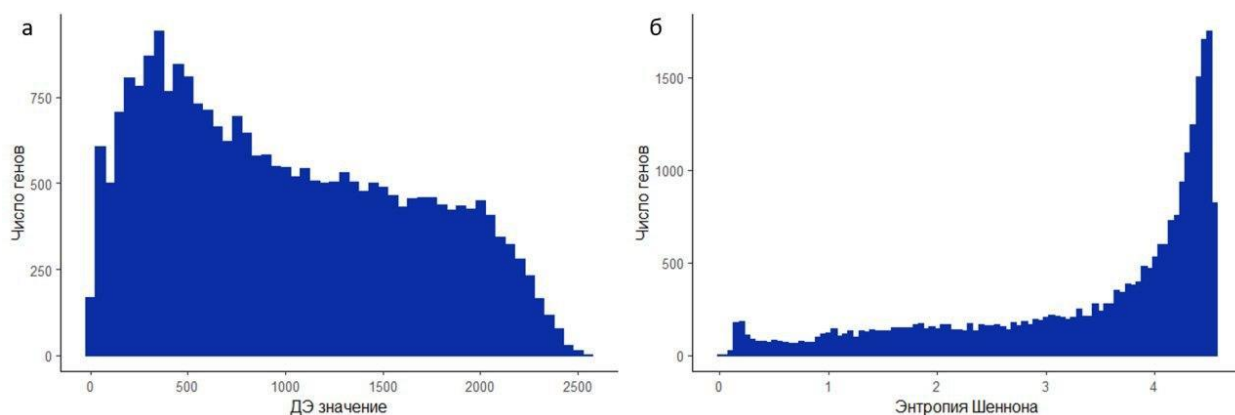


Рисунок 2. (а) Распределение ДЭ-значений генов, дифференциально экспрессирующихся хотя бы в одном сравнении образцов. (б) Распределение энтропии Шеннона генов.

Для поиска стабильно экспрессирующихся генов в качестве меры равномерности экспрессии мы использовали коэффициент вариации (CV). У 970 генов CV был меньше 0,3; эти гены были связаны с транспортом нуклеиновых кислот и белков, мембранами и процессингом РНК. При добавлении образцов, подвергнутых воздействию абиотических стрессов, 792 гена имело $CV \leq 0,3$. Наиболее стабильные гены имели CV от 0,15 до 0,2, а ДЭ-значение – от 0 до 117, тогда как энтропия Шеннона (см. ниже) варьировалась от 4,56 до 4,57.

Для решения обратной задачи – изучения тканевой специфичности экспрессии генов – мы использовали два подхода. В качестве первичного описания паттерна экспрессии гена мы использовали число образцов, в которых этот ген

транскрибируется; распределение этого значения имеет два пика: транскрипты основной части генов, кодирующих мРНК, 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). Гены, обладающие более широким паттерном экспрессии, имели более высокий уровень экспрессии.

Для более точной оценки ширины паттерна экспрессии была использована энтропия Шеннона (H). Распределение энтропии H смещено в область высоких значений, что отражает широкий паттерн экспрессии большинства генов (Рисунок 2 б). Второй локальный максимум ($H = 0,1-0,25$) образуется органоспецифичными генами. Гены с энтропией, превышающей 4,53, связаны с транспортом нуклеиновых кислот, процессингом РНК и мембранами. Гены, энтропия которых была ниже 0,15, характеризовались категориями «межклеточный сигналинг», «пектинэстеразная активность», «клеточная стенка».

Анализ паттернов экспрессии транскрипционных факторов (ТФ) важен 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 транскриптомной карты. В большинстве семейств ТФ встречались как тканеспецифичные, так и экспрессирующиеся во всех образцах гены. Наиболее узкий паттерн экспрессии имели классы 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.

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 атласа (Рисунок 1) отражает морфологию растения, и образцы преимущественно группируются по морфологическому и физиологическому сходству.

Общая характеристика транскриптома *A. thaliana* была схожа с другими растениями: 77% генов экспрессировались хотя бы в одном образце, 40% – во всех, образцы отличались друг от друга не более чем на 20%. Включение в анализ дополнительного набора образцов, подвергнутых действию абиотических стрессов, практически не увеличило число экспрессирующихся генов, показывая полноту описания транскриптома *A. thaliana* нашими образцами. Большинство образцов не только имело схожее число экспрессирующихся генов, но и близкие профили экспрессии, за несколькими 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. Многие ДЭ гены 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 растущими и зрелыми органами. Также мы оценили разнообразие экспрессии с помощью меры «ДЭ-значение гена» (Рисунок 2 а). Гены, имеющие ДЭ-значение ниже 100, были либо тканеспецифическими, либо стабильно экспрессирующимися, и участвовали в базовых биологических процессах, таких, как процессинг РНК, транспорт и локализация белков.

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 важно для понимания, какие клеточные процессы требуют стабильной экспрессии участвующих в них генов во всем организме. Используя дополнительный набор стрессовых данных совместно с транскриптомной картой, мы отобрали 27 наиболее стабильных генов ($CV \leq 0,2$). Эти гены связаны с самыми разными процессами, включая цветение, транспорт лизосом, конденсацию хроматина и адаптацию к стрессу, и, таким образом, могут служить референсными генами в широком спектре экспериментов по измерению экспрессии генов.

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 2 б) позволил обнаружить гены, обеспечивающие основы жизнедеятельности всех типов клеток и участвующих в таких процессах, как транспорт нуклеиновых кислот, метаболизм липопротеинов и мембран. Распределение H в нашей транскриптомной карте схоже с распределением u кукурузы, что показывает межвидовую универсальность соотношения общих и тканеспецифичных генов. Последние участвуют в клеточном сигналинге, формировании эндомембран и клеточной стенки; кроме того, продукты многих из них имеют пектинэстеразную активность, что соответствует большому числу генов, экспрессирующихся только 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 ранее был показан для некоторых из этих ТФ, участвующих в контроле развития органов цветка, листьев и перехода к цветению. На другом конце распределения находятся семейства SWI/SNF SWI3, SNF2, CAMTA, DDT и FAR, вовлеченных в такие универсальные клеточные процессы, 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.

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 пути инициации цветения.

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 *LEAFY (LFY)*, экспрессия которого возрастает в процессе перехода, а главным репрессором – ген *FLC (FLOWERING LOCUS C)*, инактивирующийся при зацветании; их профили экспрессии соответствуют ожидаемому (Рисунок 3), что свидетельствует об адекватности собранной нами временной серии.

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. На стадии перехода – М4-М6 – относительные уровни экспрессии генов *FLC* и *LFY* находятся на низком уровне и равны друг другу (Рисунок 3). 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 – стадия формирования первых трех примордиев цветков – экспрессия гена *FLC* продолжает уменьшаться, а экспрессия *LFY* возрастает в четыре раза. Начиная со стадии М8, экспрессия гена *APETALA1* (*API*) возрастает с нулевого уровня, активируется экспрессия гомеозисных генов, определяющих тип органов цветка (*AP3*, *PISTILLATA* и *AGAMOUS*) и начинается рост и развитие цветков.

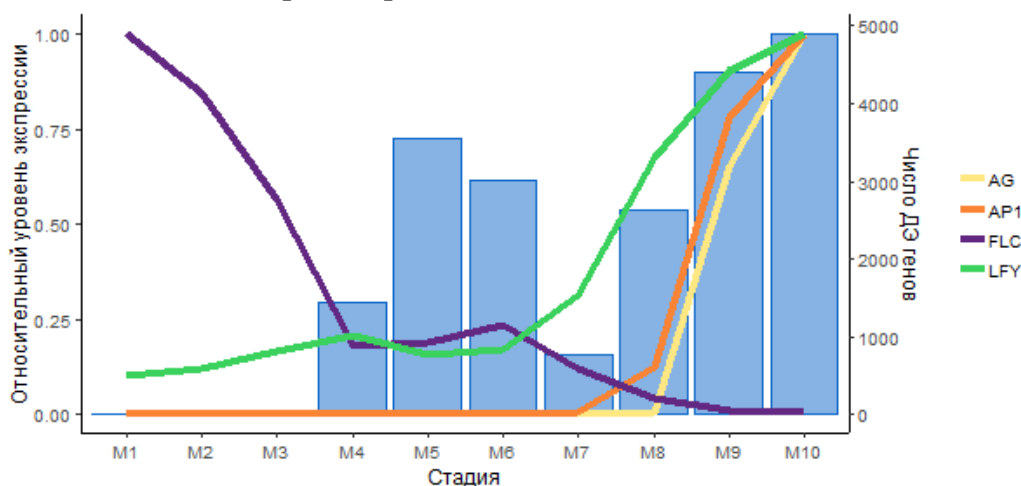


Рисунок 3. Относительный уровень экспрессии основных генов, регулирующих переход к цветению во временной серии АМ и число ДЭ генов между последовательными стадиями.

С помощью анализа дифференциальной экспрессии между последовательными стадиями и кластерного анализа мы обнаружили несколько характерных профилей экспрессии с выраженными изменениями на стадиях М4-М6 (Рисунок 4). Гены с такими профилями были связаны с модификацией хроматина, гистонами, хеликазами, биогенезом рибосом и цитоскелетом. Это позволяет предположить, что в ходе инициации цветения происходят изменения, связанные с клеточным циклом. Чтобы глубже изучить этот вопрос, мы детально рассмотрели динамику экспрессии групп генов, кодирующих элементы цитоскелета, гистоны и модификаторы гистонов, ДНК- и РНК-полимеразы, циклины, циклин-зависимые киназы и другие белки, связанные с клеточным циклом.

Мы выявили два основных паттерна 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 4). Во временной серии АМ почти все гены гистонов (маркеры S-фазы) 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): 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*.

Важным фактом является то, что паттерн типа (I) наблюдается также у генов, ассоциируемых с поздней G₂-фазой и митозом. Гены, кодирующие α- и β-тубулины,

имеют локальное повышение экспрессии в образце М5, также как и циклин-зависимые киназы (CDK) А- и В-типа.

В то же самое время 13 генов циклинов А- и В-типа, в суспензионных культурах клеток накапливающихся в поздней G₂ с пиком в раннем митозе, в нашей временной серии обладают паттерном типа (II) (Рисунок 4). Большинство генов, кодирующих компоненты пути E2F/DP/RB, который регулирует G₁/S-фазу также имеют паттерн типа (II), как и три гена, кодирующие активаторы APC (Anaphase-Promoting Complex), 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: так, для тубулинов и кинезинов ранее было показано одновременное увеличение экспрессии в G₂/М-переходе, однако во временной серии АМ, тубулины имеют профиль типа (I), а кинезины – типа (II).

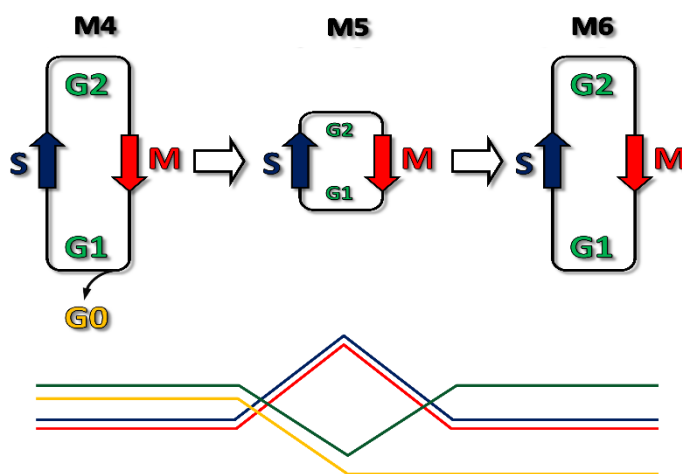


Рисунок 4. Схема, объясняющая изменения в уровнях экспрессии генов, контролирующих клеточный цикл, при переходе к цветению. На стадии М4 клеточный цикл имеет низкую скорость из-за длинных фаз G₁ и G₂; кроме того, значительная доля клеток АМ находятся вне клеточного цикла в фазе G₀. В момент перехода (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 профили 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).

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 (Рисунок 4). На стадии М4 часть клеток вегетативной АМ, которая находится 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 дифференциально экспрессируются гены, аннотированные термином «ответ на стресс» (GO:0006950). Экспрессия генов стрессовой реакции активировалась не только при сравнении органов, собранных с разных растений, но и чашелистиков и лепестков одного цветка, что исключает вероятность стрессового воздействия на часть растений. Кроме того, многие гены, участвующие в ответе на стресс, отличаются высокой тканевой специфичностью. Это позволяет предположить, что в разных органах *A. thaliana* развитие стрессовой реакции может отличаться. Хотя транскриптомный ответ *A. thaliana* на пониженную температуру неоднократно исследовался, все работы были проведены на фотосинтезирующих частях растения, и, таким образом универсальность холодовой реакции остается неизученной. В **Главе 3** мы рассматриваем изменения экспрессии генов, 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 FACTOR1-3*)), физиологической реакции растения на холод и ее интеграции с процессами роста и развития.

Мы использовали транскриптомную карту *A. thaliana*, чтобы выбрать образцы, сильно различающиеся по экспрессии генов: семядоли (далее Сд) и гипокотиль (Г) однодневного проростка, большой зеленый лист (Л), первый цветок (Ц) на стадии распускания, цветок на стадии 9 (Мц) и зрелые зеленые семена (С). 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 трех и 27 часов.

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» соответственно). Число ДЭ генов различалось от 1 022 в образце Ц-3 до 6 867 в Л-27; всего 15 459 генов были ДЭ хотя бы в одном сравнении. В дальнейших анализах образцы после трех и 27 часов воздействия +4° С рассматривались отдельно. Число генов, ДЭ во всех образцах, было крайне небольшим: в ответ на холодовое воздействие в течении трех и 27 часов возрастала экспрессия 89 и 124 генов, соответственно; участие этих генов в общем стрессовом ответе было известно ранее. Экспрессия 8 и 152 генов (три и 27 часов) снижалась, эти гены были связаны с фотосинтетическими процессами. Всего в образцах после трех часов стресса было обнаружено 4 045 органоспецифических генов и 6 780 – после 27 часов.

Для сравнения полученных нами данных с известными фактами мы рассмотрели поведение генов, входящих в регулон ТФ CBF1-3; генов COR cold-response); ТФ «первой волны» – регуляторов, активирующихся сразу после начала холодового воздействия одновременно с генами *CBF1-3*. Большинство известных участников ответа на холод активировалось в фотосинтезирующих образцах (Л, Сд, Г), но в цветках и семенах их доля была существенно ниже (36-51%).

Мы охарактеризовали органоспецифические гены стрессового ответа с помощью параметров экспрессии, основанных на транскриптомной карте *A. thaliana*. Распределение энтропии Шеннона для 15 459 ДЭ генов практически совпадает с распределением для всех экспрессирующихся генов (Рисунок 5 а и б). Универсальные гены не образовывали пика распределения на низких значениях Н (Рисунок 5 в и г), а органоспецифичные гены в большинстве случаев имели сходное с общим распределение энтропии Шеннона, за исключением нескольких образцов. Распределение энтропии генов, имеющих увеличенную экспрессию в образцах Л-3 и Ц-27, обладает локальным пиком на низких (0-0,3) значениях энтропии (Рисунок 5 д и ж); в транскриптомной карте *A. thaliana* эти гены экспрессируются исключительно в зрелых пыльниках и целых цветках, содержащих пыльники на той же стадии. В С-3 и С-27 распределение энтропии Шеннона обладает пиком в области средних значений (Рисунок 5 е и з), образующие его гены со сниженной экспрессией связаны с доменом F-box и с защитой растения, тогда как гены, экспрессия которых возрастает, ассоциированы с метаболизмом и запасанием липидов.

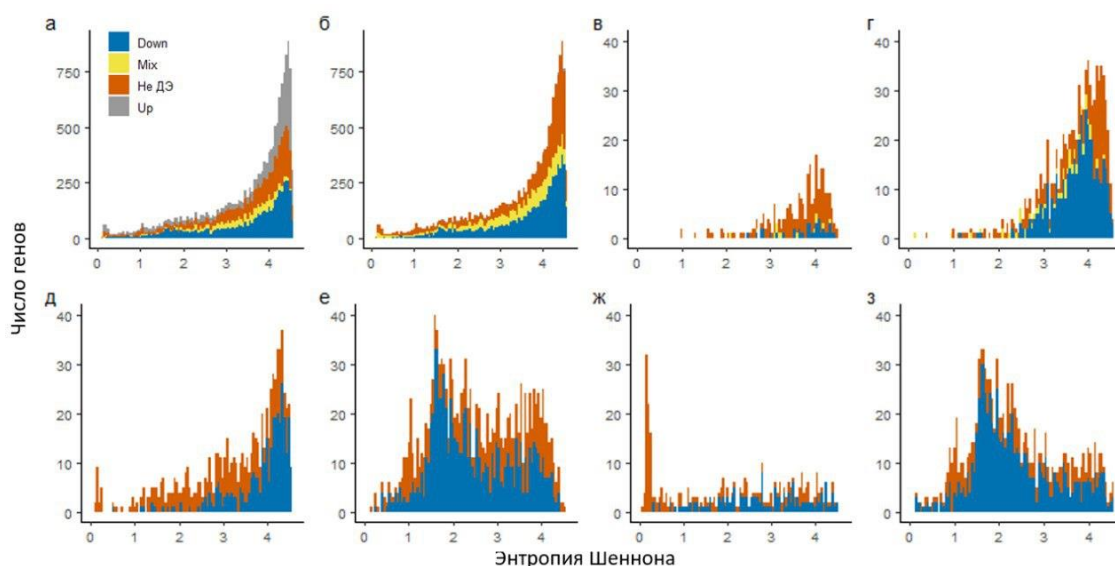


Рисунок 5. Распределение энтропии Шеннона среди всех генов, отвечающих на холодовое воздействие, с указанием типа изменения экспрессии после (а) трех и (б) 27 часов при +4°C; среди генов, ДЭ в пяти или шести образцах при (в) трех и (г) 27 часах действия холода; среди генов, специфически участвующих в холодовой акклиматизации в (д) Л-3, (е) С-3, (ж) Ц-27 и (з) С-27.

Мы проанализировали представленность регуляторных элементов различных ТФ в промоторах разных групп ДЭ генов. Как и ожидалось, в промоторных регионах универсальных ДЭ генов чаще, чем в среднем по геному, встречались регуляторные элементы ТФ CBF. Тканеспецифические гены регулировались различными ТФ; в

частности, мы нашли перепредставленность регуляторных элементов 19 ТФ из семейства NAC в промоторах ДЭ генов, экспрессия которых повышается в Мц-27.

В обсуждении Главы 3 рассматриваются основные выводы, сделанные на основании полученных результатов. 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. Изучение поведения известных участников холодовой акклиматизации показало, что полученные нами данные по фотосинтезирующим органам согласуются с литературными, при этом ответ на стресс в цветках и семенах отличается и предполагает участие регуляторов, отличных от генов *CBF*.

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.

Органоспецифические гены были связаны с разными процессами. Так, специфика транскриптомных изменений в Л-27 отражала важность торможения фотосинтеза и уменьшения вреда, наносимого накапливающимися активными формами кислорода при действии абиотического стресса, а в семенах изменялась экспрессия генов, связанных с запасанием липидов. Таким образом, помимо базовой системы генетической регуляции холодовой адаптации существует органоспецифическая акклиматизация, и выводы, сделанные при одного органа, имеют ограниченную применимость к другой части растения, что особенно необходимо учитывать в практическом применении полученных данных, например, при создании устойчивых к холоду трансгенных растений.

Для некоторых образцов органоспецифические гены имели особенности паттернов экспрессии. Так, для Л-3 был характерен локальный пик на низких значениях энтропии (Рисунок 5 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 программы защиты от холода в листе требует рекрутирования специфических для пыльцы генов. Аналогичный пик на низких значениях Н в Ц-27 также образовывали пыльцеспецифические гены (Рисунок 5 ж), вовлеченные в регуляцию роста пыльцевой трубки.

«Ядро» холодовой акклиматизации главным образом регулируются ТФ семейства *CBF*; кроме того, в каждом органе ответ на стресс модифицируется определенным набором ТФ. Так, для холодового ответа в семенах и цветке необходимо сочетание ТФ семейств *bZIP* и *NAC*, тогда как в листе в основном функционируют регуляторы из семейства *ERF/AP2*.

В Главе 4 мы использовали подробную транскриптомную карту *A. thaliana* для планирования атласа экспрессии генов распространенного по всему миру недавнего аллотетраплоида *C. bursa-pastoris*, также относящемуся к семейству Крестоцветных. В обзоре литературы Главы 4 мы рассматриваем последствия полиплоидизации: перестройку генома и транскриптома, в результате которой гены могут потерять функциональность и псевдогенезироваться, приобрести новую функцию, либо может начаться расхождение паттернов экспрессии гомеологичных генов и их субфункционализация. У многих видов растений экспрессируются на более высоком уровне гены из одного из субгеномов; это явление называется геномным доминированием. Поиск геномного доминирования у *C. bursa-pastoris* проводился неоднократно, однако были получены противоречивые результаты.

Созданный нами атлас экспрессии генов *C. bursa-pastoris* включает в себя десять наиболее различных образцов: корень, пластинка и жилка молодого листа, апикальная меристема, органы цветка, семена на двух стадиях, а также дополнительные данные: однодневные проростки, подвергнутых холодовому воздействию (+4°C) в течении 3, 15 и 27 часов, а также контроль.

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, в каждом образце было обнаружено 56-73% экспрессирующихся генов. Большинство гомеологов А и В экспрессировались во всех или почти во всех образцах; второй пик составляли гены с узким паттерном экспрессии.

Вклад гомеологов А и В в транскриптомное разнообразие образцов, оцененное с помощью анализа дифференциальной экспрессии между всеми парами образцов, был одинаков. В целом, профили экспрессии гомеологичных генов были схожи между собой ($r^2 = 0,8$ до $0,9$). Для поиска различий в уровнях транскрипции гомеологов мы проанализировали дифференциальную экспрессию внутри пар гомеологов: в каждом образце экспрессия гомеолога В сравнивалась с гомеологом А (далее ДЭАВ пары). 44% пар гомеологов было ДЭ хотя бы в одном образце. Только 867 пар гомеологов показали разное направление изменения экспрессии в двух и более образцах (Рисунок 6). Совместно с И.В. Кулаковским нами было проведено сравнение регуляторных мотивов в промоторах ДЭАВ и не-ДЭАВ гомеологов и ортологов в родительских видах, показавшее, что в промоторе одного из гомеологов в ДЭАВ парах возникают новые сайты связывания ТФ.

Обсуждение полученных результатов приведено в разделе 4.4. Общие транскриптомные характеристики *C. bursa-pastoris* находились в пределах значений, полученных для других видов растений. Как общее рассмотрение разнообразия экспрессии, так и отдельный анализ генов, участвующих в развитии холодной реакции, показал отсутствие 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.

Количество гомеологов В, экспрессия которых возрастала или снижалась по сравнению с гомеологом А, было одинаковым как в каждом образце, так и в целом (Рисунок 6). 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. Среди всех ДЭАВ пар преобладали ситуации, когда только один гомеолог экспрессировался или имел значительно более высокую экспрессию. Так как нами были собраны разнообразные органы и части растения, отсутствие детектируемой экспрессии одного из гомеологов является сильным указанием на потерю функции. Например, уровень экспрессии гомеолога А гена *FLC* 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. Изменения в профилях экспрессии гомеологов, вероятно, возникают в результате появления сайтов связывания ТФ в промоторе одного из гомеологов.

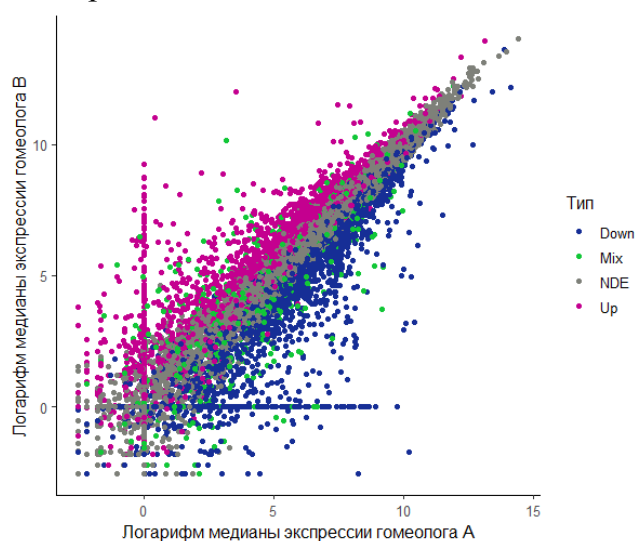


Рисунок 6. Корреляция экспрессии гомеологичных генов. Цветом отмечено направление изменения экспрессии при сравнении гомеолога В и гомеолога А: Down – гомеолог В имеет сниженную экспрессию по сравнению с гомеологом А хотя бы в одном образце и нигде не имеет повышенной; Up – гомеолог В имеет повышенную экспрессию по сравнению с гомеологом А хотя бы в одном образце и нигде не имеет пониженной; Mix – гомеолог В имеет повышенную экспрессию по сравнению с гомеологом А хотя бы в одном образце и пониженную хотя бы в одном образце; 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 и демонстрирует уровни транскрипции гомеологов в 14 образцах, а также уровень различия в их экспрессии.

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:

- 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 впервые 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* – 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. Разница в уровнях экспрессии гомеологов вероятно вызывается возникновением нового сайта посадки транскрипционного фактора в промоторе одного из гомеологов.

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.