

На правах рукописи

БОЙКО
Алексей Вячеславович

**ПОИСК ТРАНСКРИПЦИОННЫХ ФАКТОРОВ, РЕГУЛИРУЮЩИХ
ТРАНСДИФФЕРЕНЦИРОВКУ КЛЕТОК ПРИ РЕГЕНЕРАЦИИ КИШКИ
У ГОЛОТУРИИ *EUPENTASTA FRAUDATRIX***

1.5.23. Биология развития, эмбриология

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Владивосток – 2022

Работа выполнена в Федеральном государственном бюджетном учреждении науки
«Национальный научный центр морской биологии им. А.В. Жирмунского»
Дальневосточного отделения Российской академии наук

Научный руководитель:

доктор биологических наук, член-корреспондент РАН Долматов Игорь Юрьевич

Официальные оппоненты:

Ельчанинов Андрей Владимирович, доктор медицинских наук, Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова» Министерства здравоохранения Российской Федерации, зав. лабораторией регенерационной медицины

Костюченко Роман Петрович, кандидат биологических наук, доцент, Санкт-Петербургский государственный университет, и.о. зав. кафедрой эмбриологии

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт биологии развития им. Н.К. Кольцова Российской академии наук.

Защита диссертации состоится «29» ноября 2022 г. в 10 часов на заседании диссертационного совета 24.1.191.01 (Д 005.008.01) при Федеральном государственном бюджетном учреждении науки «Национальный научный центр морской биологии им. А.В. Жирмунского» Дальневосточного отделения Российской академии наук по адресу: 690041, г. Владивосток, ул. Пальчевского, д. 17.

Факс: (423)2310-900, e-mail: nscmb@mail.ru

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки «Национальный научный центр морской биологии им. А.В. Жирмунского» Дальневосточного отделения Российской академии наук: <http://www.wimb.dvo.ru/misc/dissertations/index.php/soviet-d-005-008-01/59-bojko-aleksej-vyacheslavovich>

Отзывы просим присылать на e-mail: mvaschenko@mail.ru

Автореферат разослан «___» _____ 2022 г.

Ученый секретарь
диссертационного совета

Вашченко

М.А. Вашченко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Регенерация представляет собой уникальный процесс, исследование которого входит в число приоритетных направлений современной биологии и биомедицины. Несмотря на долгую историю изучения, многие проблемы теории регенерации, такие как происхождение, эволюция и механизмы восстановительных морфогенезов до сих пор не решены. В частности, требует более детального изучения вопрос о клеточных источниках регенерации. Исследование широкого круга животных показало активное участие в морфогенезе не только стволовых (Lai, Aboobaker, 2018), но и дифференцированных клеток. Последние проходят через де- или трансдифференцировку и дают начало специализированным клеткам восстанавливающихся органов (Lavrov et al., 2018; Ereskovsky et al., 2020; Ельчанинов, Фатхудинов, 2020).

Иглокожие являются интересными модельными объектами для изучения проблемы источников регенерации. Восстановление у них протекает за счет специализированных клеток в результате их активной де- или трансдифференцировки (Garcia-Arregas, Dolmatov, 2010; Kalacheva et al., 2017). Среди всех Echinodermata, наибольший интерес в качестве моделей для изучения восстановительных морфогенезов представляют голотурии. Многие их виды обладают хорошими способностями к регенерации и проявляют широкий спектр восстановительных реакций (Dolmatov, Mashanov, 2007). Восстановление у голотурий осуществляется в достаточно короткие сроки. Кроме того, у ряда видов этих животных имеется уникальный способ аутономии – эвисцерация. В ответ на внешние раздражители голотурии способны к самопроизвольному удалению пищеварительной системы с последующим быстрым и полным ее восстановлением. Наиболее интересно происходит регенерация кишки у голотурии *Eupentacta fraudatrix* (D'yakonov & Varanova, 1958) (Holothuroidea, Dendrochirotida). У данного вида формирование передней части кишки осуществляется в результате трансдифференцировки клеток целомического эпителия (Mashanov et al., 2005). Трансдифференцировка клеток представляет собой интересное, но достаточно редкое в животном мире явление. Изучение этого феномена поможет понять механизмы работы генома клетки и их перестройки при различных процессах. Это крайне важно для разработки методов направленной дифференцировки стволовых клеток и получения искусственных тканей и органов.

Степень разработанности темы. В настоящее время хорошо исследованы только морфологические аспекты восстановления у голотурий (Долматов, 2020). Молекулярные механизмы регенерации различных систем органов изучаются, в основном, на *Apostichopus japonicus* (Li et al., 2018) и *Holothuria glaberrima* (Mashanov et al., 2015). У этих голотурий восстановление кишки происходит без трансдифференцировки клеток. Имеется только одна работа по исследованию генов при регенерации пищеварительной системы у *E. fraudatrix*, в

которой описана локализация экспрессии *sox9/10* и *sox17* (Dolmatov et al., 2021). Какого-либо анализа трансдифференцировки клеток на молекулярном уровне для данного вида не проводилось.

Цель и задачи исследования. Целью данной работы является поиск транскрипционных факторов, потенциально способных регулировать трансдифференцировку мезодермальных клеток при регенерации кишки у голотурии *E. fraudatrix*.

Задачи исследования включали: (1) разработку подхода к уменьшению сложности *de novo* сборки транскриптома без потерь биологически значимой информации и быстрому поиску ортологов между эволюционно-дистантными видами; (2) выявление генов-кандидатов на роль регуляторов трансдифференцировки с помощью анализа транскриптома на разных стадиях регенерации; (3) валидацию пространственно-временной экспрессии генов-кандидатов на разных стадиях регенерации кишки.

Научная новизна. Разработаны подходы к анализу транскриптома для видов, филогенетически далеких от имеющихся модельных видов и для которых отсутствует качественная расшифровка генома. Выработан ряд рекомендаций по транскриптомному анализу экспрессии генов и анализу сверхпредставленных процессов и сигнальных путей на примере трех видов голотурий и разных морфогенезов. Впервые для иглокожих проведен анализ транскриптома в процессе трансдифференцировки клеток и составлен список генов, участие которых в данном процессе может быть ключевым. Подтверждены некоторые закономерности, включая косвенно перестройку хроматина, характерные для регенерации и, в частности, трансдифференцировки. С помощью анализа пространственной экспрессии генов подтверждено участие 11 генов транскрипционных факторов в механизмах регенерации передней части кишки у *E. fraudatrix*. Показана низкая внутривидовая вариабельность экспрессии важных для регенерации генов.

Теоретическое и практическое значение работы. Полученные данные в первую очередь проливают свет на процессы реорганизации работы генома во время трансдифференцировки клеток одного зародышевого листка в клетки другого, что интересно не только в практическом аспекте, но и для сравнительного анализа механизмов регенерации у животных. Установление последовательностей транскриптов данного вида голотурий может быть полезным для исследований родственных видов и эволюционной биологии в целом. Также результаты пространственной и временной оценки экспрессии генов 11 транскрипционных факторов создают базу для дальнейшего изучения их функций в регенерации. В практическом смысле значение работы, в основном, заключается в

разработанных подходах к анализу немодельных и далеких от модельных видов животных в отсутствие у них расшифрованного генома.

Теоретические положения и результаты проведенных исследований использованы при подготовке научно-исследовательских отчетов по темам ННЦМБ ДВО РАН «Регенерация, бесполое размножение и иммунитет у двустворчатых моллюсков и иглокожих» (0268-2019-0001, № гос. регистрации 221022400135-6) и «Формирование фундаментальных основ устойчивого воспроизводства ценных морских гидробионтов за счет изучения их генетики, компонентов иммунной системы и особенностей морфогенеза» (0268-2018-0017, № гос. регистрации АААА-Б19-219040390019-4).

Методология и методы исследования. Для исследования динамики и состава транскриптома было использовано РНК-секвенирование на платформах Illumina HiSeq 2500, HiSeq 2000 и 454 GS FLX+. Исследование динамики и локализации экспрессии отдельных генов при регенерации кишки у голотурии *E. fraudatrix* было проведено с использованием технологии капельно-цифровой ПЦР (кцПЦР) и WMISH (whole mount in situ hybridization). Сборку и аннотацию транскриптомов проводили как с помощью классических инструментов (SPAdes, BLAST), так и с помощью специально разработанных алгоритмов (HomoloCAP, Reconciler). Анализ сверхпредставленных биологических процессов и путей был выполнен с помощью комбинации GSEA, EnrichmentMap и Cytoscape. В большинстве случаев для оценки статистической значимости использовали порог α равный 0,05.

Основные положения, выносимые на защиту:

1. Сложность *de novo* сборок транскриптомов, получаемых с помощью доступных на данный момент программ, можно уменьшить без потерь биологически значимой информации. Также возможен быстрый поиск ортологов между эволюционно-дистантными видами, более чувствительный и специфичный, чем существующий реципрокный метод.

2. Используя данные РНК-секвенирования вида, эволюционно-дистантного от модельных, возможно сократить список ключевых для процесса генов до разумных пределов, позволяющих уточнить их роль в процессе классическими методами молекулярной биологии.

3. Гены транскрипционных факторов Ef-EGR, Ef-ELF, Ef-GATA3, Ef-ID, Ef-KLF1/2/4, Ef-PRDM9, Ef-PCGF2, Ef-SNAI2, Ef-MSC, Ef-TCF24, Ef-TBX20 экспрессируются при регенерации передней части кишки у голотурии *E. fraudatrix* и могут принимать участие в трансдифференцировке клеток целомического эпителия в энтероциты.

Степень обоснованности и достоверности полученных данных. Достоверность результатов обеспечена использованием различных подходов к анализу данных секвенирования транскриптома, апробацией разработанных методов на нескольких видах

многоклеточных животных, а также сравнением получаемых данных с данными предыдущих опубликованных исследований и баз данных. Применялись актуальные программы и статистические методы обработки данных. Для подтверждения результатов исследования приведены табличные данные, рисунки и графики.

Личный вклад автора заключается в самостоятельно разработанных, реализованных и апробированных алгоритмах, анализе данных секвенирования, начиная со стадии первичных прочтений, подготовке и проведении всех экспериментов, участии в анализе пространственной экспрессии, а также в подготовке и написании публикаций.

Апробация работы. Результаты исследований были доложены на 16 Международной конференции по иглокожим (Нагоя, Япония, 2018); Международной конференции «Системная биология и биоинформатика» (Новосибирск, 2019); Международной конференции «Биоинформатика регуляции и структуры генома/системная биология» (Новосибирск, 2020); Ежегодной научной конференции ННЦМБ ДВО РАН (Владивосток, 2019) и Международной конференции «VIII Международная научно-практическая конференция молодых ученых: биофизиков, биотехнологов, молекулярных биологов и вирусологов» (Новосибирск, 2021).

Публикации. По материалу диссертации опубликовано 8 работ, в том числе 3 статьи в журналах из списка, рекомендованного ВАК.

Структура и объем диссертации. Основной текст диссертации изложен на 122 страницах и состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, выводов, списка литературы. Работа содержит 7 таблиц и 20 рисунков, а также 6 приложений на 13 страницах. Список литературы состоит из 165 наименований, из них 159 на иностранных языках.

Благодарности. Выражаю глубокую благодарность руководителю, д.б.н., чл.-корр. РАН Долматову Игорю Юрьевичу за бесконечную терпеливость в объяснении гистологических особенностей регенерации и обеспечение средств для дорогостоящих исследований, а также за помощь в интерпретации результатов биоинформационного анализа и экспериментов по оценке экспрессии исследуемых генов. Искренне благодарен А.С. Гиричу за помощь в проведении опытов по пространственной локализации экспрессии. Признателен всем сотрудникам лаборатории сравнительной цитологии ННЦМБ ДВО РАН за посильную помощь в различных вопросах и терпеливость. Кроме того, выражаю искреннюю благодарность А.Д. Кухлевскому за всегда качественные и быстрые результаты секвенирования ампликонов.

Работа выполнена при финансовой поддержке РФФИ (гранты № 20-04-00574, 19-34-90015, 17-04-01334), ДВО РАН (грант № 15-I-6-007 о) и РНФ (гранты № 14-50-00034 и 21-74-30004).

СОДЕРЖАНИЕ РАБОТЫ

1. ОБЗОР ЛИТЕРАТУРЫ

В обзоре литературы рассмотрены ключевые положения теории регенерации, клеточные источники регенерации у разных видов животных и клеточные процессы, происходящие при восстановлении. Дано описание регенерации при разных способах эвисцерации у разных видов голотурий и современного состояния исследований регенерации кишки у голотурии *E. fraudatrix*. Отдельным пунктом дан обзор современных подходов к анализу транскриптома у немодельных видов и представлен разбор связанных с этим проблем.

2. МАТЕРИАЛЫ И МЕТОДЫ

Сбор и содержание животных. Половозрелые особи голотурий *E. fraudatrix* и *A. japonicus* были выловлены в заливе Петра Великого Японского моря. Эвисцерацию у *E. fraudatrix* вызывали инъекцией 2% KCl в полость тела. Зачатки кишки для исследований брались через 3, 5–7 и 10 сут после эвисцерации (СПЭ). Эти интервалы в данной работе обозначены как первая, вторая и третья стадии регенерации, соответственно. На каждый срок использовали ткани от 10 особей. Перест *A. japonicus* вызывали температурной стимуляцией по стандартной методике. Для исследований брали по 100 особей на стадиях бластулы, гастролы, аурикулярии и пентактулы. Взрослые особи *Cladolabes schmeltzii* были собраны в заливе Ня-Чанг Южно-Китайского моря. Для исследования транскриптома были взяты ткани из средней части тела у 3 особей нормального строения и в месте формирования перетяжки у 3 особей, находящихся в процессе поперечного деления,

Выделение РНК, РНК-секвенирование и синтез кДНК. Выделение РНК и секвенирование образцов голотурий были проведены стандартными методами на Illumina. В случае выделения РНК *E. fraudatrix* для постановки количественной ПЦР (кПЦР) и капельно-цифровой ПЦР (кцПЦР) использовали особей через 3, 5, 7 и 10 СПЭ, по 5 особей на стадию. Выделение проводили для каждой особи отдельно в случае кцПЦР. Для кПЦР использовали 5 особей на стадию для каждого из 3 повторов, причем ткани стадий 5 и 7 СПЭ смешивали вместе. Синтез кДНК проводили с oligo-dT праймером набором Mint (Евроген, Россия) по протоколу производителя с 1 мкг РНК на каждый из 3 образцов на стадию с наилучшими показателями качества РНК, проверенного на Experion (Bio-Rad, США).

Разработка праймеров и зондов для кцПЦР, кПЦР и клонирования. Для подбора праймеров и зондов использовали последовательности транскриптов 17 генов, полученные при сборке транскриптома *E. fraudatrix*. Олигонуклеотиды подобраны с помощью PrimerQuest Tool (IDT, USA) в соответствии с рекомендациями для кцПЦР (Droplet Digital PCR Applications Guide, Bio-Rad, USA). Среди наборов праймеров и зондов выбирали лучшие с помощью собственного скрипта OligoAnalyse. Все упомянутые здесь и далее скрипты доступны по ссылке <https://github.com/Alteroldis/bioscripts>. Зонды и праймеры синтезировали в компании Евrogen (Россия) по технологии TaqMan.

Сборка транскриптома *de novo*. Сырые прочтения фильтровали в программе Trimmomatic v0.36. Затем корректировали ошибки секвенирования в две итерации и собирали транскриптом в программе SPAdes v3.13. Из всех транскриптов извлекали OPC (открытая рамка считывания) с минимальной длиной 30 аминокислотных остатков с помощью TransDecoder 5.5.0 с модифицированным кодом. Полученные OPC использовали для дальнейшей сборки с помощью собственной программы HomoloCAP. Далее сборку фильтровали от контаминирующих последовательностей и в соответствии с требованиями NCBI, после чего загрузили в базу TSA NCBI с индексами GHCL00000000, GFXQ00000000 и GFWR00000000 соответственно. Репрезентативность генов в транскриптоме была оценена с помощью BUSCO v3.

Анализ дифференциальной экспрессии генов. Для оценки экспрессии прочтения были картированы на транскриптом в Bowtie v2.3.4 и затем с помощью RSEM v1.3.1 было подсчитано число картирующихся прочтений и TPM (число транскриптов на миллион килобаз, англ. Transcripts Per Million). Далее, после фильтрации по числу прочтений и покрытию была оценена дифференциальная экспрессия в DESeq2 v1.18. В качестве контроля (точки отсчета) использовали образцы интактной кишки и стадию регенерации 5–7 СПЭ. Дифференциально-экспрессирующимся генами (ДЭГ) считали те, у которых экспрессия меняется более чем в два раза и вероятность ошибки $p < 0,05$.

Аннотация белок-кодирующих последовательностей транскриптома. Аннотация выполнена с BLASTp v2.7.0 со стандартным порогом величины E равным 10^{-5} по базам белков NR (non redundant) NCBI (11.12.2018), Ensembl v95, Echinobase (11.12.2018). В случае аннотации для поиска ортологов и анализа сверхпредставленных биологических процессов и путей вероятные ортологи выявляли с помощью модифицированного реципрокного метода и собственной программы Reconciler (подробнее см. «Результаты»). Среди таких ортологов затем выполняли поиск транскрипционных факторов (ТФ) на основании данных HGNC и Echinobase. Далее ТФ фильтровали по динамике экспрессии. Анализ сверхпредставленности биологических процессов и путей для *E. fraudatrix* выполнен с помощью GSEA в соответ-

ствии с протоколом плагина EnrichmentMap для анализа данных секвенирования транскриптома. Наборы генов человека, включающие хотя бы один из 11 ТФ, были взяты из базы MsigDB v6.2.

Оценка экспрессии генов с помощью кПЦР. В качестве референсного гена использовали *tubb* или *efla*. Каждая реакция в запуске представлена дважды (реплики). Каждый образец был представлен трижды – три отдельных выделения РНК из разных особей. ПЦР проводили в соответствии с инструкцией к набору ddPCR Supermix for Probes (Bio-Rad, США). Также были сделаны реакции кПЦР с интеркалирующим красителем для получения информации об отсутствии неспецифической амплификации. Далее продукты этих реакций были подготовлены к секвенированию по Сенгеру на секвенаторе ABI Prism 3130xL (Thermo Scientific, США) с помощью набора BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Scientific, США).

Валидация результатов РНК-секвенирования. Для проверки результатов секвенирования использовали 5 генов (*sox9*, *max*, *tcf24*, *foxc1* и *hes1*) и 2 референсных гена (*efla* и *tubb*). Реакции были проведены с использованием набора qPCRmix-HS SYBR (Евроген, Россия) в соответствии с протоколом производителя. Каждая реакция была проведена трижды для каждой стадии на образцах независимых выделений РНК. Каждая реакция в повторе имела 2 реплики. Эффективность праймеров была проверена серией реакций на 10 разведениях кДНК (1:2). Специфичность праймеров была проверена с помощью секвенирования ампликонов по Сенгеру, как описано выше. Далее полученные значения логарифма по основанию 2 от кратности изменения экспрессии гена на стадиях 5–7 и 10 СПЭ относительно 3 СПЭ использовали для вычисления квадрата коэффициента Пирсона между этими значениями и значениями, полученными по результатам РНК-секвенирования для данных генов.

Гибридизация *in situ*. Для синтеза РНК зонда, были амплифицированы фрагменты с размерами ампликона 400–1500 нуклеотидов и лигированы в вектор pAL2-T (Евроген, Россия), с последующей трансформацией XL1-Blue (Евроген, Россия). Вставки в плазмиды были секвенированы как описано выше. С помощью РНК-полимераз SP6 и T7 были синтезированы фрагменты для смыслового и антисмыслового РНК-зонда, используя DIG dNTP RNA labeling mix (Roche, Швейцария). Выявление продуктов генов проводили с помощью метода WMISH на фиксированных в 4% растворе параформальдегида зачатках кишки *E. fraudatrix*. Зачатки обрабатывали РНК-зондом в течение 12 ч при 65 °С и оставляли в Anti Dig Fab fragment (1:2000) на ночь. После этого красили с помощью NBT/BCIP tablet (Roche, Швейцария) при 37 °С в термостате.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1. Подбор и валидация методов сборки транскриптома *de novo* и аннотации по базам данных белков

De novo сборка транскриптома. Качество сборки транскриптома является основополагающим фактором, влияющим на качество, простоту и скорость выполнения всех дальнейших этапов, таких как анализ экспрессии, поиск ортологов, выявление статистических выбросов в массиве биологических процессов или молекулярных функций и многое другое (Hsieh et al., 2019). Во время работы с транскриптомами голотурий *C. schmeltzii* и *A. japonicus* было обнаружено, что значения статистических параметров, описывающих целостность транскриптов и избыточность сборки *de novo*, значительно хуже сборок, основанных на качественно собранном геноме. Для решения этой проблемы была разработана специальная программа HomoloCAP. В основе его алгоритма лежат два наблюдения: в сборках современных программ-сборщиков появляется избыточное число последовательностей с идентичными или вложенными ОРС, но различными нетранслируемыми регионами (НТР); OLC-сборщики вроде CAP3, которые используются для финализации сборки, работают на целой сборке долго и неэффективно с точки зрения статистических параметров, описывающих качество сборки.

Алгоритм программы работает в несколько шагов. В первую очередь с помощью TransDecoder извлекаются лучшие для каждой последовательности ОРС с длиной не менее 30 аминокислотных остатков. Далее программа работает в итеративном цикле, состоящем из двух шагов:

1. Кластеризация с помощью CD-HIT, параметры выставлены таким образом, чтобы кластер включал только последовательности, имеющие полностью идентичный с репрезентативной последовательностью кластера участок, который должен покрывать не менее 20% этой последовательности;

2. Сборка внутри каждого кластера с помощью CAP3 – OLC-сборщика. Запрещено использовать обратно-комплементарные последовательности; общий регион должен иметь не менее 100 нуклеотидов и 90% идентичность между последовательностями, входящими в него; суммарная длина регионов, лежащих около общего региона и отличающихся у последовательностей, имеющих его, должна быть не более 30% от длины общего региона; минимальная оценка общего региона снижена до 300 единиц, так как при объединении альтернативных изоформ могут возникнуть продолжительные разрывы в выравнивании.

В обычном режиме происходит 3 итерации, отличающиеся параметрами кластеризации. При этом, независимо от итерации, длина общего региона должна быть не менее 120 нуклеотидов. На последнем этапе идет кластеризация с целью объединения альтернативных изоформ общим названием последовательности.

Данный подход был вначале опробован на транскриптомах *C. schmeltzii* и *A. japonicus*. Для последнего вида он позволил достичь результатов, сравнимых с транскриптомами, полученными на основе геномных сборок. Для *C. schmeltzii* провести сравнение не было возможности, так как для этого вида других сборок не публиковалось. Наша сборка транскриптома *A. japonicus* содержит от 88 до 97% ОРС всех сборок, включая геномные, в то время как прочие покрывают 40% нашей сборки. Исключение составляет только сборка на основе генома, которая по обоим параметрам наиболее близка к таковой, полученной с использованием разработанного нами метода. Также при использовании нашего алгоритма все базовые статистические показатели получаются лучше, чем у ранее полученных сборок. То есть потеря биологически значимой информации не происходит, что также было проверено для *E. fraudatrix* на разных этапах сборки с помощью выравнивания всех последовательностей на белки человека, голотурии *A. japonicus*, морского ежа *Strongylocentrotus purpuratus* и на белки из базы SwissProt.

Фактически, основным нововведением здесь является отказ от сохранения НТР в пользу целостности транскриптов. Другим преимуществом является кластеризация последовательностей, что существенно уменьшает время вычислений, экономит необходимые ресурсы и увеличивает точность сборки по сравнению со стандартным CAP3. Подобный подход использовали и ранее (Pertea et al., 2003), но при этом кластеризацию осуществляли однократно, а не в итеративном цикле. Использование итерации является важным, поскольку она позволяет сохранить точность сборки и минимизирует вероятность появления химерных последовательностей.

Поиск ортологов. Когда нужно быстро найти пары последовательностей, которые могли бы быть ортологами в терминах выравнивания, стандартом является использование реципрокного BLAST-поиска. Однако существует два очевидных минуса данного подхода — максимальная оценка выравнивания не говорит, что последовательности выравнивания действительно ортологи; число пар, найденных таким образом, меньше реального числа ортологов, особенно если сравнение идет между эволюционно дистантными видами. В этой связи при поиске ортологов генов изучаемых видов голотурий необходимо было элиминировать, хотя бы частично, минусы реципрокного подхода. Для этого было решено включить в процедуру реципрокного метода не только лучшие совпадения, но и фракции условно лучших, то есть таких, для которых оценка выравнивания отличается от лучшей не более чем на 20–40%. Для этого нами была разработана программа Reconciler. Для его работы необходимы данные выравнивания с оценкой выравнивания (s) и индексами двух последовательностей — одна является поисковой ($a \in A$), а другая — из базы данных ($b \in B$), где a и b являются ка-

ким-либо индексом, а A и B — множеством индексов. Далее работа скрипта осуществляется в несколько этапов.

1. Создание трех словарей. В первом (словарь AB) содержится полная информация о совпадениях. Ключами являются b , а значениями — списки со всеми данными, описывающими совпадениями для b_n . Вторым (словарь B) содержит в качестве ключей b и в качестве значений имеет все совпадения с b_n в виде списков, каждый из которых включает s и a . В третьем словаре (словарь A) ключами являются a , а значениями — списки совпадений с a_n , содержащими s и b .

2. Сортировка в словарях всех значений, принадлежащих одному ключу, по убыванию s . Из словаря A и словаря B удаляются все значения, для которых верно условие $s_n < s_{n,best} * 0,8$, где $s_{n,best}$ — максимальная s для члена n множества A или B . Для словаря AB умножаем лучшую оценку на 0,7. Также мы использовали ограничение 0,6 и 0,5.

3. Итеративный обход по всем b до тех пор, пока существует хоть один b с числом записей более 1. На каждой итерации проверяем три условия:

3.1. Если нет записей для b_n , то все записи с b_n , в словаре A удаляются.

3.2. Если лучший a_n для b_n в словаре B не имеет записей в словаре A , то все записи, содержащие a_n в словаре B , удаляются. Затем применяется условие 3.1.

3.3. Если a_n и b_n в словарях A и B лучшие друг для друга, тогда пара a_n и b_n запоминается и все записи, содержащие a_n и b_n , удаляются из словаря B и словаря A , соответственно. То есть применяется условие реципрокности.

Используя пары из условия 3.3, программа находит соответствующие записи в словаре AB и формирует файл, содержащий исходные записи с парами из условия 3.3, приближенно соответствующие ортологам с точки зрения выравнивания.

Для сравнительной оценки результативности классического реципрокного и нашего подходов мы использовали известные ортологи белков 4 видов животных (*Mus musculus*, *Danio rerio*, *Ciona intestinalis*, *Drosophila melanogaster*) и белков человека из базы данных Ensembl. Как видно из Таблицы 1, с помощью Reconciler нам удалось улучшить стандартный реципрокный метод. Данная программа работает как минимум в два раза быстрее, а также находит больше ортологов, что особенно заметно при увеличении эволюционной дистанции. Таким образом, можно заключить, что использование Reconciler может применяться при отсутствии геномных данных вместо классического реципрокного BLAST-поиска. В то же время, следует учитывать, что при использовании обоих этих методов теряется большое число ортологов с множественными связями, то есть типов «один ко многим» и «многие ко многим», однако наш модифицированный метод находит больше таких ортологов.

Таблица 1. Сравнение двух подходов к поиску ортологов

Вид	OE	РГ	Р	МГ	МГ65	М	М65
<i>M. musculus</i>	20698/15815	15694/14773	15676/14755	15952/14861	15992/14877	15945/14853	15989/14873
<i>D. rerio</i>	19591/9281	9658/7089	9589/7041	10399/7492	10524/7528	10353/7450	10486/7496
<i>C. intestinalis</i>	14081/3879	4604/3091	4602/3089	4846/3148	4962/3180	4847/3145	4970/3179
<i>D. melanogaster</i>	15389/3071	4006/2579	4072/2622	4326/2660	4430/2671	4323/2657	4422/2669

Примечание. Р – стандартный реципрокный метод, М – модифицированный нами реципрокный метод. «Г» в названии столбца указывает, что при выравнивании использовали алгоритм Смита-Ватермана. 65 в названии столбца указывает, что при фильтрации выравниваний по оценке использовался пониженный порог. OE – число ортологов в геноме человека из базы Ensembl. Первое число означает все ортологи, второе – только однозначные ортологи типа «один к одному». Жирным шрифтом выделены лучшие результаты для каждого из подходов.

3.2. Поиск генов-кандидатов в регуляторы клеточной трансдифференцировки у *Eupentacta fraudatrix*

Секвенирование и сборка транскриптома *de novo*. В результате первичной сборки было получено около 2 млн. последовательностей. С помощью разработанного алгоритма NomoLoCAP это число удалось сократить до 83960 с N50, равным 1095 нуклеотидов, и средней длиной 698 нуклеотидов без потерь биологически значимой информации. Характеристики сборки соответствуют полученным на других голотуриях, за исключением числа последовательностей, которое в имеющихся сборках в лучшем случае было не больше 120 тысяч. Анализ репрезентативности генов показал, что 98,1% коровых генов Metazoa присутствует в сборке. Все параметры полученного транскриптома *E. fraudatrix* лучше или сравнимы с *de novo* сборками транскриптомов других видов иглокожих. Ясно наблюдаются более высокие статистические характеристики, описывающие полноту восстановленности последовательности транскриптов, при уменьшении числа последовательностей.

Анализ дифференциальной экспрессии генов. При поиске ДЭГ необходимо было учесть одну важную особенность регенерации у *E. fraudatrix*. У данного вида голотурий зачаток кишки после эвисцерации закладывается *de novo*, «из ничего», за счет преобразования мезодермальных производных (Mashanov et al., 2005). В этой связи более информативно было бы искать ДЭГ, сравнивая стадии регенерации с начальными этапами восстановления, например, со стадией 3 СПЭ, чем с нормой. Кроме того, поскольку нас интересовала в первую очередь ДЭГ, связанная с трансдифференцировкой клеток, которая происходит через 5–7 СПЭ, дополнительно в качестве контроля была использована эта стадия. Это позволило отсеять гены, экспрессирующиеся активнее на стадиях 3 СПЭ и 10 СПЭ, таким образом, оставив пул генов с пиком экспрессии на второй стадии.

Анализ дифференциальной экспрессии позволил выявить 17227 и 15342 ДЭГ со значимыми изменениями в экспрессии на какой-либо стадии по сравнению с образцами нормы и стадии 5–7 СПЭ соответственно. Большая часть генов, как видно на диаграммах Венна (Рисунок 1А), являются ДЭГ только по сравнению с образцами нормы.

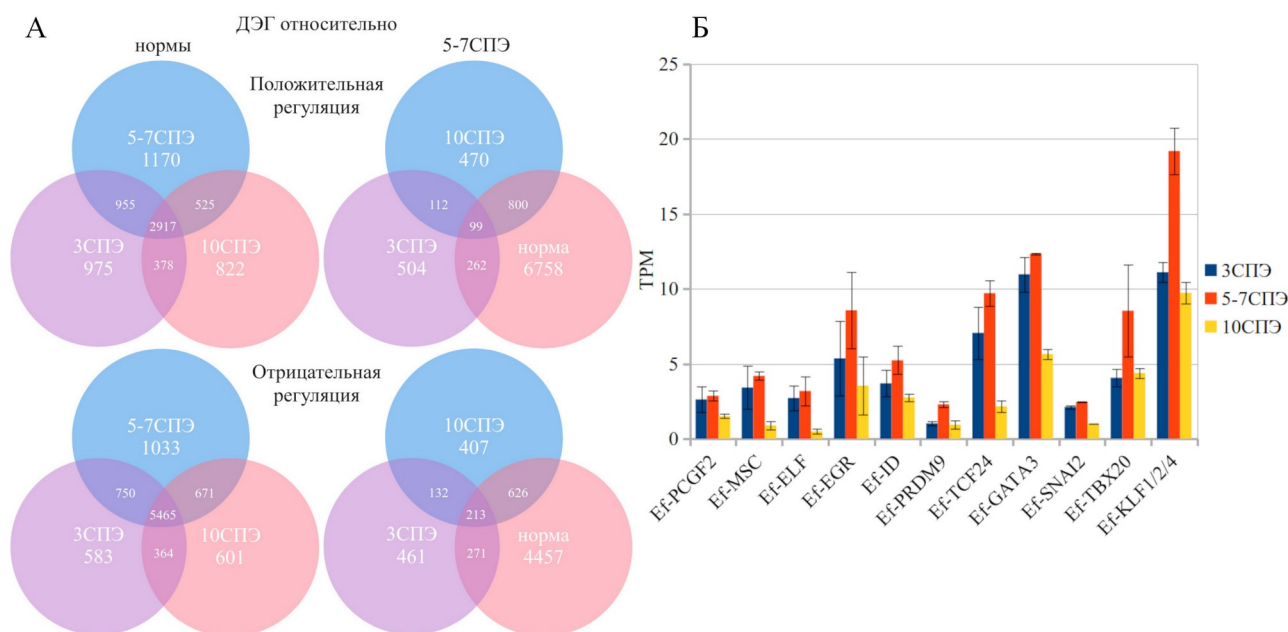


Рисунок 1. Дифференциальная активность генов при регенерации кишки у *Eupentacta fraudatrix*. А – диаграмма Венна для ДЭГ с положительной (сверху) и отрицательной (снизу) регуляцией относительно образцов нормы (слева) и 5–7 СПЭ (справа). Б – средние абсолютные оценки экспрессии (TPM) 11 генов ТФ через 3–10 СПЭ с минимумом и максимумом в отдельных повторах.

При сравнении разных стадий регенерации выделяются две закономерности. Во-первых, корреляция любого из образцов зачатков с образцами нормы значительно ниже любых других комбинаций образцов. Во-вторых, корреляция и число общих ДЭГ между образцами 3 СПЭ и 5–7 СПЭ выше, чем между каждой из них и 10 СПЭ. Таким образом, подтверждается очевидный факт, что в процессе регенерации кишки у *E. fraudatrix* происходит глобальная перестройка работы всего генома, причем ее пик приходится на 5–7 СПЭ (Рисунок 1А). Кроме того, данный анализ показывает правильность выбранной нами стратегии с использованием стадий 3 СПЭ или 5–7 СПЭ как точки отсчета.

Поиск кандидатов на роль регуляторов клеточной трансдифференцировки. Классический способ поиска важных генов-регуляторов определенного процесса заключается в выделении пула генов, относительная экспрессии которых меняется более чем в 4 раза относительно контроля и вероятностью ошибки оценки дифференциальности экспрессии ($p < 0,05$) (Mashanov et al., 2014; Mu et al., 2016). Такой классический подход в случае анализа немодельных видов имеет ряд недостатков, например, таких как отсутствие комплексного подхода к анализу процессов и разнообразных способов представления экспрессии. Это создает сложности в интерпретации результатов и построении гипотез о механизмах процесса. В

этой связи в нашей работе мы решили отойти от классической процедуры и сконцентрировались на поиске ТФ, абсолютная экспрессия которых была бы пиковой в период трансдифференцировки (5–7 СПЭ).

Всего был найден 961 ортолог ТФ человека или морского ежа, из которых 265 были идентифицированы как ТФ у обоих видов. Далее все ТФ были отфильтрованы по динамике их экспрессии так, чтобы они удовлетворяли четырем условиям: значение TPM на стадии 5–7 СПЭ больше единицы; среднее значение LogFC больше 0,5 на этой стадии, по сравнению с 3 СПЭ и 10 СПЭ; значение LogFC на второй стадии, по сравнению с первой и третьей, больше единицы; $p < 0,05$ на первой и третьей стадиях относительно второй. Третье условие позволяет обойти проблему участия одного гена в процессах не только второй, но и первой или третьей стадии, при этом гарантируя более чем двукратное увеличение экспрессии на второй стадии относительно какой-либо другой стадии регенерации. В итоге найдено 20 ТФ с пиком экспрессии через 5–7 СПЭ, 9 из которых были удалены из дальнейшего рассмотрения. Восемь ТФ не идентифицировались как осмысленные белки с доменом типа «цинковый палец». Еще одна последовательность, Ef-ELF4, являлась явной ошибкой сборки. Выявленные 11 генов являются представителями 6 классов ТФ: содержащие триптофановый кластер (Ef-ELF), C2H2-домен цинковых пальцев (Ef-PRDM9, Ef-EGR, Ef-KLF1/2/4, Ef-SNAI2), bHLH-домен (Ef-TCF24, Ef-MSC, Ef-ID), C4-домен цинковых пальцев (Ef-GATA3), белки группы Polycomb с RING-доменом (Ef-PCGF2) и T-домен (Ef-TBX20) (Рисунок 1Б).

Таким образом, данный подход позволил выделить 11 генов ТФ, для которых имелась информация об их функциях. Конечно, отсутствие исследованных функций у генов не говорит о том, что их не стоит рассматривать как возможные регуляторы того или иного процесса. Однако этот же факт не позволяет строить гипотезы об их роли в морфогенезе. Для таких генов необходимы специальные исследования их структуры, строения кодируемых ими белков, участия в регенерации и т.д.

Сеть сверхпредставленных биологических процессов и путей. Одним из важных способов проверки ТФ на возможную вовлеченность в регуляцию трансдифференцировки является построение и анализ сети биологических процессов и сигнальных путей, связанных с этими ТФ. Следует отметить, что к анализу полученных данных необходимо подходить с осторожностью, поскольку в основе генной онтологии лежат процессы, описанные у млекопитающих, и их нельзя напрямую применять для анализа морфогенезов у голотурий. Кроме того, чем больше эволюционная дистанция между видами, тем больше различий в числе генов и составе генных семейств, поэтому очень часто невозможно установить ортологию многих генов между этими группами животных.

В результате исследования было выявлено 490 сверхпредставленных процессов и путей, объединяющих 3168 ортологов генов человека. Среди этих биологических процессов и путей 204 содержали отрицательно регулируемые гены и 71 – положительно регулируемые гены на второй стадии регенерации относительно как первой, так и третьей стадий. Сеть содержит 6 крупных тесно связанных друг с другом блоков путей и биологических процессов (Таблица 2). При этом велика вероятность, что выявленные ТФ служат ключевыми факторами регуляции регенерации кишки у *E. fraudatrix*, поскольку все они являются «верховыми» регуляторами многочисленных процессов, запуская сложные каскады генов, а некоторые из них (PCGF2, PRDM9) напрямую участвуют в модификации хроматина.

Таблица 2. Блоки сети сверхпредставленных биологических процессов и путей

	Название процессов и сигнальных путей	Число генов	Уникальных генов	Число процессов	ТФ в блоке
Блок 1	Регуляция морфогенезов, детерминация клеточной судьбы	569	122	21	GATA3, ID2, KLF2, SNAI2, TBX20
Блок 2	Развитие и регуляция иммунного ответа	850	498	101	EGR1, GATA3, ID2, KLF2
Блок 3	Дифференцировка нервных клеток и ее регуляция	389	144	31	GATA3, ID2, SNAI2, TBX20
Блок 4	Развитие и морфогенезы мышечных тканей, мезенхимы, эпителиально-мезенхимная трансформация	466	150	53	EGR1, GATA3, ID2, MSC, PCGF2, SNAI2, TBX20
Блок 5	Развитие и морфогенезы соединительной ткани	222	60	14	EGR1, ID2, MSC, SNAI2
Блок 6	Развитие и морфогенез эпителиев	469	125	53	EGR1, GATA3, ID2, KLF2, SNAI2, TBX20

Блоки достаточно хорошо соответствуют морфологическим аспектам регенерации кишки у *E. fraudatrix*, таким как иммунный ответ на повреждение, миграция клеток и морфогенез, нейрогенез, ремоделирование внеклеточного матрикса и эпителиально-мезенхимная трансформация (ЭМТ) (Mashanov et al., 2005; Долматов, Машанов, 2007). В плане изучения трансдифференцировки наиболее важен блок 4, который характеризуется большим числом процессов, наиболее активных на стадии 5–7 СПЭ. Он связан с реорганизацией мезодермальных клеток, происходящих при развитии сердца у млекопитающих. Очевидно, что эти процессы у *E. fraudatrix* следует интерпретировать как преобразование целомического эпителия. По-видимому, гены и процессы именно данного блока играют определяющую роль в основных событиях регенерации кишки – в погружении целомического эпителия и трансдифференцировке его клеток. Однако активация механизмов ЭМТ при регенерации кишки у *E. fraudatrix* необычна и в какой-то мере противоречит морфологическим данным, поскольку разрушение межклеточных контактов при погружении клеток не происходит (Mashanov et al., 2005). Это можно объяснить тем, что при эпителиальных морфогенезах клетки могут

приобретать некоторые мезенхимные черты (Schöck, Pergimon, 2002). При этом, вероятно, частичная ЭМТ, как и полная, регулируется теми же генами. В частности, у *E. fraudatrix* во многих процессах с пиком активности на стадии 5–7 СПЭ принимает участие либо ID2 (*Ef-id*), либо SNAI2 (*Ef-snai2*). Эти ТФ способны провоцировать ЭМТ эпителиальных клеток (Kalinkova et al., 2018).

Таким образом, с помощью построения сети биологических процессов и сигнальных путей было показано, что выделенные нами 11 ТФ принимают участие в базовых процессах, происходящих при формировании кишки у *E. fraudatrix*. Полученные данные хорошо согласуются с морфологическими данными о механизмах регенерации пищеварительной системы у голотурий. Основу морфогенеза составляют преобразование внеклеточного матрикса (блок 5) и трансформация мезодермы (блок 4). Остальные блоки характеризуют дополнительные особенности восстановительного процесса, такие как пролиферация и миграция клеток, эпителиальный морфогенез, восстановление нервного плексуса и др.

3.3 Анализ временной и пространственной динамики экспрессии генов-кандидатов

Оценка экспрессии с помощью кцПЦР. При подборе комплектов олигонуклеотидов для постановки мультиплексной ПЦР мы столкнулись с проблемой большого числа возможных комбинаций олигонуклеотидов, которые необходимо проверить на гетеро- и гомодимеры. Так, в случае гетеродимеров, одна триплексная ПЦР-реакция с зондами требует 9 олигонуклеотидов, а значит, нужно проверить 36 сочетаний, что является нетривиальной задачей. В данной же работе необходимо было сделать 11 триплексных реакций. В связи с этим была разработана программа OligoAnalyse, позволяющая среди множества комбинаций праймеров и зондов найти лучшие, основываясь на расчете энергии Гиббса для вторичных структур, гетеро- и гомодимеров. С помощью нее были определены все необходимые олигонуклеотиды, которые работали строго специфично.

Было показано, что динамика экспрессии, определенная по методу кцПЦР, отличается от таковой, полученной методом РНК-секвенирования. Так, 4 гена (*Ef-gata3*, *Ef-pcgf2*, *Ef-msc* и *Ef-snai2*) имеют небольшой пик через 3 СПЭ (Рисунок 2А). Проведенный анализ корреляции показал, что результаты экспрессии, вычисленные по РНК-секвенированию, в основном не коррелируют с результатами кцПЦР, за исключением первого повтора образцов 10 СПЭ и 5–7 СПЭ. При этом результаты, полученные какой-либо одной методикой, показывают строгую корреляцию между повторами. Чтобы проверить, насколько низкая корреляция зависит от конкретных генов, мы вычислили парную корреляцию между всеми повторами и методами (Рисунок 2Б). Низкие значения корреляции в основном обусловлены генами *Ef-elf* и *Ef-id*,

а также, частично, *Ef-egr* и *Ef-msc*. Причем для *Ef-msc* и *Ef-id* корреляция между разными методами оценки отсутствует, в то время как между повторами одного метода значения экспрессии хорошо совпадают, особенно у *Ef-id*, где корреляция близка к абсолютной. Возможно, причиной различий в оценке экспрессии разными методами является не полностью восстановленные последовательности транскриптов этих генов, что подтверждается выравниванием с их ортологами голотурии *A. japonicus*.

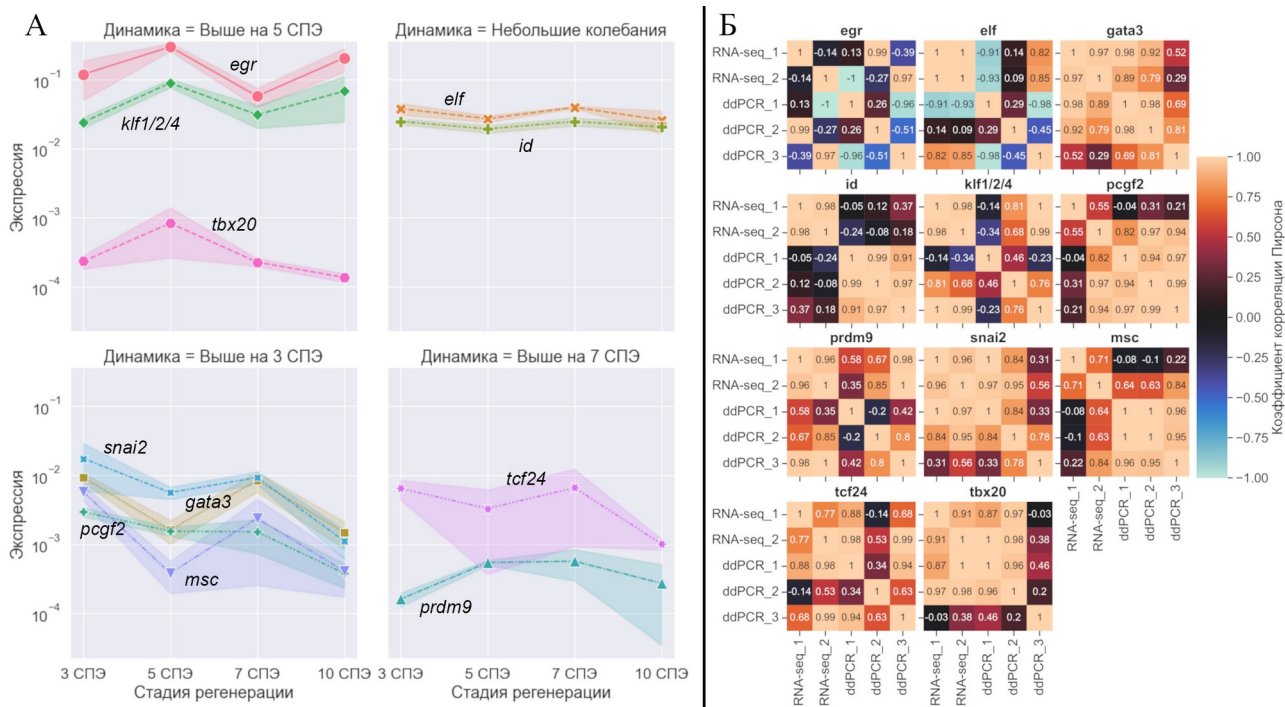


Рисунок 2. Экспрессия и корреляция 11 генов ТФ при регенерации кишки у *Eupentacta fraudatrix*. А – нормализованные на референсные гены (*tubb* и *efla*) оценки значений экспрессии 11 генов ТФ через 3–10 СПЭ, полученные с помощью кцПЦР. Точки показывают среднее значение между повторами, полупрозрачная область соответствует 90% доверительному интервалу. Масштаб логарифмический по основанию 10. Гены разбиты по характеру динамики средних значений экспрессии. Б – Тепловая карта коэффициента корреляции Пирсона между оценками экспрессии по разным повторам и методам (RNA-seq и ddPCR).

Локализация экспрессии генов транскрипционных факторов в зачатке кишки на разных стадиях регенерации. Для того чтобы подтвердить возможное участие выявленных генов в регенерации кишки у *E. fraudatrix*, была исследована локализация их транскриптов в тканях голотурии на разных стадиях восстановления. Экспрессия большинства исследованных генов отмечается в зачатке кишки и аквафарингеальном комплексе (АК) уже через 3 СПЭ. Это такие гены, как *Ef-egr*, *Ef-elf*, *Ef-gata3*, *Ef-klf1/2/4*, *Ef-pcgf2*, *Ef-tbx20*, *Ef-msc*, *Ef-id*, *Ef-prdm9* и *Ef-tcf24*. Их транскрипты обнаруживаются в кишечном мезентерии, обычно по его вентральному краю и в формирующихся структурах АК. Продукты *Ef-snai2* с помощью метода WMISH на этой стадии в зачатках не выявляются. Через 5–7 СПЭ транскрипты почти всех исследуемых генов ТФ обнаруживаются в регенерирующих органах (Рисунок 3). Интенсивно окрашивается мезентерий и соединительно-тканый зачаток кишки. Исключение со-

ставляет *Ef-id*, транскрипты которого не были выявлены. Продукты всех генов локализуются в целомическом эпителии. Однако, для *Ef-pcgf2*, *Ef-elf* и *Ef-msc* наибольшая интенсивность окраски обнаруживается в той области, где происходит трансдифференцировка и погружение клеток целомического эпителия. Транскрипты *Ef-egr* располагаются в формирующемся кишечном эпителии. Через 10 СПЭ в целомическом эпителии кишки экспрессируются все 11 генов ТФ (Рисунок 3). Продукты *Ef-egr* выявляются только на некотором расстоянии от АК. По всей длине кишки окрашивается целомический эпителий при реакции на мРНК *Ef-elf*, *Ef-id*, *Ef-gata3*, *Ef-snai2*, *klf1/2/4*, *Ef-prdm9*, *Ef-msc* и *Ef-pcgf2*. Транскрипты *Ef-egr*, *Ef-tbx20* и *Ef-tcf24* обнаруживаются не только в целомическом, но и в кишечном эпителии.

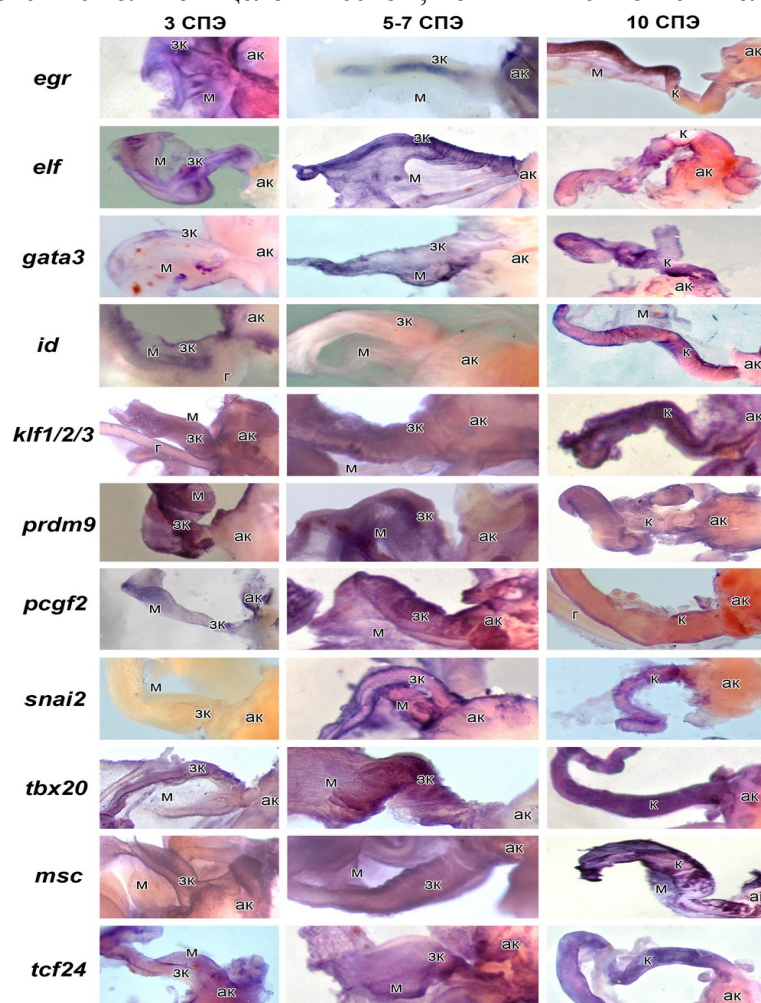


Рисунок 3. Локализация транскриптов генов *Ef-egr*, *Ef-elf*, *Ef-gata3*, *Ef-id*, *Ef-klf1/2/4*, *Ef-prdm9*, *Ef-pcgf2*, *Ef-snai2*, *Ef-tbx20*, *Ef-msc* и *Ef-tcf24* в регенерирующих органах голотурии *Eupentacta fraudatrix*. ак – аквафарингеальный комплекс, г – гонодукт, зк – зачаток кишки, к – кишка, м – мезентерий

Таким образом, все выявленные гены ТФ экспрессируются в тканях зачатка. Однако динамика экспрессии и локализация продуктов между генами различаются. Это, возможно, указывает на разные функции изучаемых ТФ в регуляции регенерации у *E. fraudatrix*. В частности, через 5–7 СПЭ транскрипты *Ef-egr* обнаруживаются в формирующемся кишечном эпителии (Рисунок 3). В этой связи можно предположить, что *Ef-egr* может быть одним из кандидатов в гены, участвующих в трансдифференцировке. Однако у *E. fraudatrix* этот ген

проявляет большую вариабельность, что видно как по результатам РНК-секвенирования, так и кцПЦР (Рисунок 2Б). Кроме того, согласно анализу сети сверхпредставленных процессов, процессы, в которых он может принимать участие, не слишком многочисленны. Тем не менее, все это может объясняться тем, что у *E. fraudatrix*, как и у планарий, он является одним из ключевых генов, запускающих регенерацию (Gehrke et al., 2019).

У морского ежа *S. purpuratus* экспрессия гена *elf* обнаруживается на стадии поздней гаструлы во всех клетках, но наибольшая концентрация наблюдается в кишке (Rizzo et al., 2006), что косвенно коррелирует с нашими данными – экспрессия увеличивается в период закладки кишечного эпителия с последующим снижением, когда кишка почти сформирована (Рисунок 1Б). Это позволяет предположить, что ортолог ELF у иглокожих может быть вовлечен в регуляцию ранних стадий формирования кишечного эпителия и, возможно, в трансдифференцировку, так как на стадии 5–7 СПЭ он экспрессируется в месте, где происходит погружение целомического эпителия в соединительную ткань зачатка (Рисунок 3).

Учитывая динамику экспрессии *Ef-gata3* (Рисунок 2А), а также тесную связь его ортолога у многоклеточных, в том числе морских ежей, с морфогенезами и трансформацией мезодермальных клеток (Materna et al., 2013; Riddle et al., 2013), можно предположить, что этот ТФ участвует не в трансдифференцировке, а в дедифференцировке и поддержании этого состояния у клеток мезотелия. Это подтверждается ранней активацией *Ef-gata3* (3 СПЭ), когда основным процессом в зачатке кишки является дедифференцировка клеток целомического эпителия, и экспрессией его только в целомическом эпителии зачатка (Рисунок 3).

Белки семейства ID, хоть и имеют HLH-домен, но, строго говоря, не являются ТФ, поскольку у них отсутствует базовый ДНК-связывающий домен. Наличие этого гена в списке ТФ-кандидатов на роль регулятора трансдифференцировки явилось следствием ошибки в базе Echinobase, где он упоминается как ТФ. Однако, несмотря на это, мы приняли решение его оставить, так как его экспрессия и функции показали нам заслуживающими внимания. Мы полагаем, что *Ef-id* у *E. fraudatrix* может выполнять функции, сходные с таковыми у млекопитающих, то есть участвовать в дедифференцировке клеток целомического эпителия и поддержании этого статуса. Одновременно он может провоцировать приобретение некоторых мезенхимных признаков за счет ко-экспрессии с геном *Ef-snai2* (Chang et al., 2013; Kamata et al., 2016). Это предположение хорошо соотносится с известными данными о функциях этих генов и морфогенетическими процессами, происходящими на начальных этапах (3–7 СПЭ) регенерации кишки у *E. fraudatrix*. Кроме того, этот ген демонстрирует исключительный уровень скоррелированности результатов оценки экспрессии между повторами одного метода, что говорит об исключительной важности гена *Ef-id* для успешного прохождения регенерации (Рисунок 2Б).

Ген *Ef-klf1/2/4*, очевидно, играет значительную роль в регенерации кишки у *E. fraudatrix*. Об этом говорят результаты WMISH, кцПЦР, РНК-секвенирования и анализ сверхпредставленных процессов (Рисунки 2, 3, Таблица 2). *Ef-klf1/2/4* может принимать участие в активации репрограммирования клеток целомического эпителия и поддержании их в дедифференцированном состоянии в процессе их трансформации в энтероциты, как это делают его гомологи у млекопитающих (Yamane et al., 2018). В этой связи, *Ef-klf1/2/4* является одним из кандидатов в регуляторы трансдифференцировки у *E. fraudatrix*.

Сложно сказать, какую функцию выполняет *Ef-PCGF2* у *E. fraudatrix*, так как модификаторы эпигенетического статуса генома действуют на широкий спектр генов. Однако исходя из функций его ортолога у человека (Lee et al., 2014), можно предположить, что *Ef-pcgf2* вовлечен в стабилизацию мезодермальных свойств клеток целомического эпителия, находящихся на поверхности зачатка кишки. В пользу этого также говорит нисходящая динамика экспрессии в течение регенерации и локализация только в целомическом эпителии (Рисунки 2А, 3).

У хордовых ТФ семейства PRDM являются важными эпигенетическими регуляторами. Например, у мыши во время развития они служат для поддержания плюрипотентного состояния клеток (Hohenauer, Moore, 2012; Yamaji et al., 2013). Мы предполагаем, что *Ef-PRDM9* у *E. fraudatrix* может быть вовлечен в дифференцировку и/или стабилизацию дедифференцированного состояния клеток при регенерации и в таком качестве участвует в трансдифференцировке. На это указывает явный пик экспрессии через 5–7 СПЭ и высокая корреляция результатов между повторами и методами (Рисунок 2).

Ef-snai2 может также участвовать в трансдифференцировке, но опосредованно. Он, вместе с *Ef-id*, возможно, способствует развитию у клеток целомического эпителия некоторых мезенхимных черт, что делает возможным их миграцию в зачаток кишки. Это согласуется с литературными данными по ортологам этих генов (Kamata et al., 2016; Kalinkova et al., 2018), данными по экспрессии, WMISH и очень большим числом сверхпредставленных процессов, в которые данный ген вовлечен (Рисунки 2, 3, Таблица 2).

В связи с отсутствием литературных данных по *Ef-tcf24* трудно предположить его функции по аналогии с другими животными. Тем не менее, по результатам WMISH через 10 СПЭ его транскрипты обнаруживаются в кишечном эпителии (Рисунок 3). Это, а также динамика экспрессии *Ef-tcf24*, указывает на его участие в регенерации. Возможно, данный ген регулирует либо поздние стадии трансформации, либо дифференцировку энтероцитов.

С учетом данных литературы об участии MSC в регуляции дифференцировки исключительно мышечных клеток и ЭМТ (Acharya et al., 2012), динамике экспрессии, а также локализации его транскриптов в месте погружения клеток целомического эпителия в зачаток че-

рез 5–7 СПЭ (Рисунки 2А, 3), можно предположить, что он либо, как и *Ef-snai2*, нужен для приобретения некоторых мезенхимных признаков погружающимися клетками целомического эпителия, либо не связан с трансдифференцировкой, а необходим для дифференцировки миоэпителиальных клеток.

Ген *Ef-tbx20* отличается от большинства исследованных генов четко видимым пиком экспрессии на стадии 5–7 СПЭ. Кроме того, результаты кцПЦР показывают строгий пик экспрессии *Ef-tbx20* через 5 СПЭ (Рисунок 2А), при этом на остальных этапах регенерации отмечаются примерно равные значения экспрессии. Также наблюдается отличная корреляция между всеми повторами и обоими методами (Рисунок 2Б). Интересно, что транскрипты *Ef-tbx20* через 10 СПЭ обнаруживаются в кишечном эпителии зачатка кишки (Рисунок 3). Все это указывает на важную роль данного ТФ в регенерации кишки у *E. fraudatrix*. Однако у позвоночных семейство ТВХ больше связано с дифференцировкой клеток экто- и мезодермы (Takashima, Suzuki, 2013). В этой связи можно предположить, что, в отличие от млекопитающих, у голотурий данный ген может участвовать не только в перестройке целомического эпителия, но и в спецификации энтероцитов.

Таким образом, предварительный анализ показал, что все 11 выявленных генов ТФ экспрессируются при регенерации пищеварительной системы у *E. fraudatrix*, что указывает на их вовлеченность в этот процесс. Среди них следует выделить 4 гена, вероятность участия которых в трансдифференцировке наиболее высока. Это *Ef-klf1/2/4*, *Ef-prdm9*, *Ef-snai2* и *Ef-id*. Все они характеризуются исключительной повторяемостью оценки экспрессии, в связи с чем должны быть важны для успешной регенерации и, вероятно, трансдифференцировки, как ключевого этапа формирования кишки у *E. fraudatrix*. *Ef-KLF1/2/4*, возможно, запускает процесс трансформации клеток целомического эпителия в энтероциты. Продукты *Ef-prdm9*, *Ef-snai2* и *Ef-id*, по-видимому, поддерживают недифференцированное состояние клеток, проходящих трансдифференцировку. *Ef-snai2* и *Ef-id* также, вероятно, играют немаловажную роль в обеспечении подвижности клеток и дальнейшего коммитирования их судьбы как энтодермальных эпителиальных клеток.

ЗАКЛЮЧЕНИЕ

Иглокожие и, в частности, голотурии, представляют собой интересные объекты для изучения регенерации, поскольку обладают высокими восстановительными способностями. Однако это преимущество нивелируется их малой изученностью, отсутствием качественно собранного генома и отсутствием информации о функциях их генов. В ходе данной работы нами был реализован ряд подходов, упрощающих анализ данных, в основном РНК-секвенирования, для таких немодельных объектов на примере голотурий. Была написана программа

для улучшения сборок, получаемых с помощью коммерческих сборщиков транскриптомов. Также был разработан и реализован новый подход к поиску ортологов генов. Кроме того, был сделан ряд важных выводов об общем алгоритме анализа транскриптомов немодельных организмов, в том числе о необходимости использовать разные способы представления данных экспрессии и разные способы выяснения функций генов, а также об объединении информации, полученной из разных источников. В нашей работе показано, что новые высокопроизводительные подходы даже в непростых случаях способны выполнять свое главное предназначение – формирование гипотез о сложных процессах. Также было ясно проиллюстрировано, что в некоторых случаях, в частности при анализе динамики экспрессии генов при регенерации, нельзя использовать в качестве контрольного образца неповрежденные ткани. Это обосновано алгоритмическими ограничениями при поиске дифференциально экспрессирующихся генов, а также тем, что у некоторых животных, как, например у *E. fraudatrix*, морфогенез начинается не с преобразования интактных структур, а с формирования зачатков *de novo*.

При анализе транскриптома было сделано два вывода об общем планировании анализа. Во-первых, в случае немодельных видов проще концентрироваться на небольшой и консервативной группе генов, например ТФ. Это позволит привлечь к анализу значительно больше информации, полученной на модельных видах. Во-вторых, как бы очевидно это не звучало, необходимо держать в голове реальные биологические процессы, происходящие в исследуемом процессе, например гистологические данные.

Следующим этапом данной работы являлась проверка гипотезы об участии 11 ТФ в регенерации, базирующейся на данных РНК-секвенирования. Здесь в первую очередь надо сделать вывод, что независимо от внутривидовой изменчивости и состояния организмов, генетическая машинерия, стоящая за регенерацией кишки работает четко и точно. В свою очередь, это означает, что результаты даже небольшого числа повторов какого-либо эксперимента можно экстраполировать с хорошей точностью на всю популяцию животных. Другим важным моментом являются результаты экспрессии в целом и результаты WMISH, в частности. Эти результаты в основном сходятся с предположениями, выдвинутыми при анализе транскриптома. Так, гены *Ef-gata3*, *Ef-pcgf2* и *Ef-tbx20*, вероятно участвующие в морфогенезах мезодермальных клеток, преимущественно имеют более варибельную динамику экспрессии, которая, возможно, является следствием разного объема ткани мезентерия, попавшей в выделение РНК, а также экспрессируются преимущественно в мезентерии или целомическом эпителии кишки. Гены *Ef-id* и *Ef-snai2*, скорее всего, участвуют в предотвращении преждевременной спецификации клеток, проходящих через трансдифференцировку, и приобретении некоторых мезенхимных признаков у клеток. *Ef-klf1/2/4*, вероятно, участвует

непосредственно в трансдифференцировке, что подтверждают все полученные данные. В то же время, осталось и часть «непонятных» генов, таких как *Ef-tcf24* и *Ef-prdm9*, данных по которым оказалось сравнительно мало и они противоречивы. В конечном счете, удалось с 11 генов уменьшить число вероятных регуляторов трансдифференцировки до приблизительно 4, что позволяет уже проводить более сложные и трудоемкие эксперименты, такие как выключение генов с помощью методов нокдауна.

ВЫВОДЫ

1. Разработан и реализован на кросс-платформенном языке Python алгоритм для улучшения и финализации *de novo* сборки транскриптомов, получаемых с помощью стандартных программ-сборщиков (SPAdes, Trinity). Данный алгоритм апробирован на 3 видах голотурий и показал значительное повышение полноты восстановленности транскриптов с одновременным уменьшением числа последовательностей и отсутствием потерь белок-кодирующих участков мРНК. Разработанный алгоритм позволяет уменьшить, по сравнению с аналогами, потребляемые во время финализации сборки вычислительные ресурсы.

2. Разработан и реализован на языке Python модифицированный метод быстрого поиска ортологов, обладающий большей чувствительностью и гибкостью по сравнению с классическим, реципрокным методом поиска.

3. Разработан скрипт для автоматического подбора лучших комбинаций олигонуклеотидов для кПЦР среди большого множества вариантов, особенно важный в случае мультиплексных реакций.

4. Выявлено 11 генов транскрипционных факторов, экспрессия которых увеличивается в период трансдифференцировки при регенерации кишки у голотурии *Eupentacta fraudatrix*: *Ef-elf*, *Ef-prdm9*, *Ef-egr*, *Ef-klf1/2/4*, *Ef-snai2*, *Ef-tcf24*, *Ef-msc*, *Ef-id*, *Ef-gata3*, *Ef-pcgf2* и *Ef-tbx20*.

5. Показано, что экспрессия генов ряда транскрипционных факторов (*Ef-prdm9*, *Ef-klf1/2/4*, *Ef-snai2*, *Ef-tcf24*, *Ef-msc*, *Ef-id*, *Ef-gata3*, *Ef-pcgf2* и *Ef-tbx20*), участвующих в регенерации, имеет достаточно малую изменчивость между разными особями одной популяции. Это говорит о том, что, несмотря на вариабельность протекания регенерации у разных особей, данные гены важны для успешного прохождения всех стадий восстановления.

6. Описаны предположительные роли всех выявленных транскрипционных факторов в регенерации кишки у *E. fraudatrix*.

7. Установлено, что *Ef-klf1/2/4*, *Ef-prdm9*, *Ef-snai2* и *Ef-id* по динамике экспрессии, распределению продуктов в тканях зачатка и предполагаемым функциям являются наиболее вероятными регуляторами процесса трансдифференцировки целомического эпителия в энтероциты при регенерации кишки у голотурии *Eupentacta fraudatrix*.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах из списка ВАК:

1. Dolmatov I.Yu., Afanasyev S.V., **Бойко А.В.** Molecular mechanisms of fission in echinoderms: transcriptome analysis // PLOS ONE. 2018. Vol. 13, № 4. Article no. e0195836.
2. **Бойко А.В.**, Girich A.S., Eliseikina M.G., Maslennikov S.I., Dolmatov I.Y. Reference assembly and gene expression analysis of *Apostichopus japonicus* larval development // Scientific Reports. 2019. Vol. 9, № 1. Article no. 1131.
3. **Бойко А.В.**, Girich A.S., Tkacheva E.S., Dolmatov I.Y. The *Eupentacta fraudatrix* transcriptome provides insights into regulation of cell transdifferentiation // Scientific Reports. 2020. Vol. 10, № 1. Article no. 1522.

В материалах конференций:

4. **Бойко А.**, Girich A., Eliseikina M., Maslennikov S., Dolmatov I. Dinamic changes of gene expression during larval development of holothurian *Apostichopus japonicus* // Proceedings of the 16th International Echinoderm Conference. Auckland: Magnolia Press, 2019. P. 150.
5. **Бойко А.**, Dolmatov I. Transcriptome assembling of non-model organisms on example holothuria *Eupentacta fraudatrix* // Systems Biology and Bioinformatics (SBB-2019): 11th International Young Scientists School: Abstracts. Novosibirsk: ICG SB RAS, 2019. P. 12.
6. **Бойко А.В.**, Долматов И.Ю. Поиск регуляторов клеточной трансдифференцировки в процессе регенерации кишки у *Eupentacta fraudatrix* (D'yakonov & Baranova, 1958) (Holothuroidea, Dendrochirotida) // Национальный научный центр морской биологии им. А.В. Жирмунского ДВО РАН. Годичная научная конференция. Сборник материалов. Владивосток: НИЦМБ ДВО РАН, 2019. С. 20–26.
7. **Бойко А.**, Dolmatov I. The first insights into regulation of cell transdifferentiation during gut regeneration in *Eupentacta fraudatrix* // Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB–2020), 12th International Multiconference, 6–10 July 2020, Novosibirsk, Russia: Abstracts. Novosibirsk: ICG SB RAS, 2020. P. 19.
8. **Бойко А.В.**, Dolmatov I.Yu. The role of transcription factors in transdifferentiation during regeneration in sea cucumbers // VIII Международная научно-практическая конференция молодых ученых: биофизиков, биотехнологов, молекулярных биологов и вирусологов – 2021: Сб. тез. / АНО «Иннов. центр Кольцово». Новосибирск : ИПЦ НГУ, 2021. С. 266.

БОЙКО
Алексей Вячеславович

**ПОИСК ТРАНСКРИПЦИОННЫХ ФАКТОРОВ, РЕГУЛИРУЮЩИХ
ТРАНСДИФФЕРЕНЦИРОВКУ КЛЕТОК ПРИ РЕГЕНЕРАЦИИ КИШКИ
У ГОЛОТУРИИ *EUPENTACTA FRAUDATRIX***

1.5.23. Биология развития, эмбриология

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук