Федеральное государственное бюджетное учреждение науки
«Национальный научный центр морской биологии им. А.В. Жирмунского»
Дальневосточного отделения Российской академии наук

На правах рукописи

Бойко Алексей Вячеславович

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

1.5.23. Биология развития, эмбриология

Диссертация на соискание учёной степени
кандидата биологических наук

Научный руководитель:

доктор биологических наук,
чл.-корр. РАН
Долматов Игорь Юрьевич

Владивосток – 2022

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ....4
1. ОБЗОР ЛИТЕРАТУРЫ....11
1.1. Понятие регенерации и клеточные источники восстановления утраченных органов. 11
1.2. Строение голотурий....12
1.3. Регенерация у голотурий....15
1.4. Трансдифференцировка клеток....16
1.5. Анализ транскриптомов немодельных объектов и связанные с этим трудности....20
1.6. Регенерация кишки у *Eupentacta fraudatrix*....27
2. МАТЕРИАЛЫ И МЕТОДЫ....32
2.1. Сбор и содержание животных....32
2.2. Выделение РНК....32
2.3. Синтез кДНК для кПЦР....34
2.4. Разработка праймеров и зондов для кцПЦР, кПЦР и клонирования....35
2.5. Секвенирование поли-А РНК....36
2.6. Сборка транскриптома *de novo*....38
2.7. Анализ дифференциальной экспрессии генов....39
2.8. Аннотация белок-кодирующих последовательностей транскриптома....40
2.9. Оценка экспрессии генов с помощью кцПЦР....42
2.10. Валидация результатов РНК-секвенирования....44
2.11. Гибридизация *in situ*....45
2.12. Применяемые математические методы и языки программирования....46
3. РЕЗУЛЬТАТЫ....47
3.1. Подбор и валидация методов сборки транскриптома *de novo* и аннотации по базам данных белков....47
3.1.1. *De novo* сборка транскриптома....47
3.1.2. Поиск ортологов....51
3.2 Поиск генов-кандидатов в регуляторы клеточной трансдифференцировки у *Eupentacta fraudatrix*....54
3.2.1. Секвенирование и сборка транскриптома *de novo*....54
3.2.2. Анализ дифференциальной экспрессии генов....57
3.2.3. Аннотация....60
3.2.4. Поиск кандидатов на роль регуляторов клеточной трансдифференцировки 61

3.2.5. Сеть сверхпредставленных биологических процессов и путей.....64
3.3 Анализ временно́й и пространственной динамики экспрессии генов-кандидатов.....69
3.3.1. Оценка экспрессии с помощью кцПЦР.....69
3.3.2. Локализация экспрессии генов транскрипционных факторов в зачатке кишки на разных стадиях регенерации.....72
4. ОБСУЖДЕНИЕ.....75
4.1. Подбор и валидация методов сборки транскриптома *de novo* и аннотации по базам данных белков.....75
4.1.1. *De novo* сборка транскриптома.....75
4.1.2. Поиск ортологов.....78
4.2. Поиск генов-кандидатов в регуляторы клеточной трансдифференцировки у *Eupentacta fraudatrix*.....80
4.2.1. Секвенирование и сборка транскриптома *de novo*.....80
4.2.2. Анализ дифференциальной экспрессии генов.....81
4.2.3. Поиск кандидатов на роль регуляторов клеточной трансдифференцировки 83
4.2.4. Сеть сверхпредставленных биологических процессов и путей.....84
4.3 Временно́е и пространственное распределение экспрессии генов-кандидатов.....89
ЗАКЛЮЧЕНИЕ.....100
ВЫВОДЫ.....104
СПИСОК ЛИТЕРАТУРЫ.....106
ПРИЛОЖЕНИЯ.....123
1. Праймеры для кПЦР.....123
2. Праймеры и зонды для кцПЦР.....124
3. Последовательности ампликонов, используемых в кцПЦР.....125
4. Последовательности праймеров для клонирования.....128
5. Последовательности отсеквенированных ампликонов, используемых в WMISH.....129
6. Корреляция оценок экспрессии отдельных генов.....135

ВВЕДЕНИЕ

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. Обычно считается, особенно для млекопитающих, что восстановление осуществляется за счет различных типов стволовых клеток [1]. Однако более детальное изучение регенерации на большом числе видов животных показало активное участие в морфогенезе не только стволовых клеток, но и специализированных клеток тканей, расположенных вблизи повреждения. Эти клетки проходят через де- или трансдифференцировку и дают начало специализированным клеткам восстанавливающихся органов. Процессы деспециализации, вероятно, возникли очень давно, поскольку даже у ряда видов Porifera регенерация осуществляется в результате трансдифференцировки дифференцированных клеток, без участия стволовых клеток (археоцитов) [2–4]. Более того, участие специализированных клеток в регенерации подтверждено и у более высокоорганизованных многоклеточных животных, таких как рыбы (регенерация сердца), земноводные (хрусталик) и млекопитающие (печень, легкие, поджелудочная железа) [5–8].

Иглокожие представляют собой интересные модельные объекты для изучения проблемы источников регенерации, поскольку наличие у них стволовых клеток (за исключением первичных половых клеток) до сих пор достоверно не установлено [9]. Многочисленные морфологические данные свидетельствуют, что регенерация у них протекает за счет специализированных клеток в результате их активной транс- или дедифференцировки [9–14]. Более того, недавние исследования показывают, что регенерация у некоторых видов иглокожих может

успешно осуществляться даже при подавлении митотической активности клеток, что также говорит против участия стволовых клеток в регенерации у данных животных [14,15].

Среди всех Echinodermata, наибольший интерес в качестве моделей для изучения регенерации представляют виды класса Holothuroidea. Эти животные распространены во всех районах Мирового океана [16] и встречаются на всех глубинах, от литорали до ультраабиссали [17,18]. Многие виды голотурий обладают хорошими способностями к регенерации и проявляют широкий спектр восстановительных реакций [10,19,20]. При этом один и тот же вид может демонстрировать несколько различающихся типов механизмов морфогенеза [21]. Восстановление у голотурий осуществляется в достаточно короткие сроки. Например, формирование кишки после ее полной утраты может занимать всего около месяца [10]. Кроме того, у этих животных имеется уникальный способ аутотомии – эвисцерация. Некоторые виды голотурий способны к самопроизвольному удалению внутренних органов в ответ на внешние раздражители. После эвисцерации происходит полное восстановление утраченных структур. Если у голотурии после выброса кишки сохраняются ткани пищеварительной системы, то регенерация кишечной трубки осуществляется за счет дедифференцировки, миграции, пролиферации и редифференцировки сохранившихся энтероцитов [22–25]. При полной утрате тканей энтодермального происхождения восстановление кишки происходит за счет трансдифференцировки мезодермальных клеток. Этот механизм описан у голотурии *Eupentacta fraudatrix* (D’yakonov & Baranova in D’yakonov, Baranova & Savel’eva, 1958) (Holothuroidea, Dendrochirotida). У данного вида формирование передней части кишки осуществляется в результате трансформации клеток целомического эпителия [12,26–28]. После эвисцерации эти клетки дедифференцируются, погружаются в соединительную ткань переднего зачатка кишки и трансформируются в энтероциты [12,13].

Трансдифференцировка представляет собой достаточно редкое явление, особенно у позвоночных животных. Чтобы спровоцировать ее, необходимо идти на различные методические ухищрения. Например, у эмбриона *Danio rerio* можно вызвать трансдифференцировку мезодермальных клеток в энтодермальные за счет индукции эктопической ко-экспрессии генов двух транскрипционных факторов – *oct4* и *sox3-2* [29]. Несмотря на теоретическую и практическую важность явления трансдифференцировки, ее механизмы до сих пор во многом не ясны. В связи с этим животные, у которых встречается «естественная» трансдифференцировка, могут быть хорошими модельными объектами для изучения механизмов трансформации клеток и перестройки работы генома.

Степень разработанности темы. В настоящее время механизмы регенерации у Echinodermata и, в частности, Holothuroidea хорошо описаны с морфологической точки зрения [9–15]. Только в последние 10 лет начали предприниматься попытки описать молекулярные механизмы регенерации различных систем органов, в основном на примере регенерации кишки у видов *Apostichopus japonicus* [30–32] и *Holothuria glaberrima* [33–35]. Однако несмотря на развитие молекулярных методов, работы все еще носят характер «вида из окна», то есть описывают не механизм, а разные аспекты его работы. Кроме того, регенерация кишки у данных видов происходит за счет дедифференцировки оставшихся частей пищеварительной системы, а не трансдифференцировки клеток других органов в отсутствие остатков кишки. Наличие трансдифференцировки в настоящий момент установлено только для одного вида голотурий, *E. fraudatrix*. У него достаточно полно описаны морфологические особенности трансформации клеток [12,19,26–28] и локализация и динамика экспрессии ряда генов, сопровождающих этот процесс. Однако достаточно глубокого и всестороннего анализа трансдифференцировки на молекулярном уровне нет [36].

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 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 внутрипопуляционная 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. 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.

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 во время регенерации кишки у голотурии *E. fraudatrix* было проведено с использованием технологии капельно-цифровой ПЦР и WMISH (whole mount in situ hybridization). Сборку и аннотацию транскриптомов проводили как с использованием 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 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 объём диссертации. Основной текст диссертации изложен на 122 страницах и состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, выводов, списка литературы. Также дано 6 приложений на 13 страницах. Работа содержит 21 иллюстрацию. 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. Кроме того, искренне благодарен Александру Гиричу за помощь в осуществлении одного из самых трудоемких этапов работы — проведению 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 Кухлевскому А.Д. 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).

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Понятие регенерации и клеточные источники восстановления утраченных органов

Регенерация как биологическое явление привлекла внимание ученых еще в 18 веке [37]. Одним из первых обобщающих трудов, посвящённых проблемам восстановления у животных, следует считать монографию Томаса Моргана «Регенерация» [38]. В своей работе он дал понятие термину «регенерация» как восстановление организмом утраченных частей тела. Несмотря на простоту формулировки, она до сих пор используется исследователями. Дальнейшее изучение регенерации привело к уточнению и детализации данного явления. Все восстановительные процессы теперь принято разделять на физиологическую и репаративную регенерацию. Первая включает случаи восстановления тканей и органов после естественного изнашивания в процессе жизнедеятельности организма, например обновление пула эритроцитов по мере их разрушения. Репаративная регенерация объединяет восстановительные морфогенезы, происходящие в ответ на воздействие деструктивных факторов внешней среды [39].

Репаративная регенерация может осуществляться несколькими способами, основными из которых являются эпиморфоз и морфаллаксис. Определение этим способам было впервые сформулировано Морганом (1901). Эпиморфоз и морфаллаксис являются наиболее распространенными способами восстановления в животном мире. Для них характерны глубокие перестройки в поврежденном органе или ткани, включающие миграцию, пролиферацию и значительные изменения в работе генных сетей клеток. Эти два способа как раз и обеспечивают самые впечатляющие проявления регенерации, такие как восстановление ампутированной конечности у земноводных или восстановлении всего животного из его части у планарий [40,41]. Здесь мы намеренно не концентрируемся на

отличительных признаках этих двух способов в связи с тем, что в чистом виде они не встречаются, дополняя в том или ином случае друг друга [42,43].

Одним из важнейших вопросов любого восстановительного процесса является вопрос об источниках материала для регенерации. Исследования 40-60-х годов XX века однозначно отвечали на него двумя словами, которые на слуху у многих в наше время — «стволовые клетки». Такое положение вещей является вполне понятным, поскольку у многих видов многоклеточных животных стволовые клетки вовлечены в процесс регенерации [1]. Самые активно регенерирующие организмы, такие как, например, виды типов Cnidaria, Porifera и Platyhelminthes в качестве клеточных источников регенерации часто используют как раз стволовые клетки (интерстициальные клетки, археоциты и необласты, соответственно). С другой стороны, параллельно накапливались данные о том, что восстановление может происходить за счет дедифференцировки специализированных клеток. Кроме того, у одних и тех же видов, в зависимости от типа повреждения, в регенерации могут участвовать как стволовые, так и дифференцированные клетки [2,3]. Так, даже у человека, имеющего стволовые клетки во многих органах и тканях, при повреждении легких, печени и поджелудочной железы задействуются механизмы дедифференцировки: клетки теряют специализированные структуры, входят в митотический цикл, а затем дифференцируются снова, формируя, таким образом, утраченные структуры [7]. У Echinodermata, несмотря на многочисленные попытки, в соматических тканях вообще не были выявлены стволовые клетки [9,19,36,44]. Регенерация у этих животных осуществляется за счет дедифференцировки или трансдифференцировки клеток оставшихся тканей [9–14].

1.2. Строение голотурий

Echinodermata и, в частности, голотурии, являются очень интересными объектами для изучения регенерации. Эти животные могут восстанавливать как

небольшие придатки, так и крупные отделы тела, например, после поперечного разрезания на две или три части. При этом регенерация осуществляется в достаточно сжатые сроки. Также удобство представителей этого типа многоклеточных состоит в широком распространении этих организмов, относительной легкости содержания и культивирования в искусственной среде и в их эволюционном положении. Echinodermata вместе с Hemichordata образуют группу Ambulacraria, которая является сестринской по отношению к хордовым животным. Следствием этого является наличие большого числа генов, генных каскадов и сигнальных путей, общих для них и млекопитающих. Важной особенностью иглокожих является широкая вариабельность способностей к восстановлению и механизмов его протекания не только в разных таксонах, но и даже на разных этапах онтогенеза у одного вида [10,19,20], что является удобным и интересным в случае изучения эволюции механизмов регенерации.

Представители класса Holothuroidea имеют вытянутое червеобразное тело с мягкими покровами. Это, в основном, донные организмы, хотя среди них встречаются виды, способные плавать [19]. Голотурии обитают на самых разных глубинах, от литорали и до 5000 и более метров. Наибольшая плотность популяций голотурий наблюдается на коралловых рифах тропических зон. Питаются голотурии планктоном и органическими остатками, собирая околоротовыми щупальцами донный ил и песок [45]. Первые голотурии появились ещё в силурийском периоде, около 40 млн. лет назад. На данный момент класс представлен более чем 1400 видами, разделёнными на 6 отрядов – Apodida, Elasipodida, Holothuriida, Persiculida, Molpadida, Synallactida и Dendrochirotida [46] (Рисунок 1А). В водах Российской Федерации обитает порядка 100 видов. Скелет голотурий не выражен, он представлен мелкими спикулами и окологлоточным известковым кольцом, что отличает этих животных от других Echinodermata. Тело голотурий разделяется на 10 сегментов – 5 радиальных (амбулакральных) и столько же интеррадиальных сегментов. Вдоль амбулакральных сегментов, или радиусов, расположено множество

амбулакральных ножек, служащих для передвижения. У разных видов часть амбулакральных ножек преобразована в различные выросты, шипы. Вокруг ротового отверстия располагается венчик щупалец (Рисунок 1Б). С внутренней стороны стенки тела по радиусам проходят нервные тяжи, радиальные гемальные сосуды, радиальные каналы амбулакральной системы и продольные мышечные ленты. Эти структуры формируют амбулакры. В переднем отделе животных расположен аквафарингеальный комплекс (АК), окружающий глоточный отдел пищеварительной трубки. АК свойственен только голотуриям [19]. В состав АК входят органы амбулакральной, нервной и гемальной систем. На кольцевом амбулакральном канале, ограничивающем АК сзади, располагаются каменистые каналы с мадрепоритами и мешкообразные полиевы пузыри. Кроме того, у представителей отряда Dendrochirotida есть мускулы-ретракторы, втягивающие АК внутрь тела [19].

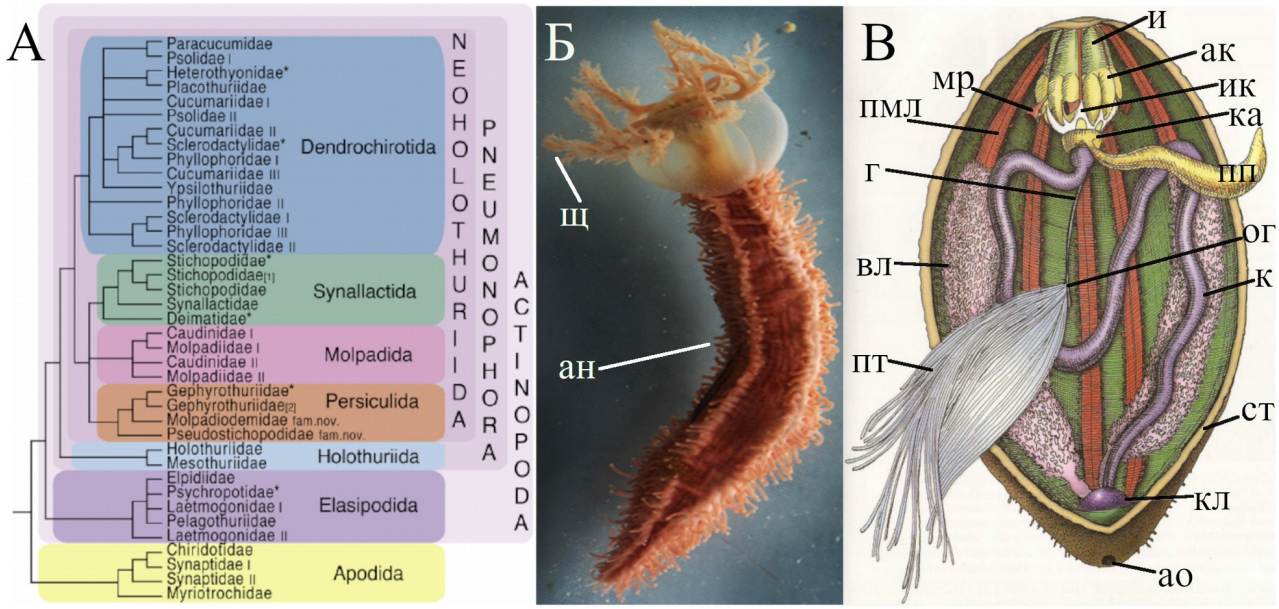

Рисунок 1. Филогенетическое дерево и морфология голотурий. А: Филогенетические отношения видов Holothuroidea (по: [46]). Б: Внешний вид голотурии *Eupentacta fraudatrix*. В: Схема внутреннего строения голотурий (по: [19]). щ - щупальца, ан - амбулакральные ножки, ак -аквафарингеальный комплекс, ао - анальное отверстие, вл - водное легкое, г - гонодукт, и - интроверт, ик - известковое окологлоточное кольцо, к - кишечник, ка - кольцевой амбулакральный канал, кл - клоака, мр - мускул-ретрактор, ог - основание гонады, пмл - продольная мышечная лента, пп - полиев пузырь, пт - половые трубочки, ст - стенка тела

Голотурии имеют обширный целом, вмещающий органы пищеварительной, половой и дыхательной систем. С внутренней стороны интеррадиусы выстланы

целомическим эпителием, в состав которого входят миоэпителиальные клетки, формирующие кольцевую мускулатуру тела. Глотка и часть пищевода лежат внутри АК, кишечник располагается в полости тела, прикреплённый к стенке тела мезентерием. На заднем конце тела кишечник заканчивается клоакой, от которой отходят парные водные легкие (Рисунок 1В). У представителей отряда Holothuriida здесь же имеются и Кювьеровы трубочки – специфические защитные органы.

1.3. Регенерация у голотурий

Многие виды голотурий обладают способностью к редкому виду аутотомии – эвисцерации. Она заключается в удалении внутренних органов в ответ на стрессовые воздействия, такие как общее ухудшение условий окружающей среды и/или нападение хищников. Данный процесс у разных видов протекает по-разному. У представителей Synallactida и Holothuriida выброс внутренностей происходит через разрыв в стенке клоаки (задняя эвисцерация). В результате животные теряют бóльшую часть кишечника и одно или оба водных легких, а АК, глотка, пищевод и клоака сохраняются [22,23]. После повреждения происходит развитие двух зачатков – на клоаке и оборванном конце пищевода. Регенерация кишечника осуществляется за счет роста этих зачатков навстречу друг другу по краю мезентерия и объединения в единую пищеварительную трубку. Кишечный эпителий формируется за счет энтероцитов сохранившихся частей пищеварительной системы, клоаки и пищевода. Основным механизмом регенерации кишечной выстилки является эпителиальный морфогенез. Энтероциты дедифференцируются, митотически делятся и мигрируют в составе эпителия в соединительную ткань соответствующего зачатка [47].

У ряда видов Dendrochirotida внутренности удаляются через передний конец тела (передняя эвисцерация). В результате животные теряют АК и всю пищеварительную систему, за исключением клоаки [26–28]. Регенерация кишки

осуществляется, как и в случае задней эвисцерации, за счет формирования и роста двух зачатков. Задний зачаток отрастает от клоаки, а передний – от развивающегося АК. Задняя часть кишки развивается, как и у представителей Synallactida и Holothuriida, за счет энтероцитов клоаки. В переднем зачатке закладка кишечной выстилки происходит по-иному. Поскольку в передней части голотурий утрачиваются все ткани энтодермального происхождения, формирование кишечного эпителия здесь происходит из мезодермальных клеток (целомического эпителия) за счет их трансдифференцировки. Клеточные механизмы регенерации описаны только для одного вида – *Eupentacta fraudatrix* [12] (см. ниже).

1.4. Трансдифференцировка клеток

Любой морфогенез характеризуется изменением специализации клеток. В зависимости от процесса и исходного состояния, клетки могут подвергаться дедифференцировке, редифференцировке или трансдифференцировке. Все эти процессы изменения клеточной судьбы тесно связаны между собой и их можно объединить аналогией, предложенной Уоддингтоном в 1942 году, названной впоследствии «эпигенетический ландшафт» (Рисунок 2) [48].

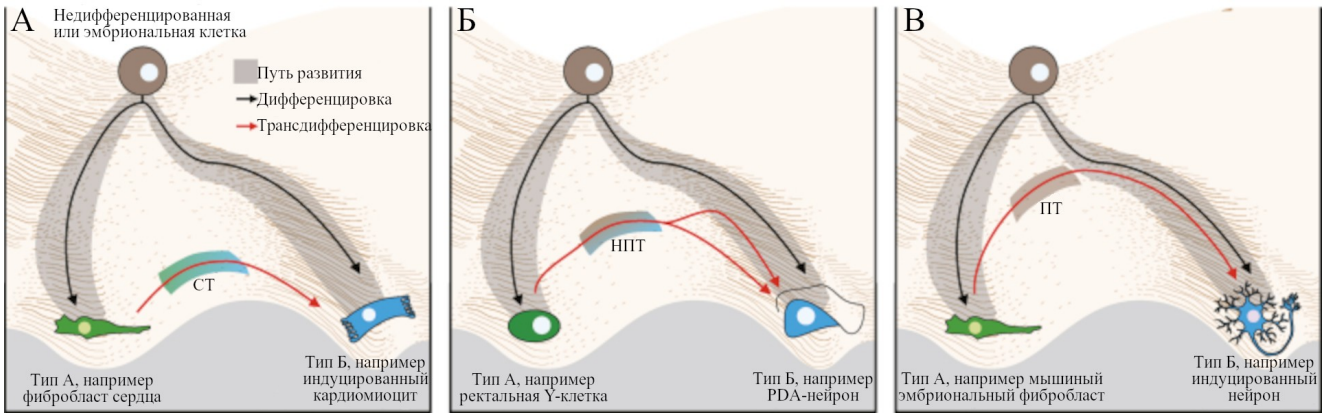

Рисунок 2. Эпигенетический ландшафт применительно к дифференцировке клеток (по: [48]). СТ - смешанный тип, НПТ - неспецифический промежуточный тип, ПТ - прогениторный тип.

Данная аналогия хоть и была предложена изначально для визуализации процесса развития, но также подходит для понимания клеточной судьбы и

дифференцировки клеток. Суть ее сводится к влиянию на некий объект, например, клетку, внешних факторов, предопределяющих путь по холмистому рельефу. Данный процесс детерминации пути и движения по нему и является дифференцировкой, в то время как клеточная судьба — установленная окончательно специализация клетки, то есть информация о типе клетки, в которую превратится клетка-предшественник в результате дифференцировки. В ходе развития организма можно выявить своеобразные точки равновесия, в которых меняется «потентность» клеток, то есть их потенциал к производству разных типов клеток.

Гипотеза эпигенетического ландшафта также предполагает важную особенность дифференцировки — однонаправленность и необратимость. Как шарик в конце пути не может снова подняться на вершину холма, так и терминально дифференцированные клетки не способны к де-специализации. Однако это верно не везде и не всегда. Специализированное состояние клеток может быть изменено при трансдифференцировке — процессе прохождения терминально дифференцированной клеткой пути развития в обратном направлении с последующим превращением в клетки другого типа. Например, трансформирующиеся в нейроны фибробласты мыши на определенном этапе демонстрируют профиль экспрессии генов, сходный с таковым нейробластов [49]. В то же время, такое переходное состояние может быть похоже на нечто промежуточное между начальным и конечным типами клеток, или вовсе не быть похожим по профилю экспрессии генов на какой-то определенный тип клеток [50].

Кроме того, описана и прямая трансдифференцировка. Для нее характерно отсутствие выраженных промежуточных состояний между начальным и конечным состояниями клетки. Так, в случае индукции эктопической экспрессии транскрипционного фактора (ТФ) ELT-7 у нематоды *Caenorhabditis elegans* краевые клетки глотки (pharyngeal margin cells) претерпевают трансдифференцировку в клетки кишечника [51]. У млекопитающих гепатоциты

после повреждения печени могут трансформироваться в холангиоциты [52]. Нечто подобное наблюдается и при трансдифференцировке фибробластов в мышечные клетки при экспрессии MyoD [53].

Клетка, в зависимости от глубины дедифференцировки, меняется в широких пределах не только морфологически, но и в плане экспрессии генов и пространственной организации генома. Так, исследования с помощью метода Hi-C, позволяющего выявлять пространственные отношения между участками генома, показали, что структура топологически-ассоциированных доменов (topologically associating domain), участков хроматина с общими свойствами, у плюрипотентных клеток значительно отличается от соматических и меняется в процессе дифференцировки [54].

Важнейшую роль в специализации, де- и трансдифференцировке клеток играют ТФ [55]. Например, с помощью четырех ТФ — OCT4, SOX2, KLF4 и MYC, была вызвана дедифференцировка фибробластов мыши, трансформировавшая специализированную клетку в близкое к плюрипотентному состояние [56]. Для запуска трансдифференцировки необходимы один или несколько ТФ, причем набор ТФ для каждого направления трансформации является специфичным [49,57–60].

В то же время, модификация уровня или времени экспрессии генов ТФ не является обязательным условием для репрограммирования клеток и трансдифференцировки. Как упоминалось ранее, конформация хроматина сильно зависит от состояния клетки, однако логично предположить, что верно и обратное. При этом, конформация хроматина мало того, что в какой-то степени стохастически изменчива [55], так еще и находится под контролем ряда белков, которые могут быть активированы воздействием на сигнальные пути с помощью различных малых молекул в обход или с вовлечением ТФ. Известным примером является 5-азацитидин, который ингибирует ДНК-метилтрансферазу, вызывая уменьшение уровня метилирования некоторых участков генома, в том числе так называемых генов плюрипотентности, что приводит к их активации и

репрограммированию клетки. Тем не менее, имеющиеся на сегодняшний момент данные говорят о небольшом влиянии малых молекул самих по себе на репрограммирование клеток по сравнению с ТФ. В конечном счете, малые молекулы лишь запускают некие генные сети, в которых все равно участвуют те же специфичные ТФ, под чьим контролем уже и происходит репрограммирование и трансдифференцировка [61].

Трансдифференцировка вовлечена в регенерацию у достаточно широкого спектра видов. Так, даже в классическом примере регенерации конечности у амфибий, бластема, вероятно, сформирована клетками, проходящими через трансдифференцировку [50]. Другим примером является регенерация хрусталика после его полного удаления у тритонов, во время которой пигментные эпителиальные клетки радужки превращаются в эпителиальные клетки хрусталика [6]. Достаточно разнообразны механизмы трансдифференцировки у иглокожих [21]. У морских лилий надсемейства Himerometroidea клетками-предшественниками являются мезенхимные клетки эктодермального происхождения [62]. Они могут трансформироваться в двух направлениях и давать начало эпидермису и энтероцитам кишки [14,63]. У морских лилий надсемейства Antedonoidea и голотурий трансдифференцировке могут подвергаться клетки целомического эпителия [12,64].

Удивительно, но даже у губок встречается трансдифференцировка клеток, хотя долгое время считалось, что клеточным источником регенерации у них являются стволовые клетки, археоциты. Причем это подтверждено для разных видов из разных классов Porifera [2–4]. Наличие механизмов трансдифференцировки у представителей базальных групп многоклеточных животных указывает на древность этого явления. Таким образом, можно заключить, что трансдифференцировка клеток во время регенерации является важным процессом и задействуется у различных таксонов и для разных типов повреждений.

1.5. Анализ транскриптомов немодельных объектов и связанные с этим трудности

В настоящее время секвенирование, в том числе секвенирование транкриптома, плотно вошло в методологию многих направлений биологии [65]. Это неудивительно, так как это единственный способ выяснить последовательности транскриптов определенных генов. Кроме того, подобные методы позволяют оценить экспрессию всего пула активных генов, не имея никакой информации ни об их последовательностях, ни даже семействах, к которым эти гены принадлежат. При этом, что немаловажно, полученные оценки экспрессии генов подтверждаются с высокой точностью другими, более устоявшимися подходами, такими как количественная полимеразная цепная реакция (кПЦР) или ДНК-микрочипы [66–69]. Также стоит заметить, что трудоемкость подготовки образца к секвенированию и стоимость секвенирования кардинально снизилась за последние 5-10 лет [70]. Все это привело к тому, что в настоящее время секвенировано большое число образцов широкого спектра видов животных. Многие из них являются «немодельными», то есть, в контексте биоинформатики, не имеющих хорошо изученного, хотя бы на уровне генов, генома или даже транскриптома. В связи с этим, у исследователей, причем, как это ни парадоксально, не только тех, что секвенировали транскриптом немодельного вида, но и тех, кто занимается сравнительной геномикой, филогенетикой или просто использует для своей работы данный транскриптом, возникает ряд проблем на разных этапах его анализа.

В первую очередь стоит остановиться на процессе сборки транскриптома, то есть получения последовательностей транскриптов из сырых данных секвенирования. Несмотря на появление секвенаторов от Oxford Nanopore, способных прочесть целиком транскрипт, причем даже в форме РНК, то есть без необходимости обратной транскрипции, основным способом секвенирования РНК являются технологии Illumina и, изредка, Ion Torrent, результатом чего является

большое число — порядка десятков и сотен миллионов — короткий прочтений, длиной около 50-150 нуклеотидов. После этапов очистки этих прочтений от технических участков и регионов плохого качества, существует три возможных пути дальнейшей работы. Самый простой путь в основном подходит для модельных организмов, где уже известны последовательности транскриптов, в результате чего сборка в принципе не нужна и можно сразу переходить к оценке экспрессии. Второй путь требует сборки. При этом необходимо привлечение дополнительных данных в виде известного генома исследуемого вида, причем с размеченными участками генов и их экзон-интронной структуры. Такой вариант сборки также называют в англоязычной литературе «genome-guided assembly» или «alignment-based assembly», так как в большинстве случаев идет простое картирование прочтений на гены и дальнейшее их объединение в целый транскрипт, включая различные сплайс-варианты.

Последний и самый сложный путь — так называемая сборка *de novo*. Этот способ применяется, когда нет готового генома исследуемого вида. В настоящее время чаще всего применяются программы-сборщики на основе графов де Брюйна [71]. Также используется гибридная сборка, когда в наличии не только короткие прочтения, но и полные транскрипты, как, например, с секвенаторов третьего поколения Oxford Nanopore или PacBio. Основная проблема сборки *de novo* — это фрагментарные транскрипты, поскольку вместо получения ожидаемых нескольких десятков тысяч полных транскриптов, формируется более сотни тысяч, а то и несколько сотен тысяч последовательностей. Многие из них либо являются ошибкой сборки, то есть «мусором», либо фрагментами одного и того же транскрипта [72,73]. Этот момент затрудняет исследование немодельных видов, так как влечет за собой сложности в оценке экспрессии генов, поиске гомологов и влияет на все дальнейшие этапы анализа транскриптома [74].

С появлением секвенаторов третьего поколения, то есть тех, что могут считывать фрагменты более нескольких тысяч нуклеотидов, эта проблема должна уйти в прошлое. Однако и тут есть сложность, поскольку покрытие при

секвенировании РНК очень неравномерное в связи с разным уровнем экспрессии генов. Так, обычно около 20% генов дают до 80% всей мРНК [75,76]. В итоге с учетом того, что производительность секвенаторов третьего поколения не слишком высока (например, у Oxford Nanopore — до 15 гигабаз на одну ячейку), мы получаем крайне малое число прочтений на бóльшую часть генов, а низкоэкспрессирующиеся гены могут вообще не иметь прочтений. Эту проблему можно решить, применив ферментативную нормализацию кДНК [77], но с некоторыми допущениями. Так, нормализация требует обратной транскрипции, а значит, и считывать интактную РНК не получится. Из этого следует также, что мы сможем получить данные только по мРНК, а также какой-то процент привнесенных нуклеотидных замен. К тому же будет опять много фрагментированных транскриптов, особенно если они были достаточно длинные, так как в процессе синтеза кДНК не всегда происходит обратная транскрипция мРНК целиком [78]. Кроме того, этот метод требователен к качеству РНК.

Аннотация является вторым по важности и сложности после сборки этапом анализа данных секвенирования транскриптома *de novo*. Она включает поиск ортологов, различных структурно-функциональных элементов (гены, регуляторные участки, сайты мутаций, теломерные и центромерные регионы, CpG-островки и т.д.) в геноме, нетранслируемые регионы (НТР) в мРНК, домены, повторы и сигнальные участки в белках. Аннотация, в конечном итоге, сводится к присуждению какому-либо участку биологической последовательности «ярлыка» и функции на основе некой эмпирической информации. Однако в контексте анализа транскриптома *de novo*, последовательности которого уже избавлены от НТР, мы будем пользоваться более узким понятием аннотации, которое подразумевает только выявление ортологов, то есть поиск белков, гомологичных или ортологичных анализируемым последовательностям, в доступных базах данных белков с целью получения информации о вероятном названии и функциях этих последовательностей,

Такая аннотация может быть выполнена с помощью большого числа программ, разница которых, в основном, сводится к типу выравнивания (локальное, глобальное или профильное) и типу базы. Локальное выравнивание представляет собой поиск сходных участков у двух анализируемых последовательностей, тогда как при глобальном требуется поиск оптимального выравнивания, включающего в себя целиком обе последовательности. Профильное выравнивание может быть как локальным, так и глобальным и производится с использованием вероятностных моделей Маркова, которые построены по множественному выравниванию гомологичных последовательностей, чаще всего белковых. Далее, с помощью алгоритма разбора Витерби и алгоритма «прохода вперед-назад» проводится поиск соответствия участка или целой последовательности модели Маркова [79]. Типы базы отличаются природой последовательностей (белки, нуклеиновые кислоты), формой (алфавит или вероятности). Такое разделение не всегда безусловно. Например, существуют вариации алгоритма BLAST, комбинирующие разные типы выравнивания и разные типы баз.

Наиболее простым способом является поиск гомологов с помощью выравнивания BLAST [80] или поиск доменов, с целью получения информации о гомологии по доменной архитектуре последовательности [81]. Для простой аннотации обычно оставляют самое лучшее совпадение по параметру e-value или bitscore. E-value равен вероятности найти последовательность данной длины в базе данного объема при условии случайной аминокислотной композиции всех участвующих последовательностей. Второй параметр является более информативным, так как включает в себя метрику сходства последовательностей по вероятностной матрице аминокислотных замен, например BLOSUM62 [82], и значение e-value. Такой способ подходит для поиска гомологов в терминах выравнивания, но не для поиска ортологов, так как зачастую, особенно при отсутствии в базе полного набора белков близкого вида, на одну последовательность из базы может приходится несколько лучших совпадений с

анализируемыми белками. Также путаницу вносят фрагментированные транскрипты или изоформы. Естественно также, что чем полнее восстановлены транскрипты, тем достовернее можно установить гомологию. Это указывает на прямую зависимость анализа транскриптома от качества сборки. При наличии большого числа неполных последовательностей одному и тому же гену могут быть гомологичны несколько фрагментов, в результате чего затем будет неясно по какому транскрипту оценивать экспрессию данного гена.

Вообще более ценной является информация об ортологичности, а не гомологичности генов, так как она позволяет более точно судить о функциях белковых продуктов. При этом бывают разные типы ортологий. База Ensembl использует следующие три типа — ортолог «один к одному», ортолог «один ко многим» и ортолог «многие ко многим» (Рисунок 3). Однако ортологические отношения нельзя установить с помощью простого выравнивания последовательностей, даже после сортировки или фильтрации по какому-то показателю. Для получения такой информации требуются либо специальные программы типа OrthoMCL или OrthoFinder, либо простые способы, вроде реципрокного BLAST поиска.

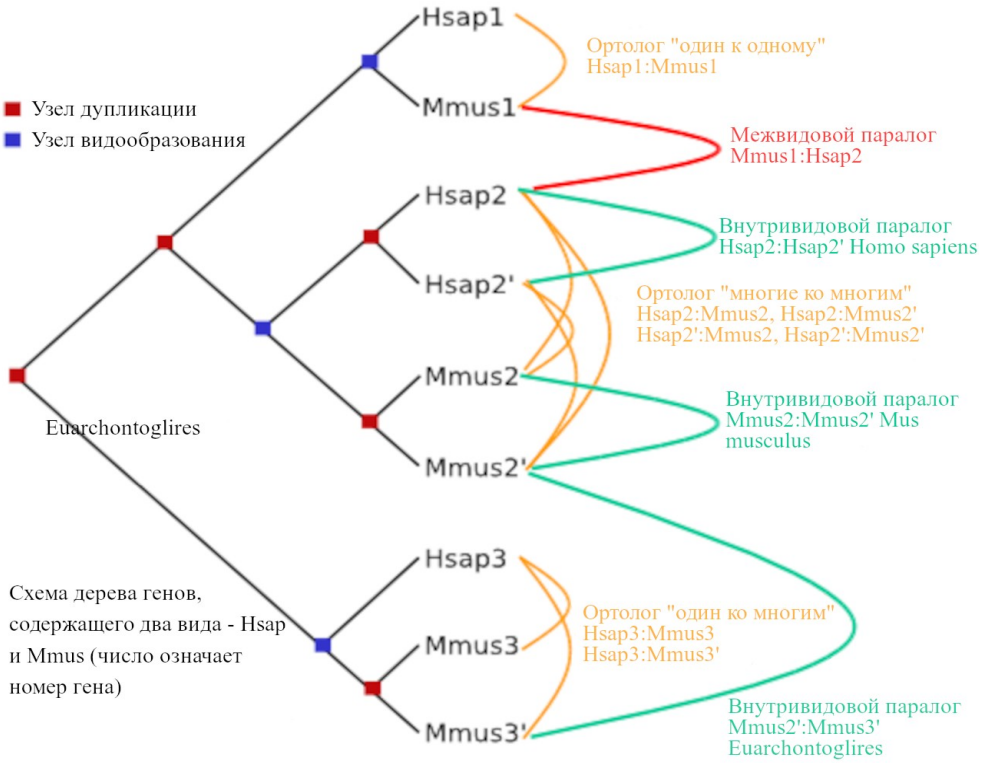

Рисунок 3. Типы гомологических отношений генов, принятые в базе Ensembl

В последнем случае выравнивают белки некоего вида на транскриптом и обратно, затем берут лучшие совпадения по e-value или оценке выравнивания и, если в обоих направлениях поиска совпадения идентичны, то считают такую пару ортологами. Основной недостаток данного подхода тот же, что и у простого BLAST-поиска, так как лучшая оценка выравнивания не говорит о том, что последовательности являются ортологами, но добавляется и еще один минус — большие потери информации [83]. В конечном счете, ошибочная аннотация, с одной стороны, ведет к неправильным выводам исследования, а с другой стороны, приводит к появлению в базах последовательностей неправильных названий генов, что сильно сказывается на других исследователях, занимающихся родственными видами. По этой причине аннотацию необходимо проводить по нескольким базам данных белков, а лучше даже белкам видов, причем основным должен быть некий модельный вид. При большой эволюционной дистанции между модельным и исследуемым видами ухудшается точность определения гомологов [84,85], но становится возможным получить информацию о процессах, в которых принимают участие гомологичные белки, а также ассоциированным болезням, что может помочь сформировать гипотезу и подкрепить ее.

Третьим основным этапом анализа транскриптома немодельных видов является оценка экспрессии генов. Этот этап напрямую зависит от сборки и аннотации, как упоминалось ранее. После получения тем или иным способом информации о числе прочтений, приходящихся на каждую последовательность, во многих исследованиях проводят оценку дифференциальной экспрессии и затем работают только с информацией об изменении экспрессии какого-то гена в каком-либо образце относительно выбранного контрольного образца [86]. Такой подход позволяет дать также и оценку достоверности изменения. Однако часто забывают, что такая относительная оценка экспрессии скрывает информацию о числе прочтений на ген. В результате этого в анализе возникает большое число генов, изменение экспрессии которых может быть очень большим, но в то же время базируется на том, что в каком-то из образцов было 2 прочтения на образец, а в

другом 10. Несмотря на 5-кратное изменение экспрессии, такое малое число прочтений на ген в принципе должно быть отфильтровано перед анализом дифференциальной экспрессии [30,87].

Хорошей практикой является разбиение генов на группы, например по процессам (GO аннотация), в которых они принимают участие. Такую информацию можно также совместить с экспрессией генов, например с помощью программы GSEA [88], что позволяет значительно уменьшить объем информации и строить гипотезы, не говоря уже, что поддержка у таких гипотез будет выше за счет вовлечения большого числа генов и их экспрессии. В случае с немодельными видами, особенно эволюционно дистантными от модельных, GO аннотация и дальнейший анализ должны быть выполнены максимально аккуратно, так как многие процессы нельзя воспринимать дословно. При этом оценка достоверности при поиске сверхпредставленных процессов или других GO терминов может быть низка, когда для исследуемого вида найдено мало точных ортологов генов модельного вида [83]. Кроме того, способ представления результатов такого анализа оказывает большое влияние на его понимание. Так, во многих работах результаты представлены в виде столбчатых диаграмм [86,89], в то время как более информативным является сетевой вид, позволяющий видеть взаимосвязи между разными процессами [83].

Таким образом, анализ имеющейся литературы показал, что существует ряд проблем анализа данных секвенирования немодельных организмов, включающих сборку, аннотацию и оценку экспрессии. Наличие или отсутствие качественно собранного и аннотированного генома в какой-то степени даже дает новое определение модельного вида. Отсутствие данных по геномам и транскриптомам изучаемых видов затрудняет молекулярные исследования этих животных, в том числе и поиск ортологов.

1.6. Регенерация кишки у *Eupentacta fraudatrix*

Голотурия *E. fraudatrix* представляет собой уникальный модельный объект для изучения различных аспектов регенерации [21]. Как и многие представители отряда Dendrochirotida она способна к передней эвисцерации, в результате чего происходит удаление АК и всей пищеварительной системы [26]. Было показано, что кишечная выстилка переднего зачатка кишки восстанавливается за счет клеток целомического эпителия в результате их трансдифференцировки [12]. В настоящее время это единственный вид голотурий, для которого показано наличие трансдифференцировки.

Морфологические особенности регенерации внутренних органов у *E. fraudatrix* после эвисцерации хорошо изучены [12,19,26–28]. Наиболее подробно процесс восстановления кишки на клеточном уровне был описан в работах Лейбсон [27] и Машанова и др. [12]. Регенерация внутренних органов у *E. fraudatrix* начинается с формирования соединительно-тканного утолщения на переднем конце животного, которое представляет собой зачаток АК. Он объединяет структуры амбулакров. Первая стадия регенерации кишки (3 сут после эвисцерации, СПЭ) характеризуется сформированным утолщением на краю мезентерия в передней части особи (Рисунок 4А). Данное утолщение представляет собой соединительно-тканный зачаток, покрытый клетками целомического эпителия. Преобразование и синтез внеклеточного матрикса происходит за счет тесного взаимодействия различных матриксных металлопротеиназ и их ингибиторов [36].

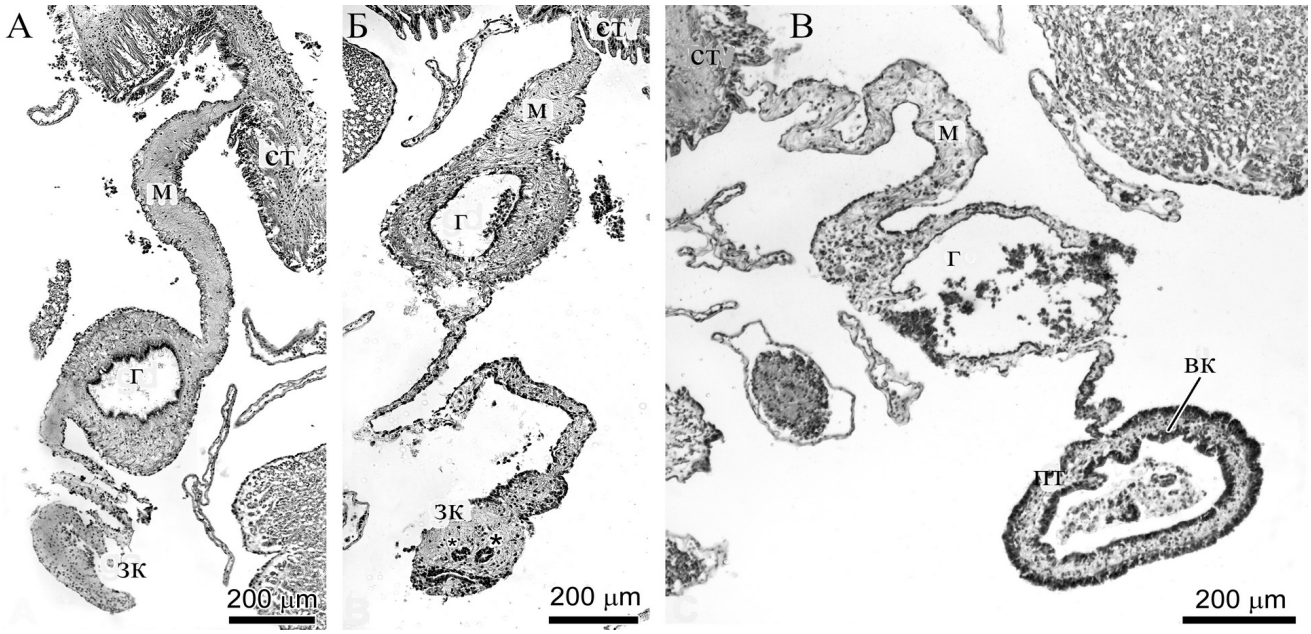

Рисунок 4. Структура кишки голотурии *Eupentacta fraudatrix* на разных стадиях регенерации. А: Соединительнотканное утолщение на краю мезентерия через 3 СПЭ. Б: Зачаток кишки через 6 СПЭ; звездочкой обозначены погрузившиеся в зачаток кишки клеток. В: Пищеварительная трубка через 12 СПЭ. ст - стенка тела, пт - пищеварительная трубка, зк - зачаток кишки, г - гонодукт, вк - выстилка кишки, м - мезентерий.

Целомический эпителий различается по своему строению в зависимости от расположения на зачатке. Клетки, покрывающие боковые стороны зачатка, сходны по строению с таковыми целомического эпителия интактных животных. Имеется хорошо выраженная базальная мембрана, контактирующая с собственно соединительной тканью зачатка. В эпителии здесь выявляются перитонеальные и миоэпителиальные клетки, встречаются нейроны и их отростки, формирующие базиэпителиальный нервный плексус. Перитонеоциты – это высокие клетки с развитыми микроворсинками и апикально расположенной ресничкой. Их ядро содержит крупное ядрышко. В цитоплазме выявляются развитый аппарат Гольджи, множество секреторных гранул, митохондрий, цистерн шероховатого эндоплазматического ретикулума (ШЭР), липидных гранул. В базальной части клеток располагаются пучки опорных промежуточных филаментов. Миоэпителиальные клетки содержат в базальной части хорошо развитый сократительный аппарат, состоящий из пучков миофиламентов. В их цитоплазме встречаются митохондрии, фагосомы, секреторные и липидные гранулы, свободные рибосомы.

На вентральной стороне зачатка целомический эпителий представлен уплощенными дедифференцированными клетками. Перитонеоциты утрачивают пучки промежуточных филаментов, а миоэпителиальные клетки – миофиламенты. Последние часто формируют специфические веретеновидные тела, которые могут достаточно долго сохраняться в цитоплазме. Эти веретеновидные тела являются хорошим маркером клеток целомического эпителия [12]. Базальная мембрана истончается и фрагментируется. Несмотря на дедифференцировку клетки целомического эпителия сохраняют межклеточные контакты [19]. В течение регенерации соединительно-тканный зачаток постепенно удлиняется и растет назад по краю мезентерия.

Следующая стадия регенерации (5-7 СПЭ) характеризуется активной трансдифференцировкой клеток (Рисунок 4Б). На вентральной стороне зачатка часть клеток целомического эпителия формирует складки и начинает погружаться в соединительно-тканное утолщение. В результате этого в зачатке кишки образуются многочисленные скопления клеток. Одновременно с погружением происходит изменение структуры клеток. У поверхности зачатка клетки сходны с таковыми целомического эпителия. Они сохраняют микроворсинки и реснички, а цитоплазма содержит крупные фагосомы и липидные гранулы. По мере погружения и трансдифференцировки ультраструктура клеток меняется. В цитоплазме таких клеток исчезают опорные тонофиламенты и миофиламенты. Реснички и микроворсинки исчезают. В цитоплазме увеличивается количество свободных рибосом, иногда встречаются веретеновидные структуры из миофиламентов.

На данной стадии трансдифференцировки в месте погружения происходит экспрессия генов *Ef-sox9/10* и *Ef-sox17* [36]. При этом транскрипты *Ef-sox9/10* выявляются как в погружающихся клетках целомического эпителия, так и в клетках, мигрировавших внутрь зачатка кишки. Ген *Ef-sox17* экспрессируется только в погружающихся клетках.

Через 8-9 СПЭ в апикальной части погрузившихся клеток появляются микровезикулы и секреторные вакуоли, характерные для энтероцитов кишки. Трансформирующиеся клетки сохраняют межклеточные контакты, то есть не претерпевают эпителио-мезенхимную трансформацию (ЭМТ) [12,19]. Соединительно-тканное утолщение продолжает расти назад по краю мезентерия, в результате чего передний зачаток кишки удлиняется. Скопления клеток утрачивают связь с целомическим эпителием. Они объединяются и формируют кишечный эпителий. На данной стадии он представляет собой короткую трубку, слепо замкнутую с обеих сторон. Клетки продолжают дифференцироваться.

Через 10 СПЭ в кишечном эпителии начинают формироваться неглубокие складки, клетки приобретают вытянутую форму, появляются апикальные микроворсинки. В цитоплазме равномерно распределены гранулы средней электронной плотности, фагосомы, цистерны ШЭР, митохондрии. Появляется базальная мембрана, подстилающая кишечный эпителий [19]. В этот период в клетках кишечного эпителия выявляются транскрипты *Ef-sox9/10* и *Ef-tensilin3*. В целомическом эпителии формирующейся кишки экспрессируются гены матриксных металлопротеиназ и *Ef-sox17* [36].

Через 12 СПЭ имеется хорошо оформленный кишечный эпителий, который занимает всю длину переднего зачатка кишки (Рисунок 4В). Передний зачаток продолжает расти назад. Клетки кишечного эпителия постепенно дифференцируются и приобретают структуры характерные для энтероцитов, такие, как многочисленные секреторные вакуоли и апикальнаые микроворсинки.

Одновременно с развитием переднего зачатка формируется и задний зачаток кишки. Он отрастает от клоаки и удлиняется по краю мезентерия вперед. Его выстилка формируется за счет врастания в него кишечного эпителия клоаки. Объединение зачатков происходит через 15-20 СПЭ, в результате чего образуется единая пищеварительная трубка [12].

Таким образом, в настоящий момент молекулярные механизмы регенерации кишки у *E. fraudatrix* практически не описаны. Установлены места локализации

экспрессии только нескольких генов, большинство из которых связаны с регуляцией обновления внеклеточного матрикса и, очевидно, не задействованных в механизмах трансдифференцировки. *Ef-sox9/10*, хотя и экспрессируется в месте погружения целомического эпителия, но, скорее всего, участвует в дифференцировке энтероцитов [36]. Гены, которые могли быть задействованы в процессе трансдиффернцировки, не установлены.

2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Сбор и содержание животных

Половозрелые особи голотурий *E. fraudatrix* и *A. japonicus* были выловлены в заливе Петра Великого Японского моря и содержались в аквариумах с аэрируемой морской водой объемом 3 кубических метра. Температура воды составляла 16 °С. Эвисцерацию у *E. fraudatrix* вызывали инъекцией 2% KCl в полость тела [83]. После удаления внутренностей животных помещали в тот же аквариум, где их содержали до полной регенерации внутренних органов. Вода в аквариуме менялась ежедневно. Нерест *A. japonicus* был вызван температурной стимуляцией [90]. Зиготы переносили в 370-литровый аквариум с водой комнатной температуры. Личинок кормили микроводорослями *Dunaliella salina*. Взрослые особи *Cladolabes schmeltzii* были собраны в заливе Ня-Чанг Южно-Китайского моря вблизи южной части острова Хон-Че [89].

2.2. Выделение РНК

Для выделения РНК *E. fraudatrix* и дальнейшего ее секвенирования были взяты зачатки кишки через 3 СПЭ, 5-7 СПЭ и 10 СПЭ, а также участок передней части интактной кишки (норма). Во всех случаях использовали ткани от 10 особей, объединенных в две группы по 5 штук в каждой. Материал фиксировался в 3 мл RNAlater (Sigma, USA) в течение 24 час при 4 °С, и затем ткани хранились в RNAlater при –20 °С до выделения РНК. Перед выделением РНК ткань была промыта морской водой. Затем образцы были помещены в 1 мл монофазного водного раствора фенола и гуанидинизотиоцианата ExtractRNA (Евроген, Россия) вместе с пятью металлическими шариками. Измельчение ткани и лизирование клеток проводили в гомогенизаторе TissueLyserLT (Quagen, Germany) в течение 5 мин на максимальной скорости. Затем, следуя протоколу ExtractRNA (Евроген,

Россия), в течение 10 мин инкубировали при комнатной температуре для диссоциации нуклеопротеидных комплексов и окончательного лизирования клеток. Далее к лизату было добавлено 100 мкл 1-бром-3-хлорпропан, смесь перемешивали встряхиванием пробирки в течение 10 сек и держали при комнатной температуре 3 мин с периодическим, раз в минуту, встряхиванием пробирки. После этого суспензию центрифугировали при 4 °C с ускорением 14000g в течение 15 мин для разделения фаз. Далее верхнюю, водную, фазу перемещали в чистую пробирку и добавляли к ней 500 мкл 100% изопропанола с последующим встряхиванием и инкубацией при комнатной температуре в течение 10 мин. На этом этапе в смесь был добавлен 1 мкл соосадителя SatelliteRed (Евроген, Россия). Затем образец центрифугировали при комнатной температуре и ускорении 14000g в течение 10 мин. Далее, после удаления супернатанта, к осадку был добавлен 1 мл 70% этанола и затем образец центрифугировали. как описано выше. Далее супернатант был удален и осадок высушен при 40 °C в течение 5 мин с последующим растворением осадка при 37 °C в течение 20 мин в 20 мкл смеси, содержащей 10 мМ Tris-HCl (pH 7,5), 2,5 мМ $MgCl_2$, 0,1 мМ $CaCl_2$ и 0,1 ед./мкл ДНКазы I (Thermo Scientific, США) с последующей чисткой на магнитных частицах Ampure XP (Beckman-Coulter, США) по протоколу производителя.

Выделение РНК *A. japonicus* и *C. schmeltzii* проводили по той же методике. При этом было взято по 100 особей на 4-х личиночных стадиях *A. japonicus* — бластула, гаструла, аурикулярия и пентактула. Для исследования транскриптома у вьетнамской голотурии были взяты ткани из средней части тела у трех особей для контроля и также у трех особей в процессе поперечного деления при бесполом размножении.

В случае выделения РНК для постановки кПЦР и капельно-цифровой ПЦР (кцПЦР) использовали предыдущий протокол с небольшими модификациями. Зачатки кишки получали от особей через 3, 5, 7 и 10 СПЭ, по 5 особей на каждый срок регенерации. Ткани предварительно промывали буфером CMFSS (10,7 mM KCl, 436,3 mM NaCl, 21,1 mM Na_2HPO_4, 16,7 mM глюкозы, 12 mM HEPES, pH 7,7)

и гомогенизировали при 36 °С в течение 10 мин в 100 мл лизирующего раствора с помощью стерильного пестика. Лизирующий раствор содержал 10 мМ Tris-HCl (pH 8,0), 10 мМ NaCl, 30 мМ $MgCl_2$, 20 ед. коллагеназы I типа (Gibco, США). Выделение проводили для каждой особи отдельно в случае кцПЦР. Для кЦПР использовали 5 особей на стадию для каждого из 3 повторов, причем ткани 5 и 7 СПЭ смешивали вместе.

РНК чистили с помощью магнитных частиц Ampure XP (Beckman-Coulter, США) по протоколу производителя в соотношении объемов частиц к раствору РНК 1/1 и элюировали при 35 °С в течение 5 мин с помощью 20 мкл 1X Tris-EDTA буфера (pH 8,0). Перед синтезом кДНК были проверены концентрация каждого образца на флуориметре Qubit 3.0 (Thermo Scientific, США), качество РНК с помощью капельного электрофореза Experion (Bio-Rad, США) с чипами HighSens и чистота на спектрофотометре BioSpec-nano (Shimadzu, Япония).

2.3. Синтез кДНК для кПЦР

Среди 5 образцов РНК *E. fraudatrix* для каждой стадии регенерации были выбраны 3 с наилучшими показателями качества, чистоты и количества РНК. Синтез проведен с использованием 1 мкг тотальной РНК и набором Mint (Евроген, Россия) по стандартному протоколу производителя на амплификаторе C1000 Touch (Bio-Rad, США) с oligo-dT праймером. Полученный раствор одноцепочной кДНК разводили трехкратно водой I типа, смешивали с 10X реакционным буфером и инкубировали с 1 ед. РНКазы H (Thermo Scientific, США) в течение 1 часа при 15 °С. Далее раствор чистили магнитными частицами AmpureXP, как указано в предыдущем разделе, и элюировали в 60 мкл 1X Tris-EDTA буфера (pH 8,0).

2.4. 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*. Среди них было 15 генов ТФ и гены *tubb* и *ef1a* в качестве референсных. Олигонуклеотиды подобраны в соответствии с рекомендациями по оценке экспрессии генов с помощью метода капельно-цифровой ПЦР (Droplet Digital PCR Applications Guide, Bio-Rad, USA) — длина ампликона в диапазоне 60-190 нуклеотидов, содержание GC 45-65% для праймеров и 45-75% для зондов, расчетная температура плавления праймеров 57-63 °C и зондов на 5-10 °C выше, длина праймеров 18-24 нуклеотида, длина зонда 16-24 нуклеотидов, отсутствие повторов цитозина или гуанина с длиной более 3, преобладание содержания цитозина над гуанином, наличие гуанина или цитозина на 3`-конце праймера, отсутствие гуанина на 5`-конце зонда. Подбор праймеров и зондов проводили с помощью веб-сервиса PrimerQuest Tool (IDT, USA). Для корректировки расчетной температуры плавления с помощью формулы из статьи SantaLucia [91] введены поправки на концентрацию моновалетных катионов солей (50 мМ), дивалентных катионов солей (3,8 мМ), концентрацию дНТФ (дезоксинуклеозидтрифосфат) (0,8 мМ), концентрацию каждого праймера (900 нМ) и концентрацию зонда (250 нМ). Когда это было возможно, предполагаемые ампликоны располагали на известной последовательности открытой рамки считывания (ОРС) мРНК не ближе 100 нуклеотидов от 3`-конца, но при этом как можно ближе к нему. В случае кПЦР праймеры подбирали по тем же параметрам, за исключением концентрации праймера, равной в данном случае 250 нМ. В случае праймеров для клонирования и синтеза зонда для WMISH длина ампликона ограничена диапазоном 400-1500 нуклеотидов, концентрация праймеров снижена до 250 нМ (Приложение 4).

Полученные наборы праймеров и зондов (Приложение 1, 2 для кПЦР и кцПЦР, соответственно) проверяли с помощью собственного скрипта OligoAnalyse (все упомянутые здесь и далее скрипты доступны по ссылке

https://github.com/Alteroldis/bioscripts) на отсутствие вторичных структур, имеющих изменение энергии Гиббса меньше -2 ккал/моль для шпилек, меньше -8 ккал/моль для гомодимеров, меньше -8 ккал/моль для гетеродимеров и меньше -10 ккал/моль для гетеродимеров между олигонуклеотидами обоих референсных и таргетного генов, или двух референсных генов, в случае подбора праймеров на один из них. В случае расчета изменения энергии Гиббса для гетеродимеров вводилась поправка на расположение димера: если длина структуры больше 2 нуклеотидов, то в случае расположения структуры на 3`-конце обоих праймеров происходило умножение длины димера на 2, в случае расположения структуры на 3`-конце только одного из праймеров, длина димера умножалась на 1,5. При расчете изменения энергии Гиббса вводились те же поправки, что и при исходном подборе праймеров, а температура симуляции реакции для расчета изменения энергии выставлена на 30 °C. Далее среди наборов праймеров и зондов выбирались те, что уникально картируются на целевую последовательность в транскриптоме, имеют максимально схожие температуры плавления внутри пары праймеров и максимально отличную от нее температуру плавления зонда, стабильное содержание гуанина и цитозина, наиболее близкое к положительному изменение энергии Гиббса без явных перекосов в пороговые значения для хотя бы одного типа вторичных структур и максимально соответствуют вышеупомянутому требованию к расположению ампликона на известной ОРС мРНК. Зонды и праймеры синтезировали в компании Евроген (Россия) по технологии TaqMan [92], где гасителем флуоресценции являлся BHQ-1 на 3`-конце зонда, а флуорофором — FAM (6-карбоксифлуоресцеин), для таргетных генов, или HEX (гексахлорофлуоресцеин), для референсных генов, расположенный на 5`-конце.

2.5. Секвенирование поли-А РНК

Подготовка образцов и секвенирование было проведено компанией Евроген (Россия). Концентрация 8 образцов *E. fraudatrix* (по два повтора на стадию) была

определена с помощью флуориметра Qubit (Thermo Scientific, США) и подготовка библиотек осуществлялась с помощью набора Illumina TruSeq Stranded mRNA Library Prep Kit (Illumina, США), позволяющего точно знать в каком положении находились прочтения на исходном транскрипте (смысловая или антисмысловая цепь). Далее были отобраны фрагменты готовой библиотеки длиной 250-450 нуклеотидов, что было проверено с помощью капиллярного электрофореза на приборе Agilent TapeStation (Agilent, США) с высокочувствительными чипами. Библиотеки были смешаны эквимолярно, что было проверено с помощью кПЦР и секвенированы на двух дорожках чипа прибора Illumina HiSeq 2500 (Illumina, США) с использованием реактивов TruSeq Sequencing Kit v4 (Illumina, США) и длиной прочтения 101 нуклеотид с каждого конца фрагмента. Файлы прочтений в формате FASTQ получены с помощью bcl2fastq v2.17.1.14 (Illumina, США) с форматом записи качества нуклеотидов Phred33. Прочтения были загружены в базу данных SRA NCBI с индексами SRR8297983-SRR8297990 для трех стадий регенерации и нормы, соответственно. РНК *A. japonicus* была подготовлена и секвенирована также, прочтения были загружены с индексами SRR6075435-SRR6075438.

РНК *C. schmeltzii* была секвенирована схожим способом, но после синтеза кДНК с набором Mint (Евроген, Россия) по стандартному протоколу производителя на амплификаторе C1000 Touch (Bio-Rad, США) с oligo-dT праймером проводили DSN-нормализацию кДНК [93]. Далее полученную нормализованную кДНК использовали для приготовления библиотек с NEBNext DNA Library Prep (NEB, США) и секвенировали на HiSeq 2000 по 100 нуклеотидов с каждого конца фрагмента. Также данная кДНК была секвенирована на 454 GS FLX+ (Roche, Швейцария).

2.6. Сборка транскриптома *de novo*

Для *E. fraudatrix* сырые прочтения с 8 библиотек в FASTQ формате подвергли очистке от технических последовательностей, поли-N и низкокачественных участков, а также прочтений с длиной менее 21 нуклеотида в программе Trimmomatic v0.36 [94] с параметрами «LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 AVGQUAL:25 MINLEN:21». Затем все прошедшие фильтрацию прочтения использовали для коррекции ошибок секвенирования в две итерации и сборке в программе SPAdes v3.13 [95] с тремя значениями длины k-мера — 25, 33 и 49. Затем из всех получившихся последовательностей извлекали ОРС с минимальной длиной 30 аминокислотных остатков (аа) с помощью TransDecoder 5.5.0 [96]. Программа была модифицирована так, чтобы за ОРС принималась последовательность от старт-кодона, либо, при отсутствии его, от начала последовательности, а не от остатка метионина (триплета ATG). В процессе получения ОРС использовали также информацию о выравнивании последовательностей с белками баз SwissProt (11.12.2018) и Echinobase (11.12.2018) [97,98] как описано в инструкции TransDecoder. Здесь и далее в скобках указывается дата релиза базы данных, когда речь идет о базах, не имеющих версии.

Затем, полученные ОРС использовали для дальнейшей сборки с помощью собственной программы HomoloCAP [90] как описано в подразделе 1 раздела «Результаты». Далее были удалены все последовательности, для которых выполняются три условия: найдено совпадение по результатам BLAST-поиска [80] против базы данных белков NR NCBI с белками, не принадлежащими видам группы Deuterostomia; с оценкой такого совпадения не менее 200; нет совпадения среди белков видов *A. japonicus*, *Parastichopus parvimensis*, *C. schmeltzii*, *Patiria miniata*, *Lytechinus variegatus*, и *Strongylocentrotus purpuratus* [89,90,98] с оценкой не менее 80% от лучшего совпадения с белками, не принадлежащими видам группы Deuterostomia. Далее сборку фильтровали в соответствии с требованиями

NCBI и загрузили в базу TSA NCBI с индексом GHCL00000000. Репрезентативность генов в транскриптоме была оценена с помощью BUSCO v3 [99] в режиме «protein» с использованием набора данных Metazoa v9. Также была проведена сборка транскриптомов *A. japonicus* и *C. schmeltzii*, последовательности которых загружены в NCBI под индексами GFXQ00000000 и GFWR00000000, соответственно.

2.7. Анализ дифференциальной экспрессии генов

Для оценки экспрессии прочтения были картированы на транскриптом в Bowtie v2.3.4 [100] с параметрами «-no-mixed -no-discordant -gbar 50 -end-to-end -k 200 -very-sensitive -minins 50 -q -maxins 450 -fr» и затем с помощью RSEM v1.3.1 было подсчитано число прочтений, картирующихся на последовательность [101]. Далее для анализа экспрессии были сохранены только последовательности, для которых число картирующихся прочтений было не менее 50 и с не менее чем 10-кратным покрытием. Дифференциальная экспрессия была оценена в DESeq2 v1.18 [87] с использованием фильтрованной как описано выше матрицы числа картирующихся на последовательности в 8 образцах прочтений. При этом в любом образце допускалось нулевое число картирующихся прочтений. В качестве контроля (точки отсчета) использовали образцы интактной кишки и стадию регенерации 5-7 СПЭ. Ген признавался дифференциально-экспрессирующимся, если имел уровень экспрессии в два раза отличающийся от контрольного и вероятность ошибки оценки экспрессии была ниже 0,05. Кроме того использовали нормированные на размер библиотеки (число прочтений) значения TPM (число транскриптов на миллион килобаз, англ. Transcripts Per Million), рассчитанные RSEM в процессе подсчета числа картирующихся прочтений.

2.8. Аннотация белок-кодирующих последовательностей транскриптома

Все варианты аннотации по разным базам данных белков выполнены с использованием программы BLASTp v2.7.0 со стандартным порогом вероятности случайного совпадения последовательностей равным 10^{-5}. Первичная аннотация проведена по базе данных белков NR (non redundant) NCBI (11.12.2018). Аннотация для анализа сверхпредставленности выполнена по белкам человека из базы Ensembl v95 [102]. Аннотация для поиска транскрипционных факторов (ТФ) выполнена по белкам человека (Ensembl v95) и морского ежа *S. purpuratus* из базы проекта Echinobase (11.12.2018), при этом далее анализировались только те гомологи известных ТФ, предсказанная длина белок-кодирующей последовательности которых более 600 нуклеотидов. Далее результаты аннотации подвергались этапу, различному для первичной аннотации и аннотации по белкам человека или морского ежа.

В случае первичной аннотации выполняли фильтрацию совпадений, по оценке выравнивания. Для каждой последовательности сборки транскриптома проводили сортировку совпадений по убыванию оценки выравнивания bitscore. Далее извлекали только совпадение, имеющее лучшую оценку. Также учитывалась биологическая значимость, которая в данном случае определяется как наличие у белка в базе уникального названия, предполагающего, по мнению авторов записи в базе, наличие известной функции или гомологичности с белками, имеющими таковую. Одновременно проверяли наличие в названии белка хотя бы одного слова или словосочетания из списка - "hypothetical", "uncharacterized", "predicted protein", "Predicted protein", "Uncharacterized", "Hypothetical", "unknown protein", "Unknown protein", "unnamed protein", "Unnamed protein". Если совпадение было найдено, то искали следующее лучшее совпадение с белком, оценка выравнивания с которым не менее 80% от максимальной, а название является биологически значимым.

В случае аннотации с целью поиска ортологов и анализа сверхпредставленных биологических процессов и путей, вероятные ортологи извлекали из данных о совпадениях среди белков человека и морского ежа. Этот процесс проводили на основе оценки выравнивания с использованием модифицированного реципрокного метода. Последний был реализован с помощью собственного скрипта Reconciler на языке Python3 (подробно описан в подразделе 1 раздела «Результаты»). Среди выявленных ортологов затем выполняли поиск ТФ, основываясь на информации об ортологичных белках человека и морского ежа, для чего использовали данные HGNC [103] и Echinobase, соответственно. Далее ТФ фильтровали по трем критериям: значение TPM выше 1 на второй стадии; среднее значение логарифма по основанию 2 от изменения экспрессии (LogFC, Logarithm of Fold Change) на второй стадии относительно первой и третьей стадий более 0,5; значение логарифма по основанию 2 от изменения экспрессии на второй стадии относительно первой или третьей стадий более 1. Также из рассмотрения были убраны ТФ, не идентифицированные как осмысленные белки с доменом zinc finger, то есть не имеющие значащего названия и описанных функций.

Анализ сверхпредставленности биологических процессов и путей для *E. fraudatrix* выполнен с помощью GSEA [88] в соответствии с протоколом плагина EnrichmentMap [104] для анализа данных секвенирования транскриптома. Наборы генов человека для анализа были взяты из базы MsigDB v6.2 [88], однако из них сохранены были только те наборы, которые включали хотя бы один из 11 ТФ, являющихся предполагаемыми генами-регуляторами клеточной трансдифференцировки в процессе регенерации кишки (гены перечислены в подразделе 2 раздела «Результаты»). Затем результаты анализа были визуализированы в виде сети с помощью Cytoscape [105] и плагина EnrichmentMap, где процессы были представлены как точки, а связи между ними отображали наличие и число общих генов.

2.9. Оценка экспрессии генов с помощью кцПЦР

Для валидации уровня и динамики экспрессии генов, полученной по результатам секвенирования транскриптомов зачатков кишки на 3 стадиях регенерации (3, 5-7 и 10 СПЭ) и нормальной кишки, был проведен анализ экспрессии 10 генов ТФ с помощью кцПЦР. В качестве референсного гена использовали *tubb* или *ef1a*. В каждом запуске было 10 генов с одним референсом и 4 образца (3, 5, 7 и 10 сутки регенерации), каждая реакция в запуске представлена дважды (реплики). Также каждый образец был представлен трижды — три отдельных выделения РНК из разных особей. Итоговое число реакций на каждый таргетный ген и каждый образец — 6. ПЦР смесь готовили в соответствии с инструкцией к набору ddPCR Supermix for Probes (Bio-Rad, США) с 900 нМ каждого праймера референсного и таргетного генов, 250 нМ каждого зонда референсного и таргетного генов и 1 мкл кДНК в конечном объеме 20 мкл. Далее смесь загружали в DG8-картридж (Bio-Rad, США), в него же помещали 70 мкл масла для генерации капель (Bio-Rad, США) и создавали эмульсию на генераторе капель QX200 (Bio-Rad, США). Затем 40 мкл эмульсии переносили в 96-луночный планшет (Bio-Rad, США) и запаивали его фольгой в приборе PX1 (Bio-Rad, США) при 180ºС в течение 5 сек. Затем выполняли амплификацию по рекомендованной программе на амплификаторе C1000 Touch (Bio-Rad, США):

1. 95º С, 10 мин — активация полимеразы;
2. 94º С, 30 сек — денатурация дцДНК;
3. 60º С, 60 сек — отжиг олигонуклеотидов и элонгация;
4. Повтор шагов 2-3 40 раз;
5. 98º С, 10 мин — деактивация полимеразы.

Скорость охлаждения/нагрева выставлена на 2ºС/сек., температура крышки амплификатора – на 105 ºС и объем образца – на 40 мкл. Далее проводили считывание числа положительных, имеющих продукт амплификации, и

отрицательных капель в каждой реакции с помощью прибора QX200 Droplet Reader (Bio-Rad, США).

Перед рабочими запусками был сделан ряд калибровочных, чтобы подобрать температуру отжига олигонуклеотидов на основе уровня флуоресценции позитивных и негативных капель. Также, по заданной программе были сделаны реакции кПЦР с интеркалирующим красителем для получения информации об отсутствии неспецифической амплификации. Далее продукты этих реакций были очищены от компонентов ПЦР-смеси и 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). Полученные последовательности искали с помощью BLAST в транскриптоме, чтобы подтвердить принадлежность к целевым транскриптам. Полученные последовательности ампликонов, а также индекс в базе NCBI TSA и идентичность ей указаны в Приложении 3.

При анализе результатов использовали вычисленное программой для QX200 QuantaSoft (Bio-Rad, США) значение числа копий на микролитр реакции. Для этих значений между репликами внутри каждого образца брали среднее, затем делили на аналогичное значение для референсных генов, после чего, между полученными нормализованными на ef1a и tubb значениями вычисляли среднее геометрическое. Далее, для анализа корреляции между значениями экспрессии по результатам РНК-секвенирования и кцПЦР, брали среднее арифметическое между полученными значениями из образцов 5 и 7 СПЭ. В случае РНК-секвенирования, в качестве оценки экспрессии использовали число выравненных на ген прочтений, после чего нормализовали на длину в килобазах и затем, как и в случае с кцПЦР, на *tubb* и *ef1a*, с вычислением среднего геометрического для оценок, полученных в результате нормализации на референсные гены. Для отображения динамики экспрессии использовали среднее арифметическое между 3 запусками с вычислением 90% доверительного интервала. При этом использовали

логарифмическую шкалу отображения для сохранения видимости изменений между сильно и слабо экспрессирующимися генами.

2.10. Валидация результатов РНК-секвенирования

Для проверки результатов РНК-секвенирования использовали 5 генов (*sox9*, *max*, *tcf24*, *foxc1* и *hes1*) и 2 референсных гена (*ef1a* и *tubb*). Реакции были проведены с использованием набора qPCRmix-HS SYBR (Евроген, Россия) на амплификаторе CFX96 Touch (Bio-Rad, США) 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), полученной со всех образцов в программе CFX Manager v3.1 software (Bio-Rad, USA). Специфичность праймеров была проверена с помощью секвенирования ампликонов по Сенгеру, как описано выше.

Значения экспрессии вычисляли по модифицированному методу $2^{-\Delta\Delta Ct}$ [106]. Однако при этом, вместо вычисления среднего геометрического между разными повторами, использовали среднее арифметическое, так как все вычисления проводились в логарифмическом пространстве, а $2^{-\Delta\Delta Ct}$ метод предполагает перед вычислением среднего геометрического получение антилогарифма для каждого повтора. Метод реализован в виде скрипта qpcr на языке Python3. В качестве контрольного образца использовали зачатки кишки *E. fraudatrix* на стадии 3 СПЭ. 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.

2.11. Гибридизация *in situ*

Для синтеза РНК зонда были амплифицированы фрагменты с размерами ампликона 400-1500 нуклеотидов, как указано выше (Приложение 4). Полученные фрагменты были лигированы в вектор pAL2-T (Евроген, Россия), после чего проводили химическую трансформацию в компетентные клетки XL1-Blue (Евроген, Россия). Далее использовали секвенирование с M13 форвардного праймера для выяснения положения вставки в плазмиде, как описано выше для кПЦР (Приложение 5). На следующем этапе с помощью SP6 или T7 РНК-полимераз (в зависимости от положения вставки в плазмиде) были синтезированы фрагменты для смыслового и антисмыслового РНК зонда, используя DIG dNTP RNA labeling mix (Roche, Швейцария). После обработки ДНКазой I (Thermo Scientific, США), РНК зонд был очищен с помощью магнитных частиц AmpureXP (Beckman Coulter, США) и растворен в 50 мкл mQ. По одному мкл зонда анализировали на спектрофотометре Biospec nano (Shimadzu, Япония) и с помощью электрофореза. РНК зонд был разведен в Hyb mix (50% Formamide; 1xSSC; 0.15мг/мл Heparin; 5мг/мл Torula RNA; 0.1% Tween20) до концентрации 12-14 нг/мкл.

Выявление продуктов генов в зачатках кишки *E. fraudatrix* проводили с помощью метода WMISH (whole mount in situ hybridization). Животных на всех сроках регенерации предфиксировали инъекцией 4% параформальдегида (ПФА) на 1X PBS (137 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1,8 mM KH_2PO_4, pH 7,4) в полость тела. Далее зачатки АК и кишки вырезались и фиксировались 2 час в том же растворе ПФА при комнатной температуре. Образцы хранили в метаноле при -20^{o} C. Перед использованием материал отмывали от метанола 15 мин в растворе Reduction solution (0.5% SDS, 10мМ DTT на 1xPTW). Затем их помещали в 1% раствор Triton X100 с протеиназой К (100 мкг/мл) при 37 ºС на 5 мин для ранних сроков регенерации (3-7 СПЭ), 10 мин для поздних (10СПЭ) и 15 мин для нормы. После этого образцы обрабатывали при комнатной температуре 15 мин H_2O_2,

затем 15 мин Acetylation solution (0,1 M triethanolmine-HCl; 0,25% acenic anhydride) и 1 час смесью HybMix (0,5% Torula RNA, 0,25 мМ Heparin, 0,1% Tween 20, 50% Formamid на 4X SSC) при 45 ºС. Между всеми обработками делали 10-минутные отмывки в 1x PTW. Зачатки обрабатывали РНК зондом в течение 12 час при 65 ºС. После этого их промывали 50% Formamid/SSC, 2xSSC, 0,2xSSC при 65 ºС, проводили блокировку в 5% растворе овечьей сыворотки и оставляли в 1:2000 Anti Dig Fab fragment на ночь. После этого делали 10 промывок в PBS, две промывки в Stayning buffer (0,1 M Tris-HCl (pH 9,5), 0,1 M NaCl, 0,1% Tween 20) и красили с помощью NBT/BCIP tablet (Roche, Швейцария) при 37 ºС в термостате. Окрашенные зачатки промывали в PBS и фиксировали в PFA, в котором и хранили при +4 ºС.

2.12. Применяемые математические методы и языки программирования

В ходе выполнения работы использовали языки программирования Bash и Python3, включая библиотеки Python Statistics, NumPy, SciPy, Seaborn и Matplotlib для статистических расчетов и построения графиков и тепловых карт.

3. РЕЗУЛЬТАТЫ

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

3.1.1. *De novo* сборка транскриптома

В процессе работы над транскриптомами трех видов голотурий был разработан метод улучшения сборки, получаемой с помощью доступных на данный момент программ-сборщиков (Рисунок 5). Основное преимущество состоит в использовании для финализации сборки только предсказанных белок-кодирующих областей, что облегчает дальнейшую работу при наличии в транскриптоме нескольких изоформ или внутривидовой вариабельности последовательностей.

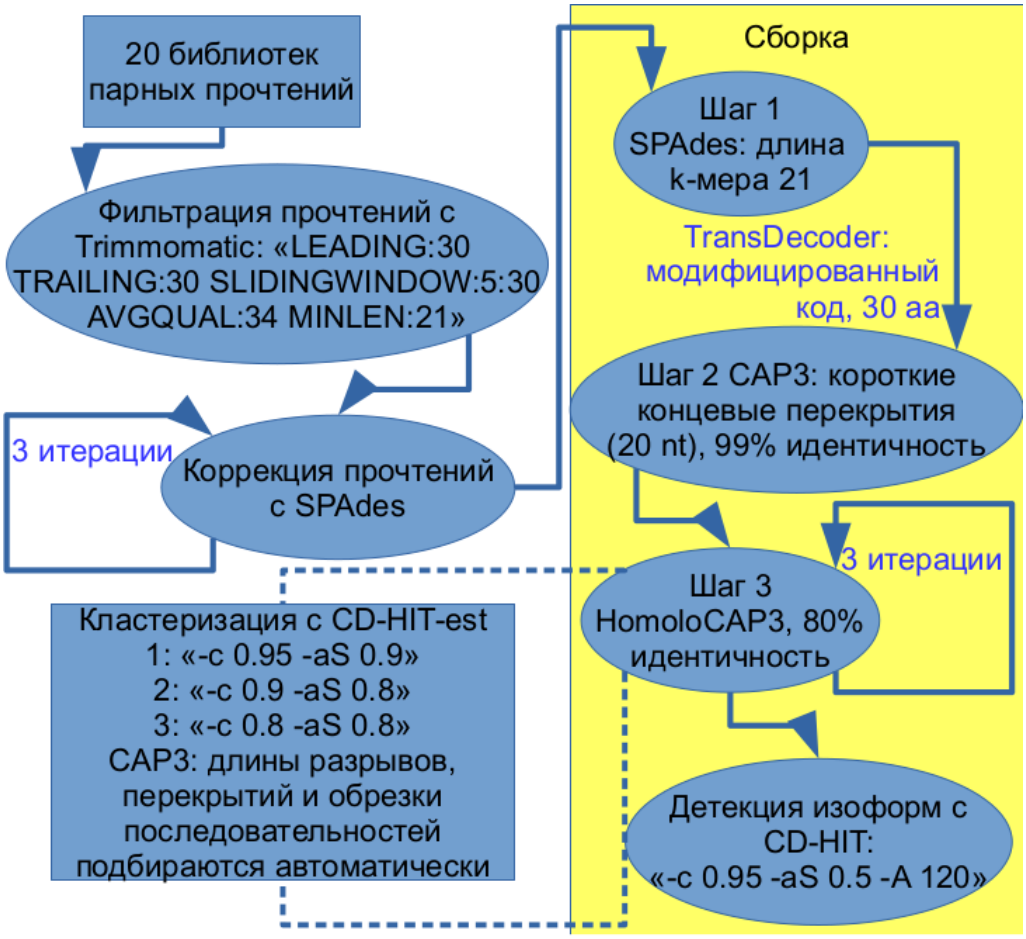

Рисунок 5. Пример схемы сборки транскриптома с помощью HomoloCAP

Алгоритм программы работает в несколько шагов:

1. В первую очередь с помощью TransDecoder извлекаются лучшие для каждой последовательности ОРС с длиной не менее 30 аминокислотных остатка, а также те, что имеют значимые совпадения при поиске схожих последовательностей с помощью BLAST в выбранной базе данных белков. База данных и значимость, соответствующая параметру «e-value» BLAST, так же, как и сама необходимость использовать дополнительную проверку с помощью BLAST, остается на усмотрение пользователя;
2. Далее программа работает в итеративном цикле, состоящем из двух шагов:
   2.1. Проводится кластеризация нуклеотидных последовательностей с помощью CD-HIT в режиме локального выравнивания [107]. Параметры кластеризации выставлены таким образом, чтобы кластер включал только последовательности, имеющие полностью идентичный с репрезентативной последовательностью кластера участок, который должен покрывать не менее 20% этой последовательности;
   2.2. Сборка, которая проводится для каждого кластера последовательностей, с помощью CAP3 [108] — OLC сборщика с параметрами, позволяющими чуть более гибкое объединение последовательностей при более строгом, относительно параметров по умолчанию, ограничении на общие регионы. Так, запрещено использовать обратно-комплементарную 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. Так, на первой итерации кластер может включать только

последовательности, имеющие полностью идентичный с репрезентативной последовательностью кластера участок, который должен покрывать не менее 100% этой последовательности, то есть последовательности кластера должны быть целиком идентичны участку самой большой последовательности кластера. На второй — схожесть понижается до 90% и перекрытие до 90%. На третьей — процент перекрытия понижен до 80. При этом, независимо от итерации, длина общего региона должна быть не менее 120 нуклеотидов.

3. На последнем этапе идет кластеризация с целью объединения альтернативных изоформ общим названием последовательности. Изоформы в данном случае понимаются как последовательности, имеющие общий с репрезентативной последовательностью кластера участок, длиной не менее 300 нуклеотидов, схожестью не менее 95% и покрывающий не менее 30% этой последовательности. То есть в данном случае изоформы могут быть как действительно существующим результатом альтернативного сплайсинга, так и результатом ошибки сборки или просто невозможности объединения этих форм в одну последовательность с помощью данного алгоритма.

Данный подход был опробован сначала на транскриптомах *C. schmeltzii* и *A. japonicus* [89,90] и позволил для последнего вида достичь результатов, сравнимых с транскриптомами, полученными на основе геномных сборок, что видно из Таблицы 1. Для первого вида провести сравнение не было возможности, так как по нему других сборок не публиковалось. Последние три столбца Таблицы 1 характеризуют полноту сборки, причем видно, что наша сборка содержит от 88 до 97% ОРС всех сборок, включая геномные, в то время как прочие покрывают только около 40% нашей. Исключение составляет только сборка на основе генома [109] (последняя строка Таблицы 1), которая по обоим параметрам наиболее близка к таковой, полученной с использованием разработанного нами метода. При этом при использовании нашего алгоритма все базовые статистические показатели получаются лучше, чем у ранее полученных сборок. Исключением являются

только данные, полученные на основе секвенирования генома, где показатель средней длины ОРС чуть выше.

Таблица 1. Сравнение общих статистических показателей сборки транскриптома голотурии *Apostichopus japonicus*

Сборка	**Год**	**Число ОРС**	**Средняя длина, н.о.**	**N10, н.о.**	**N30, н.о.**	**N50, н.о.**	**Покрытие нашей, %**	**Покрытие другими, %**	**Картирование, %**
Наши данные	2019	27598	1274	**5865**	**3090**	**1946**	100	100	**28,6**
NCBI ESTs	-	2816	569	1173	735	606	95	4	14
Du et al., 2012	2012	13724	776	1824	1224	912	97	22	16,2
Zhou et al., 2014	2014	26174	947	3513	1938	1317	98	46	25,8
Reich et al., 2015 (GAVS01.1)	2015	31611	922	3015	1743	1206	93	35	23,4
Jo et al., 2016 (HADD01.1)	2016	27670	991	3648	2097	1410	95	43	24,9
Jo et al., 2016 (HADE01.1)	2016	27445	856	2889	1725	1194	96	40	24,6
Jo et al., 2016 (HADF01.1)	2016	27396	855	3027	1746	1203	96	39	24,6
Jo et al., 2017 (геном)	2017	17111	1059	3531	1971	1380	91	35	17,7
Zhang et al., 2017 (геном)	2017	22643	**1281**	3891	2259	1551	88	70	21,4

Примечание. Жирным шрифтом выделены лучшие значения для каждого столбца, н.о. - нуклеотидное основание

Для нашей сборки базовые статистические показатели на каждом этапе незначительно ухудшаются, выравниваясь к финальной сборке, при этом число последовательностей уменьшается в 4,4 раза при почти полном (<1%) отсутствии потерь биологически значимой информации, вычисленной как сумма оценок

выравнивания всех последовательностей на белки человека, *A. japonicus*, *S. purpuratus* и базы SwissProt (Таблица 2).

Таблица 2. Базовые статистические показатели сборки *Eupentacta fraudatrix* на разных этапах применения нашего алгоритма

Этап	Число ОРС	Средняя длина, н.о.	N10, н.о.	N30, н.о.	N50, н.о.	БИ-AJ, %	БИ-SP, %	БИ-HS, %	БИ-SW, %
До	370067	**716**	**3630**	**1836**	**1155**	100/100	100/100	100/100	100/100
Итерация 1	239076	684	3558	1755	1086	99,981/ 100,02	99,988/ 100,007	99,999/ 99,999	99,988/ 100,025
Итерация 2	163376	632	3582	1728	1023	99,655/ 99,94	99,564/ 100,055	99,756/ 99,922	99,564/ 100,237
Итерация 3	142529	562	3264	1524	835	99,474/ 99,99	99,366/ 100,145	99,653/ 99,933	99,366/ 100,434
Финальная	83960	698	3624	1794	1095	98,795/ 99,634	98,978/ 100,013	99,320/ 99,805	98,978/ 100,777

Примечание. БИ - процент биологически значимой информации. AJ, SP, HS, SW - база белков, по которой проводили выравнивание для оценки БИ, *Apostichopus japonicus*, *Strongylocentrotus purpuratus*, *Homo sapiens*, SwissProt соответственно. Числа через косую черту относятся к разным способам фильтрации совпадений — в первом случае фильтрация только по лучшему совпадению для белков из базы, во втором случае проводилась еще фильтрация по лучшему совпадению для последовательности из сборки

3.1.2. Поиск ортологов

При поиске ортологов была поставлена задача элиминировать минусы реципрокного подхода за счет включения в процедуру реципрокного метода не только лучших совпадений, но и фракции условно лучших, то есть таких, для которых оценка выравнивания отличается от лучшей не более чем на 20-40%. Такой порог позволяет, во-первых, убрать совпадения с низкой оценкой выравнивания, что снижает вероятность ошибки I рода. Во-вторых, он увеличивает чувствительность, позволяя рассматривать больше вариантов, чем реципрокный подход. Для этого нами был разработан скрипт Reconciler на языке Python v3.6. Для его работы необходим файл, содержащий результаты поиска совпадений в какой-либо базе данных, независимо от природы

последовательностей. Эти данные должны быть записаны в виде строк, в каждой из которых есть несколько полей. Необходимыми из них являются оценка выравнивания (s) и индексы двух последовательностей, между которыми имеется выравнивание — одна является поисковой ($a \in A$), а другая — из базы данных ($b \in B$), где a и b являются каким-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, в сумме описывающие все совпадения для b_n (нижним индексом далее обозначены конкретные члены множества). Каждый список по отдельности включает все данные из первичного файла, описывающие какое-либо совпадение a_n с b_n, включая s_n (оценка выравнивания), a_n и полную строку из первичного файла, соответствующую данному совпадению. Второй (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 удаляем все значения, для которых верно условие $s_n < s_{n,best}$*0,8, где $s_{n,best}$ — самая высокая оценка выравнивания для члена n множества A или B. Для словаря AB выполняется та же процедура, но умножение лучшей оценки происходит на 0,7. Также мы использовали ограничение в 40 и 50 процентов, соответственно.
3. Итеративный обход по всем b до тех пор, пока существует хоть один b с числом записей более 1. В течение каждой итерации происходит проверка трех условий:

   3.1. Если нет записей для b_n, тогда все записи, которые имеют b_n, в словаре A удаляются.

3.2. Если лучший (по оценке) a_n, соответствующий 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 лучший a_n, соответствующий b_n в словаре B, имеет в качестве лучшего b_n в словаре A, 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, здесь применяется непосредственно условие реципрокности из соответствующего метода.

4. Используя пары из условия 3.3, программа находит соответствующие записи в словаре AB и формирует файл того же формата, что исходный, но содержащий только 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 разных животных и белков человека из базы данных Ensembl (Таблица 3).

Таблица 3. Сравнение двух подходов к поиску ортологов

Вид	**OE**	**РГ**	**Р**	**МГ**	**МГ65**	**М**	**М65**
Mus musculus	20698/ 15815	**15694/ 14773**	15676/ 14755	15952/ 14861	**15992/ 14877**	15945/ 14853	15989/ 14873
Danio rerio	19591/ 9281	**9658/ 7089**	9589/ 7041	10399/ 7492	**10524/ 7528**	10353/ 7450	10486/ 7496
Ciona intestinalis	14081/ 3879	**4604/ 3091**	4602/ 3089	4846/ 3148	4962/ 3180	4847/ 3145	**4970/ 3179**
Drosophyla melanogaster	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 «один к одному». Жирным выделены лучшие результаты для каждого из подходов

Поиск ортологов с помощью BLAST провели шестью способами, из которых 2 приходится на реципрокный метод поиска (столбцы 3-4) и 4 — на

модифицированный (столбцы 5-8). При использовании обоих подходов BLAST поиск проводили двумя способами — стандартный поиск и поиск с активной фильтрацией участков последовательности низкой сложности (то есть, относительно гомогенных) и итоговым выравниванием с помощью алгоритма Смита-Ватермана [110]. Модифицированный метод поиска, кроме того, представлен в двух вариантах — со стандартным ограничением оценки выравнивания (0,8 и 0,7) и пониженным (0,6 и 0,5). Анализ полученных данных показал, что в целом наблюдается увеличение числа найденных ортологов с помощью модифицированного метода по сравнению с реципрокным. Причём в пересчёте на ожидаемое число ортологов, действительно наблюдается тенденция роста разницы между двумя методами при увеличении эволюционной дистанции.

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

3.2.1. Секвенирование и сборка транскриптома *de novo*

Качество РНК, выделенной для анализа транскриптома пищеварительной системы в процессе регенерации у голотурии *E. fraudatrix* было высоким (Рисунок 6А). Большая часть фрагментов при подготовке библиотек находилась в диапазоне длин 250-350 нуклеотидов (Рисунок 6Б).

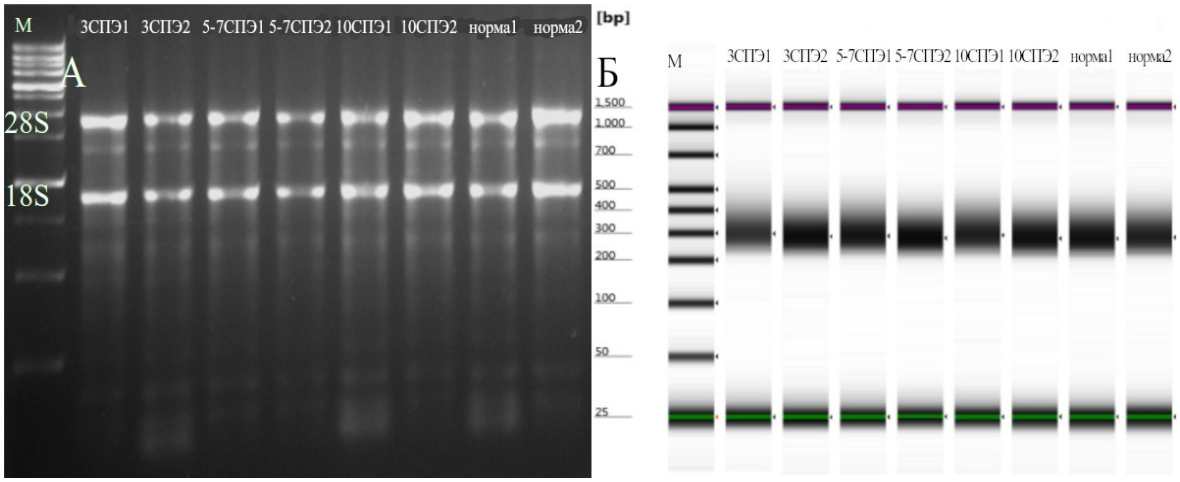

Рисунок 6. Оценка качества мРНК голотурии *Eupentacta fraudatrix*, используемой для секвенирования. А: Электрофорез тотальной РНК, выделенной из интактной кишки и ее зачатков на разных сроках регенерации Б: Капиллярный электрофорез готовых к секвенированию библиотек.

В результате секвенирования 8 библиотек кДНК на приборе Illumina HiSeq 2500 было получено почти 413 миллионов парных прочтений — по 100 нуклеотидов с каждого конца фрагмента (Таблица 4), что эквивалентно 82,6 гигабазам (млрд. нуклеотидов).

Таблица 4. Статистические показатели результатов секвенирования 8 библиотек РНК интактной кишки и передних зачатков пищеварительной трубки на 3 сроках регенерации у *Eupentacta fraudatrix*.

SRA индекс	Образец	Исходные прочтения			Корректированные прочтения			Карти-руется (%)
		Число прочтений	Среднее качество	Длина	Число прочтений	Среднее качество	Длина	
SRR8297983	3 СПЭ (повтор 1)	49057381	35.46	100	46864415	35.62	99.28	19.19
SRR8297984	3 СПЭ (повтор 2)	50908678	35.58	100	47942974	35.78	99.77	18.77
SRR8297985	5-7 СПЭ (повтор 1)	57251433	35.55	100	54536553	35.78	99.76	19.42
SRR8297986	5-7 СПЭ (повтор 2)	55567180	35.56	100	52385442	35.79	99.8	19.29
SRR8297987	10 СПЭ (повтор 1)	42239523	35.56	100	40063136	35.79	99.78	18.85
SRR8297988	10 СПЭ (повтор 2)	50834156	35.54	100	47960668	35.78	99.75	19.94
SRR8297989	Норма (повтор 1)	52394208	35.52	100	49408813	35.78	99.79	27.17
SRR8297990	Норма (повтор 2)	54747365	35.52	100	51978341	35.76	99.7	28.42

Примечание. SRA - Sequence Read Archive, база данных секвенирования

Перед коррекцией ошибок и сборкой транскриптома все прочтения были отфильтрованы с помощью Trimmomatic по качеству и длине, которая менялась в результате удаления из прочтений технических последовательностей. После данной процедуры было сохранено 95% парных прочтений со средним качеством 35,7 единиц по шкале Phred 33 и средней длиной 99,4 нуклеотида. Кроме того,

около 5% от исходного числа прочтений было сохранено в виде непарных прочтений, которые также использовали на этапах коррекции ошибок и сборки. Этап коррекции ошибок в прочтениях, выполненный в две последовательные итерации, не дал значительного сдвига в соотношении парных прочтений к непарным — доля непарных прочтений от изначального числа всех прочтений увеличилась менее чем на две сотые процента.

Сборка транскриптома проходила в два этапа — первичная сборка с помощью доступных транскриптомных сборщиков и вторичная — с помощью собственных скриптов. Перед первичной сборкой была проведена коррекция ошибок прочтений в две последовательные итерации с помощью модуля SPAdes — BayesHammer. Сборка была проведена в двух режимах работы SPAdes — автоматический выбор длины к-мера, составившей 33 и 49 нуклеотидов, и пользовательский выбор длины к-мера, равной 23 нуклеотида. В общей сложности, было получено более 2 миллионов последовательностей. Затем, после фильтрации по наличию ОРС с длиной более 30 аминокислотных остатков и совпадении в базах данных белков SwissProt и Echinobase, число последовательностей было сокращено до 370067. Далее, после фильтрации по k-мерному покрытию, еще 8259 последовательностей было удалено из транскриптома. После каждого шага фильтрации изменение статистических параметров, таких как распределение длин последовательностей, N50, N30, N10 и средняя длина (Таблица 2), указывает на ухудшение значений данных параметров при одновременном уменьшении числа последовательностей. В то же время, потерь биологически значимой информации, взятой в данном контексте как суммарная оценка выравниваний с белками трех видов и базы SwissProt, обнаружено не было.

Вторичная сборка была реализована с помощью алгоритма, описанного в подразделе 1 данного раздела. Результатом этой сборки, при сравнении с отфильтрованными данными первичной сборки, явилось кардинальное уменьшение числа последовательностей — 83960, а также сохранение

статистических параметров сборки и отсутствие потерь биологически значимой информации. При учете только самой длинной изоформы (условно далее такие изоформы будут упоминаться как гены) число последовательностей снижается до 70538. Обратно на сборку выравнивается 21% парных ридов, прошедших фильтрацию.

Кроме того, была проведена оценка полноты сборки с помощью программы BUSCO. В результате такого анализа в нашей сборке была найдена 98,1% коровых генов Metazoa, из которых 83% были представлены одно-копийными ортологами.

3.2.2. Анализ дифференциальной экспрессии генов

После выравнивания парных прочтений на сборку с помощью Bowtie2 и подсчета числа успешно выровненных на каждую последовательность прочтений, как описано в разделе «Материалы и методы», было обнаружено, что в среднем на сборку выравнивается 21,4% всех парных прочтений. Далее данные о числе выровненных прочтений подверглись фильтрации по числу картированных прочтений для удаления заведомо низкоэкспрессирующихся генов или с ошибочно выровненными прочтениями. Полученные матрицы использовались во всех дальнейших этапах анализа, включая анализ корреляции (Рисунок 7А), анализ главных компонент, поиск дифференциально экспрессирующихся генов (ДЭГ) и поиск кандидатов на роль регуляторов трансдифференцировки.

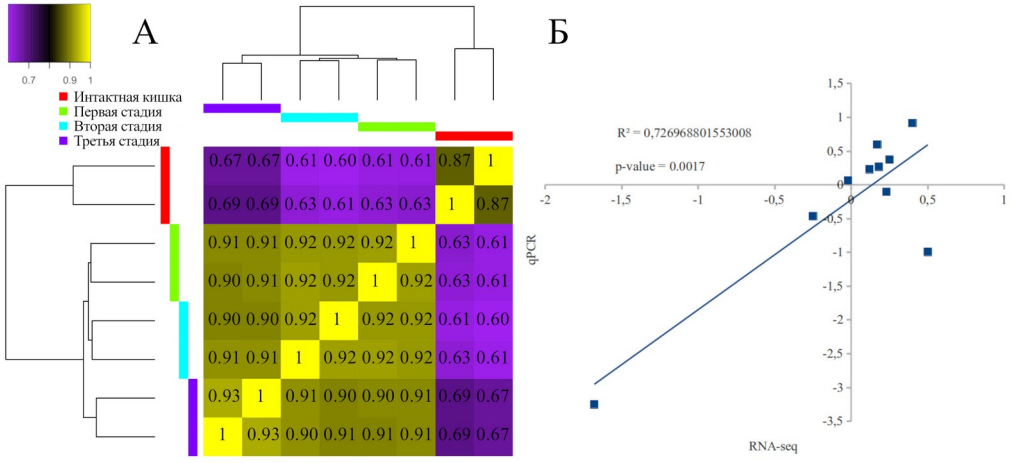

Рисунок 7. Анализ качества секвенирования мРНК *Eupentacta fraudatrix*. А: Тепловая карта квадрата коэффициента корреляции Пирсона экспрессии РНК-секвенирования. Б: Квадрат коэффициента корреляции Пирсона между оценками относительной экспрессии 5 генов ТФ на двух стадиях регенерации по результатам РНК-секвенирования (RNA-seq) и кПЦР (qPCR).

Анализ корреляции между образцами и репликами, а также анализ главных компонент был выполнен с помощью DESeq2 и компонентов пакета Trinity. Квадрат коэффициента корреляции Пирсона для реплик на всех стадиях регенерации был выше 0,92, в то время как для образцов, полученных из неповрежденной кишки, это значение составляет 0,87 (Рисунок 7А). 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 между образцами 3 СПЭ и 5-7 СПЭ выше, чем между каждой из них и 10 СПЭ, а также равна корреляции между репликами этих стадий. Анализ первых трех компонент показывает близкие результаты — высокая схожесть реплик во время регенерации и низкая — в норме, а также большое сходство всех 4 образцов с первых двух стадий регенерации (Рисунок 7А).

Другим этапом анализа качества секвенирования, сборки и выравнивания прочтений на нее в терминах достоверности оценки экспрессии и ее корреляции между репликами одного состояния является проверка значений дифференциальной экспрессии генов с помощью постановки кПЦР с несколькими генами. В данном случае оценкой является корреляция значений ДЭГ, полученных с помощью РНК-секвенирования и кПЦР. Была проведена кПЦР для 5 генов (*sox9*, *max*, *tcf24*, *foxc1* и *hes1*) на 3 стадиях регенерации. Корреляция была вычислена как квадрат коэффициента Пирсона и была равна 0,73 единицы с уровнем значимости, равным 0,0017 (Рисунок 7Б).

Анализа дифференциальной активности генов в DESeq2 позволил выявить 17227 и 15342 гена со значимыми изменениями в экспрессии на какой-либо стадии по сравнению с образцами нормы и стадии 5-7 СПЭ, соответственно. Из них 13234 и 11942 последовательности имели значимые совпадения среди белков базы NR NCBI. Бóльшая часть генов, как видно на диаграммах Венна (Рисунок 8), являются ДЭГ только по сравнению с образцами нормы. Такое соотношение сохраняется и при делении генов на группы с отрицательным и положительным,

относительно контроля (нормы и стадии 5-7 СПЭ), изменением экспрессии. Так, левая половина рисунка 7 показывает большее сродство стадий 3 СПЭ и 5-7 СПЭ друг другу, чем к норме или стадии 10 СПЭ, четкое увеличение числа уникальных для 5-7 СПЭ генов, меняющих относительно нормы свой уровень экспрессии, а также значительное число генов, меняющих свою экспрессию относительно нормы на всех стадиях регенерации кишки. Правая половина рисунка подтверждает эти наблюдения, но в контексте одной из стадий регенерации (5-7 СПЭ), а также показывает, что стадия 10 СПЭ имеет большее сходство с нормой, чем две другие.

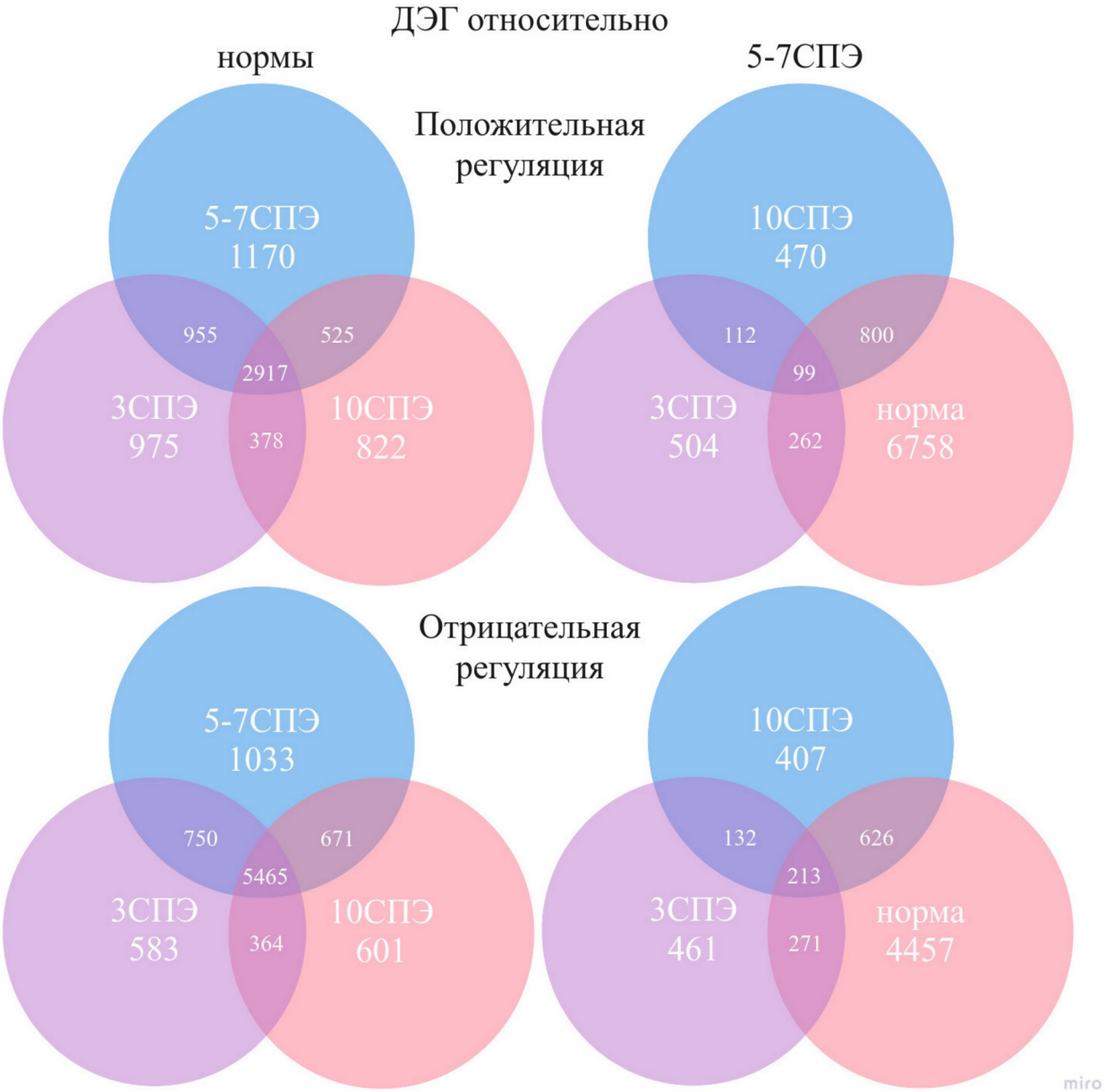

Рисунок 8. Диаграмма Венна для ДЭГ с положительной (сверху) и отрицательной (снизу) регуляцией относительно образцов нормы (слева) и 5-7 СПЭ (справа)

Таким образом, сравнение стадий регенерации с нормой показывает очевидный результат глобальной перестройки работы генома клеток при восстановительном процессе. Такой анализ не позволяет вычленить гены, активные на стадии трансдифференцировки. В связи с этим в дальнейшем анализировались только образцы регенерирующих тканей. Кроме того, оценка изменения экспрессии в течение восстановительного процесса проводилась относительно стадии 5-7 СПЭ.

3.2.3. Аннотация

Поиск гомологов известных белков среди 83960 последовательностей итоговой сборки был проведен с помощью BLASTp и базы белков NR NCBI и позволил найти 48677 изоформ и 37790 генов со значимыми совпадениями, включая 7118 последовательностей, имеющих совпадения только с белками, не имеющими осмысленного названия (подробную таблицу см. в 83). Белки лучших совпадений в базе принадлежали 1022 видам, из которых почти 80% генов принадлежали различным видам Echinodermata (Таблица 5).

Таблица 5. Филогенетическая принадлежность белков, которые являются лучшими совпадениями для белков *Eupentacta fraudatrix*.

Тип	**ОРС, №**	**Идентичность, %**	**Покрытие, %**	**Надцарство**	**ОРС, №**	**Идентичность, %**	**Покрытие, %**
Echinodermata	30047	55,3	80,99	Eukaryota	37426	53,94	80,32
Chordata	2637	47,79	78,19	Bacteria	341	42,42	73,79
Cnidaria	1391	46,85	78,06	**Царство**	**ОРС, №**	**Идентичность, %**	**Покрытие, %**
Arthropoda	863	49,3	77,07	Metazoa	36760	54,03	80,39
Mollusca	653	46,33	76,94	Fungi	179	45,76	70,37
Hemichordata	601	55,94	77,69	Viridiplantae	82	43,37	69,48
Apicomplexa	213	53,61	82,08	none	770	47,25	77,53

Примечание. Проценты показателей идентичности и покрытия являются средними значениями для всех выравниваний

При этом имелось некоторое число совпадений с белками бактерий, а также белками видов из царств Fungi и Viridiplantae, покрывающих суммарно 1,6% генов. Распределение совпадений по таксонам ранга «тип» не учитывает последние три группы организмов, так как включает в себя только те таксоны, которые имеют более чем 200 совпадений с генами *E. fraudatrix*. Суммарно такие таксоны покрывают 96,3% генов со значимыми совпадениями, из которых 8,3% принадлежат таксонам Protostomia.

Аннотация по белкам человека и морского ежа *S. purpuratus*, заключавшаяся в выявлении пар ортологов с помощью модифицированного реципрокного поиска, описанного в подразделе 3.1.2, позволила найти 10358 и 14617 вероятных ортологов, соответственно. Кроме того, был осуществлен поиск ТФ. За ТФ принимались те белки, которые обозначались как ТФ в базах HGNC (человек) и Echinobase (*S. purpuratus*). В результате данной аннотации были найдены 918 и 308 ТФ, соответственно. Всего было выявлено 961 разных ТФ. 265 белков были общими для морского ежа и человека.

Онтологическая аннотация выполнена схожим образом — для каждой пары генов голотурии и человека были извлечены аннотации биологических процессов и сигнальных путей из базы MsigDB. Из них были сохранены исключительно те биологические процессы и пути, которые включают хотя бы один из 11 ТФ, выбранных как вероятные кандидаты в регуляторы трансдифференцировки (процедура поиска данных ТФ описана ниже). Таких ассоциированных процессов и путей обнаружилось 790, и они объединяют 9545 генов.

3.2.4. Поиск кандидатов на роль регуляторов клеточной трансдифференцировки

Как указано в предыдущем разделе, 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 в

течение рассматриваемых стадий регенерации, так, чтобы прошедшие фильтрацию ТФ удовлетворяли четырем условиям:

1. Значение TPM на второй стадии регенерации больше единицы;
2. Среднее значение LogFC больше 0,5 на второй стадии, по сравнению с первой и третьей;
3. Значение LogFC на второй стадии, по сравнению с первой или третьей, больше единицы;
4. Значение p-value меньше 0,05 на первой и третьей стадиях относительно второй.

Сумма таких условий дает возможность убрать гены с низкой абсолютной экспрессией, недостоверно малым изменением экспрессии на второй стадии относительно первой и третьей, то есть не критичных для трансдифференцировки. Третье условие позволяет обойти проблему участия одного гена в процессах не только второй стадии, но и первой или третьей, при этом гарантируя более чем двукратное увеличение экспрессии на второй стадии относительно какой-либо другой стадии регенерации. Последнее условие дополнительно проверяет достоверность оценки экспрессии. В итоге было найдено 20 ТФ, 9 из которых были удалены из дальнейшего рассмотрения (Таблица 6). Восемь ТФ не идентифицировались как осмысленные белки с доменом zinc finger. Еще одна последовательность, Ef-ELF4, являлась ошибкой сборки. Она была полностью идентична белку Ef-ELF2 в области домена ETS, однако не имела значимых совпадений в базе данных белков NRP NCBI. Кроме того, у иглокожих гомолог генов человека *ELF 1*, *2* и *4* представлен в единственном числе [111]. В связи с этим Ef-ELF2 далее будет упоминаться как Ef-ELF. Дополнительно, Ef-KLF2 был переименован в Ef-KLF1/2/4, так как у голотурий он является ортологом KLF1, KLF2 и KLF4 человека [35]. Также Ef-ID2 и Ef-EGR1 далее упоминаются как Ef-ID и Ef-EGR в связи с наличием только по одному гену *id* и *egr* у иглокожих [97,98].

Таблица 6. Оставшиеся после фильтрации по экспрессии гены ТФ голотурии *Eupentacta fraudatrix*.

	Ортолог		Значения TPM			
Название	**человека**	**морского ежа**	**3 ДПЭ**	**5-7 ДПЭ**	**10 ДПЭ**	**Норма**
Ef-PCGF2	PCGF2	none	2,62	2,86	1,5	12,55
Ef-MSC	MSC	Myor2	3,41	4,19	0,88	0,04
Ef-ELF	ELF2	ElfA, ElfB	2,72	3,18	0,48	1,81
Ef-EGR	EGR1	Egr	5,36	8,57	3,54	3,06
Ef-ID	none	Id	3,69	5,23	2,75	3,39
Ef-PRDM9	PRDM9	none	1,02	2,29	0,93	0,31
Ef-TCF24	TCF24	Myor4	7,06	9,71	2,16	0,33
Ef-GATA3	GATA3	GataC	10,97	12,34	5,65	1,92
Ef-SNAI2	SNAI2	none	2,12	2,46	1,01	0,19
Ef-TBX20	TBX20	Tbx20_1	4,06	8,54	4,38	0,76
Ef-KLF1/2/4	KLF2	Klf2/4	11,11	19,2	9,73	18,71
	ELF4	ElfA, ElfB	2,45	3,52	0,78	1,97
	ZBTB47	none	1,08	1,98	1,01	2,29
	ZNF320	none	1,19	2,33	2,08	1,97
	ZNF300	none	1,29	2,67	2,58	2,07
	ZNF136	none	3,36	4,4	2,1	2,26
	OVOL3	none	1,48	2,41	1,24	0,84
	ZNF205	none	0,31	2,14	0,86	0,72

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). Для большинства выбранных генов динамика экспрессии схожа (Рисунок 9). Согласно методике отбора, все они демонстрируют максимальное значение TPM на второй стадии регенерации. При этом оно не сильно отличается от таковой на первой стадии. В то же время на третьей стадии экспрессия большинства генов в разы меньше, чем на второй или первой. Так, в зависимости от гена, на первой стадии экспрессия выше, чем на третьей в 1,1-5,7 раз. Наименьшая разница наблюдается для гена *Ef-prdm9*, а самая большая — для *Ef-elf*. Исключением является *Ef-tbx20*, у которого экспрессия на третьей стадии в 1,1 раза больше, чем на первой. Экспрессия через 5-7 СПЭ выше, чем через 3 СПЭ в 1,09-2,45 раза. При этом минимальная разница наблюдается для гена *Ef-pcgf2*, а

наибольшая — для гена *Ef-prdm9*. Экспрессия на второй стадии выше, чем на третьей в 1,9-6,63 раз, где наименьшая разница выявлена у гена *Ef-id*, а самая большая — у гена *Ef-elf*. Самые высокие показатели экспрессии через 5-7 СПЭ демонстрирует *Ef-klf1/2/4*, который наряду с *Ef-tbx20*, *Ef-prdm9* и *Ef-id*, входит в четверку генов, с самой схожей экспрессией на первой и третьей стадиях, что делает более явным пик экспрессии на второй стадии регенерации.

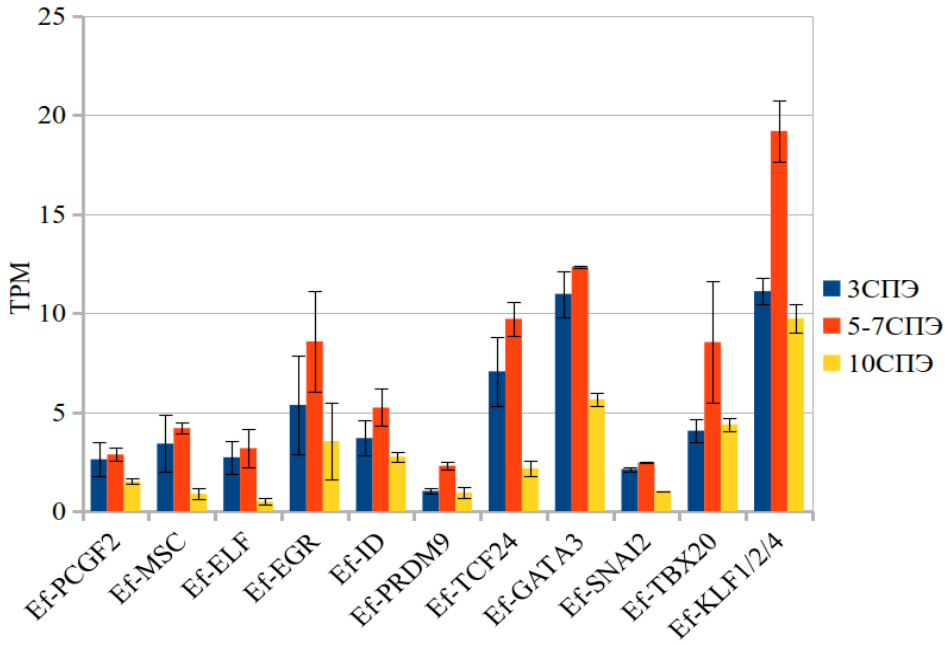

Рисунок 9. Средние значения экспрессии TPM для 11 генов ТФ на 3 стадиях регенерации *Eupentacta fraudatrix*. Погрешность показывает максимальное и минимальное значения в каком-либо повторе

3.2.5. Сеть сверхпредставленных биологических процессов и путей

Как уже упоминалось выше, было найдено 790 аннотаций (процессов и путей), однако 300 из них содержали либо слишком мало (меньше 5), либо слишком много (больше 100) генов. Такие аннотации не представляют явного интереса, так как большие аннотации обычно имеют более высокий ранг, то есть включают в себя другие аннотации, а малое число генов в аннотации связано с отсутствием этих генов среди обнаруженных у *E. fraudatrix*. Таким образом, число биологических процессов и путей сократилось до 490. Они объединяют 3168 ортологов генов человека (Рисунок 10).

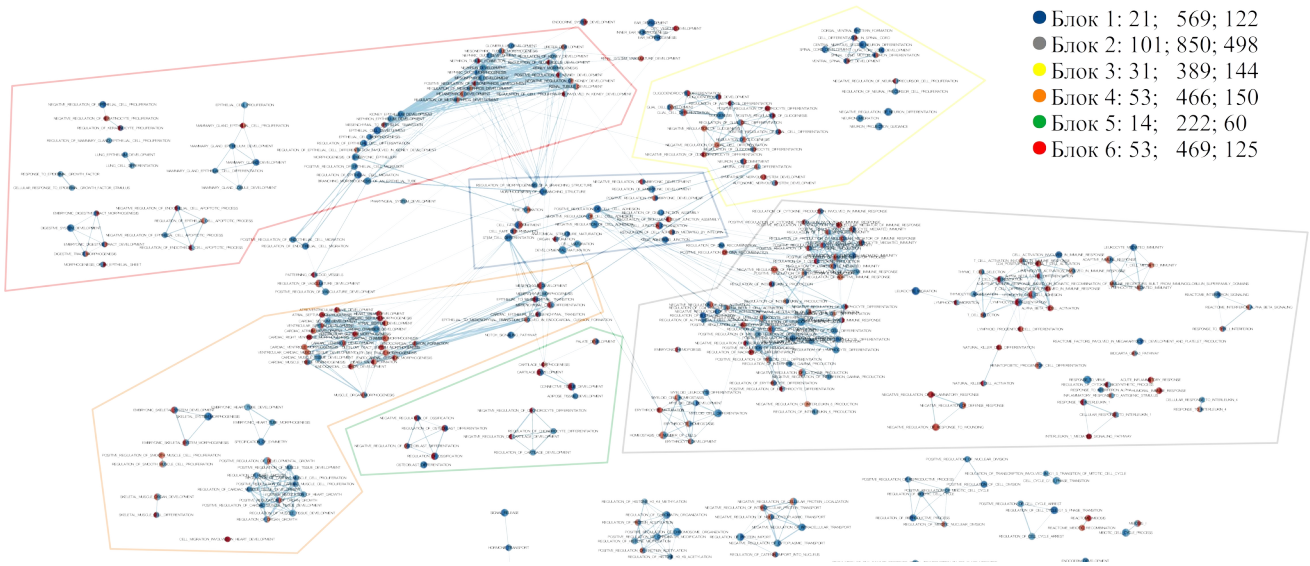

Рисунок 10. Сеть сверхпредставленных процессов и сигнальных путей, ассоциированных с 11 ТФ. Узел представляет процесс (набор участвующих генов), ребра показывают размер набора общих генов. Размер узла и толщина ребра зависят от числа генов. Градиент цвета указывает на повышение (красный) и понижение (синий) уровня экспрессии (оценка сверхпредставленности) на второй стадии относительно первой (правая половина) или третьей (левая половина) стадии регенерации. Описание блока сильно связанных процессов включает число процессов в блоке, число генов в нем и число уникальных генов блока

Среди этих биологических процессов и путей 204 содержали отрицательно регулируемые гены и 71 — положительно регулируемые гены на второй стадии относительно и первой, и третьей (Таблица 7).

Таблица 7. 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, ЭМТ	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

Сеть содержит 6 крупных, тесно связанных друг с другом блоков путей и биологических процессов. Эти блоки различаются числом генов, а также процессов и сигнальных путей. Кроме того, некоторые из этих блоков имеют большое число общих генов с другими блоками.

Первый блок включает такие биологические процессы, как регуляция морфогенезов, детерминация клеточной судьбы, регуляция клеточной адгезии и регуляция эмбрионального развития (Таблица 7, Рисунок 11). Динамика экспрессии генов, ассоциированных с процессами блока достаточно разнородна, однако большая часть генов имеет спад экспрессии на второй стадии. Лишь один процесс, регулирующий образование плотных межклеточных контактов, содержит гены с пиком экспрессии на второй стадии регенерации. По числу аннотаций данный блок не выделяется на общем фоне, однако он является вторым по числу генов.

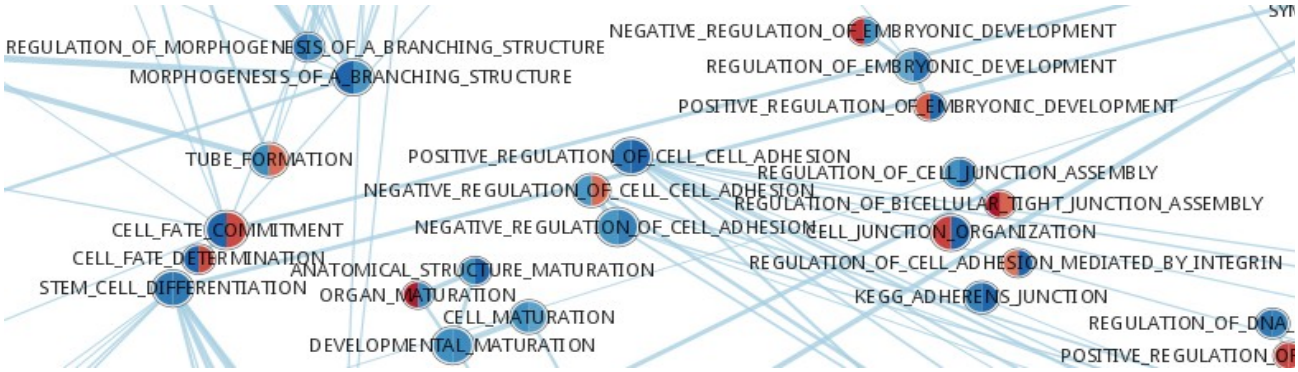

Рисунок 11. Блок 1. Включает термины, связанные с регуляцией морфогенезов, детерминацией клеточной судьбы, регуляцией клеточной адгезии и регуляцией эмбрионального развития

Второй блок включает процессы, связанные с развитием и регуляцией иммунного ответа на клеточном и молекулярном уровнях, включая дифференцировку иммунных клеток (Таблица 7, Рисунок 12). Кроме того, это самый большой блок по числу аннотаций, участвующих генов и уникальных для данного блока генов. Бóльшая часть процессов содержит гены с максимумом экспрессии на первой стадии и постепенным ее снижением к третьей.

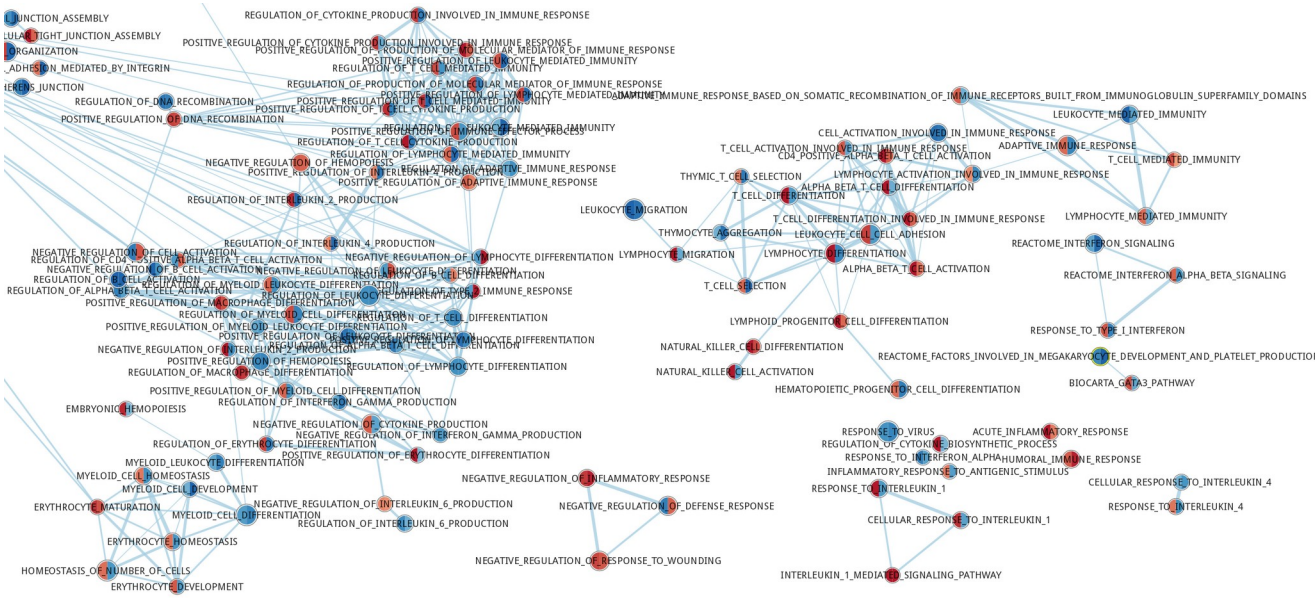

Рисунок 12. Блок 2. Включает процессы, связанные с развитием и регуляцией иммунного ответа на клеточном и молекулярном уровнях, включая дифференцировку иммунных клеток

Третий блок связан с нейрогенезом и, в основном, содержит процессы дифференцировки нервных клеток и ее регуляции (Таблица 7, Рисунок 13). Для большинства процессов здесь характерна градиентная динамика участвующих в них генов, с максимумом на третьей стадии регенерации, либо экспрессия с пиком на второй стадии.

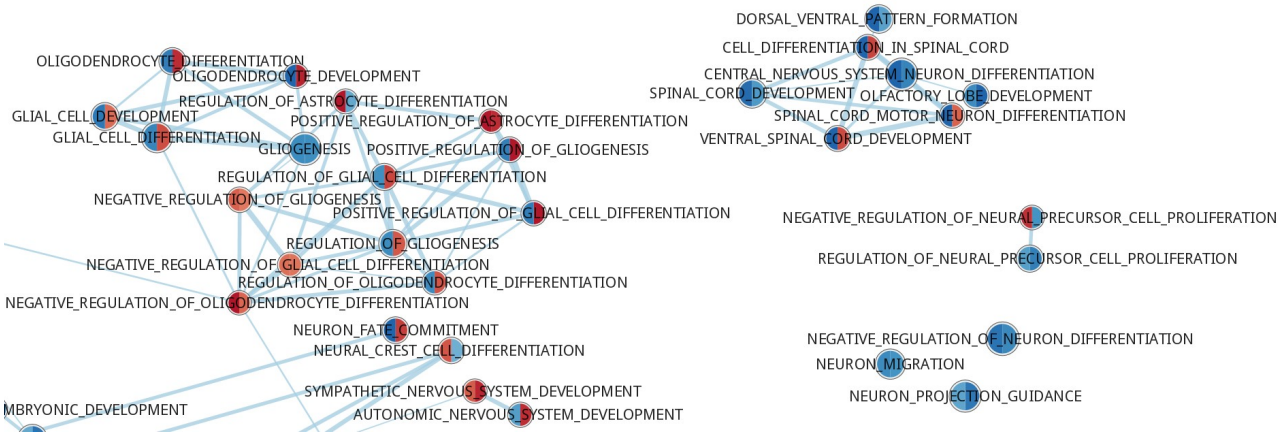

Рисунок 13. Блок 3. Связан с нейрогенезом и содержит процессы дифференцировки нервных клеток и ее регуляции

Четвертый блок объединяет процессы развития сердца, ЭМТ, развития и морфогенеза мышечных тканей и мезенхимы, дифференцировки и пролиферации мезенхимных и мышечных клеток (Таблица 7, Рисунок 14). Это второй по числу аннотаций и уникальных генов блок, а также третий по количеству участвующих генов. Процессы данного блока в большинстве содержат гены с максимумом

экспрессии на второй стадии, либо постепенно снижающейся экспрессией от первой к третьей стадиям.

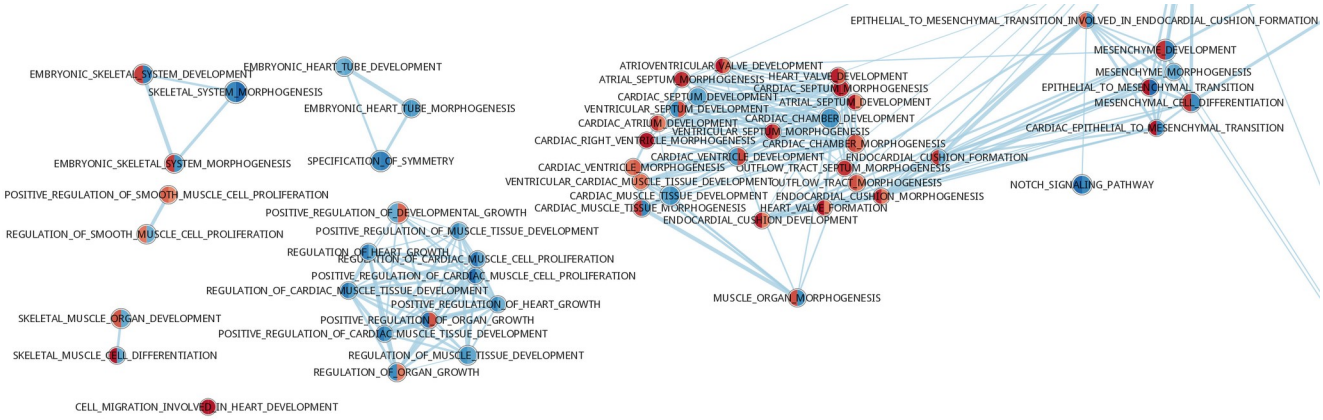

Рисунок 14. Блок 4. Объединяет процессы развития сердца, ЭМТ, развития и морфогенеза мышечных тканей и мезенхимы, дифференцировки и пролиферации мезенхимных и мышечных клеток

Пятый блок самый небольшой по числу процессов и генов. Он включает процессы развития и морфогенеза соединительной ткани, дифференцировки остеобластов и хондроцитов, а также регуляции внеклеточного матрикса (Таблица 7, Рисунок 15). Подавляющая часть процессов содержит гены с нисходящим характером экспрессии. Исключение составляет только морфогенез хрящевой ткани, который наиболее активен на второй стадии. Большая часть процессов данного блока связана с генами *snai2* и *id2*.

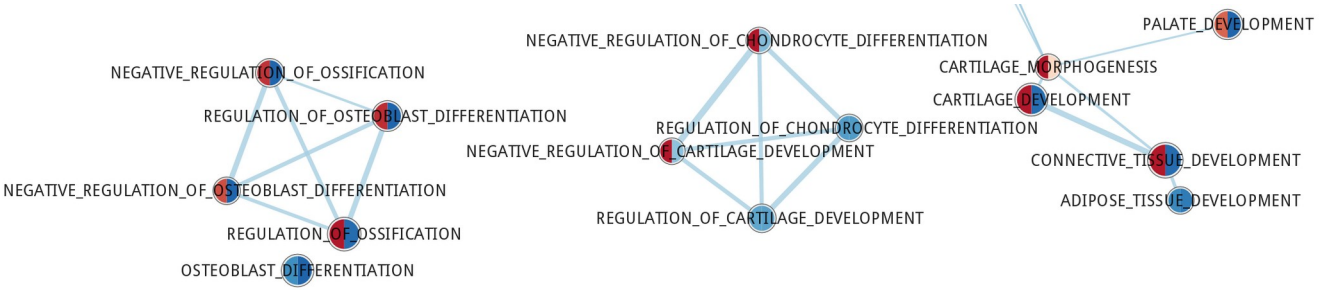

Рисунок 15. Блок 5. Включает процессы развития и морфогенеза соединительной ткани, дифференцировки остеобластов и хондроцитов, а также регуляции внеклеточного матрикса

Последний, шестой блок, связан с процессами развития почек, развития и морфогенеза различных эпителиев, дифференцировкой, пролиферацией и миграцией эпителиальных клеток (Таблица 7, Рисунок 16). Динамика экспрессии генов здесь достаточно разнообразна, но превалирующим остается уменьшение уровня экспрессии на второй стадии. Многие процессы данного блока по

динамике делятся на две примерно равные группы с постепенным увеличением или уменьшением уровня экспрессии генов при регенерации. И только два процесса, связанные с пролиферацией эпителиальных клеток гланд и репрессией пролиферации кератиноцитов, имеют гены с пиком экспрессии на второй стадии.

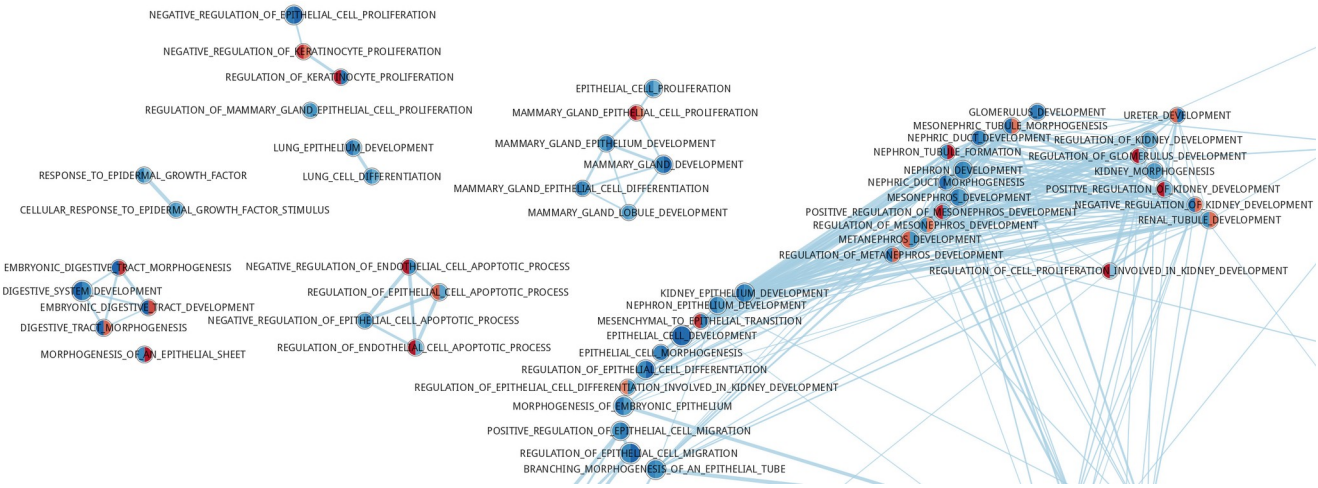

Рисунок 16. Блок 6. Связан с процессами развития почек, развития и морфогенеза различных эпителиев, дифференцировкой, пролиферацией и миграцией эпителиальных клеток

Помимо перечисленных шести блоков, есть небольшое число процессов и путей, не входящих ни в один блок по причине слабой связности их между собой, как на уровне общих генов, так и по «смыслу», определяющемуся в данном случае как процессы, действующие в основном в пределах клеточных потомков какого-либо одного зародышевого листка, ткани, органа или, как в случае со вторым блоком — сложного процесса. Оставшиеся процессы связаны с регуляцией и прохождением клеточного цикла, транспортом различных молекул и модификацией белков.

3.3 Анализ временно́й и пространственной динамики экспрессии генов-кандидатов

3.3.1. Оценка экспрессии с помощью кцПЦР

Для подбора комплектов олигонуклеотидов для постановки мультиплексной ПЦР был разработан скрипт OligoAnalyse, который помогает выбрать из множества возможных комплектов праймеров и зондов, или просто праймеров,

лучшие, минимизируя вероятность возникновения вторичных структур, гетеро- и гомодимеров.

Специфичность отжига праймеров была проверена с помощью секвенирования продукта ПЦР для каждого гена (последовательности в Приложении 3). Процент идентичности полученных последовательностей и соответствующих последовательностей из транскриптома составил в среднем 96,7, максимум у гена *Ef-klf1/2/4* - 99,2%, а минимум у гена *Ef-tbx20* — 92,4%.

При анализе результатов кцПЦР в первую очередь была проверена попарная корреляция между всеми стадиями, повторами и данными РНК-секвенирования. При этом, в случае РНК-секвенирования, брали не нормализованные значения числа выровненных прочтений на ген, а нормализовали на длину в килобазах и затем, как и в случае с кцПЦР, на *tubb* и *ef1a*, с вычислением среднего геометрического для оценок, полученных в результате нормализации на референсные гены. Это было необходимо в связи с тем, что обычно данные РНК-секвенирования нормализуются с использованием всех оценок экспрессии образца. Такой подход хоть и дает более точную оценку, но при этом вносит дополнительную вариабельность в случае сравнения экспрессии между данными секвенирования и кПЦР, так как последний использует нормализацию на референсные гены.

Проведенный анализ 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 (Рисунок 17). При этом результаты, полученные какой-либо одной методикой, показывают строгую корреляцию между повторами. Также наблюдается более высокая корреляция внутри образца 5-7 СПЭ, чем во всех прочих. Также был проведен анализ корреляции для каждого гена в отдельности, который показал, что наиболее вариабельными генами являются *Ef-egr*, *Ef-elf*, *Ef-msc* и *Ef-id* (Приложение 6). Причем для последних двух корреляция плохая между разными методами оценки,

в то время как между повторами одного метода значения экспрессии отлично скоррелированы, особенно у *Ef-id*, где корреляция близка к абсолютной.

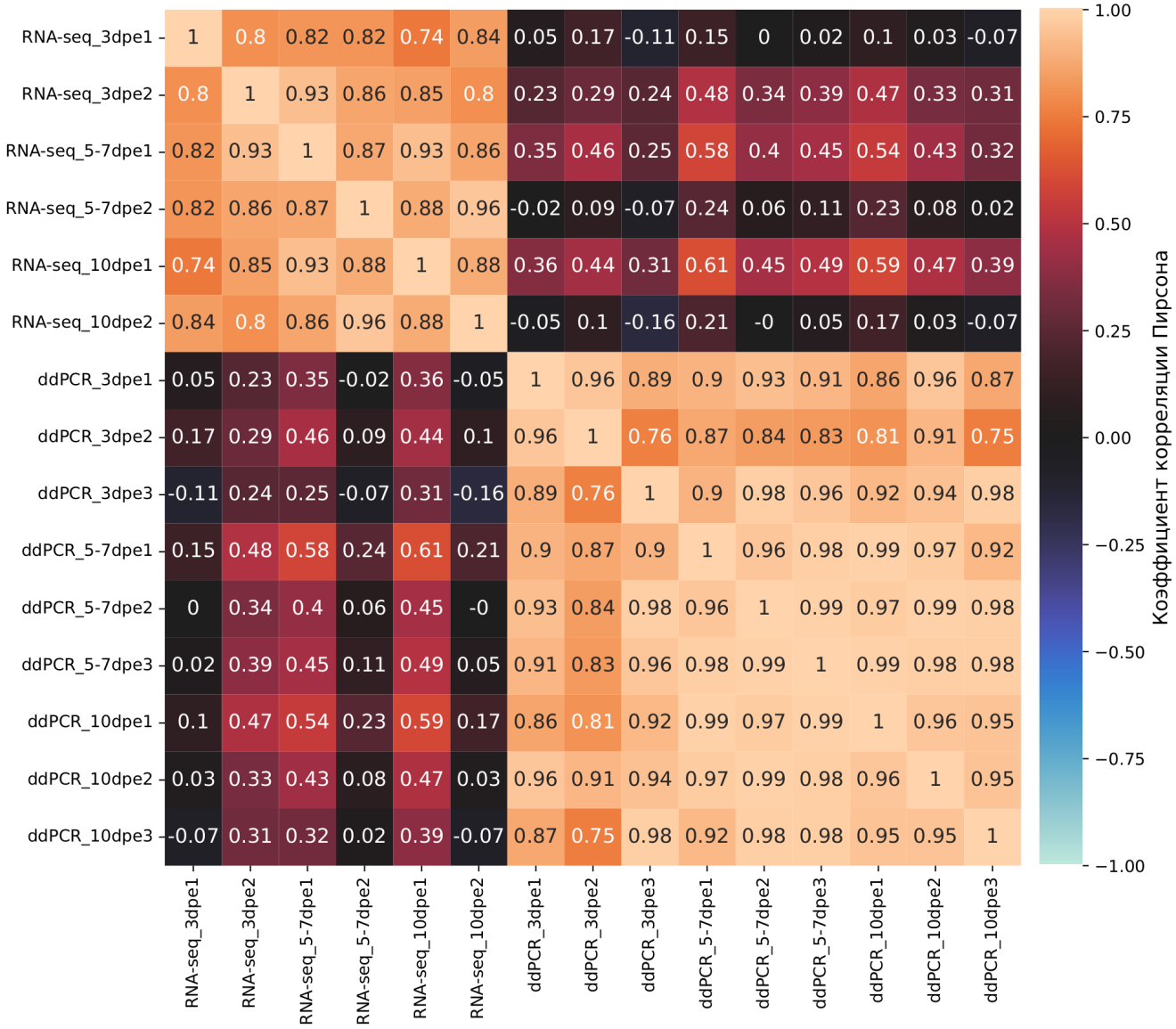

Рисунок 17. Тепловая карта коэффициентов корреляции Пирсона между оценками экспрессии 11 генов ТФ во всех имеющихся образцах, включая РНК-секвенирования (RNA-seq) и кцПЦР (ddPCR). Цифра после названия образца указывает на номер повтора

Динамика экспрессии по результатам кцПЦР также отличается для большинства генов (Рисунок 18). Так, заметный пик на стадии 5-7 СПЭ наблюдается у 5 генов — *Ef-egr*, *Ef-klf1/2/4*, *Ef-prdm9, Ef-tbx20* и *Ef-tcf24*. Еще 2 гена, хоть и имеют самую высокую экспрессию через 7 СПЭ, но в целом меняются слабо — *Ef-elf* и *Ef-id*. Другие 4 гена — *Ef-gata3*, *Ef-pcgf2*, *Ef-snai2* и *Ef-msc* — имеют в среднем пик через 3 СПЭ, но при этом в одном (*Ef-snai2*) или двух (*Ef-gata3*) повторах имеют пик через 7 СПЭ.

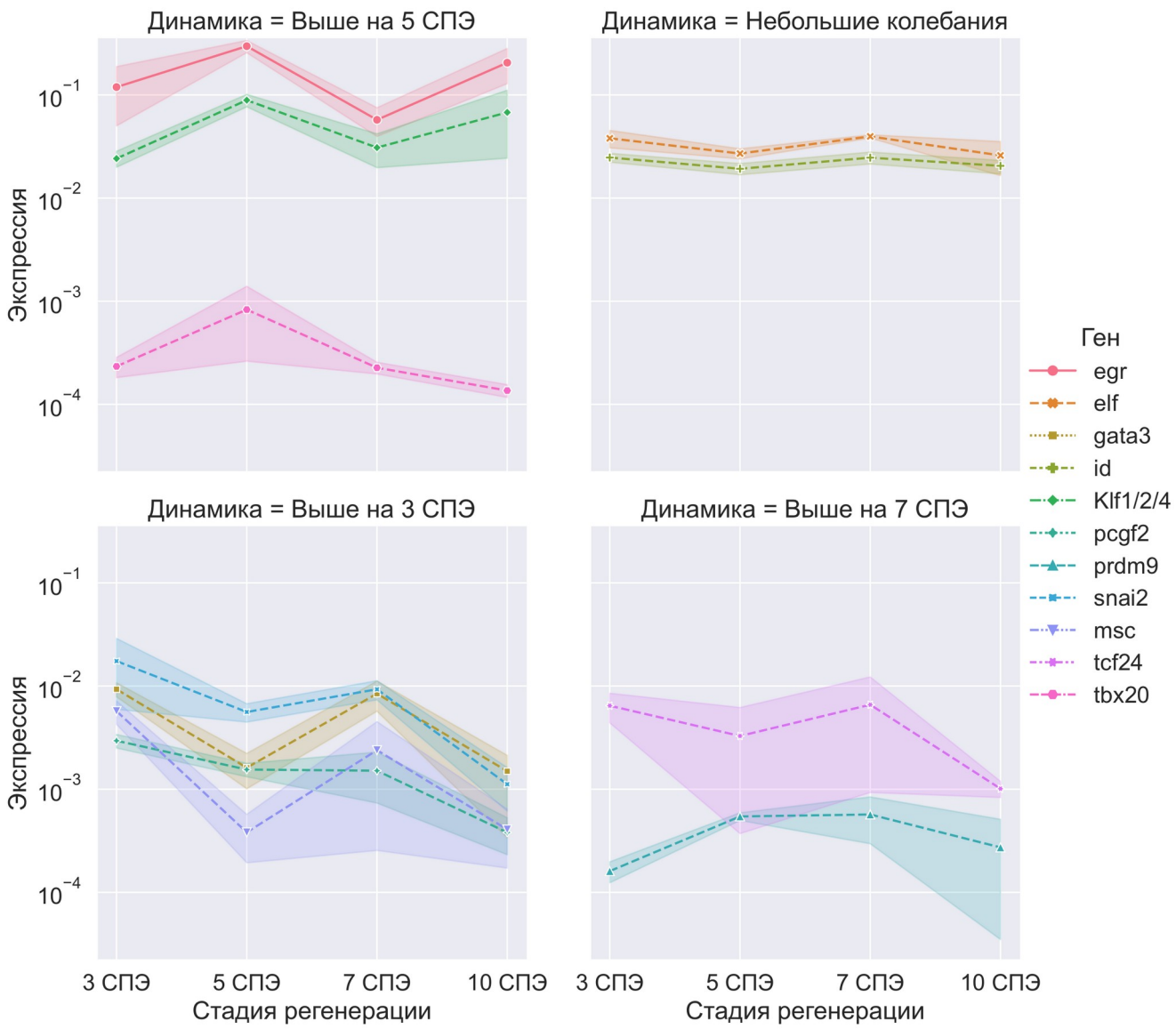

Рисунок 18. Нормализованные на референсные гены (tubb и ef1a) оценки значений экспрессии 10 генов ТФ через 3-10 спэ, полученные с помощью кцПЦР. Точки показывают среднее значение между повторами, полупрозрачная область показывает 90% доверительный интервал. Масштаб логарифмический по основанию 10. Гены разбиты по характеру динамики средних значений экспрессии.

3.3.2. 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. Было показано, что гены всех 11 ТФ экспрессируются в зачатке пищеварительной системы, однако распределение их транскриптов зависит от стадии регенерации (Рисунок 19, 20).

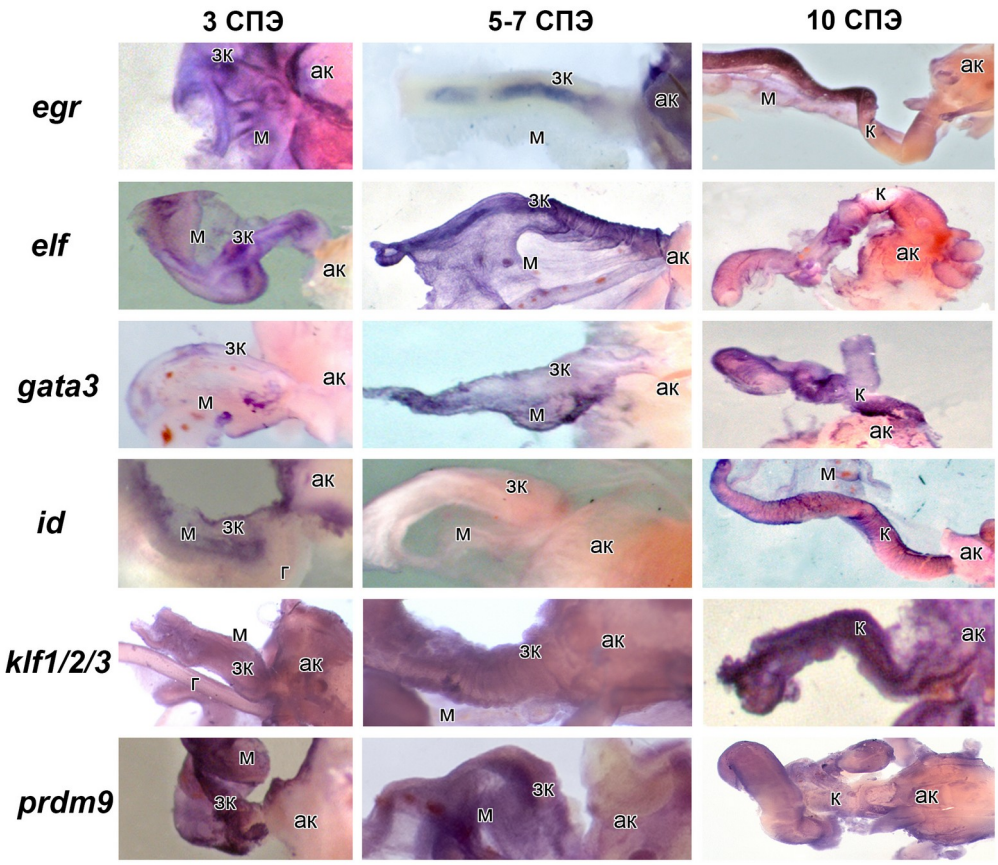

Рисунок 19. Локализация транскриптов генов *Ef-egr*, *Ef-elf*, *Ef-gata3*, *Ef-id*, *Ef-klf1/2/4* и *Ef-prdm9* в регенерирующих органах голотурии *Eupentacta fraudatrix*. ак – аквафарингеальный комплекс, г – гонодукт, зк – зачаток кишки, к – кишка, м – мезентерий

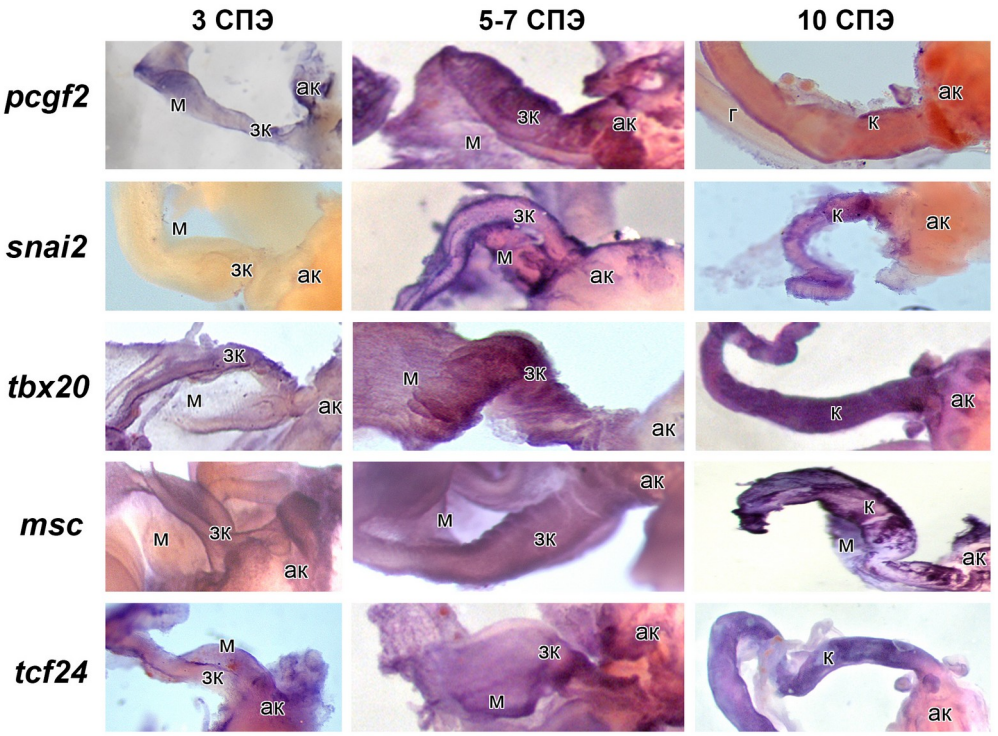

Рисунок 20. Локализация транскриптов генов *Ef-pcgf2*, *Ef-snai2*, *Ef-tbx20*, *Ef-msc* и *Ef-tcf24* в регенерирующих органах голотурии *Eupentacta fraudatrix*. ак – аквафарингеальный комплекс, г – гонодукт, зк – зачаток кишки, к – кишка, м – мезентерий

Экспрессия большинства исследованных генов отмечается в зачатке кишки и АК уже через 3 СПЭ. Это такие гены, как *Ef-egr*, *Ef-elf*, *Ef-gata3*, *Ef-klf1/2/4*, *Ef-pcgf*, *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 (Рисунок 19, 20). Продукты *Ef-snai2* с помощью метода WMISH на этой стадии в зачатках не выявляются.

Через 5-7 СПЭ заметна экспрессия почти всех исследуемых генов ТФ (Рисунок 19, 20). Интенсивно окрашивается мезентерий и соединительно-тканный зачаток кишки. Исключение составляет *Ef-id*, транскрипты которого методом WMISH не выявляются в регенерирующих органах в этот период. Транскрипты всех генов локализуются в целомическом эпителии. Однако, для *Ef-pcgf2*, *Ef-elf* и *Ef-msc* наибольшее количество мРНК (наибольшая интенсивность окраски) обнаруживается на вентральной стороне зачатка кишки, то есть 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 19, 20). Однако распределение транскриптов зависит от гена. Продукты *Ef-egr* выявляются только на некотором расстоянии от АК. По всей длине кишки окрашивается целомичекский 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.

4. ОБСУЖДЕНИЕ

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

4.1.1. *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 [112]. В большой степени оно зависит от предварительных условий, таких как изученность объекта, наличие аннотированных транскриптомов или геномов. Во время работы с транскриптомами голотурий *C. schmeltzii* и *A. japonicus*, мы обнаружили, что значения статистических параметров, описывающих целостность транскриптов и избыточность сборки *de novo*, значительно хуже сборок, основанных на качественном геноме [73,86,90,113–116]. Так, в случае со сборкой транскриптома *A. japonicus* в процессе развития, в результате работы геномного сборщика SPAdes было найдено 371845 последовательностей, содержащих ОРС длиной более 30 аминокислотных остатков и гомологией с белками морского ежа *S. purpuratus*, в то время как транскриптом, построенный на основе генома, содержал только 22643 подобных последовательности [90].

Такая проблема наблюдается независимо от вида животного или программы, используемой для сборки транскриптома *de novo.* В этой связи мы разработали метод улучшения сборки, получаемой с помощью доступных на данный момент программ. Данный подход к улучшению сборки был протестирован на наборах данных секвенирования двух голотурий – *C. schmeltzii* и *A. japonicus*. Несмотря на различия в количестве данных секвенирования и способах подготовки библиотек к секвенированию (с нормализацией кДНК и без нее), наш метод позволил

значительно уменьшить избыточность сборки при одновременном улучшении целостности транскриптов и сохранении биологически значимой информации.

Данное утверждение может показаться ошибочным, так как оценка БИ падает на каждом новом этапе сборки, проведенной по нашему алгоритму (Таблица 2). В частности, на первом этапе, когда происходит удаление исключительно одинаковых или являющихся частью других последовательностей, показатель БИ уменьшаться не должен. Однако оценка информации, тем не менее, падает. По нашему мнению, это происходит в связи как раз с уменьшением числа одинаковых последовательностей, которые тоже могут иметь какое-то выравнивание в базе, даже несмотря на фильтрацию по лучшему совпадению не *в* базе, а *для* какого-либо белка из базы. Наше предположение подтверждается при сравнении показателей информации на разных этапах сборки, указанных в последних двух столбцах Таблицы 2. Наибольшее падение БИ происходит при сравнении с базой SwissProt, а не с белками человека. Это объясняется тем, что SwissProt содержит в 20,2 раза больше последовательностей, а значит и выше вероятность того, что две очень схожие (>95%) последовательности из сборки будут являться лучшими совпадениями для двух разных белков из данной базы. Таким образом, получается завышение показателя информации просто из-за большего числа схожих последовательностей.

К сожалению, нам не удалось придумать как качественно обойти этот момент. Обычная нормализация на число последовательностей в сборке на разных этапах будет также вводить в заблуждение из-за того, что по факту это будет средняя оценка выравнивания, которая вполне могла бы быть выше после использования нашего алгоритма из-за увеличения длины последовательностей при одновременной потере некоторых из них. Впрочем, данная нормализованная оценка также приведена в Таблице 2 и, как видно, она почти неизменна.

В основе алгоритма работы нашей программы HomoloCAP лежат два наблюдения:

1. Доступные программы-сборщики при поиске Эйлерова пути в графе де Брюйна, представляющем отдельные транскрипты, совершают ошибку, переходя к области НТР раньше реального начала НТР, или же завершая ее там, где уже должна идти собственно белок-кодирующая область. Таким образом, в сборке появляется избыточное число последовательностей, белок-кодирующая область которых идентична более чем на 95% или последовательностей, белок-кодирующая область которых пересекается или же просто должна принадлежать одному транскрипту. НТР таких последовательностей различны частично или же полностью. Результатом этого становится увеличение избыточности сборки и фрагментированности транскриптов. Это наблюдение может быть легко подтверждено уменьшением числа ОРС в полтора раза после первой итерации алгоритма, которая состоит только лишь в объединении полностью идентичных ОРС (Таблица 2).
2. OLC сборщики, которые имеет смысл использовать после первичной сборки с помощью графов де Брюйна для улучшения целостности транскриптов и схлопывания схожих последовательностей, работают на целой сборке долго и неэффективно с точки зрения статистических параметров, описывающих качество сборки.

Алгоритм нашей программы учитывает данные наблюдения и предусматривает несколько шагов для их реализации, описанных подробно в разделе «Результаты». Фактически, основным нововведением здесь является отказ от сохранения НТР в пользу целостности транскриптов. За счет этого становится возможным объединение изоформ транскрипта и его частей в одну последовательность. Другим преимуществом является кластеризация последовательностей, что в дальнейшем позволяет не только использовать многопоточные процессоры, в отличие от стандартного CAP3, который этого не умеет, но и исключить из сборки те последовательности, которые не похожи. Все это существенно уменьшает время вычислений, экономит необходимые ресурсы и

увеличивает точность сборки. Последнее происходит за счет уменьшения вероятности включить в итоговую последовательность ту, которой там не должно быть, то есть получить химерный транскрипт.

Данный подход, то есть кластеризация и затем сборка внутри кластера, использовали и ранее [117], но при этом кластеризация осуществлялась однократно, а не в итеративном цикле. Использование итерации является важным, поскольку она, хотя и требует дополнительного времени, но зато позволяет сначала объединить максимально схожие последовательности, и только затем переходить к более отличающимся. Это не только сохраняет точность сборки, но, что также важно, минимизирует вероятность появления кластера последовательностей, члены которого нельзя собрать единственным способом. Последнее может приводить к сбою программы-сборщика и, соответственно, пропуску данного кластера.

Конечно, как было упомянуто выше, для работы нашего алгоритма приходится отказаться от информации об НТР. Последняя, несомненно, важна для характеристики данного гена или транскрипта. В НТР могут располагаться различные регуляторные участки, например, рибопереключателя, AU-богатых элементов или короткой ОРС, которые влияют на стабильность и уровень трансляции мРНК. Тем не менее, судя по наблюдаемой нами картине наличия большого числа последовательностей с разными НТР, но идентичными ОРС, в случае отсутствия сборки генома использовать эту информацию нужно крайне осторожно, так как неизвестно, насколько можно доверять последовательностям НТР, полученным путем сборки транскриптома *de novo*.

4.1.2. Поиск ортологов

Когда нужно быстро найти пары последовательностей, которые могли бы быть ортологами в терминах выравнивания (то есть заключение об ортологичности строится только на основе выравнивания), стандартом является использование реципрокного BLAST-поиска. В подразделе 3.1.2 был

продемонстрирован (Таблица 3) ряд очевидных минусов 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, например, принадлежащих разным типам.

Последнее хорошо иллюстрируется почти полным отсутствием разницы между двумя методами в случае ортологов мыши и человека, что ожидаемо, так как состав генных семейств, структура генома и сами последовательности близки у данных видов, соответственно и правильно определить ортологи с помощью BLAST проще. Однако, чем больше увеличивается эволюционная дистанция, тем больше становится разница между реципрокным и модифицированным методами, особенно относительно известного числа ортологов по данным базы. Кроме того, если сравнивать методы по всем ортологам, а не только по типу ортологов «один к одному», как ранее, то, во-первых, видно, что разница подходов становится значительно более очевидной. Например, для *D. rerio* модифицированный метод позволяет найти на 866 ортологов больше (Таблица 3). Во-вторых, сохраняется тенденция увеличения разницы между методами при увеличении эволюционной дистанции, за исключением *D. rerio*, видимо по причине частой дупликации генов у рыб [118,119]. В-третьих, если сравнить первое и второе число среди разных методов, то можно заметить, что основная разница между ними складывается из двух типов ортологий — «один ко многим» и «многие ко многим». Это очень важный вывод, так как при увеличении эволюционной дистанции происходит и увеличение доли этих двух типов ортологий в связи с увеличением разницы между строением генных семейств. Однако стоит заметить, что оба метода дают достаточно высокую оценку ложноположительных результатов, то есть числа ортологов, которые по данным Ensembl ортологами не являются.

Также мы попробовали провести пост-фильтрацию по различным квантилям в массиве оценок выравнивания, предсказанных ортологов, но результат был хуже, чем при вышеописанном варианте снижения числа ошибочно

предсказанных ортологов. В то же время, использование информации о доменах должно увеличить точность поиска ортологов. При этом поиск доменов проводится быстрее, чем BLAST-поиск, что важно для ускорения работы. Такой подход применяется во многих методах определения типов гомологий между группой видов. Например, базы Ensembl и EggNOG, кроме привлечения информации о расположении генов в геноме и построения филогенетических деревьев в пределах группы родственных белков, используют информацию о доменном строении последовательностей. Также важным является тот факт, что модифицированный метод, в отличие от реципрокного, не требует двух кругов BLAST-поиска, что вдвое уменьшает время расчетов. Кроме того стоит заметить, что нет необходимости проводить BLAST-поиск с окончательным выравниванием по алгоритму Смита-Ватермана как предлагается в работе Ward et al [120]. Результаты с применением этого алгоритма и без него отличаются действительно незначительно во всех вариантах нашего анализа, в то время как скорость выполнения BLAST-поиска падает в связи с тем, что алгоритм Смита-Ватермана нельзя масштабировать на несколько вычислительных потоков.

Таким образом, можно заключить, что подход к быстрому поиску ортологов, разработанный и апробированный в данной работе, работает быстрее и несколько лучше классического реципрокного и может применяться при отсутствии геномных данных. 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».

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

4.2.1. Секвенирование и сборка транскриптома *de novo*

Полученная сборка транскриптома по всем параметрам является хорошей при сравнении с показателями *de novo* сборок транскриптомов других видов

иглокожих [73,86,113,114,116,121,122] (подробнее в подразделе 4.1.1). Ясно наблюдаются более высокие статистические характеристики, описывающие полноту восстановленности последовательности транскриптов, при уменьшении числа последовательностей, что было достигнуто благодаря описанному в подразделе 3.1.1 подходу к улучшению первичной сборки. В то же время, может выглядеть странным, что на сборку выравнивается лишь 21% прочтений, что является крайне низким значением по сравнению с обычным для транскриптомов [73,113,114]. Однако данный показатель объясняется строгими параметрами выравнивания и отсутствием НТР. Влияние НТР на показатель процента обратно-картирующихся прочтений можно наблюдать, если сравнить его у сборок *A. japonicus* [73,114], где он равен 80-85% и наших сборок без НТР [90], где это значение равнялось 16-26%. При этом сформированный нами транскриптом [90] имеет лучшие статистические показатели без потерь информации относительно прочих сборок.

4.2.2. Анализ дифференциальной экспрессии генов

При анализе дифференциальной экспрессии в первую очередь можно заметить, что существенная часть генов является ДЭГ при сравнении уровня экспрессии на какой-либо из стадий регенерации с нормой. Так, среди трех изученных стадий найдено 2917 общих ДЭГ с положительной динамикой относительно нормы и 5465 — с отрицательной. Эту картину дополняет поиск ДЭГ относительно стадии 5-7 СПЭ, при котором на положительную динамику приходится 6758 уникальных для нормы генов и 4457 — на отрицательную динамику. Все это говорит о том, что в процессе регенерации экспрессия многих генов, имеющаяся в неповрежденной кишке, подавляется. При этом происходит увеличение экспрессии генов, не экспрессирующихся или экспрессирующихся значительно слабее (в 4 раза) в интактной кишке. То есть подтверждается очевидный факт, что в процессе регенерации кишки у *E. fraudatrix* происходит глобальная перестройка работы всего генома, причем пик этого изменения

приходится на 5-7 СПЭ. Затем происходит постепенное возвращение к тому шаблону работы генома, который наблюдается в норме, что подтверждается соотношением положительно и отрицательно регулируемых генов на разных стадиях регенерации (Рисунок 8).

При анализе активности генов в регенерирующей кишке *E. fraudatrix* наибольшее сходство наборов ДЭГ характерно для стадий 3 СПЭ и 5-7 СПЭ, в отличие от 3 СПЭ и 10 СПЭ. Этот факт коррелирует с ранее полученными морфологическими данными [12,19,26,27]. В первые 7 сут после эвисцерации основными процессами являются преобразование внеклеточного матрикса, дедифференцировка клеток целомического эпителия и их миграция. Трансдифференцировка затрагивает ограниченное число клеток, поэтому вклад ее в наборы ДЭГ незначителен. Естественно, в течение этого периода времени профили экспрессии генов большинства клеток будут близки между собой. Через 10 СПЭ основные структуры кишки уже сформированы и начинается активная дифференцировка составляющих их клеток. Такая кардинальная смена работы генома, естественно, отражается и на составе ДЭГ.

Кроме того, полученные нами данные показывают, что в случае с *E. fraudatrix* сравнивать экспрессию в нормальных тканях с экспрессией в измененных тканях можно, но только для получения общей, описательной картины происходящего. Дело в том, что у этого вида после эвисцерации в переднем отделе тела удаляются все клетки пищеварительной системы [12]. Зачаток кишки закладывается *de novo*, «из ничего», за счет преобразования мезодермальных производных – клеток целомического эпителия. Соответственно, именно это состояние и эти клетки являются «нулевой точкой отсчета» в процессе морфогенеза, а не неповрежденная пищеварительная трубка с дифференцированными энтероцитами. Таким образом, в случае регенерации передней части кишки у *E. fraudatrix* информация о наборе генов и изменении их экспрессии на стадии 5-7 СПЭ относительно нормы может показать лишь то, что работа генома кардинально реорганизована. Поскольку же целью данной работы

являлся поиск возможных регуляторов трансдифференцировки клеток целомического эпителия в энтероциты, соответственно и сравнение должно происходить с клетками мезентерия. В нашей работе мы дополнительно использовали в качестве контроля саму стадию 5-7 СПЭ, что в дальнейшем позволило нам не только отсечь гены, экспрессирующиеся активнее на стадии 3 СПЭ, но и на 10 СПЭ, таким образом оставив пул генов с пиком экспрессии на стадии 5-7 СПЭ, то есть в период трансдифференцировки.

4.2.3. Поиск кандидатов на роль регуляторов клеточной трансдифференцировки

Классический способ поиска важных генов-регуляторов определенного процесса заключается в выделении пула генов, относительная экспрессия которых на тех или иных этапах пиковая. Отбираются гены с изменением экспрессии более чем в 4 раза относительно контроля и вероятностью ошибки оценки дифференциальности экспрессии (p-value или p-adj) менее 0,05 [31,86,123,124]. Такой классический подход в случае анализа немодельных видов, как указывалось в разделе «Обзор литературы», имеет ряд недостатков, таких как отсутствие комплексного подхода к анализу процессов и разнообразных способов представления экспрессии. Что, в свою очередь, создает сложности в интерпретации результатов и построении гипотез о механизмах процесса.

В этой связи в нашей работе мы решили отойти от такой классической процедуры и сконцентрировались на поиске ТФ, абсолютная экспрессия которых была бы пиковой на второй стадии регенерации. Для этого в первую очередь мы нашли ортологи генов человека и морского ежа, используя ранее описанный модифицированный реципрокный подход. Далее среди найденных ортологов мы выделили те, которые являются ТФ по данным баз HGNC и Echinobase. Из них только 20 проявляли искомую нами динамику экспрессии и только 11 из этих генов были ортологами генов ТФ, для которых имелись исследования их функций — 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. Конечно, отсутствие исследованных функций у ряда генов не говорит, что их не стоит рассматривать как возможные регуляторы трансдифференцировки. Однако этот же факт не позволяет строить гипотезы об их функциях при регенерации. Кроме того, ТФ человека крайне активно изучаются на предмет вовлечения в различные заболевания, развитие и дифференцировку клеток, в связи с чем низка вероятность того, что важный в таких процессах ген до сих пор не попал в поле зрения исследователей. Также для этих генов не найдены ортологи среди ТФ морского ежа и значения их экспрессии почти всегда с трудом проходят пороговые значения при фильтрации.

4.2.4. Сеть сверхпредставленных биологических процессов и путей

Другим этапом валидации списка ТФ как возможных регуляторов трансдифференцировки являлось построение и анализ сети биологических процессов и сигнальных путей, связанных с выявленными 11 ТФ. Полученные данные подтверждают участие этих ТФ в восстановлении кишки у *E. fraudatrix*. В построенной сети выявляются большие, связанные аннотации (блоки), описывающие те или иные морфогенезы на клеточном, тканевом и органном уровнях. При этом велика вероятность, что выявленные ТФ служат ключевыми факторами регуляции регенерации кишки у *E. fraudatrix*. Все они являются «верховыми» регуляторами многочисленных процессов, запуская сложные каскады генов, а некоторые из них напрямую участвуют в модификации хроматина, о чем будет подробнее упомянуто ниже.

К анализу полученных данных необходимо подходить с осторожностью и принимать во внимание два соображения. Во-первых, в основе генной онтологии лежат процессы, описанные у человека и других млекопитающих. В этой связи они не могут быть напрямую применены для анализа морфогенезов у животных, эволюционно дистантных позвоночным. Особенно это касается интерпретации таких групп терминов, как «biological process» и «molecular function». У беспозвоночных может быть совсем иной набор органов, чем у млекопитающих, а

возможные «гомологичные» органы могут иметь другое строение и функции. В частности, у иглокожих отсутствует выделительная система, однако в сети сверхпредставленных процессов и сигнальных путей, ассоциированных с изучаемыми ТФ *E. fraudatrix*, имеется большой кластер генов (блок 6), в который входят такие термины, как «glomerulus development», «mesonephric tubule morphogenesis» и др. Анализ и интерпретация таких процессов у *E. fraudatrix* (и у других животных) возможен только в контексте стоящих за этими терминами общебиологическими функциями. Например, «mesonephric tubule morphogenesis» у *E. fraudatrix*, вероятно, характеризует процессы, связанные с преобразованием эпителиев регенерирующей кишки и формирования трубчатого зачатка кишечной выстилки.

Во-вторых, чем больше эволюционная дистанция между видами, тем больше различий в числе генов и, что важнее, составе генных семейств. В этой связи очень часто невозможно установить ортологию многих генов между этими видами животных. Основной массив информации о функциях генов базируется на функциях генов человека, вследствие чего точная функция гомологов человеческих генов у беспозвоночных с помощью такого анализа не может быть определена.

Блок 1 топографически располагается в центре сети и связан со всеми основными блоками. Такое расположение соотносится с процессами, которые в нем сосредоточены: регуляция морфогенезов, детерминация клеточной судьбы, регуляция клеточной адгезии и регуляция эмбрионального развития. Это базовые процессы, происходящие при любом типе морфогенеза, в том числе и при регенерации. В совокупности, первый блок достаточно точно характеризует события, происходящие при формировании кишки у *E. fraudatrix* [12,19]. Например, среди узлов, содержащих гены с положительной динамикой экспрессии, имеются такие процессы, как «cell fate commitment» и «cell fate determination». Увеличение экспрессии генов, участвующих в процессах детерминации клеточной судьбы и ее закрепления соответствует специализации

клеток целомического и кишечного эпителиев, происходящей в период 3-11 СПЭ. Наличие в этом блоке KLF2 (Ef-KLF1/2/4), SNAI2, ID2 (Ef-ID), GATA3 и TBX20 указывает на их участие в базовых механизмах регенерации кишки у *E. fraudatrix*.

Второй блок содержит гены и процессы, связанные с развитием и регуляцией иммунного ответа. Большое число аннотаций, входящих в данный блок, указывает на значительную активацию иммунной системы после эвисцерации. Вероятно, это один из важнейших процессов, позволяющий голотуриям противостоять бактериальному заражению после значительных повреждений. Кроме того, выделяющиеся иммуноцитами активные формы кислорода и антиоксидантные ферменты, по-видимому, являются триггерами запуска таких процессов, как дедифференцировка и миграция клеток [44].

Третий блок, связанный с нейрогенезом, по-видимому, отражает процессы восстановления базиэпителиального нервного сплетения. В норме оно представлено большим количеством нейронов и их отростков, расположенных в целомическом эпителии мезентерия и кишки. После эвисцерации они сохраняются только в мезентерии. Новые аксоны появляются в зачатке кишки только через 10-15 СПЭ [125]. Наши результаты подтверждают морфологические данные. Активность нейрогенеза, судя по экспрессии участвующих генов, возрастает по мере регенерации.

Четвертый блок объединяет процессы, большинство из которых связаны с развитием сердца. Поскольку сердце отсутствует у иглокожих, данный блок следует рассматривать как совокупность процессов преобразования мезодермы и миоэпителиальных клеток. Аннотации здесь связаны с реорганизацией мезодермальных клеток, таких как пролиферация, миграция и ЭМТ, происходящих при развитии сердца у млекопитающих [126]. По-видимому, это следует интерпретировать как отражение сходных преобразований миоэпителиальных клеток при регенерации кишки у *E. fraudatrix*. Это единственный блок, который характеризуется большим числом процессов,

наиболее активных на стадии 5-7 СПЭ. Это указывает, что гены и процессы данного блока играют определяющую роль в регенерации кишки у *E. fraudatrix*.

Активация механизмов ЭМТ при регенерации кишки *E. fraudatrix* необычна и в какой-то мере противоречит морфологическим данным, поскольку межклеточные контакты при погружении целомического эпителия в соединительно-тканное утолщение не разрушаются и ЭМТ не происходит. Это можно объяснить тем, что при эпителиальных морфогенезах клетки могут приобретать некоторые мезенхимные черты [127]. Они сохраняют межклеточные контакты, но уплощаются и формируют ламеллоподии [128,129]. Сходные процессы могут происходить во время погружения целомического эпителия в соединительно-тканный зачаток кишки у *E. fraudatrix* и клетки могут принимать некоторые мезенхимные свойства. 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. Эти ТФ способны провоцировать ЭМТ эпителиальных клеток [130–136]. Также это может объясняться трансформацией части целомического эпителия в мезенхиму, которая не связана с формированием энтероцитов, как описано для голотурии *H. glaberrima* [33].

Тем не менее большая часть процессов, ожидаемо связана с TBX20, который как указывалось выше, критичен для развития сердца и мезодермальных тканей. Данный ген, как и GATA3, в четвертом блоке у млекопитающих в основном ассоциирован с различными морфогенезами и миграции клеток при развитии сердца. У *E. fraudatrix* данные процессы, возможно, участвуют в обеспечении миграции целомического эпителия. Активизация многих процессов, объединенных в четвертый блок, именно на второй стадии указывает на важность этого периода в регенерации кишки.

Пятый блок также представляет большое значение при анализе механизмов восстановления пищеварительной системы, поскольку связан с ремоделированием внеклеточного матрикса и дифференцировкой клеток соединительной ткани.

Постепенно затухающая активность этих процессов находится в соответствии с морфологическими данными. Формирование соединительнотканного зачатка кишки наиболее интенсивно происходит в первую неделю регенерации [12,19], что отражается в наибольшей экспрессии генов на стадии 3 СПЭ. Дифференцировка хондроцитов и остеобластов, отсутствующих как таковые у иглокожих, по-видимому, отражает изменения фибробластов, также как и активное формирование внеклеточного матрикса стенки кишки. Интересно, что процесс морфогенеза хрящевой ткани, единственный в блоке с наибольшей активностью через 5-7 СПЭ, имеет много общих генов с процессами из четвертого блока, связанными с регуляцией и прохождением ЭМТ. Это указывает на большое значение ко-экспрессии *Ef-snai2* и *Ef-id* во время регенерации вообще и в процессе трансдифференцировки, в частности.

Шестой блок, по-видимому, описывает процессы эпителиальной трансформации, поскольку голотурии не имеют почек или каких-либо специализированных выделительных органов [137]. У *E. fraudatrix* гомологи этих генов, вероятно, вовлечены в реорганизацию мигрирующих кластеров эпителиальных клеток в трубчатый кишечный эпителий, что отражают такие процессы как «формирование канальца нефрона», «морфогенезы мезонефрических канальцев» и т. д. Большая часть процессов ассоциирована с GATA3, но также в этом блоке имеются KLF2, TBX20, EGR1, SNAI2 и ID2. Большая часть процессов здесь, по-видимому, не связана с самой трансдифференцировкой, но важна для правильного становления структуры регенерирующей кишки.

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.

4.3 Временно́е и пространственное распределение экспрессии генов-кандидатов

При подборе комплектов олигонуклеотидов для постановки мультиплексной ПЦР мы столкнулись с проблемой большого числа возможных комбинаций олигонуклеотидов, которые надо проверить на гетеро- и гомодимеры. Так, в случае гетеродимеров, наличие которых наиболее сложно проверить при разработке олигонуклеотидов для мультиплексных ПЦР, даже дуплексная ПЦР с интеркалирующим красителем потребует 4 олигонуклеотида, то есть 6 разных сочетаний праймеров из 4 по 2, которые надо проверить на энергию Гиббса. Когда же проводится та же дуплексная ПЦР, но с зондами, число олигонуклеотидов вырастает до 6, а значит сочетаний получится 15. В случае триплексной реакции, которая и была изначально задумана в данной работе, сочетаний для проверки становится уже 36. Таким образом, даже для одного таргетного гена и двух референсных (триплексная реакция) сделать все проверки в бесплатных программах для подбора олигонуклеотидов становится сложно. В данной же работе необходимо было сделать 11 триплексных реакций. В связи с этим был разработан скрипт OligoAnalyse, позволяющий среди множества комбинаций праймеров и зондов найти лучшие, основываясь на расчете энергии Гиббса для вторичных структур, гетеро- и гомодимеров.

Карта корреляций образцов в целом нам показывает значительно более низкие значения при сравнении образцов с разными методами оценки экспрессии (Рисунок 17). Однако при этом видно, что один из повторов, как через 5-7 СПЭ, так и 10 СПЭ, имеет значительно лучшую корреляцию с результатами кцПЦР.

Возможно, это обусловлено высокой вариабельностью экспрессии ТФ через 3 СПЭ. Действительно, если посмотреть на Рисунок 18, то видно, что некоторые гены сильно вариабельны в этот период.

Для того чтобы проверить, насколько в целом низкая корреляция зависит от конкретных генов, мы вычислили парную корреляцию между всеми повторами, включая повторы РНК-секвенирования, для каждого гена по отдельности (Приложение 6). Полученные данные показывают, что, по-видимому, низкие значения корреляции в основном обусловлены генами *Ef-elf* и *Ef-id*, а также, частично *Ef-egr* и *Ef-msc*. Возможно, причиной различий в оценке экспрессии разными методами является не полностью восстановленные последовательности транскриптов этих генов. Предварительные данные показывают, что Ef-elf, Ef-id и Ef-msc, судя по выравниванию с их ортологами у голотурии *A. japonicus* не имеют полностью собранного транскрипта. Решением данной проблемы может быть либо более глубокое секвенирование, либо формирование генома исследуемого вида и сборка на основе геноманых данных.

Транскрипты всех 11 ТФ выявляются в клетках зачатка кишки начиная с 3 СПЭ, что указывает вовлеченность этих ТФ в процесс регенерации. Однако распределение продуктов разных генов несколько меняется в зависимости от стадии восстановления. Кроме того, отличается и интенсивность окраски, что, в какой-то мере, может указывать на изменение уровня экспрессии данного гена. Все это дает основание предположить, что функции изучаемых ТФ в регенерации различаются и каждый из генов участвует в разных морфогенетических процессах. По результатам WMISH невозможно установить роль конкретного гена в восстановлении кишки, однако в некоторых случаях распределение транскриптов позволяет сделать предположение о функциях гена и возможном участии в трансдифференцировке.

На всех изученных стадиях регенерации все 11 ТФ экспрессируются в целомическом эпителии мезентерия и зачатка кишки. Мезотелий играет большую роль в механизмах регенерации у голотурий [138]. Его клетки на ранних этапах

после эвисцерации дедифференцируются, мигрируют и митотически делятся. Позднее, происходит редифференцировка, в результате чего формируются две популяции клеток – перитонеоциты и миоэпителиальные клетки [11,13]. Одновременно начинает восстанавливаться базиэпителиальный нервный плексус. Все эти процессы требуют тонкой и скоординированной регуляции, в которой, вероятно, и принимают участие выявленные нами ТФ. Однако это не исключает участие их в трансдифференцировке и формировании кишечного эпителия. Ранее нами было показано, что ген ТФ *Ef-sox9/10* также экспрессируется в целомическом эпителии при регенерации кишки у *E. fraudatrix* [36]. Однако на стадии 5-7 СПЭ его транскрипты обнаруживаются также в зоне погружения клеток в соединительно-тканное утолщение, а затем, через 10 СПЭ, в кишечном эпителии.

Транскрипты *Ef-egr* детектируется на всех исследованных стадиях, что совпадает с результатами кцПЦР. Через 5-7 СПЭ хорошо видно, что продукты этого гена располагаются в формирующемся кишечном эпителии. В этой связи можно предположить, что *Ef-egr* может быть одним из кандидатов в гены, участвующих в трансдифференцировке. У млекопитающих он участвует в прохождении клеточного цикла в различных типах рака, а также в регенерации печени [139,140]. Кроме того, недавно у бескишечной турбеллярии *Hofstenia miamia* было обнаружено, что ген *egr* экспрессируется вблизи места повреждения и, вероятно, является первичным регулятором генной сети, активация которой приводит к регенерации отсеченной части тела [141]. Также экспрессия этого гена отмечена у позвоночных, книдарий, планарий и морских звезд на ранних стадиях заживления ран [141] и при регенерации личинок морской звезды *P. miniata* после поперечного разрезания [142]. Однако у *E. fraudatrix egr* проявляет большую вариабельность, что видно как по результатам РНК-секвенирования, так и кцПЦР. 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.

По результатам WMISH мРНК *Ef-elf* через 5-7 СПЭ обнаруживается на вентральной стороне зачатка кишки, то есть в той области, где происходит трансдифференцировка и погружение клеток целомического эпителия. В дальнейшем транскрипты *Ef-elf* локализуются только на поверхности формирующейся пищеварительной трубки. Члены семейства Ets, которому принадлежит Ef-ELF, являются регуляторами клеточной пролиферации, ангиогенеза, гематопоэза, опухолевой трансформации и дифференцировки [143,144]. Однако функции ELF1, ELF2 и ELF4 у позвоночных связаны с дифференцировкой лимфоидных и эндотелиальных клеток, выживанию стволовых и прогениторных гематопоэтических клеток и супрессии опухолей [145–148]. У морского ежа *S. purpuratus* экспрессия гена *elf* обнаруживается на стадии поздней гаструлы во всех клетках, но наибольшая концентрация наблюдается в кишке [111]. В целом, наши данные косвенно коррелируют с поведением *elf* в развитии. У *E. fraudatrix Ef-elf* экспрессируется на первых стадиях регенерации в период закладки кишечного эпителия, в то время как позднее (10 СПЭ), когда пищеварительная трубка уже сформирована, экспрессия резко уменьшается почти до нуля (Рисунок 9, 18). Это позволяет предположить, что ортолог ELF у иглокожих может быть вовлечен в регуляцию ранних стадий формирования кишечного эпителия и, возможно, в трансдифференцировку.

GATA3 является широко известным ТФ, который у многоклеточных тесно связан с мезодермальными клетками, их морфогенезами и трансформацией [149–151]. Сходные функции демонстрирует данный ген и в эмбриогенезе морских ежей [150]. *Ef-gata3* – это один из самых хорошо коррелирующих между повторами и оценками разных методов ген. Судя по данным анализа сверхпредставленных процессов, он также участвует в большом числе процессов. Однако по результатам WMISH он обнаруживается только в целомическом эпителии зачатка кишки. Анализ сети сверхпредставленных процессов показал его

вовлеченность в преобразование мезодермы. В этой связи, учитывая динамику экспрессии *Ef-gata3* в норме и при регенерации, можно предположить, что этот ТФ участвует не в трансдифференцировке, а в дедифференцировке и поддержании этого состояния у клеток мезотелия. Это предположение подтверждается ранней активацией *Ef-gata3* (3 СПЭ), когда основным процессом в зачатке кишки является дедифференцировка клеток целомического эпителия.

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 показались нам заслуживающими внимания. Несмотря на отсутствие ДНК-связывающего домена, этот белок оказывает модифицирующее действие на ТФ с bHLH-доменом, образуя с ними гетеродимер [134]. Экспрессия гена *id2* у млекопитающих тесно связана с эпителиальными клетками, он может частично восстанавливать эпителиальный фенотип у мезенхимных клеток. Кроме того, он является антагонистом ЭМТ и предотвращает преждевременную дифференцировку предшественников эпителиальных клеток в процессе развития кишечника у мышей. Уменьшение экспрессии гена *id2* в эпителиальных клетках вызывает их дедифференцировку, а также ингибирует дифференцировку разных типов клеток [133–136]. Мы полагаем, что у *E. fraudatrix Ef-id* может выполнять сходные функции, 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*. Это предположение хорошо коррелирует с известными данными о функциях этих генов и морфогенетическими процессами, происходящими на начальных стадиях (3-7 СПЭ) регенерации кишки.

Интересно, что данный ген демонстрирует исключительный уровень скоррелированности результатов оценки экспрессии между повторами одного метода. Разница же в оценках разных методов, возможно, является результатом неполноты транскрипта, в связи с чем пострадала точность оценки РНК-секвенирования. Тем не менее, такая уникальная степень повторяемости оценки говорит об исключительной важности гена *Ef-id* для успешного прохождения регенерации.

Ef-KLF1/2/4, вероятно, является гомологом человеческих генов KLF1, KLF2 и KLF4 [35]. KLF участвуют в регуляции клеточной пролиферации, дифференцировки и развития [152]. В частности, KLF2 и KLF4 — ключевые ТФ в поддержании близкого к стволовому состояния клеток и репрограммировании соматических клеток [153]. Ген *Ef-klf1/2/4*, очевидно, играет значительную роль в регенерации кишки у *E. fraudatrix*. Среди всех изученных генов *Ef-klf1/2/4* показывает наибольший уровень экспрессия на всех трех исследованных стадиях регенерации. Как и у *Ef-egr*, его экспрессия в несколько раз выше на стадии 5 СПЭ относительно 3 СПЭ и 7 СПЭ. Также для данного гена характерны высокие значения корреляции при любых комбинациях повторов и методов. Результаты WMISH также сходятся с данными РНК-секвенирования и кцПЦР, по которым транскрипты *Ef-klf1/2/4* детектируется в основном через 5-7 СПЭ. Результаты анализа сети сверхпредставленных процессов показывают, что *Ef-klf1/2/4* связан с активирующимися на второй стадии процессами детерминации клеточной судьбы (Таблица 7, Рисунок 11). Таким образом, *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. В этой связи, данный ген является одним из кандидатов в регуляторы трансдифференцировки у *E. fraudatrix*.

Косвенным подтверждением этого предположения является работа Машанова и др. [35]. Данные авторы не зафиксировали каких-либо изменений в экспрессии *klf1/2/4* при регенерации кишки и нервного тяжа у голотурии *H.*

glaberrima. На наш взгляд такое отличие является следствием разных механизмов регенерации кишки у *E. fraudatrix* и *H. glaberrima*. Известно, что у *H. glaberrima* кишечный эпителий формируется за счет дедифференцировки энтероцитов оставшейся части пищевода [23]. То есть, у данного вида при регенерации нет необходимости в кардинальном репрограммировании работы генома клеток, как это происходит у *E. fraudatrix* или у других животных [56].

Экспрессия гена *Ef-pcgf2* в течение регенерации сравнительно мала и имеет нисходящий характер динамики. Также невелика и разница в экспрессии между стадиями, особенно 3 и 5-7 СПЭ. Это указывает на то, что кодируемый этим геном белок регулирует некие процессы, происходящие в течение всей регенерации. Продукты *Ef-pcgf2* локализуются исключительно в целомическом эпителии зачатка кишки. Ген связан с 4 блоком сети, объединяющий процессы преобразования мезодермы и миоэпителиальных клеток, а также ЭМТ. ТФ семейства PCGF являются хроматин-модифицирующими белками, регулируя опосредованно экспрессию ряда генов. Показано, что PCGF2 способен к ингибированию ЭМТ через репрессию генов *zeb1* и *zeb2* при раке груди у человека, а также при раке желудка негативно регулирует свойства стволовых клеток [154,155]. Сложно сказать какую функцию выполняет этот ТФ у *E. fraudatrix*, так как модификаторы эпигенетического статуса генома действуют на широкий спектр генов. Однако исходя из функций его ортолога у человека, можно предположить, что *Ef-pcgf2* вовлечен в стабилизацию мезодермальных свойств клеток целомического эпителия, находящихся на поверхности зачатка кишки.

В противоположность *Ef-klf1/2/4*, экспрессия *Ef-prdm9* находится на самом низком уровне по сравнению с другими генами ТФ. Однако *Ef-prdm9* с точки зрения динамики экспрессии достаточно интересен, так как пик экспрессии у него на стадии 5-7 СПЭ более явный, чем у других. Также данные кцПЦР хорошо коррелируют между собой и РНК-секвенированием. Однако относительно его функции сложно сказать что-то конкретное, так как в блоки сети сверхпредставленных процессов он не попал. По результатам WMISH его

экспрессия обнаруживается только в целомическом эпителии зачатка кишки. У хордовых ТФ семейства PRDM являются важными эпигенетическими регуляторами, работающими во время развития и клеточной дифференцировки. Они нужны для поддержания плюрипотентного состояния клеток у мыши во время развития [156–159]. Естественно, что разбалансировка работы этих генов приводит к развитию различных болезней, включая несколько типов рака. При этом одни члены семейства являются онкогенами, а другие — супрессорами онкогенеза [160]. 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.

Ef-snai2, как и *Ef-id*, судя по анализу сети сверхпредставленных процессов, участвует в очень большом числе процессов. При этом его экспрессия снижается в процессе регенерации. Однако один из повторов кцПЦР демонстрирует низкие показатели и выбивается при вычислении попарной корреляции, что снизило среднее значение экспрессии в этот период. Остальные повторы отлично коррелируют как между собой, так и между кцПЦР и РНК-секвенированием и, возможно, в действительности данный ген имеет увеличение экспрессии в период 5-7 СПЭ, как показало РНК-секвенирование. Также транскрипты *Ef-snai2* обнаруживаются в целомическом эпителии зачатка кишки. ТФ семейства SNAI играют важную роль в ЭМТ и репрессии мезенхимо-эпителиальной трансформации [130–132]. При регенерации кишки у *E. fraudatrix* как таковой ЭМТ во время трансдифференцировки не происходит, клетки целомического эпителия во время трансформации не теряют межклеточные контакты [12]. Тем не менее они могут приобретать во время погружения в соединительную ткань зачатка некоторые мезенхимные признаки. Схожие процессы происходят при развитии трубчатых органов в эмбриогенезе и при регенерации у широкого спектра многоклеточных животных [127–129]. При регенерации губок трансформация эпителия может протекать как без утраты межклеточных

контактов, так и с их частичным разрушением [4]. Наличие заметной экспрессии *Ef-snai2* может объясняться и взаимодействием Ef-SNAI2 с ID. Последний способен образовывать гетеродимерный комплекс с ТФ семейства SNAI, подавляющего репрессионную активность SNAI на регулируемые им гены [135,136]. Таким образом, можно предположить, что *Ef-snai2* может 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.

Ef-msc и *Ef-tcf24* принадлежат семейству TCF. Это достаточно обширная и хорошо изученная группа ТФ, члены которого принимают активное участие в процессах канцерогенеза, эмбриогенеза и регенерации [161]. *Ef-msc* и *Ef-tcf24* имеют характерную динамику экспрессии с пиком на стадии 5-7 СПЭ и мало отличающуюся от уровня экспрессии на стадии 3 СПЭ. И если по MSC (TCF21) можно найти немало литературы, то по другому гену, TCF24, в настоящее время отсутствует информация о его участии в регенерации, развитии или канцерогенезе. *Ef-tcf24* имеет достаточно вариабельную экспрессию как по результатам РНК-секвенирования, так и кцПЦР, особенно в период 5-7 СПЭ. При этом *Ef-tcf24* в норме почти не детектируется, что говорит о его участии в регенерации. Более того, экспрессия этого гена через 10 СПЭ снижается в 5 раз, из чего можно сделать вывод о его возможном участии в трансдифференцировке. Также, данный ген не попал в какие-либо блоки сверхпредставленных процессов, так как он отсутствует в базах биологических процессов GO или MSigDB. В этой связи трудно предположить его функции по аналогии с другими животными. Тем не менее по результатам WMISH через 10 СПЭ его транскрипты обнаруживаются в кишечном эпителии. Этот факт, а также динамика экспрессии *Ef-tcf24*, указывает на его участие в регенерации. Возможно, данный ген регулирует либо поздние стадии трансформации, либо дифференцировку энтероцитов.

Сходная картина экспрессии характерна и для *Ef-msc*, ортолог которого у млекопитающих вовлечен в ЭМТ, регуляцию миогенеза и функционирование

стволовых клеток, репрессируя миогенез и действуя как линий-специфичный репрессор развития эмбриональной скелетной мускулатуры [162,163]. Также он регулирует LIF-индуцированную экспрессию некоторых ренопротекторных факторов в сторонней популяции клеток почек взрослого организма, которые находятся в плюрипотентном состоянии [162]. Продукты *Ef-msc* обнаруживаются в целомическом эпителии через 3 СПЭ. А в период трансдифференцировки его транскрипты имеются на вентральной стороне зачатка кишки, в месте погружения целомического эпителия. Однако в дальнейшем, в кишечном эпителии продукты *Ef-msc* не выявляются. С учетом данных литературы об участии этого гена в регуляции дифференцировки исключительно мышечных клеток и ЭМТ, 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.

Последний из исследованных генов, *Ef-tbx20*, отличается от большинства исследованных генов четко видимым пиком экспрессии на стадии 5-7 СПЭ. Это могло бы послужить отличной основой для гипотезы об участии его в трансдифференцировке, однако семейство TBX больше связано с дифференцировкой клеток экто- и мезодермы. В частности TBX20 у млекопитающих важен для кардиогенеза, а также экспрессируется при развитии нервной системы и в латеральной мезодерме зародышей мыши [164,165]. По данным РНК-секвенирования его экспрессия увеличивается на стадии 5-7 СПЭ. Результаты кцПЦР показывают строгий пик экспрессии на 5 СПЭ, при этом на остальных стадиях примерно равные значения экспрессии. Также наблюдается отличная корреляция между всеми повторами и обоими методами. Интересно, что транскрипты *Ef-tbx20* через 10 СПЭ обнаруживаются в кишечном эпителии. В этой связи можно предположить, что в отличие от млекопитающих, у голотурий данный ген может участвовать 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 участие в этом процессе. 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 регенерации кишки у *E. fraudatrix*. Ef-KLF1/2/4, вероятно, 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 модельные объекты для изучения проблемы источников регенерации, поскольку наличие у них стволовых клеток (за исключением первичных половых клеток) до сих пор достоверно не установлено [9]. Многочисленные морфологические данные свидетельствуют, что восстановление у них протекает за счет специализированных клеток в результате их активной транс- или дедифференцировки [9–14]. При этом у иглокожих в широких пределах варьирует скорость регенерации, потенциал восстановления у разных видов, а также механизмы регенерации утраченных структур [10,12,27,28,13,19,20,22–26].

В то же время, несмотря на очевидные преимущества Echinodermata в качестве модельных объектов для изучения механизмов регенерации и их эволюции, на данный момент хоть сколько-то модельным, с точки зрения наличия качественного генома и знания о функциях генов, можно считать только морского ежа *S. purpuratus* и, в значительно меньшей степени, дальневосточного трепанга *A. japonicus*. При этом морские ежи имеют наименьший потенциал к регенерации среди всех классов Echinodermata [10]. Кроме того, существует обширный ряд сложностей, связанных с анализом данных секвенирования немодельных видов. В связи с этим, в первую очередь нам необходимо было хотя бы частично решить эти проблемы.

В ходе данной работы нами был реализован ряд подходов, упрощающих анализ данных, в основном РНК-секвенирования. Была написана программа для улучшения сборок, получаемых с помощью коммерческих сборщиков транскриптомов, не требующая доступа к высокопроизводительным вычислительным системам и реализованная средствами кросс-платформенных программ и языка программирования. Результаты ее работы с одной стороны обладают точностью, с другой — улучшенными показателями сборки, в том числе таким важным параметром, как полнота восстановленности транскриптов. Также

был разработан и реализован новый подход к поиску ортологов генов. Для него была показана лучшая чувствительность, чем для классических «быстрых» способов поиска ортологов. 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. Такое объединение разных типов информации позволяет выделять ключевые гены и формировать некие предположения, проверку которых уже можно реализовать средствами классической молекулярной биологии. В нашей работе показано, что новые высокопроизводительные подходы даже в непростых случаях способны выполнять свое главное предназначение — формирование гипотез о сложных процессах, что позволяет ускорить исследования и уменьшить их обширность за счет прицельных экспериментов. Также было ясно проиллюстрировано, что в некоторых случая, в частности при анализе динамики экспрессии генов при регенерации, зачастую нельзя использовать в качестве контрольного образца неповрежденные ткани. Например, у *E. fraudatrix* при эвисцерации удаляется вся пищеварительная система и клетки кишки не участвуют в ее восстановлении. В этой связи сравнивать экспрессию генов клеток неповрежденной и регенерирующей кишки бессмысленно. Кроме того, все алгоритмы оценки дифференциальной экспрессии генов в своей основе предполагают, что большая часть генов между образцами не будет менять свою экспрессию, чего не наблюдается при сравнении профиля экспрессии генов в интактном органе и регенерате.

При анализе транскриптома было сделано два вывода об общем планировании анализа. Во-первых, зачастую не имеет смысла ограничиваться только генами, сильно меняющими экспрессию относительно контроля. В приведенных в разделах 3.2 и 4.2 работах такие гены в основном были связаны с некими, безусловно важными процессами, но не являлись непосредственными его

регуляторами, а скорее генами, экспрессия которых необходима для приобретения клеткой неких признаков, то есть генами структурных белков и ферментов. В случае же 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. Обсуждению данного момента и того, к чему может привести абстрагированность от таких данных посвящена значительная часть раздела 4.2.

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, с первого взгляда кажущийся не сильно важным — регенерация в первые 10 суток у разных особей идет крайне схоже, если базироваться на данных экспрессии оцененной кцПЦР. Не исключено, что это следствие выборки генов, но даже такая информация важна. Она показывает, что независимо от внутрипопуляционной 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 в целом. Другим важным моментом являются результаты экспрессии в целом и результаты WMISH, в частности. Эти результаты в основном сходятся с предположениями, выдвинутыми в подразделах 4.2.3 и 4.3.3. Так, гены *Ef-gata3*, *Ef-pcgf2* и *Ef-tbx20*, вероятно участвующие в морфогенезах мезодермальных клеток, преимущественно имеют более вариабельную 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*, как и было предположено, скорее участвуют в предотвращении преждевременной

спецификации клеток, проходящих через трансдифференцировку, и приобретении некоторых мезенхимных признаков у клеток. А *Ef-klf1/2/4*, скорее всего, 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 ЛИТЕРАТУРЫ

1. Lai A.G., Aboobaker A.A. EvoRegen in animals: Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes // Dev. Biol. 2018. Vol. 433, № 2. P. 118–131.
2. Borisenko I.E., Adamska M., Tokina D.B., Ereskovsky A. V. Transdifferentiation is a driving force of regeneration in *Halisarca dujardini* (Demospongiae, Porifera) // PeerJ. 2015. Vol. 3.
3. Lavrov A.I., Bolshakov F. V, Tokina D.B., Ereskovsky A. V. Sewing up the wounds : The epithelial morphogenesis as a central mechanism of calcaronean sponge regeneration // J. Exp. Zool. Part B Mol. Dev. Evol. 2018. Vol. 330, № 6–7. P. 351–371.
4. Ereskovsky A. V., Tokina D.B., Saidov D.M., Baghdiguian S., Le Goff E., Lavrov A.I. Transdifferentiation and mesenchymal-to-epithelial transition during regeneration in Demospongiae (Porifera) // J. Exp. Zool. B. Mol. Dev. Evol. 2020. Vol. 334, № 1. P. 37–58.
5. Jopling C., Sleep E., Raya M., Martí M., Raya A., Belmonte J.C.I. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation // Nature. 2010. Vol. 464, № 7288. P. 606.
6. Maki N., Suetsugu-Maki R., Tarui H., Agata K., Del Rio-Tsonis K., Tsonis P.A. Expression of stem cell pluripotency factors during regeneration in newts // Dev. Dyn. 2009. Vol. 238, № 6. P. 1613.
7. Ahmed E., Sansac C., Assou S., Gras D., Petit A., Vachier I., et al. Lung development, regeneration and plasticity: From disease physiopathology to drug design using induced pluripotent stem cells // Pharmacol. Ther. 2018. Vol. 183. P. 58–77.
8. Ельчанинов А.В., Фатхудинов Т.Х. Регенерация печени млекопитающих: межклеточные взаимодействия. Москва: Наука, 2020. 126 с.

9. Vogt G. Hidden Treasures in stem cells of indeterminately growing bilaterian invertebrates // Stem Cell Rev. Reports. 2012. Vol. 8, № 2. P. 305–317.
10. Dolmatov I.Yu. Regeneration in echinoderms // Russ. J. Mar. Biol. 1999. Vol. 25, № 3. P. 225–233.
11. Dolmatov I.Yu., Ginanova T.T. Muscle regeneration in holothurians // Microsc. Res. Tech. 2001. Vol. 55, № 6. P. 452–463.
12. Mashanov V.S., Dolmatov I.Yu., Heinzeller T. Transdifferentiation in holothurian gut regeneration // Biol. Bull. 2005. Vol. 209, № 3. P. 184–193.
13. Garcia-Arraras J.E., Dolmatov I.Yu. Echinoderms: potential model systems for studies on muscle regeneration // Curr. Pharm. Des. 2010. Vol. 16, № 8. P. 942–955.
14. Kalacheva N. V., Eliseikina M.G., Frolova L.T., Dolmatov I.Yu. Regeneration of the digestive system in the crinoid *Himerometra robustipinna* occurs by transdifferentiation of neurosecretory-like cells // PLoS One. 2017. Vol. 12, № 7. P. 1–28.
15. Mashanov V.S., Zueva O.R., García-Arrarás J.E. Inhibition of cell proliferation does not slow down echinoderm neural regeneration // Front. Zool. 2017. Vol. 14, № 1. P. 1–9.
16. Wolff T. The concept of the hadal or ultra-abyssal fauna // Deep Sea Res. Oceanogr. Abstr. 1970. Vol. 17, № 6. P. 983–1003.
17. Rogacheva A. V. Revision of the Arctic group of species of the family Elpidiidae (Elasipodida, Holothuroidea) // Mar. Biol. Res. 2007. Vol. 3, № 6. P. 367–396.
18. O'Loughlin P.M., Whitfield E. New species of Psolus Oken from Antarctica (Echinodermata: Holothuroidea: Psolidae) // Zootaxa. 2010. Vol. 2528, № 1. P. 61–68–61–68.
19. Dolmatov I.Yu., Mashanov V.S. Regeneration in holothurians / Vladivostok: Dalnauka, 2007.
20. Mashanov V.S., García-Arrarás J.E. Gut regeneration in holothurians: A snapshot of recent developments // Biol. Bull. 2011. Vol. 221, № 1. P. 93–109.

21. Dolmatov I.Yu. Variability of Regeneration Mechanisms in Echinoderms // Russ. J. Mar. Biol. 2020. Vol. 46, № 6. P. 391–404.
22. Shukalyuk A.I., Dolmatov I.Yu. Regeneration of the digestive tube in the holothurian *Apostichopus japonicus* after evisceration // Russ. J. Mar. Biol. 2001. Vol. 27, № 3. P. 168–173.
23. García-Arrarás J.E., Estrada-Rodgers L., Santiago R., Torres I.I., Díaz-Miranda L., Torres-Avillán I. Cellular mechanisms of intestine regeneration in the sea cucumber, *Holothuria glaberrima* Selenka (Holothuroidea:Echinodermata) // J. Exp. Zool. 1998. Vol. 281, № 4. P. 288–304.
24. Dolmatov I.Yu., Khang N.A., Kamenev Y.O. Asexual reproduction, evisceration, and regeneration in holothurians (Holothuroidea) from Nha Trang Bay of the South China Sea // Russ. J. Mar. Biol. 2012. Vol. 38, № 3. P. 243–252.
25. Dolmatov I.Yu. New data on asexual reproduction, autotomy, and regeneration in holothurians of the order Dendrochirotida // Russ. J. Mar. Biol. 2014. Vol. 40, № 3. P. 228–232.
26. Leibson N.L., Dolmatov I.Yu. Evisceration and regeneration of the inner complex in a sea cucumber *Eupentacta fraudatrix* (Holothuroidea, Dendrochirota) // Zool. Zhurnal. 1989. Vol. 68, № 8. P. 67–74.
27. Leibson N. Regeneration of digestive tube in holothurians *Stichopus japonicus* and *Eupentacta fraudatrix* // Keys for Regeneration / Basel: Karger, 1992. № 23. P. 51–61.
28. Dolmatov I.Yu. Regeneration of the aquapharyngeal complex in the holothurian *Eupentacta fraudatrix* (Holothuroidea, Dendrochirota) // Keys for Regeneration / Basel: Karger, 1992. P. 40–50.
29. Campbell C., Lancman J.J., Palazon R.E., Matalonga J., He J., Graves A., et al. In vivo lineage conversion of vertebrate muscle into early endoderm-like cells // bioRxiv. 2019. P. 722967.
30. Sun L., Chen M., Yang H., Wang T., Liu B., Shu C., et al. Large scale gene expression profiling during intestine and body wall regeneration in the sea

cucumber *Apostichopus japonicus* // Comp. Biochem. Physiol. Part D Genomics Proteomics. 2011. Vol. 6, № 2. P. 195–205.

31. Sun L., Yang H., Chen M., Ma D., Lin C. RNA-Seq Reveals Dynamic Changes of Gene Expression in Key Stages of Intestine Regeneration in the Sea Cucumber *Apostichopus japonicus* // PLoS One. 2013. Vol. 8, № 8. P. e69441.
32. Li Y., Wang R., Xun X., Wang J., Bao L., Thimmappa R., et al. Sea cucumber genome provides insights into saponin biosynthesis and aestivation regulation // Cell Discov. 2018. Vol. 4, № 1.
33. García-Arrarás J.E., Valentín-Tirado G., Flores J.E., Rosa R.J., Rivera-Cruz A., San Miguel-Ruiz J.E., et al. Cell dedifferentiation and epithelial to mesenchymal transitions during intestinal regeneration in *H. glaberrima* // BMC Dev. Biol. 2011. Vol. 11, № 1. P. 61.
34. Ortiz-Pineda P.A., Ramírez-Gómez F., Pérez-Ortiz J., González-Díaz S., Santiago-De Jesús F., Hernández-Pasos J., et al. Gene expression profiling of intestinal regeneration in the sea cucumber // BMC Genomics. 2009. Vol. 10. P. 1–21.
35. Mashanov V.S., Zueva O.R., García-Arrarás J.E., Lin G.G., Scott J.G. Expression of pluripotency factors in echinoderm regeneration // Cell Tissue Res. 2015. Vol. 359, № 2. P. 521–536.
36. Dolmatov I.Yu., Kalacheva N. V., Tkacheva E.S., Shulga A.P., Zavalnaya E.G., Shamshurina E. V., et al. Expression of piwi, mmp, timp, and sox during gut regeneration in holothurian *Eupentacta fraudatrix* (Holothuroidea, dendrochirotida) // Genes. 2021. Vol. 12, № 8.
37. Dinsmore C.E. The foundations of contemporary regeneration research: historical perspectives // Monogr. Dev. Biol. 1992. Vol. 23. P. 1–27.
38. Morgan T.H. Regeneration. New-York: Macmillan, 1901. 316 p.
39. Воронцова М.А., Лиознер Л.Д. Бесполое размножение и регенерация. Москва: Советская наука, 1957. 413 с.
40. Карлсон Б.М. Регенерация. Москва: Наука, 1986. 296 с.

41. Saló E., Abril J.F., Adell T., Cebrià F., Eckelt K., Fernández-Taboada E., et al. Planarian regeneration: Achievements and future directions after 20 years of research // Int. J. Dev. Biol. 2009. Vol. 53, № 8–10. P. 1317–1327.
42. Светлов П.Т. Процессы морфогенеза на клеточном и организменном уровнях // Физиология (механика) развития. Ленинград: Наука, 1978. 279 с.
43. Лиознер Л.Д. Регенерация и развитие. Москва: Наука, 1982. 167 с.
44. Dolmatov I.Yu., Zhirmunsky A. V. Molecular Aspects of Regeneration Mechanisms in Holothurians // Genes. 2021. Vol. 12, № 2. P. 250.
45. Hyman L.H. The Coelomate Bilateria // The Invertebrates: Echinodermata. New York: McGraw-Hill Book Company, 1955. P. 763.
46. Miller A.K., Kerr A.M., Paulay G., Reich M., Wilson N.G., Carvajal J.I., et al. Molecular phylogeny of extant Holothuroidea (Echinodermata) // Mol. Phylogenet. Evol. 2017. Vol. 111. P. 110–131.
47. Долматов И.Ю. Регенерация пищеварительной системы у голотурий // Журнал общей биологии. 2009. Т. 70, № 4. С. 316–327.
48. Goldberg A.D., Allis C.D., Bernstein E. Epigenetics: a landscape takes shape // Cell. 2007. Vol. 128, № 4. P. 635–638.
49. Treutlein B., Lee Q.Y., Camp J.G., Mall M., Koh W., Shariati S.A.M., et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq // Nature. 2016. Vol. 534, № 7607. P. 391.
50. Reid A., Tursun B. Transdifferentiation: do transition states lie on the path of development? // Curr. Opin. Syst. Biol. 2018. Vol. 11. P. 18.
51. Riddle M.R., Spickard E.A., Jevince A., Nguyen K.C.Q., Hall D.H., Joshi P.M., et al. Transorganogenesis and transdifferentiation in *C. elegans* are dependent on differentiated cell identity // Dev. Biol. 2016. Vol. 420, № 1. P. 136–147.
52. Merrell A.J., Stanger B.Z. Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style // Nat. Rev. Mol. Cell Biol. 2016. Vol. 17, № 7. P. 413–425.

53. Eguizabal C., Carlos J., Belmonte I. Reprogramming: Future Directions in Regenerative Medicine // Semin. Reprod. Med. 2013. Vol. 31. P. 82–94.
54. Denholtz M., Plath K. Pluripotency in 3D: Genome organization in pluripotent cells // Curr. Opin. Cell Biol. 2012. Vol. 24, № 6. P. 793.
55. Stadhouders R., Filion G.J., Graf T. Transcription factors and 3D genome conformation in cell-fate decisions // Nature. 2019. Vol. 569, № 7756. P. 345–354.
56. Takahashi K., Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors // Cell. 2006. Vol. 126, № 4. P. 663–676.
57. Liu Z., Wang L., Welch J.D., Ma H., Zhou Y., Vaseghi H.R., et al. Single Cell Transcriptomics Reconstructs Fate Conversion from Fibroblast to Cardiomyocyte // Nature. 2017. Vol. 551, № 7678. P. 100.
58. Kagias K., Ahier A., Fischer N., Jarriault S. Members of the NODE (Nanog and Oct4-associated deacetylase) complex and SOX-2 promote the initiation of a natural cellular reprogramming event in vivo // Proc. Natl. Acad. Sci. U. S. A. 2012. Vol. 109, № 17. P. 6596–6601.
59. Di Tullio A., Vu Manh T.P., Schubert A., Månsson R., Graf T. CCAAT/enhancer binding protein α (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation // Proc. Natl. Acad. Sci. U. S. A. 2011. Vol. 108, № 41. P. 17016–17021.
60. Patel T., Tursun B., Rahe D.P., Hobert O. Removal of Polycomb Repressive Complex 2 makes *C. elegans* germ cells susceptible to direct conversion into specific somatic cell types // Cell Rep. 2012. Vol. 2, № 5. P. 1178.
61. Qin H., Zhao A., Fu X. Small molecules for reprogramming and transdifferentiation // Cell. Mol. Life Sci. 2017. Vol. 74, № 19. P. 3553–3575.
62. Wilkie I.C. Autotomy as a prelude to regeneration in echinoderms // Microsc. Res. Tech. 2001. Vol. 55, № 6. P. 369–396.

63. Kalacheva N. V., Kamenev Y.O., Dolmatov I.Yu. Regeneration of the digestive system in the crinoid *Lamprometra palmata* (Mariametridae, Comatulida) // Cell Tissue Res. 2021.
64. Mozzi D., Dolmatov I.Yu., Bonasoro F., Carnevali M.D.C. Visceral regeneration in the crinoid *Antedon mediterranea*: basic mechanisms, tissues and cells involved in gut regrowth // Cent. Eur. J. Biol. 2006. Vol. 1, № 4. P. 609–635.
65. Frias-De-Diego A., Jara M., Pecoraro B.M., Crisci E. Whole Genome or Single Genes? A Phylodynamic and Bibliometric Analysis of PRRSV // Front. Vet. Sci. 2021. Vol. 8. P. 658512.
66. Zhao X., Li J., Lian B., Gu H., Li Y., Qi Y. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA // Nat. Commun. 2018. Vol. 9, № 1. P. 5056.
67. Bhattarai U.R., Li F., Katuwal Bhattarai M., Masoudi A., Wang D. Phototransduction and circadian entrainment are the key pathways in the signaling mechanism for the baculovirus induced tree-top disease in the lepidopteran larvae // Sci. Rep. 2018. Vol. 8, № 1. P. 17528.
68. Sõber S., Reiman M., Kikas T., Rull K., Inno R., Vaas P., et al. Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes // Sci. Rep. 2015. Vol. 5. P. 13336.
69. Knapp D., Schulz H., Rascon C.A., Volkmer M., Scholz J., Nacu E., et al. Comparative Transcriptional Profiling of the Axolotl Limb Identifies a Tripartite Regeneration-Specific Gene Program // PLoS One. 2013. Vol. 8, № 5.
70. Wetterstrand K.A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [Электронный ресурс]. 2021. Режим доступа: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
71. Pevzner P.A. l-Tuple DNA Sequencing: Computer Analysis // J. Biomol. Struct. Dyn. 1989. Vol. 7, № 1. P. 63–73.

72. Jo J., Oh J., Lee H.-G.G., Hong H.-H.H., Lee S.-G.G., Cheon S., et al. Draft genome of the sea cucumber *Apostichopus japonicus* and genetic polymorphism among color variants // Gigascience. 2017. Vol. 6, № 1. P. 1–6.
73. Zhou Z.C., Dong Y., Sun H.J., Yang A.F., Chen Z., Gao S., et al. Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers // Mol. Ecol. Resour. 2014. Vol. 14, № 1. P. 127–138.
74. Eldem V., Zararsiz G., Taşçi T., Duru I.P., Bakir Y., Erkan M. Transcriptome Analysis for Non-Model Organism: Current Status and Best-Practices // Appl. RNA-Seq Omi. Strateg. - From Microorg. to Hum. Heal. 2017.
75. O'Neil S.T., Dzurisin J.D.K., Carmichael R.D., Lobo N.F., Emrich S.J., Hellmann J.J. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon* // BMC Genomics. 2010. Vol. 11, № 1. P. 1–15.
76. Vella F. Molecular biology of the cell (third edition): by B Alberts, D Bray, J Lewis, M Raff, K Roberts and J D Watson. pp 1361. Garland Publishing, New York and London. 1994 // Biochem. Educ. 1994. Vol. 22, № 3. P. 164.
77. Zhulidov P.A., Bogdanova E.A., Shcheglov A.S., Vagner L.L., Khaspekov G.L., Kozhemyako V.B., et al. Simple cDNA normalization using kamchatka crab duplex-specific nuclease // Nucleic Acids Res. 2004. Vol. 32, № 3. P. e37–e37.
78. Piché C., Schernthaner J.P. Optimization of in Vitro Transcription and Full-Length cDNA Synthesis Using the T4 Bacteriophage Gene 32 Protein // J. Biomol. Tech. 2005. Vol. 16, № 3. P. 239.
79. Durbin R., Eddy S.R., Krogh A., Mitchison G. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids // Cambridge: Cambridge University Press, 1998. 356 p.
80. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., et al. BLAST+: architecture and applications // BMC Bioinformatics. 2009. Vol. 10, № 1. P. 421.

81. Blum M., Chang H.Y., Chuguransky S., Grego T., Kandasaamy S., Mitchell A., et al. The InterPro protein families and domains database: 20 years on // Nucleic Acids Res. 2021. Vol. 49, № D1. P. D344–D354.
82. Pearson W.R. Selecting the Right Similarity-Scoring Matrix // Curr. Protoc. Bioinformatics. 2013. Vol. 43, № SUPL.43. P. 3.5.1.
83. Boyko A.V., Girich A.S., Tkacheva E.S., Dolmatov I.Yu. The *Eupentacta fraudatrix* transcriptome provides insights into regulation of cell transdifferentiation // Sci. Rep. 2020. Vol. 10, № 1.
84. Seret M.-L., Baret P. V. IONS: Identification of Orthologs by Neighborhood and Similarity-an Automated Method to Identify Orthologs in Chromosomal Regions of Common Evolutionary Ancestry and its Application to Hemiascomycetous Yeasts. // Evol. Bioinform. Online. 2011. Vol. 7, № 7. P. 123–133.
85. Nichio B.T.L., Marchaukoski J.N., Raittz R.T. New tools in orthology analysis: A brief review of promising perspectives // Front. Genet. 2017. Vol. 8. P. 165.
86. Mashanov V.S., Zueva O.R., García-Arrarás J.E. Transcriptomic changes during regeneration of the central nervous system in an echinoderm // BMC Genomics. 2014. Vol. 15, № 1. P. 1–21.
87. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 // Genome Biol. 2014. Vol. 15, № 12. P. 550.
88. Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles // Proc. Natl. Acad. Sci. U. S. A. 2005. Vol. 102.
89. Dolmatov I.Yu., Afanasyev S.V., Boyko A.V. Molecular mechanisms of fission in echinoderms: transcriptome analysis // PLoS One. 2018. Vol. 13, № 4.
90. Boyko A. V., Girich A.S., Eliseikina M.G., Maslennikov S.I., Dolmatov I.Yu. Reference assembly and gene expression analysis of *Apostichopus japonicus* larval development // Sci. Rep. 2019. Vol. 9, № 1. P. 1–11.

91. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics // Proc. Natl. Acad. Sci. U. S. A. 1998. Vol. 95, № 4. P. 1460–1465.
92. Holland P.M., Abramson R.D., Watson R., Gelfand D.H. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase. // Proc. Natl. Acad. Sci. U. S. A. 1991. Vol. 88, № 16. P. 7276.
93. Zhulidov P.A., Bogdanova E.A., Shcheglov A.S., Shagina I.A., Wagner L.L., Khazpekov G.L., et al. A method for the preparation of normalized cDNA libraries enriched with full-length sequences // Russ. J. Bioorganic Chem. 2005. Vol. 31, № 2. P. 170–177.
94. Bolger A.M., Lohse M., Usadel B. Trimmomatic : a flexible trimmer for Illumina sequence data // Bioinformatics. 2014. Vol. 30, № 15. P. 2114–2120.
95. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing // J. Comput. Biol. 2012. Vol. 19, № 5. P. 455–477.
96. Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity // Nat. Protoc. 2013. Vol. 8, № 8.
97. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge // Nucleic Acids Res. 2019. Vol. 47, № D1. P. D506–D515.
98. Kudtarkar P., Cameron A. Echinobase : an expanding resource for echinoderm genomic information // Database. 2017. Vol. 2017. P. 1–9.
99. Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E. V., Zdobnov E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs // Bioinformatics. 2015. Vol. 31, № 19. P. 3210–3212.
100. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2 // Nat. Methods. 2012. Vol. 9, № 4. P. 357–359.

101. Li B., Dewey C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome // BMC Bioinformatics. 2011. Vol. 12, № 1. P. 323.
102. Zerbino D.R., Achuthan P., Akanni W., Amode M.R., Barrell D., Bhai J., et al. Ensembl 2018 // Nucleic Acids Res. 2017. Vol. 46, № D1. P. D754–D761.
103. Tweedie S., Braschi B., Gray K., Jones T.E.M., Seal R.L., Yates B., et al. Genenames.org: the HGNC and VGNC resources in 2021 // Nucleic Acids Res. 2021. Vol. 49, № D1. P. D939–D946.
104. Merico D., Isserlin R., Stueker O., Emili A., Bader G.D. Enrichment Map: a network-based method for gene-set enrichment visualization and interpretation // PLoS One. 2010. Vol. 5, № 11. P. e13984.
105. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks // Genome Res. 2003. Vol. 13.
106. Rao X., Huang X., Zhou Z., Lin X. An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis // Biostat. Bioinforma. Biomath. 2013. Vol. 3, № 3. P. 71.
107. Fu L., Niu B., Zhu Z., Wu S., Li W. CD-HIT : accelerated for clustering the next-generation sequencing data // Bioinformatics. 2012. Vol. 28, № 23. P. 3150–3152.
108. Huang X., Madan A. CAP3: A DNA Sequence Assembly Program // Genome Res. 1999. Vol. 9, № 9. P. 868–877.
109. Zhang X., Sun L., Yuan J., Sun Y., Gao Y., Zhang L., et al. The sea cucumber genome provides insights into morphological evolution and visceral regeneration // PLOS Biol. 2017. Vol. 15, № 10. P. 1–31.
110. Smith T.F., Waterman M.S. Identification of common molecular subsequences // J. Mol. Biol. 1981. Vol. 147, № 1. P. 195–197.
111. Rizzo F., Squarzoni P., Archimandritis A., Arnone M.I. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*) // Dev. Biol. 2006. Vol. 300. P. 35–48.

112. Hsieh P.H., Oyang Y.J., Chen C.Y. Effect of de novo transcriptome assembly on transcript quantification // Sci. Reports. 2019. Vol. 9, № 1. P. 1–12.
113. Zhou X., Cui J., Liu S., Kong D., Sun H., Gu C., et al. Comparative transcriptome analysis of papilla and skin in the sea cucumber, *Apostichopus japonicus* // PeerJ. 2016. Vol. 4. P. e1779.
114. Du H., Bao Z., Hou R., Wang S., Su H., Yan J., et al. Transcriptome Sequencing and Characterization for the Sea Cucumber *Apostichopus japonicus* (Selenka, 1867) // PLoS One. 2012. Vol. 7, № 3. P. e33311.
115. Luttrell S.M., Gotting K., Ross E., Alvarado A.S., Swalla B.J. Head regeneration in hemichordates is not a strict recapitulation of development // Dev. Dyn. 2016. Vol. 245, № 12. P. 1159–1175.
116. Zhang P., Li C., Zhu L., Su X., Li Y., Jin C., et al. De Novo Assembly of the Sea Cucumber *Apostichopus japonicus* Hemocytes Transcriptome to Identify miRNA Targets Associated with Skin Ulceration Syndrome // PLoS One. 2013. Vol. 8, № 9. P. e73506.
117. Pertea G., Huang X., Liang F., Antonescu V., Sultana R., Karamycheva S., et al. TIGR Gene Indices clustering tools (TGICL): a softwaresystem for fast clustering of large EST datasets // Bioinformatics. 2003. Vol. 19, № 5. P. 651–652.
118. Lu J., Peatman E., Tang H., Lewis J., Liu Z. Profiling of gene duplication patterns of sequenced teleost genomes: Evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications // BMC Genomics. 2012. Vol. 13, № 1. P. 1–10.
119. Howe K., Clark M.D., Torroja C.F., Torrance J., Berthelot C., Muffato M., et al. The zebrafish reference genome sequence and its relationship to the human genome // Nature. 2013. Vol. 496, № 7446. P. 498–503.
120. Ward N., Moreno-Hagelsieb G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? // PLoS One. 2014. Vol. 9, № 7. P. 1–6.

121. Gildor T., Cary G.A., Lalzar M., Hinman V.F., Ben-Tabou de-Leon S. Developmental transcriptomes of the sea star, *Patiria miniata*, illuminate how gene expression changes with evolutionary distance // Sci. Reports. 2019. Vol. 9, № 1. P. 1–12.
122. Ordoñez J.F.F., Galindez G.G.S.T., Gulay K.T., Ravago-Gotanco R. Transcriptome analysis of growth variation in early juvenile stage sandfish *Holothuria scabra* // Comp. Biochem. Physiol. Part D Genomics Proteomics. 2021. Vol. 40. P. 100904.
123. Mu C., Wang R., Li T., Li Y., Tian M., Jiao W., et al. Long Non-Coding RNAs (lncRNAs) of Sea Cucumber: Large-Scale Prediction, Expression Profiling, Non-Coding Network Construction, and lncRNA-microRNA-Gene Interaction Analysis of lncRNAs in *Apostichopus japonicus* and *Holothuria glaberrima* During LPS Challeng // Mar. Biotechnol. 2016. Vol. 18, № 4. P. 485–499.
124. Mashanov V.S., Zueva O.R., García-Arrarás J.E. Posttraumatic regeneration involves differential expression of long terminal repeat (LTR) retrotransposons // Dev. Dyn. 2012. Vol. 241, № 10. P. 1625–1636.
125. Nieves-Ríos C., Alvarez-Falcón S., Malavez S., Rodriguez-Otero J., García-Arrarás J.E. The nervous system component of the mesentery of the sea cucumber *Holothuria glaberrima* in normal and regenerating animals // Cell Tissue Res. 2020. Vol. 380, № 1. P. 67–77.
126. Moorman A.F.M., Christoffels V.M. Cardiac chamber formation: development, genes, and evolution // Physiol. Rev. 2003. Vol. 83, № 4. P. 1223–1267.
127. Schöck F., Perrimon N. Molecular mechanisms of epithelial morphogenesis // Annu. Rev. Cell Dev. Biol. 2002. Vol. 18, № 1. P. 463–493.
128. Schöck F., Perrimon N. Cellular processes associated with germ band retraction in Drosophila // Dev. Biol. 2002. Vol. 248, № 1. P. 29–39.
129. Radice G.P. The spreading of epithelial cells during wound closure in *Xenopus larvae* // Dev. Biol. 1980. Vol. 76, № 1. P. 26–46.
130. Kalinkova L., Zmetakova I., Smolkova B., Minarik G., Sedlackova T., Horvathova Kajabova V., et al. Decreased methylation in the SNAI2 and ADAM23 genes

associated with de-differentiation and haematogenous dissemination in breast cancers // BMC Cancer. 2018. Vol. 18, № 1. P. 1–12.

131. Zhou Y., Liu Q., Dai X., Yan Y., Yang Y., Li H., et al. Transdifferentiation of type II alveolar epithelial cells induces reactivation of dormant tumor cells by enhancing TGF-β1/SNAI2 signaling // Oncol. Rep. 2018. Vol. 39, № 4. P. 1874–1882.
132. Chen Y., Wang K., Qian C.N., Leach R. DNA methylation is associated with transcription of Snail and Slug genes // Biochem. Biophys. Res. Commun. 2013. Vol. 430, № 3. P. 1083–1090.
133. Nigmatullina L., Norkin M., Dzama M.M., Messner B., Sayols S., Soshnikova N. Id2 controls specification of Lgr5 + intestinal stem cell progenitors during gut development // EMBO J. 2017. Vol. 36, № 7. P. 869–885.
134. Veerasamy M., Phanish M., Dockrell M.E.C. Smad Mediated Regulation of Inhibitor of DNA Binding 2 and Its Role in Phenotypic Maintenance of Human Renal Proximal Tubule Epithelial Cells // PLoS One. 2013. Vol. 8, № 1. P. 1–8.
135. Kamata Y., Sumida T., Kobayashi Y., Ishikawa A., Kumamaru W., Mori Y. Introduction of ID2 enhances invasiveness in ID2-null oral squamous cell carcinoma cells via the SNAIL axis // Cancer Genomics and Proteomics. 2016. Vol. 13, № 6. P. 493–498.
136. Chang C., Yang X., Pursell B., Mercurio A.M. Id2 complexes with the SNAG domain of Snai1 inhibiting Snai1-mediated repression of Integrin 4 // Mol. Cell. Biol. 2013. Vol. 33, № 19. P. 3795–3804.
137. Smiley S. Holothuroidea // Microscopic anatomy of invertebrates. Volume 14: Echinodermata / New York: Wiley-Liss, 1994. P. 401–471.
138. Mashanov V.S., Zueva O.R., Rojas-Catagena C., Garcia-Arraras J.E. Visceral regeneration in a sea cucumber involves extensive expression of survivin and mortalin homologs in the mesothelium // BMC Dev. Biol. 2010. Vol. 10, № 1. P. 117.

139. Lai S., Yuan J., Zhao D., Shen N., Chen W., Ding Y., et al. Regulation of mice liver regeneration by early growth response-1 through the GGPPS/RAS/MAPK pathway // Int. J. Biochem. Cell Biol. 2015. Vol. 64. P. 147–154.
140. Yan L., Wang Y., Liang J., Liu Z., Sun X., Cai K. MiR-301b promotes the proliferation, mobility, and epithelial-to-mesenchymal transition of bladder cancer cells by targeting EGR1 // Biochem. Cell Biol. 2017. Vol. 95, № 5. P. 571–577.
141. Gehrke A.R., Neverett E., Luo Y.-J., Brandt A., Ricci L., Hulett R.E., et al. Acoel genome reveals the regulatory landscape of whole-body regeneration // Science. 2019. Vol. 363, № 6432. P. eaau6173.
142. Cary G.A., Wolff A., Zueva O., Pattinato J., Hinman V.F. Analysis of sea star larval regeneration reveals conserved processes of whole-body regeneration across the metazoa // BMC Biol. 2019. Vol. 17, № 1. P. 16.
143. Hsu T., Trojanowska M., Watson D.K. Ets proteins in biological control and cancer // J. Cell. Biochem. 2004. Vol. 91, № 5. P. 896–903.
144. Oikawa T., Yamada T. Molecular biology of the Ets family of transcription factors // Gene. 2003. Vol. 303, № 1–2. P. 11–34.
145. Sashida G., Bazzoli E., Menendez S., Liu Y., Nimer S.D. The oncogenic role of the ETS transcription factors MEF and ERG // Cell Cycle. 2010. Vol. 9, № 17. P. 3457–3459.
146. Guan F.H.X., Bailey C.G., Metierre C., O'Young P., Gao D., Khoo T.L., et al. The antiproliferative ELF2 isoform, ELF2B, induces apoptosis in vitro and perturbs early lymphocytic development in vivo // J. Hematol. Oncol. 2017. Vol. 10, № 1. P. 1–17.
147. Suico M.A., Shuto T., Kai H. Roles and regulations of the ETS transcription factor ELF4/MEF // J. Mol. Cell Biol. 2017. Vol. 9, № 3. P. 168–177.
148. Xie Y., Koch M.L., Zhang X., Hamblen M.J., Godinho F.J., Fujiwara Y., et al. Reduced Erg dosage impairs survival of hematopoietic stem and progenitor cells // Stem Cells. 2017. Vol. 35, № 7. P. 1773–1785.

149. Lentjes M.H., Niessen H.E., Akiyama Y., de Bruïne A.P., Melotte V., van Engeland M. The emerging role of GATA transcription factors in development and disease // Expert Rev. Mol. Med. 2016. Vol. 18. P. 1–15.
150. Materna S.C., Ransick A., Li E., Davidson E.H. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos // Dev. Biol. 2013. Vol. 375, № 1. P. 92–104.
151. Riddle M.R., Weintraub A., Nguyen K.C.Q., Hall D.H., Rothman J.H. Transdifferentiation and remodeling of post-embryonic *C. elegans* cells by a single transcription factor. // Development. 2013. Vol. 140, № 24. P. 4844–4849.
152. Ilsley M.D., Gillinder K.R., Magor G.W., Huang S., Bailey T.L., Crossley M., et al. Krüppel-like factors compete for promoters and enhancers to fine-tune transcription // Nucleic Acids Res. 2017. Vol. 45, № 11. P. 6572–6588.
153. Yamane M., Ohtsuka S., Matsuura K., Nakamura A., Niwa H. Overlapping functions of Krüppel-like factor family members: targeting multiple transcription factors to maintain the naïve pluripotency of mouse embryonic stem cells // Development. 2018. Vol. 145, № 10. P. dev162404.
154. Lee J.Y., Park M.K., Park J.H., Lee H.J., Shin D.H., Kang Y., et al. Loss of the polycomb protein Mel-18 enhances the epithelial-mesenchymal transition by ZEB1 and ZEB2 expression through the downregulation of miR-205 in breast cancer // Oncogene. 2014. Vol. 33, № 10. P. 1325–1335.
155. Wang X.-F., Zhang X.-W., Hua R.-X., Du Y.-Q., Huang M.-Z., Liu Y., et al. Mel-18 negatively regulates stem cell-like properties through downregulation of miR-21 in gastric cancer // Oncotarget. 2016. Vol. 7, № 39.
156. Hohenauer T., Moore A.W. The Prdm family: expanding roles in stem cells and development // Development. 2012. Vol. 139, № 13. P. 2267–2282.
157. Chu L.F., Surani M.A., Jaenisch R., Zwaka T.P. Blimp1 expression predicts embryonic stem cell development in vitro // Curr. Biol. 2011. Vol. 21, № 20. P. 1759–1765.

158. Yamaji M., Ueda J., Hayashi K., Ohta H., Yabuta Y., Kurimoto K., et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells // Cell Stem Cell. 2013. Vol. 12, № 3. P. 368–382.

159. Grabole N., Tischler J., Hackett J.A., Kim S., Tang F., Leitch H.G., et al. Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation // EMBO Rep. 2013. Vol. 14, № 7. P. 629–637.

160. Vervoort M., Meulemeester D., Béhague J., Kerner P. Evolution of Prdm genes in animals: Insights from comparative genomics // Mol. Biol. Evol. 2016. Vol. 33, № 3. P. 679–696.

161. Skinner M.K., Rawls A., Wilson-Rawls J., Roalson E.H. Basic Helix-Loop-Helix Transcription Factor Gene Family Phylogenetics and Nomenclature // Differentiation. 2010. Vol. 80, № 1. P. 1.

162. Hishikawa K., Marumo T., Miura S., Nakanishi A., Matsuzaki Y., Shibata K., et al. Musculin/MyoR is expressed in kidney side population cells and can regulate their function // J. Cell Biol. 2005. Vol. 169, № 6. P. 921–928.

163. Acharya A., Baek S.T., Huang G., Eskiocak B., Goetsch S., Sung C.Y., et al. The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors // Development. 2012. Vol. 139, № 12. P. 2139–2149.

164. De Benedittis P., Jiao K. Alternative splicing of T-box transcription factor genes. // Biochem. Biophys. Res. Commun. 2011. Vol. 412, № 4. P. 513–517.

165. Takashima Y., Suzuki A. Regulation of organogenesis and stem cell properties by T-box transcription factors // Cell. Mol. Life Sci. 2013. Vol. 70, № 20. P. 3929–3945.

ПРИЛОЖЕНИЯ

1. Праймеры для кПЦР

Ген	Тип	Последовательность	Начало	Длина	Tm	GC%	Ампликон
GHCL01011038.1 (FOXC1)	Forward	GGCCAACACAGTCGCTAATAAC	1527	22	63.699	50	
GHCL01011038.1 (FOXC1)	Reverse	GTCATGTCGTATGGCGTGTAATG	1640	23	63.627	47.826	
GHCL01011038.1 (FOXC1)	Product						114
GHCL01017253.1 (HES1)	Forward	CCAGCAGCATACGATAACCATAC	651	23	63.114	47.826	
GHCL01017253.1 (HES1)	Reverse	GGATGACGAGTGCGATCTCT	811	20	63.527	55	
GHCL01017253.1 (HES1)	Product						161
GHCL01042978.1 (MAX)	Forward	GACAAGAGAGCCCACCACAAT	4	21	64.616	52.381	
GHCL01042978.1 (MAX)	Reverse	TTCTGTCTTTTGAGGTCGTCTATGT	218	25	64.117	40	
GHCL01042978.1 (MAX)	Product						215
GHCL01041911.1 (SOX9)	Forward	ACGGATACGACTGGTCAACGA	272	21	65.299	52.381	
GHCL01041911.1 (SOX9)	Reverse	ATTCGGCGTTATGGATGTTCG	430	21	63.258	47.619	
GHCL01041911.1 (SOX9)	Product						159
GHCL01042546.1 (TCF24)	Forward	ACAAACGAACAGCGGCATCAAC	189	22	66.422	50	
GHCL01042546.1 (TCF24)	Reverse	GTCACCATCGTCCAATGTCTTC	414	22	63.194	50	
GHCL01042546.1 (TCF24)	Product						226

2. Праймеры и зонды для кцПЦР

Ген	Тип	Последовательность	Начало	Длина	Tm	GC%	Ампликон
tubb (GHCL01014613.1)	Forward	GATCCCGAACAACGTTAAG	1089	19	60.951	47.368	
tubb (GHCL01014613.1)	Probe	CCTCGACCTTCATCGGCAACA	1148	21	67.041	57.143	
tubb (GHCL01014613.1)	Reverse	CTTTCGCCTGAACATAGC	1233	18	60.989	50	
tubb (GHCL01014613.1)	Product						145
EF1a (GHCL01012521.1)	Forward	GGTAATCATCCTGAACCATC	1191	20	60.446	45	
EF1a (GHCL01012521.1)	Probe	CACATCGCCTGCAAGTTCGC	1261	20	66.913	60	
EF1a (GHCL01012521.1)	Reverse	CTTCACCATCTTGGGATTC	1347	19	60.555	47.368	
EF1a (GHCL01012521.1)	Product						157
TBX20 (GHCL01014615.1)	Forward	TGACATCTAGCCATACCAG	917	19	61.03	47.368	
TBX20 (GHCL01014615.1)	Probe	TCACTGCCTACCAGAACCAGC	971	21	66.642	57.143	
TBX20 (GHCL01014615.1)	Reverse	TGCCACTTTCTCTCTCTC	1083	18	61.05	50	
TBX20 (GHCL01014615.1)	Product						167
EGR (GHCL01011410.1)	Forward	CCATCTGACAACGCATATC	1269	19	60.86	47.368	
EGR (GHCL01011410.1)	Probe	TCTCGCACTGGAACGGCTTC	1324	20	67.167	60	
EGR (GHCL01011410.1)	Reverse	TCTCGCTTAACCTTCTGG	1403	18	61.393	50	
EGR (GHCL01011410.1)	Product						135
ELF (GHCL01018819.1)	Forward	CTTTCCTTCACACCACAG	577	18	60.522	50	
ELF (GHCL01018819.1)	Probe	TGCAGTGGTCAGCTTTCCCT	670	20	66.894	55	
ELF (GHCL01018819.1)	Reverse	GGTCAAAGGTCACATCATAG	733	20	60.595	45	
ELF (GHCL01018819.1)	Product						157
GATA3 (GHCL01013574.1)	Forward	GAGTACCCGATGTCGAAC	1255	18	61.783	55.556	
GATA3 (GHCL01013574.1)	Probe	CGTCGATCTGCAACACCCTGT	1328	21	67.204	57.143	
GATA3 (GHCL01013574.1)	Reverse	ACTAAGATTGATCCCTGTGG	1413	20	61.685	45	
GATA3 (GHCL01013574.1)	Product						159
ID (GHCL01037074.1)	Forward	GGTCACAAATCCAGTGAAG	63	19	60.981	47.368	
ID (GHCL01037074.1)	Probe	TACCACAAGCGGCAGAGCAT	94	20	67.124	55	
ID (GHCL01037074.1)	Reverse	CTTGTAGCAATCGCTCATC	201	19	61.09	47.368	
ID (GHCL01037074.1)	Product						139
KLF1/2/4 (GHCL01020067.1)	Forward	CCTACCTGCTGACTACTG	369	18	60.934	55.556	
KLF1/2/4 (GHCL01020067.1)	Probe	AAGAGGAACCACCATCGCCA	437	20	66.656	55	
KLF1/2/4 (GHCL01020067.1)	Reverse	CTCTCGCATCGTGAAATG	525	18	60.942	50	
KLF1/2/4 (GHCL01020067.1)	Product						157
PCGF2 (GHCL01014412.1)	Forward	CTGCGAAAGTCTGTGTTAAG	572	20	61.479	45	
PCGF2 (GHCL01014412.1)	Probe	ACGGGTTCTCGTCATCATCTGC	679	22	66.838	54.545	
PCGF2 (GHCL01014412.1)	Reverse	ATGGACCAATCCTTCTCC	742	18	61.336	50	
PCGF2 (GHCL01014412.1)	Product						171
PRDM9 (GHCL01005242.1)	Forward	GCAGCAGTGAGCTTAATAAC	1256	20	61.689	45	
PRDM9 (GHCL01005242.1)	Probe	ACTGACAAGAGCCTTCTTCTGGC	1315	23	66.955	52.174	
PRDM9 (GHCL01005242.1)	Reverse	GTTCACTTGTGGCAGAATC	1423	19	61.555	47.368	
PRDM9 (GHCL01005242.1)	Product						168
SNAI2 (GHCL01015842.1)	Forward	TCCTCCCGTTACCTCTAC	626	18	62.08	55.556	
SNAI2 (GHCL01015842.1)	Probe	TTCAGTCGTGCCTGCCTTCA	710	20	66.964	55	
SNAI2 (GHCL01015842.1)	Reverse	CTCTTCTCCTGATGGGTTTC	800	20	62.248	50	
SNAI2 (GHCL01015842.1)	Product						175
MSC (GHCL01033790.1)	Forward	CAAAGTTCATGTCGCTACC	10	19	61.258	47.368	
MSC (GHCL01033790.1)	Probe	ACGGTTGTTCCTACCCGAGC	63	20	66.758	60	
MSC (GHCL01033790.1)	Reverse	CAGTGTCGTAGGTCATCTC	160	19	61.493	52.632	
TCF21 (GHCL01033790.1)	Product						151
TCF24 (GHCL01042546.1)	Forward	CCCAGATACGAAACTGTCC	327	19	62.18	52.632	

3. Последовательности ампликонов, используемых в кцПЦР

>1 — соответствует GHCL01011410.1, идентичность 96.9%, ген Ef-egr, длина сиквенса - 108, длина расчетная - 135

GAGGCAGCCGTTCAGTGCGAGACGTGCGGGCGCAAATTCGCGCGCAGCGAC
GAAAGGAAGCGCCACACAAAGATCCACCAGCGCCAGTAAGGTTAAGCGAGG
AAAAAA

>2 — соответствует GHCL01018819.1, идентичность 95.1%, ген Ef-elf, длина сиквенса - 125, длина расчетная - 157

TAGTCGAATCAGATATGAGTTGTTATACCAAATTCCACAACAACAGAAGAGT
GCCGATTGCGAGTGGTCAGCTTTCCCTAGGTGAACACACATTCCCAGGATAC
CTATGATGTGACCTTTGACCA

>3 — соответствует GHCL01013574.1, идентичность 97.6%, ген Ef-gata3, длина сиквенса - 125, длина расчетная - 159

CGTCATTCGTACACTCGCCCTCGATGGGCTACCCCCAGTCGTCGATCTGCAAC
ACCCTGTACCATGGGCCGAGCCCCCCGCAGCACCACATGTCAATGCAGGGTC
CCACAGGGATCAATCTTAGT

>4 — соответствует GHCL01037074.1, идентичность 96.0%, ген Ef-id, длина сиквенса - 108, длина расчетная - 139

CCAGAGGACAGAGCATTGCAGCGATGCTACACCGAGCACAGAGTCTCTCGTT
CGACAGCCGCTCTTCAGCGCTGGACAATTTCTCGATGAGGCGATTGCTACAA
GAAA

>5 — соответствует GHCL01014412.1, идентичность 97.8%, ген Ef-pcgf2, длина сиквенса - 138, длина расчетная - 171

CCTCATGCACAAGTTTGATTGAAGCAAAGTATCAGATTGAACTAACTCATGCA
GATGATGACGAGAACCCGTTATCAGATGACTATACACTCATGGATGTGGCATA
CATATATGAATGGAGAAGGATTGGTCCATAAA

>6 — соответствует GHCL01005242.1, идентичность 97.0%, ген Ef-prdm9, длина сиквенса - 136, длина расчетная - 168

AGCTTAGCCACTGGTCAGAGTGTTTCACTGACAAGAGCCTTCTTCTGGCAGA
AAAACAAAGTCCACTTCATGCTCATTCTCCTGATGATATGATTGAATGTACATC
CAGTTTGAAGATTCTGGCCACAAGTGAACC

>7 — соответствует GHCL01020067.1, идентичность 99.2%, ген Ef-klf1/2/4, длина сиквенса - 124, длина расчетная - 157

CTCCGATGCTGGATCGTACCCAAGATGGAGATCAAAGAGGAACCACCATCGC
CACCCAGCTACAACCATGAGCTGTCTATGGCTGCTGATATCAAAGACTTTGGT
CATTTCACGATGCGAGAGA

>8 — соответствует GHCL01015842.1, идентичность 98.6%, ген Ef-snai2, длина сиквенса - 143, длина расчетная - 175

TCCCCAACCTCCCCGCTCGCTACCGGTCTCCGTATGTCATGTGGGCGATCTTC
AGTCGTGCCTGCCTTCACCGAGCTCTGACGGGGAATCGCGCCTACCGTCACC
CGGCTCTGACGGCGGGGAAACCCATCAGGAGAAGAGAA

>9 — соответствует GHCL01014615.1, идентичность 92.4%, ген Ef-tbx20, длина сиквенса - 137, длина расчетная - 167

CGGGGGGGCTTTCACAGTCAGTCGCTGCCTACCAAAACCAACTGATTACCAG
ACGTCAAGATAGATAGTAACCCATTCGCCAAGGGGATTCCGAGACTCGACAC
GGCTGACCGATTTCGAGAGAGAGAAAGTGGCAA

>10 — соответствует GHCL01033790.1, идентичность 94.2%, ген Ef-msc, длина сиквенса - 125, длина расчетная - 151

GTCATGTACCGTATCCACCACGGTGTGTTGCCTACCCGAGCAGTTCGGTGTAC
GAGGACCTGCAAAGTGTCGCCGTTTGTCCGGAGCTGACGATACAATTATCGA
GATGGGCCTACGACACTGAA

>11 — соответствует GHCL01042546.1, идентичность 99.0%, ген Ef-tcf24, длина сиквенса - 109, длина расчетная - 131

CTGGTCTCGCTACACCTATATTTCACATTTAATGAAGACATTGGACGATGGTG
ACGTCATGGACCCTTCCGAATCCAAGCTATCGGAGAAACAGCACCAAAGAAT
ATGA

>1 — соответствует ???, идентичность ???, ген ???, длина сиквенса - 108, длина расчетная - ???

>1 — соответствует ???, идентичность ???, ген ???, длина сиквенса - 108, длина расчетная - ???

4. Последовательности праймеров для клонирования

Ген	Тип	Последовательность	Начало	Длина	Tm	GC%	Ампликон
EGR (GHCL01011410.1)	Forward	GGTCGAGGATATAGTCACATCCAT	204	24	63.348	45.833	
EGR (GHCL01011410.1)	Reverse	GCTGGTGGATCTTTGTGTGG	1384	20	63.396	55	
EGR (GHCL01011410.1)	Product						1181
ELF (GHCL01018819.1)	Forward	CTCCTAGACATCTCACCAACAAC	70	23	62.171	47.826	
ELF (GHCL01018819.1)	Reverse	GGCAAGAATACCTCGCTGATAG	690	22	62.325	50	
ELF (GHCL01018819.1)	Product						621
PRDM9 (GHCL01005242.1)	Forward	CTGGTGTACTATGGCAAGGATTA	961	23	61.825	43.478	
PRDM9 (GHCL01005242.1)	Reverse	AGTGTGGATCCGTTCATGTC	1820	20	61.835	50	
PRDM9 (GHCL01005242.1)	Product						879
KLF1/2/4 (GHCL01020067.1)	Forward	CCTGGACTTCATCCTGAACAAC	342	22	62.705	50	
KLF1/2/4 (GHCL01020067.1)	Reverse	GGTGATCGGAGCGAGAGAA	1096	19	62.976	57.895	
KLF1/2/4 (GHCL01020067.1)	Product						755
SNAI2 (GHCL01015842.1)	Forward	CCACTACCAGTGTACGATCC	370	20	61.094	55	
SNAI2 (GHCL01015842.1)	Reverse	CTATAGCGCTTGACGTGAGA	1178	20	61.218	50	
SNAI2 (GHCL01015842.1)	Product						809
ID (GHCL01037074.1)	Forward	CAGTGAAGTTACAGGAAACATACC	74	24	61.313	41.667	
ID (GHCL01037074.1)	Reverse	CCTGTTTGTCGAATGCTCTG	496	20	61.05	50	
ID (GHCL01037074.1)	Product						423
GATA3 (GHCL01013574.1)	Forward	CAGTCAGGTCTCGGTAACAA	58	20	61.133	50	
GATA3 (GHCL01013574.1)	Reverse	GCATTGCACAGGTAGTGTC	857	19	60.768	52.632	
GATA3 (GHCL01013574.1)	Product						800
PCGF2 (GHCL01014412.1)	Forward	ACTAACAGAATGTCTCCACTCA	108	22	60.984	40.909	
PCGF2 (GHCL01014412.1)	Reverse	CGGTATGTGTTTGATCTTCTGG	804	22	60.986	45.455	
PCGF2 (GHCL01014412.1)	Product						697
TBX20 (GHCL01014615.1)	Forward	GCAGACGAATGTTTCCAGCTCTT	548	23	65.465	47.826	
TBX20 (GHCL01014615.1)	Reverse	CTATCTTCGGACAGCCTTCCA	1372	21	63.552	52.381	
TBX20 (GHCL01014615.1)	Product						825
MSC (GHCL01033790.1)	Forward	GCAGTTCGGTTACGAGGA	81	18	61.09	55.556	
MSC (GHCL01033790.1)	Reverse	CCTGTAGATGACCGATATAGCA	559	22	60.792	45.455	
MSC (GHCL01033790.1)	Product						479
TCF24 (GHCL01042546.1)	Forward	CTGATGTCATGACTACCCGAT	2	21	61.177	47.619	
TCF24 (GHCL01042546.1)	Reverse	TCAACCGGTGCTGTTTCT	463	18	61.565	50	
TCF24 (GHCL01042546.1)	Product						462

5. Последовательности отсеквенированных ампликонов, используемых в WMISH

>1 — соответствует GHCL01011410.1, идентичность 96.7%, ген Ef-egr

GGTCGAGGATATAGTCACWTCCATGTCGATGGSGGTTCCAAATAATCAGGATA
TTTTCTCGAACGTGAACACACTGCAGAGCATAGCCGGTAGCATGGAAGAATT
TGAGTACAAACCTCAGGTGACCGAGGGTGGCTTCGTGGACGTGAACCATAA
CGTGAACCCGTACCCGACAGGCGGTCCGATTGCCATGTCACCTTCGACGACC
CACAGTGCCTCGCCGGTACCCAGCACCAGCCCTGACAGTGGGTGCTCCCCTA
GCTCGGGGTGGTGGTCCGGCCCATCGACCCCTGTCGAGCGACCTATGACCTC
ACTAGCCGATACTATCGGCGGAATCATCACTAGCTCAAGCAGCTTGATTAACG
ACACCGCCTGCAGCCAGTCCATTATTGGCGGTAGCCCAGTGCCTCCTATGGTC
ACTGACGCCTTCACGCCCCCATCGACATGCTACGACGCCATCACTACCGCGG
ACCTGTCTGGCGTCCATGCCATGCCATCCTTCTCATGTTGCTACACGCTACCG
ACCACTGCCTCGATGGAAACGAATGCCTTCGATGGAATGCAGCCAATGAACC
TTCAGCAACCGTCGCTGAGCAACACGAACATTCCCTGCACGCAATACGAGCC
ACAGCCACCAATCAAAGTGGAAGTCATGAGGTATAACTGGGCRACCCGYYW
YMRRACCCCAGCCCCCTGGGGGTTCGGGCGGCCAGAAGAGCTGCTCGGGCA
GCCGTCCCTCTCATGCTCGCAAATGCCGCAGACAACCATCGTCGCCCGTCAC
CAAAGGCATCGGTGACAGCCTGATGATGCAGATCACGACCAGCGCGGACGC
AGTGCGTCAGCCGTAACACGTGGACCCCGAGCGTCGTCGGAAGAACGAACT
ACGACGAAGCCAAGCGGTATAGCCGCACGAGAGCACCYCACCACACGTAGA
GGTC

>2 — соответствует GHCL01018819.1, идентичность 97.5%, ген Ef-elf

ACATTGGCAAGAATACCTGCGATAGRAGKACCTTAGCGCTCTACCCATCGTCT
CGTAKGYYATGTCTGGTTTGTTCTTGTGAAGCCCCCAGAGTTTTGAGACAGC
CTTGGAGTCAACCAGTTTGAAGATTCCTTGCTCTCGGCTGGTCCACTTGATG
AACTTCGGACACGTTTCCTCATTCTGTAGCGGGTCTAGAAGGAACTCCCAGA
GGTATGTTGTGTGACTGTCTTTACTTTTTCTTTCCTTCTTGATGCCTCGAGGTG

ACTGAAGTGGGATATCGGATTTGGACCTTGATTTTTTATCTRGTTTCTTGCTTT
TCTTTGTAGGTGTCCTTGATATAGACTCTGTTGTGAGGCTGACATGATCTTCTG
GGTCAAAGGTCACATCATAGGTACCTGGGAATGTGTGTTCACCTAGGGAAAG
CTGACCACTGCAATCGGACTCTTCTGTTGTTGTMGAATTTGGTATAACAACTT
CATATTCTGATTTGATGGAATCTTCTGATTTTGACTGTGGTGTGAAGGAAAGG
CCTAAGGGATGTGTATCATCGATACAGTCCGTAGTGACAGAATCTGTGGGTGG
GAACAGTTTATCTTCCCCAGATGTTGTTGGTGAGATGTCAGGAG

>3 — соответствует GHCL01013574.1, идентичность 97.7%, ген Ef-gata3

TAGACAGGTCTCGGTAACAACCAAGTCTACCGCCCACACTTCCACMSTCCCA
GCAGCTTCGCCCAGTGGGTGGACACCTCTGCAAGAGCCGTCGGACTACACA
GTATCCAGACATCCCCAGCTGGCMCCACGTGGTATAACATGCAGCAGACACC
CCCCACCAACCCCTCGCACGAGAACAGCGCCAACAGCCCTCCAGCCACTGG
CAACCTACCTAACTACTCCCCCGGTTACGGCACCTTCCGTGGGTCCCCCGTAG
CGACCCCGACAAACGCCCTGCCGCAGCTCCCAGCCCACCCCGCGACGGGAG
GGGTACCCTTGCATCAACATACGCCATCACATCACTACTTCAACTACCCACCA
ACCCCGCCCAAGGATTCTCCGGAAATCGGCGCAACGTCACCTCAGCGATTTG
CGTCACATTGCATGCAGAAGTCACCCGTCCTCGGGAGCGAAGAAAAGGACG
GAAGCCCGAAGGAGTCATCCGACCCCACGAAGACGATGAGTATGTACTCGG
AATCTACGGGACCAGGGTCGTTCCCGCAGACGCGTCCGTCTTTCCCCACGTG
CGGGTCCACGTTCCCGGGGTTCCTGCCAGATCTTGGTGCATCGCTGAACTTC
CCGACCACGGGGGTCCTGACGACCGGTAATAACAGCAACAAGAGCCTCTAC
CCGTCGACCAAGCCGAGGGCCAAGAACAGATCGAGCACTGAGGGCAGAGA
GTGCGTTAACTGCGGTGCCACGTCGACGCCCCTTTGGCGTCGCGACGGGAAC
GGACACTACCTGTGCAATGC

>4 — соответствует GHCL01037074.1, идентичность 96.7%, ген Ef-id

CAGTGAAGTTACAGGAAACATACCACAAGCGACAGAGCATTGCAGCGATGC
TACWGCGAGCACAGAGTCTCTCGTTCGACAGCCGCTCCTCAGCGCTGGACA
ATTTCTCGATGAGCGATTGCTACAAGAAACTGGTCGAGATAGTTCCGACGATT
CCTCGAGATCGGAAGGTCTCGAAGGTGGAGATTTTACAGCACGTTATTGACT

ATATCCAGGACTTACAAACGGCGCTGGATAACCGGGCGATGAGGCACCTCCA
CCACCATCACGAGKCCACGGTATCGACCCCCAACAGGACCCCCTTCAGCACA
TTACAGACCACTACCTCATCGGGACCATCCAGGACACCTTCATCCACATCATC
AGTCACAACAGACGAACAACAGAGGATACCCAACAGAGCATTCGACAAACA
GG

>5 — соответствует GHCL01014412.1, идентичность 79.8%, ген Ef-pcgf2

ACTAACAGAATGTCCCACTCATTTTGTCGAAGCTGTCTCGTTCGGTACTTTCA
CACCAGTGAAAACTGCCCAACCTGCGATACTTTGGTCCATAAGTCGAGACCC
CTCTTGAATGCAAGGCCGGATAAAGCTCTCCAAAGTATCGTTTACAAAGCAG
TTCCCGGACTGTTTAAAGATGAAATGAAAAAGAGGCGCGACTTCTATTCTGA
ACCGCCAGTTTTAGGTGAGTTGGAGGAAGGAGAACAAGCAGATTATCGAGC
TGTTGTAATCCATACTGATGATGAGTGTATAAACCTCTCATTGCAYMTACCATC
TTCCGTTCGATGGTGCGAACTCTGCTGATGGAACGAGCATTGAAGACGAGTC
GAGGCATGTGGCTGATGTGATAGAAGATATGGCTGGCGACGATGAGACCCAT
GTGCGATACCTTCGTTGTCCTGCGGCAGTTCGAGTTAGGCAGATCAGGAAAC
TTATCAAGCACAAGTTCGACTTAAAGCCAGAGTATCAGGTTCAAATAACTCAT
GCAGACAATGACGAGGACCCRYYTATCAGATGACTATACACTCATGGATGTG
GCGTACATACATGAATGGAGAAGGATTGGACCATTACCTCTGGTGTTCAAAGT
CTTTCAAGACACCCGTAAGCGCCAGAAGATCAAACACATACCG

>6 — соответствует GHCL01005242.1, идентичность 89.0%, ген Ef-prdm9

CTGGTGTACTATGGCAAGGACATGCTCGKGATCTTGGGATTGAGAGGGTATTC
TCTGGGGATCCTAAGAAGGTGGAAATGACCCAAGGTGGACCAAGCTTCAGG
TGTCAGGAGTGTGGCCGTTTGTACACCATACACATCTACCTGTTGAAGCATCT
CAAGCATTCTCACGGCCATGCKATATATTTCCCGAGGGAGTCAAACACACCTA
ACAGAGGTATCATGTCAACCCAGGGCATTCCATATATAATTGGATTGAAGAGG
CCCAAAACACTTGCCCATGGTAAGCAGATATGCAGCAGTGAGCTTAATAACA
GTGTACATCCTGTTAAAGCCACTGGTCAGAGTGTTTCAACTGACAAGAGCCT
TCTTCTGGCAGAAGAACAAAGTCCACTTCATGATCATTCTCCTGATGATTTGA
ATGGATGTACATCCAGTGTGAACGATTCTGCCACAAGTGAACTTATGTCTAAT

RCKGRKGAGGWGWGAWACATAGWGGTATCYTCCATCCTGCAGGGAAGTGA
ACTGAAAATGAGTGACCATGCTGACCCCTGTGTAGGCAATGTAGTAATTGTA
GTAATCATTGGAGCCAAATAGGACGTKGATGAGCACGAGCAGATTCCMCTCC
TGGGTGGKGAAMAGCATGKGTGKGCAACATGKGGAAAGCCTTCTTACAAMR
GCTGGGTAGTCTGRARGTACATKTAAGGATTCCCMTCTGGKGTCCAACYTCA
TCCAATGTAAAMATKGGAGGCTTTTACCATGCTGTACTCGAAGAGTCATTGG
AASTGATGTCACCTC

>7 — соответствует GHCL01020067.1, идентичность 93.8%, ген Ef-klf1/2/4

CCTGGACTTCATCCTGAACAACAMWAGSCTACCTGCTGACTACTGCAGCCAG
GCCCCCTCCCCGCCTGCTGGACTCGTACCCAAGATGGAGATCAAAGAGGAAC
CACCATCGCCACCCAGCTACAACCCTGAGCTGTCTATGGCTGCTGATATCAAA
GACTTTGGTCATTTTACGATGCGAGAGCAGATCTCTCCGCCGCCAGCGTACCC
CTACCATGGAGCGTCCGCCAGACCCCTCGGTGCCGCATTCCAGTACAACTCA
CTACCCGTACAGCACCTTCAGGGCCTCGTCGCTGCACAGAGCCAGATGTCGC
CCTTACAGCTGAACATGGGCCCACACATGGCCCTCCTCAATGGCCCACTGCC
AACGGGGATGATGTCACCACCACAGACGCCACCCAGTGCTTCWCCTCTTCTG
GACCTGCTCCGCAGCACCCCGGTCGAGAACCCATCAGCGATCAACCCCGCTG
GTCCGCAGGTTGTSCGTRAGCRMRRAMMRAMRCAMCCTGGGGGGAGGAGA
CGCTCAACGACGCACACCTGCAGCTACGCTGGGATGTAACAAGACATACACC
AAGTCCTCTCACCTGAAGGCTCATGTCCCGTACACATACAGGAGAGAAGCCA
TACCCACTGTAACTGGAAAGGGATGCGGCTGGGAAGTTCGCCCGATCAGAC
GAAGCTCAMCKCGTCACTACCGCAAGCAWTACTGGAAGACCGTCCAATTCC
AGGKGCCAMCTTTTGKGAGARGAGCCCTCCTCCTCGCTCCGAATCACC

>8 — соответствует GHCL01015842.1, идентичность 98.8%, ген Ef-snai2

CTATAGCGCTTGACGTGAGAGTGCGTCTGAAGATGTGCGCGTAGATTGGACC
TATCAGCAAAGGCCCTCCGACAGTGTGCGCAAGAGAATGGCTTCTCACCAGT
GTGTGTACGGATGTGTCCCTGGAGGAGCCATGGCCTCGAGAAAGCCTTACCG
CAGATGGTACACTTACATGGTAGCGTATGCGTACGCAGATGCATCTTGAGCGC
ACCCAACGCAGTGTACTCTTTACCGCAGTACTACAGGCGAATGCGCGTGCGT

TCTGCCCGGCGGAGCATTGCATCTGCTGATGCTTGCTCAGCCCTCCTGCCGTA
CAGTACTCCTTGCCACAGTCTGGACATTTCAGCTTCTGGGACGATTTTCTGGG
GTTCGACGCCCTCTTCTCCTGATGGGTTTCCCCGCCGTCAGAGCCGGGTGAC
GGTAGGCGCGATTCCCCGTCAGAGCTCGGTGAAGGCAGGCACGACTGAAGA
TCGCCCACATAACATACGGAGACCGGTAGCGAGCGGGGAGGTGGGTGGTAG
GAGTTTCTCTGGGGTAGAGGTAACGGGAGGAACGGTCTGAAGAACGGCTTG
GTTGTTACAGGCGGAGCTGGTGCGCTGGTGTTGACACGTAGCCCCTCGGTGG
TAGGGTTCGACCGAGGTAAAACTGGACCGCACGACGTGGTGATTGACCGGG
GCTCCCAGGGACGCAGACGGGGTTGAGGGTGAACAGTGGCCTGGACGAAA
GTGATGACTGGTCGTCTCGGCTCCGGCTTGGCCAAAGTGGTGATCACTGTGG
CTGCACGGGTCGGATCGTACACTGGTAGTGG

>9 — соответствует GHCL01014615.1, идентичность 96.1%, ген Ef-tbx20

GCAGACGAATGTTTCCAGCTCTTCGCGTTTCCTTCAGTGGGTTGGACCCCCA
ACTCGCGGTATCTGATTCTACTCGATATATTACCGGTCGATAACAAGCGATACC
GGTATGCGTACCGCCGATCATCGTGGCTCGTTGCTGGGAAGGCAGACCCACC
GATCAAGAACCGAGTTTATCTTCACCCTGATTCACCGTTCCTTGGAGAACATT
TATCGAAGCAGATCATCTCATTCGAGAAGGTCAAGCTCACAAACAACACACT
AGGCAGTCAAAAACATCAGATCGTTCTGAACTCGATGCATCGCTACCAACCT
AGGATCCATCTCGTACGTGATAGTGGTGTTAATCTAGAGAACCTGGAAGATCA
CCTGACATCCAGCCATACCAGATCATTCGTCTTCAAGGAGACGGCCTTCACA
GCGGTCACTGCCTACCAAAACCAACTGATAACCAGACTCAAGATAGATAGTA
ATCCATTCGCCAAGGGATTCCGAGACTCGACACGGCTGACAGATTTCGAGAG
AGAGAAAGTAGCAAGTCTCCTGAGTAAACCGGAGGAGCCATCCCCGCTGTC
GAACACATGCCTATATATGACGTCAGATCAGCGGTCACACCAACTCTCACCAT
ACGGATCATTAGTATTAGATAGATTGTCGTGTACGTCGTCTCTGGACAGTGTG
ACGTCATTGATGTTCCCTTTCTTCAAACGCCCTCAGTTAGCTGACGAACATGA
GGTCCCTTCGACGTTTGGAACACTCTGCAATATTCATCGCACATCCTTGAGCT
ATCCAAGAAGGAGGACACTGGAAGGCTGTCCGAAGATAG

>10 — соответствует GHCL01033790.1, идентичность 97.4%, ген Ef-msc

GACCTGCAAGGTTCGCCGTTGTCCGGAGCTGACGATACAATTATCGAGATGA
CCTACGACACTGATTTCGGCTTTCTCGAAATGACTAGCGTGAACGACTTTTCG
AACTCGAAGTTTACCCAGGAGGGGAATGAGGAGGCGATTATTGGTCAAGTG
AACAAAGTCGCCAACACTGAACGCCGCAAAGGAAACGGCAAAACGCAAAC
GCACAGGAGGAGGCAGCGACAAACGGACGCAGACGATGCCTCCGTGAAGA
ACACGTCATCATGCAGGCGGTCGCGGAACCGGGCACCGATAAAACCGTCGC
AGCGCAACGCGGCTAATCAACGGGAGCGCAGCCGGATGAGAGTGCTCAGTA
AGGCGTTCACAAGACTGAAGACCAGTCTCCCGTGGGTACCGGCGGACACGA
AACTCTCGAAGCTTGATACCCTCAGACTCGCTGCCTGCTATATCGGTCATCTA
CAGG

>11 — соответствует GHCL01042546.1, идентичность 99.0%, ген Ef-tcf24

CTGGTCTCGCTACACCTATATTTCACATTTAATGAAGACATTGGACGATGGTG
ACGTCATGGACCCTTCCGAATCCAAGCTATCGGAGAAACAGCACCAAAGAAT
ATGA

6. Корреляция оценок экспрессии отдельных генов

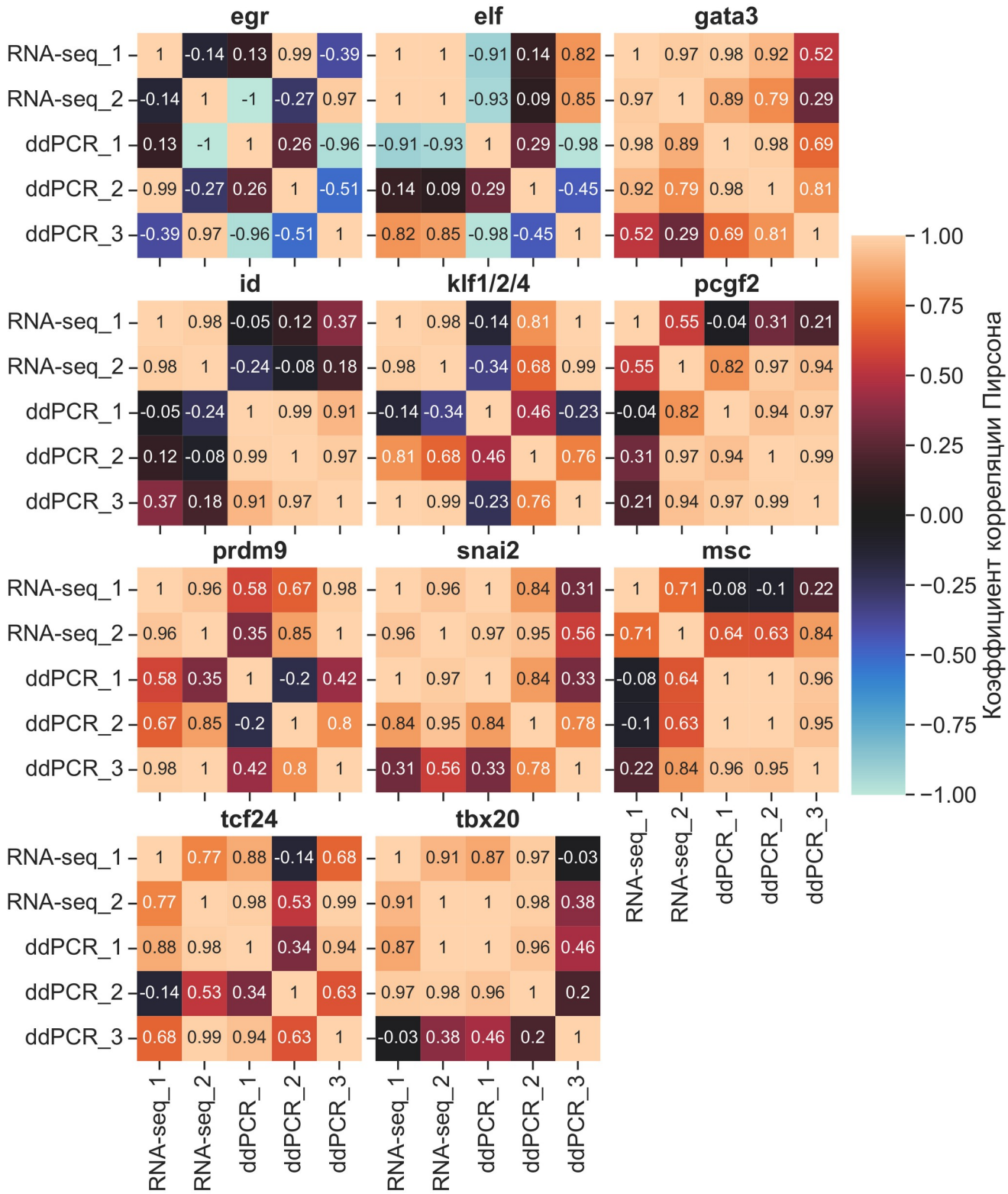

Рисунок 21. Корреляция оценок экспрессии генов 11 ТФ между разными повторами и методами. ddPCR — повторы кцПЦР, RNA-seq — повторы РНК-секвенирования. Число означает номер повтора