

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
GRADUATION THESIS**

**Поиск дифференциально экспрессирующихся генов по данным РНК секвенирования
Metridia Longa**

Обучающийся / Student Баранова Алёна Ивановна

Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа
ИТМО интердисциплинарного инжиниринга

Группа/Group A4250

Направление подготовки/ Subject area 19.04.01 Биотехнология

Образовательная программа / Educational program Прикладная геномика 2023

Язык реализации ОП / Language of the educational program Русский

Квалификация/ Degree level Магистр

Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного инжиниринга, доцент (квалификационная категория "ординарный доцент")

Обучающийся/Student

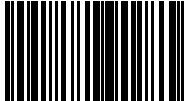
Документ подписан	
Баранова Алёна Ивановна	
26.05.2025	

(эл. подпись/ signature)

Баранова Алёна
Ивановна

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Райко Михаил Петрович	
26.05.2025	

(эл. подпись/ signature)

Райко Михаил
Петрович

(Фамилия И.О./ name
and surname)

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**ЗАДАНИЕ НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ /
OBJECTIVES FOR A GRADUATION THESIS**

Обучающийся / Student Баранова Алёна Ивановна

Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа
ИТМО интердисциплинарного инжиниринга

Группа/Group A4250

Направление подготовки/ Subject area 19.04.01 Биотехнология

Образовательная программа / Educational program Прикладная геномика 2023

Язык реализации ОП / Language of the educational program Русский

Квалификация/ Degree level Магистр

Тема ВКР/ Thesis topic Поиск дифференциально экспрессирующихся генов по данным
РНК секвенирования *Metridia Longa*

Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических
наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного
инжиниринга, доцент (квалификационная категория "ординарный доцент")

Характеристика темы ВКР / Description of thesis subject (topic)

Тема в области фундаментальных исследований / Subject of fundamental research: да /
yes

Тема в области прикладных исследований / Subject of applied research: нет / not

Основные вопросы, подлежащие разработке / Key issues to be analyzed

Выпускная квалификационная работа посвящена биоинформатическому анализу на основе данных РНК секвенирования, полученных от морского планктонного ракообразного *Metridia longa*. Цель исследования — выявление дифференциально экспрессирующихся генов между различными фракциями организма (целые особи, фракции с железами и без желез), с последующей аннотацией и функциональной интерпретацией выявленных транскриптов, потенциально связанных с биолюминесценцией. Исходными данными служили RNA-seq риды, полученные после секвенирования образцов *M. longa*, собранных в Баренцевом и Карском морях.

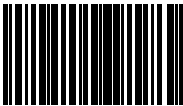
В рамках работы были выполнены: сборка транскриптома, биоинформатическая обработка данных, анализ дифференциальной экспрессии, поиск и структурная аннотация генов-кандидатов, функциональная аннотация, а также проведена интерпретация результатов.

Дата выдачи задания / Assignment issued on: 20.01.2025

Срок представления готовой ВКР / Deadline for final edition of the thesis 27.05.2025

СОГЛАСОВАНО / AGREED:

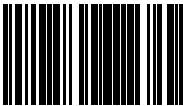
Руководитель ВКР/
Thesis supervisor

Документ подписан	
Райко Михаил Петрович	
21.04.2025	

(эл. подпись)

Райко Михаил
Петрович

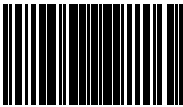
Задание принял к
исполнению/ Objectives
assumed BY

Документ подписан	
Баранова Алёна Ивановна	
24.04.2025	

(эл. подпись)

Баранова Алёна
Ивановна

Руководитель ОП/ Head
of educational program

Документ подписан	
Райко Михаил Петрович	
22.05.2025	

(эл. подпись)

Райко Михаил
Петрович

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**АННОТАЦИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
SUMMARY OF A GRADUATION THESIS**

Обучающийся / Student Баранова Алёна Ивановна
Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа
ИТМО интердисциплинарного инжиниринга
Группа/Group A4250
Направление подготовки/ Subject area 19.04.01 Биотехнология
Образовательная программа / Educational program Прикладная геномика 2023
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Магистр
Тема ВКР/ Thesis topic Поиск дифференциально экспрессирующихся генов по данным
РНК секвенирования *Metridia Longa*
Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических
наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного
инжиниринга, доцент (квалификационная категория "ординарный доцент")

**ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
DESCRIPTION OF THE GRADUATION THESIS**

Цель исследования / Research goal

Цель исследования заключалась в идентификации генов, потенциально вовлечённых в процессы биолюминесценции у *Metridia longa* на основе данных транскриптомного анализа.

Задачи, решаемые в ВКР / Research tasks

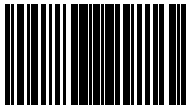
1) Провести de novo сборку транскриптома на основе данных РНК секвенирования *Metridia longa* и оценить его качество. 2) Идентифицировать транскрипты, которые содержат консервативный FYY-мотив. 3) Провести анализ дифференциальной экспрессии между светящимися и несветящимися фракциями организма. 4) Провести функциональную аннотацию отобранных последовательностей с применением биоинформатических инструментов.

Краткая характеристика полученных результатов / Short summary of results/findings

В ходе исследования был выполнен комплексный анализ транскриптомных данных копеподы *Metridia longa* с целью выявления генов, потенциально вовлечённых в процессы биолюминесценции. Сборка транскриптома de novo позволила получить широкое и полное представление о наборе экспрессируемых транскриптов у *Metridia longa* в условиях отсутствия референсного генома, обеспечив достаточное покрытие для дальнейших биоинформатических анализов. С использованием фильтрации по FYY-мотиву — трипептидной последовательности, присутствующей в предполагаемых предшественниках люциферина согласно литературным данным — было выявлено 85 транскриптов, из

которых три кандидата продемонстрировали различия в экспрессии между светящимися и несветящимися тканями. Результаты анализа дифференциальной экспрессии показали значимое разделение экспериментальных групп и наличие транскриптов, обогащённых в светоизлучающих фракциях. Один из кандидатов (NODE_255237) демонстрирует статистически значимое повышение экспрессии и сходство с белками других копепод, что позволяет рассматривать его как перспективный объект для дальнейшего функционального изучения. Два других транскрипта, несмотря на отсутствие достоверных структурных аналогов, также могут быть связаны с биolumинесцентной функцией, включая потенциальное участие в транспорте или в модификации предшественников люциферина. Кроме того, в ходе анализа был рассмотрен транскрипт предполагаемой оксидоредуктазы с FYU-мотивом, однако он не продемонстрировал значимых различий в экспрессии, что приводит к необходимости дальнейшего изучения его связи с механизмом свечения. Полученные данные формируют основу для последующих исследований, включая структурное моделирование, экспериментальную валидацию биolumинесцентной активности и расширенный поиск кандидатов вне рамок FYU-мотива.

Обучающийся/Student

Документ подписан	
Баранова Алёна Ивановна	
26.05.2025	

(эл. подпись/ signature)

Баранова Алёна
Ивановна

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Райко Михаил Петрович	
26.05.2025	

(эл. подпись/ signature)

Райко Михаил
Петрович

(Фамилия И.О./ name
and surname)

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	8
1 ОСНОВНАЯ ЧАСТЬ	13
1.1 Род <i>Metridia</i> и особенности <i>Metridia longa</i>	13
1.2 Биолюминесценция как биологическое явление.....	16
1.3 Биолюминесценция как инструмент в биотехнологии.....	17
1.4 Молекулярная природа биолюминесценции	20
1.5 Гипотезы о происхождении люциферина у <i>Metridia</i>	22
1.6 Биолюминесценция у копепод	24
1.7 Транскриптомный анализ и поиск дифференциально экспрессируемых генов.....	27
1.8 FYY-мотив и функциональная аннотация последовательностей.....	29
2. МАТЕРИАЛЫ И МЕТОДЫ	31
2.1 Сбор и подготовка биологического материала	31
2.2 Выделение РНК и секвенирование	32
2.3 Сборка и первичный анализ транскриптома de novo.....	33
2.4 Предсказание ORF, поиск доменных структур и фильтрация по FYY-мотиву.....	34
2.5 Анализ дифференциальной экспрессии (DEG)	35
2.6 Функциональная аннотация белков.....	37
3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ.....	39
3.1 Дифференциальная экспрессия генов между светящимися и несветящимися фракциями <i>Metridia longa</i>	39
3.2 Функциональная аннотация транскриптов, содержащих FYY-мотив	45
3.3 Локализация оксидоредуктазы с FYY-мотивом.....	50

ЗАКЛЮЧЕНИЕ	53
ВЫВОДЫ.....	58
СПИСОК ЛИТЕРАТУРЫ	61

ВВЕДЕНИЕ

В последние годы существенно возрос интерес к вопросам диагностики и лечению заболеваний, а также к фундаментальным и прикладным исследованиям в области медицины, фармакологии и биотехнологий. Эффективная диагностика и своевременная разработка инновационных лекарственных препаратов являются важными условиями успешного лечения и профилактики заболеваний. В результате активно разрабатываются и внедряются различные диагностические тест-системы для медицинских исследований *in vitro* и современные технологии для неинвазивного наблюдения молекулярных процессов в живых организмах, называемые биоимиджингом.

Одним из наиболее перспективных направлений биоимиджингом является использование технологий, основанных на биолюминесцентных реакциях. Биолюминесцентная визуализация обеспечивает возможность эффективного мониторинга и оценки различных биологических процессов, таких как активность опухолей, метастазирование, распределение и активность лекарственных препаратов, экспрессия генов и процессы генной терапии. Такие подходы широко применяются в лабораториях молекулярной биологии и биомедицины, предлагая безопасные и чувствительные инструменты для исследований.

Биолюминесценция представляет собой одно из уникальных природных явлений, которое проявляется в способности живых организмов испускать свечение [1]. Такое явление встречается у широкого спектра организмов – от бактерий и грибов до насекомых и морских организмов, и выполняет важные функции, включая защиту, маскировку, коммуникацию и привлечение пищи. Биолюминесценция может быть вызвана различными факторами у разных организмов. В самом распространённом случае, биолюминесценция обусловлена ферментативными реакциями между специфическими белками-ферментами (люциферазами) и их органическими субстратами

(люциферинами). Ответом на раздражение, при возвращении в основное состояние, сопровождается испусканием фотона света.

С каждым годом возрастает потребность в новых биолюминесцентных белках с заданными спектральными и биофизическими характеристиками. В связи с этим, во многих научных группах ведётся активный поиск ранее неописанных природных белков, способных к светоизлучению, а также совершенствуются уже известные белки-репортеры с помощью методов белковой инженерии. Это открывает широкие перспективы для их применения в медицинской диагностике, терапии и мониторинге физиологических процессов.

Сегодня широкое распространение в качестве биолюминесцентных репортёров получили люциферазы копепод – планктонных ракообразных, одной из основных групп зоопланктона. Ключевыми преимуществами этих ферментов являются их высокая чувствительность, нетоксичность и способность естественным образом секретироваться, что делает их удобными инструментами для исследований физиологических процессов. Тем не менее, их физико-химические характеристики и биолюминесцентные свойства до сих пор остаются недостаточно изученными, что ограничивает их широкое использование в прикладных исследованиях [2].

Особенный интерес вызывает род морских копепод *Metridia*, в частности вид *Metridia longa*, характеризующийся уникальной биолюминесцентной активностью с высокой интенсивностью и импульсным характером свечения. *Metridia longa* обитает преимущественно в холодных водах Арктики и Субарктики, что подразумевает адаптацию к низким температурам и высокому давлению, предположительно обусловившим появление уникальных структурных и функциональных характеристик люцифераз данного вида. Важной особенностью свечения у данного вида является ее импульсный характер, а также высокая интенсивность.

Несмотря на значительное количество исследований биolumинесцентных систем у таких известных объектов, как гидромедузы *Aequorea victoria*, кораллы *Renilla reniformis* и светлячки *Photinus pyralis* и других организмов, генетические основы свечения копепод остаются недостаточно изученными. До сих пор не полностью раскрыты последовательности и регуляторные элементы генов, участвующих в биосинтезе и секреции люциферина у *Metridia longa*. Ограниченность данных по генетическим аспектам обусловлена сложностью получения достаточного количества биологического материала, а также недостаточным развитием методик и инструментов для анализа транскриптомов и геномов немодельных организмов. Помимо этого, структурные особенности белков, такие как наличие большого числа дисульфидных связей, создают трудности для эффективного фолдинга и получения люцифераз в активной форме, особенно при использовании прокариотических систем экспрессии. Недостаточная изученность этих аспектов существенно ограничивает наше понимание молекулярных механизмов, регулирующих биolumинесцентные реакции, и не позволяет в полной мере использовать копеподные люциферазы для решения практических задач в биотехнологии и диагностике.

Цель настоящего исследования заключалась в идентификации генов, которые могли бы быть вовлечены в процессы биolumинесценции у вида *Metridia longa* на основе данных транскриптомного анализа.

Для достижения поставленной цели было необходимо выполнить следующие задачи:

- 1) Провести de novo сборку транскриптома на основе данных RNA-секвенирования *Metridia longa* и оценить его качество.
- 2) Идентификация транскрипты, которые содержат консервативный FYY-мотив, провести анализ дифференциальной экспрессии между светящимися и несветящимися тканями организма.

- 3) Провести функциональную аннотацию транскриптов-кандидатов с повышенной экспрессией в светящихся тканях и наличием FYY-мотива, с использованием современных биоинформатических инструментов.

Научная новизна данной работы заключается в том, что впервые выполнен комплексный транскриптомный анализ вида *Metridia longa* с целью выявления молекулярных маркером процессов биолюминесценции. Полученные результаты позволяют расширить представления о роли FYY-мотива в светящейся и несветящейся фракциях, а также позволяет предположить новые гипотезы относительно возможных механизмов эндогенного синтеза люциферина у морских копепод.

Практическая значимость настоящей исследования проявляется в спектре возможных применений полученных данных. Во-первых, белки-люциферазы *Metridia longa*, отличаются большой яркостью и стабильностью свечения при низких температурах. Это может быть использовано в биомедицинских исследованиях для визуализации клеточных процессов в реальном времени без разрушения образца. Во-вторых, благодаря большой чувствительности этих биолюминесцентных систем может появиться возможность создания простых и быстрых экотоксикологических тестов. В-третьих, новые биосенсоры на основе *Metridia*-люцифераз могут быть способны непрерывно мониторить важные метаболические параметры (кислород, pH, концентрацию ионов) в живых культурах или органах живых существ без необходимости взятия больших образцов. И наконец, применение таких «свелящихся» инструментов в диагностике состояния здоровья человека может значительно повысить точность и скорость неинвазивных исследований — например, обнаружение маркеров заболеваний на ранних стадиях по флуоресценции. Разработка и внедрение этих систем откроет новые возможности как для фундаментальной науки, так и для прикладных технологий в медицине и экологии.

В этом исследовании были использованы современные методы, которые позволяют всесторонне изучить гены и белки у вида *Metridia longa*. В первую очередь, с помощью данных RNA-seq, были получены короткие фрагменты РНК. Затем эти фрагменты были «склеены» вместе в транскриптом de novo. Для выявления отличий в активности генов между группами световоспринимающих и контрольных фракций был применен пакет DESeq2. Это позволило понять, какие транскрипты работают сильнее, а какие слабее. После этого, было выяснено, какую функцию отобранные белки могут выполнять. Был проведен поиск на основе сходства с известными структурами в базах данных, определяли домены и мотивы.

В целом, изучение процессов биolumинесценции на основе данных транскриптомного анализа и поиск дифференциально экспрессирующихся генов по данным РНК-секвенирования у вида *Metridia Longa* имеет огромный потенциал. Данное исследование позволяет глубже понять молекулярные механизмы биolumинесценции, а также формирует платформу для дальнейших функциональных исследований. Полученные результаты уточняют представления о возможной роли консервативных мотивов и структурных особенностей белков, вовлечённых в биolumинесцентные процессы. Эти данные могут быть использованы в будущих исследованиях, направленных на разработку биосенсоров и методов визуализации для биомедицинских и экологических задач.

1 ОСНОВНАЯ ЧАСТЬ

1.1 Род *Metridia* и особенности *Metridia longa*

Копеподы (*Copepoda*) — одна из самых многочисленных групп зоопланктона, которая доминирует как в морских, так и в пресных экосистемах. В глобальном масштабе копеподы составляют основную часть планктонных сообществ, зачастую достигая 55–95% от общего числа особей [3]. К данному классу относятся ракообразные, как правило, небольшого размера, это свободноживущие и паразитические виды, обитающие в морских и пресных водах [4]. Тело копепод овально-продолговатое, состоит из трёх отделов: головогруды (цефалона), груди (торакса) и брюшка (абдомена). Головогрудь включает акрон и пять сегментов, несёт антенны, мандибулы, максиллы и максиллипеды. Грудь состоит из пяти сегментов с двуветвистыми плавательными конечностями; последний сегмент имеет одноветвистые. Брюшко четырёхсегментное, конечностей не имеет; первый сегмент — половой, последний — анальный, с каудальной вилочкой (фуркой). Жабры отсутствуют, кровеносная система часто без сердца. Выделение обеспечивают максиллярные железы, нервная система включает головной ганглий и брюшную нервную цепочку. Органы чувств — науплиальные глаза и чувствительные щетинки. Копеподы — раздельнополые, с выраженным половым диморфизмом; гонады расположены в тораксе, отверстия открываются на половом сегменте [5].

Род *Metridia* относится к отряду *Calanoida*. Данный отряд включает ряд видов, которые представлены преимущественно пелагическими морскими. К представителям относится и *Metridia longa*.

Metridia longa — это морской планктонный рачок из отряда *Calanoida*, класса *Copepoda*, обитающий в холодных водах Арктики, северной части Атлантического и Тихого океанов. Он играет важную роль в зоопланктонных сообществах, являясь важным компонентом пищевых цепей. Самки этого вида достигают в среднем 4,2 мм в длину, самцы немного мельче — около 3,5 мм.

Тело *Metridia longa* почти полностью прозрачно, что помогает организму оставаться незаметным для хищников в толще воды. Форма тела вытянутая, с чётко различимыми отделами: головной частью (цефалоном), грудным отделом (тораксом) и брюшком (уросомой). У самок уросома состоит из трёх сегментов, а у самцов — из пяти, что используется как диагностический признак. Уросома составляет около 65% длины передней части тела (просомы). Усики у представителей этого вида длиннее просомы, но короче всего тела, на их базальных сегментах расположены крепкие шипы. Ротовые части массивные, что позволяет предположить способность питаться не только микроскопическим фитопланктоном, но и мелкими зоопланктонными формами [6].

Как и другие копеподы, *Metridia longa* — раздельнополый организм с хорошо выраженным половым диморфизмом. Его анатомия включает двуветвистые грудные плавательные конечности, анальную фурку на последнем брюшном сегменте и органы чувств, такие как науплиальные глаза и чувствительные щетинки. Этот вид совершает суточные вертикальные миграции — ночью поднимается в верхние слои океана для кормления, а днём опускается в глубину, что играет важную роль в вертикальном переносе углерода и питательных веществ в морских экосистемах [7].

Одной из наиболее интересных особенностей *Metridia longa* является его способность к биолюминесценции. Это организм, способный самостоятельно испускать свет в ответ на раздражение, будь то механическое или химическое воздействие. Биолюминесценция носит импульсный характер — вспышки короткие, но яркие, и испускаются в сине-голубом диапазоне (480–485 нм), что оптимально для распространения света в морской воде. Свечение обусловлено активностью секретируемых белков — люцифераз, которые катализируют реакцию окисления молекулы люциферина. Светящиеся железы расположены в антеннах, уросоме и ректальной области. Предполагается, что

у *Metridia longa* может использоваться целентеразин как субстрат для свечения, как и у других морских организмов. [8].



Рисунок 1 — *Metridia longa* (Copepoda: Calanoida): внешняя морфология [9].

Биолюминесценция *Metridia longa* играет важную биологическую роль: она используется для защиты от хищников, а также может быть вовлечена в внутривидовую коммуникацию, особенно в условиях глубоководной темноты. Высокая интенсивность свечения и его температурная стабильность делают биолюминесцентную систему этого копепода объектом интереса для биотехнологии. Свечение у *Metridia longa* объясняется активностью секретируемых белков — люцифераз [10]. Это делает данный вид интересным для изучения механизмов свечения, в частности у морских беспозвоночных.

Белки люциферазы, выделенные из *Metridia longa*, обладают высокой активностью при низких температурах и высокой светоотдачей, что делает их перспективными для разработки биосенсоров, молекулярных маркеров и

инструментов визуализации в медицине, в том числе для неинвазивной диагностики и мониторинга физиологических процессов в реальном времени.

В молекулярной основе свечения у *Metridia longa* лежит активность секретируемых белков, а именно — секреция люциферазы, которая использует целентеразин в качестве субстрата. Отличительной особенностью от других биолюминесцентных систем является содержание у люциферазы *Metridia longa* сигнального пептида. Это делает данный вид интересным для изучения механизмов секреции биолюминесцентных белков [11]. Кроме того, у *Metridia longa* были идентифицированы как минимум четыре изоформы люциферазы (MLuc1–4). Эти изоформы отличаются по длине, аминокислотной последовательности, спектральным свойствам и чувствительности к ионам, что свидетельствует о диверсификации системы биолюминесценции вида [12].

1.2 Биолюминесценция как биологическое явление

Биолюминесценция — это способность живых организмов испускать свет в результате ферментативной окислительной реакции, в которой участвуют субстрат — люциферин, молекулярный кислород и специфический фермент — люцифераза. В ходе этой реакции образуется возбужденный продукт (оксилуциферин), испускающий фотон света при возвращении в основное состояние [13]. Этот процесс протекает при минимальном тепловыделении и представляет собой особый тип хемилюминесценции, в отличие от накаливания, где свечение обусловлено тепловым возбуждением электронов.

Несмотря на кажущееся сходство, биолюминесценцию нельзя отождествлять с фосфоресценцией или флуоресценцией — формами светоизлучения, при которых возбуждённое состояние создается за счёт поглощения внешнего света. При фосфоресценции излучение сохраняется в течение некоторого времени после прекращения облучения, тогда как при флуоресценции свечение прекращается практически мгновенно [14]. В

биолюминесценции же возбуждение происходит в результате химической реакции, а не под действием света, и свечение может происходить в полной темноте без внешнего источника энергии. Хотя некоторые организмы способны к совмещению биолюминесценции и флуоресценции, это не является универсальным свойством, и наличие флуоресцентных участков не свидетельствует о наличии биолюминесцентных клеток или органелл [15].

Самым поразительным аспектом биолюминесценции является её широкое филогенетическое распространение и независимое происхождение различных светящихся систем. Данное явление зафиксировано у различных групп живых организмов. Наибольшее молекулярное и таксономическое разнообразие наблюдается у морских организмов. Более 76% пелагических видов имеют способность к свечению [16].

Предполагается, что биолюминесценция выполняет разнообразные функции: маскировку (встречная иллюминация), межвидовую и внутривидовую коммуникацию, привлечение пищи и отпугивание хищников. В частности, резкие яркие вспышки могут дезориентировать или отпугнуть хищника, повышая шансы на выживание. Тем не менее, роль биолюминесценции в жизненных процессах многих организмов остаётся до конца не изученной [17].

Эволюционно биолюминесценция представляет собой яркий пример конвергентной эволюции, поскольку способность к свечению независимо возникала у филогенетически удалённых групп организмов. Несмотря на внешнее сходство светящихся систем, они основаны на различных люциферинах, люциферазах и механизмах регуляции, что делает биолюминесценцию особенно интересным объектом для сравнительных и эволюционных исследований [18].

1.3 Биолюминесценция как инструмент в биотехнологии

Биолюминесценция широко применяется в биотехнологии благодаря своей высокой чувствительности, специфичности и возможности безвредного

использования в живых клетках. Основными объектами интереса являются люциферазы и фотопротеины — белки, способные катализировать светоизлучающую реакцию, а также их субстраты, люциферины. Эти системы нашли применение в фундаментальных и прикладных исследованиях, включая медицину, фармакологию, молекулярную биологию, экологию и даже инженерные науки

Одним из наиболее известных примеров является фотобелок aequorin, выделенный из гидромедузы *Aequorea victoria*, который светится в присутствии ионов кальция. Данный белок стал классическим инструментом для изучения кальциевой сигнализации в клетках из-за своей способности генерировать свечение, прямо пропорциональное концентрации Ca^{2+} [19]. Другим, не менее важным белком является зеленый флуоресцентный белок (GFP), используемый как маркер экспрессии генов, белков и клеток *in vivo* [20].

Люциферазы бактерий, светлячков (*Photinus pyralis*), морских организмов (*Renilla*, *Gaussia*) успешно используются в составе биосенсоров. Например, люцифераза светлячков, активируемая АТФ, применяется в диагностике и скрининге фармакологических агентов, так как она позволяет детектировать клеточную жизнеспособность и уровень метаболической активности [21].

В современной биотехнологии биолюминесценция используется в следующих направлениях:

- 1) Биолюминесцентная визуализация физиологических процессов *in vivo*: одно из ключевых применений. С помощью генно-модифицированных организмов, экспрессирующих люциферазу, возможно наблюдать в режиме реального времени пролиферацию опухолей, активность иммунных клеток (например, CAR-T), процессы регенерации, экспрессию интересующих генов и распространение инфекции. Системы на основе D-люциферина и её

аналогов являются стандартом для визуализации у млекопитающих благодаря высокой яркости и низкой токсичности, хотя их эффективность ограничивается поглощением света в тканях [22].

- 2) Высокопроизводительный скрининг: Биолюминесцентные системы применяются в HTS-форматах для оценки активности лекарств, ингибиторов, антибактериальных и противовирусных соединений. Система D-люциферин/FLuc позволяет точно измерять уровни АТФ — индикатора клеточной жизнеспособности и метаболической активности [23]. Благодаря высокой чувствительности, возможна регистрация сигнала от одной клетки.
- 3) В медицине и фармакологии люциферазы используются в составе репортерных систем для скрининга новых препаратов, оценки активности промоторов, визуализации опухолей и процессов апоптоза. Системы, основанные на световых сигналах, являются менее инвазивными и позволяют проводить длительное наблюдение за биологическими процессами в живом организме [24].

Особый интерес представляют целентеразин-зависимые биолюминесцентные системы (*Renilla*, *Metridia*, *Gaussia*). Целентеразин обладает высокой проникающей способностью и может свободно диффундировать в ткани, что делает его особенно подходящим для *in vivo* визуализации. Кроме того, излучение этих систем находится в синем-зеленом спектральном диапазоне, а значит минимально поглощается тканями. Это и позволяет регистрировать сигналы с глубины [22].

Биолюминесцентные репортеры также активно применяются в экотоксикологических исследованиях. Они используются для экспресс-оценки токсичности проб, загрязнённости воды или почвы, а также в микробиологических тестах на антибиотикорезистентность.

Исходя из вышесказанного, биолюминесценция представляет собой мощный инструмент не только в фундаментальной биологии, но и в

прикладных отраслях — от молекулярной диагностики до фармакологического скрининга и мониторинга окружающей среды.


1.4 Молекулярная природа биолюминесценции

Светоизлучающая способность живых организмов обусловлена взаимодействием двух основных компонентов: люциферина — специфического органического субстрата, и люциферазы — фермента, катализирующего его окисление. Данная реакция протекает в присутствии молекулярного кислорода. Продуктом является возбужденное соединение, возвращающееся в основное состояние с излучением фотона видимого света. Биолюминесцентные системы различных организмов уникальны по своим биохимическим характеристикам, но в основе большинства случаев лежит ферментативное окисление субстрата со свечением [25].

У представителей различных таксономических групп независимо сформировались специфические комбинации люциферазы и люциферина. Так, например, у наземных организмов наиболее подробно изучены биолюминесцентные системы светлячков (*Photinus pyralis*), использующих люциферин, активируемый АТФ, и бактерий (*Vibrio fischeri*), у которых свечение обеспечивается ферментом люциферазой, окисляющей FMNH₂ и алифатические альдегиды [13]. Из более чем 30 известных биолюминесцентных систем на сегодняшний день полностью охарактеризованы лишь 11 пар «люциферин–люцифераза» [22].

Среди морских организмов известны такие пары, как *Gaussia princeps* и *Gaussia*-люциферин, *Renilla reniformis* и *Renilla*-люциферин. Кроме того, ракообразные рода *Metridia*, включая *Metridia longa*, используют целентеразин в качестве люциферина [8].

Целентеразин (*coelenterazine*) — это органическое соединение, обладающее способностью излучать свет. Его химическая структура включает три ароматических кольца и боковые цепи с остатками аминокислот фенилаланина и тирозина [26]. Данная молекула распространена среди



Целентеразин $\xrightarrow[\text{O}_2]{\text{Люцифераза}}$ Целентерамид + CO₂ + свет

Пути поступления целентеразина в организм могут отличаться. Некоторые организмы, как было доказано на примере гидромедузы *Aequorea victoria*, неспособны к его синтезу и получают люциферин исключительно с пищей [28]. В то время как для глубоководного ракообразного *Metridia pacifica* экспериментально доказана способность к синтезу целентеразина из двух молекул тирозина и одной фенилаланина, что указывает на возможность эндогенного биосинтеза этого соединения [10].

Молекулярная природа биолюминесценции представляет собой разнообразную и адаптивную систему. Она включает ферментативные и

структурные компоненты. Их комбинация варьируется у разных таксонов и может зависеть как от генетических, так и от экологических факторов.

1.5 Гипотезы о происхождении люциферина у *Metridia*

В отличие от большинства морских организмов, которые получают целентеразин (люциферин) исключительно с пищей, у ракообразного *Metridia pacifica* была экспериментально подтверждена способность к его эндогенному синтезу. Было доказано, путём кормления животных изотопно-мечеными аминокислотами, что *Metridia pacifica* синтезирует целентеразин из двух молекул тирозина и одной фенилаланина [29]. Это открытие делает представителей рода *Metridia* одними из немногих известных организмов с синтезом целентеразина de novo.

Наибольшее внимание, в контексте происхождения люциферина, уделяется трипептидной последовательности FYY (фенилаланин–тирозин–тирозин). Предполагается, что она является ключевым предшественником целентеразина [29]. Такой интерес обусловлен тем, что химическая структура целентеразина может быть интерпретирована как результат циклизации и модификации именно этого трипептида [30]. FYY-мотив, локализованный на С-конце белков (C-terminus), был обнаружен в ряде транскриптов у светящихся организмов, в том числе у гребневиков (Ctenophore), что послужило дополнительным аргументом в пользу гипотезы о его универсальной роли в биосинтезе целентеразина [29].

Происхождение целентеразина у *Metridia*, на сегодняшний день, остаётся предметом научных обсуждений. Существует две основные гипотезы, объясняющие возможный путь биосинтеза FYY-мотива, лежащего в основе молекулы целентеразина:

- 1) Рибосомальный синтез. Согласно этой гипотезе, FYY-последовательность кодируется в составе обычного пептида-прекурсора, где С-концевой участок содержит три аминокислоты: $\text{H}_2\text{N}-\text{F}-\text{Y}-\text{Y}-\text{COOH}$. Возможная схема биосинтеза представлена на

рисунке 3. Этот пептид подвергается ферментативному посттрансляционному модификациям с последующим образованием циклического люциферина. На данный момент, гомологии с известными ферментами, которые способны катализировать подобные превращения, не выявлены [1].

- 2) Нерибосомальный синтез. Альтернативная гипотеза предполагает, что FYY-мотив может формироваться из свободных аминокислот (фенилаланина и тирозина) с участием ферментов нерибосомального синтеза пептидов (NRPS, Non-Ribosomal Peptide Synthetases). Подобные ферменты известны у бактерий и грибов, где они формируют биоактивные молекулы, включая антибиотики. Однако их наличие у копепод не подтверждено, и в геномах *Metridia* соответствующих генов пока не обнаружено [13].

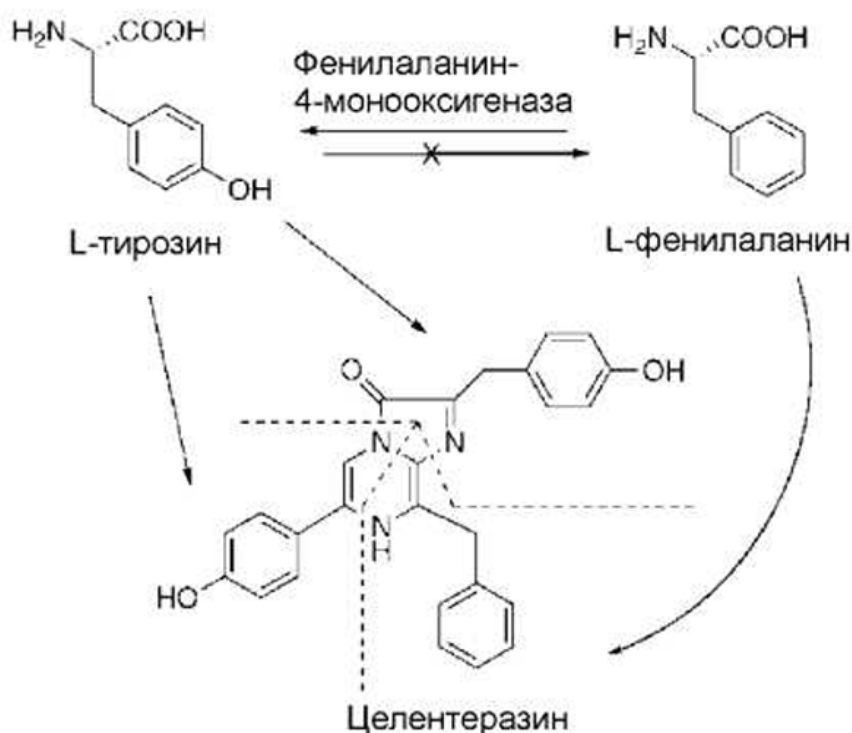


Рисунок 3 — Возможная схема биосинтеза целентеразина копеподами из природных аминокислот. Пунктиром помечено разделение целентеразина на исходные части [31].

Следует подчеркнуть, что обе эти гипотезы остаются на уровне теоретических моделей и пока что не имеют прямых экспериментальных доказательств. Попытки идентифицировать гены или ферменты, непосредственно вовлечённые в циклизацию FYU и синтез целентеразина у *Metridia*, продолжаются. Транскриптомные исследования указывают на наличие кандидатов, содержащих FYU-мотив, но их биологическая функция и каталитическая активность на данный момент не охарактеризованы.

1.6 Биolumинесценция у копепод

Копеподы (*Copepoda*) — один из ключевых компонентов морского зоопланктона, в который входит тысячи видов, обитающих в различных водных экосистемах. Биolumинесценция у копепод является широко распространённым и важным адаптивным механизмом, особенно у видов, которые живут в средних и больших глубинах. Способность к свечению используется для отпугивания хищников, маскировки, участвует в межвидовой коммуникации и, возможно, для привлечения партнёров [32].

Биolumинесценция копепод носит импульсный характер. Световой сигнал обычно проявляется как короткая, но яркая вспышка, которая провоцируется раздражением. Такой тип биolumинесценции отличается от, например, постоянного свечения у бактерий или коротких вспышек у кальмаров [1].

Свелящиеся железы у представителей рода *Metridia* локализованы в основном на антеннулах (чувствительных придатках головы), на последнем торакальном сегменте, а также вдоль брюшных сегментов (ближе к заднему концу тела). Именно в этих участках происходит выброс люминесцентного секрета [33]. На рисунке 4 показаны свечение особи *Metridia pacifica*, на котором стрелками обозначены участки тела с биolumинесцентными железами.

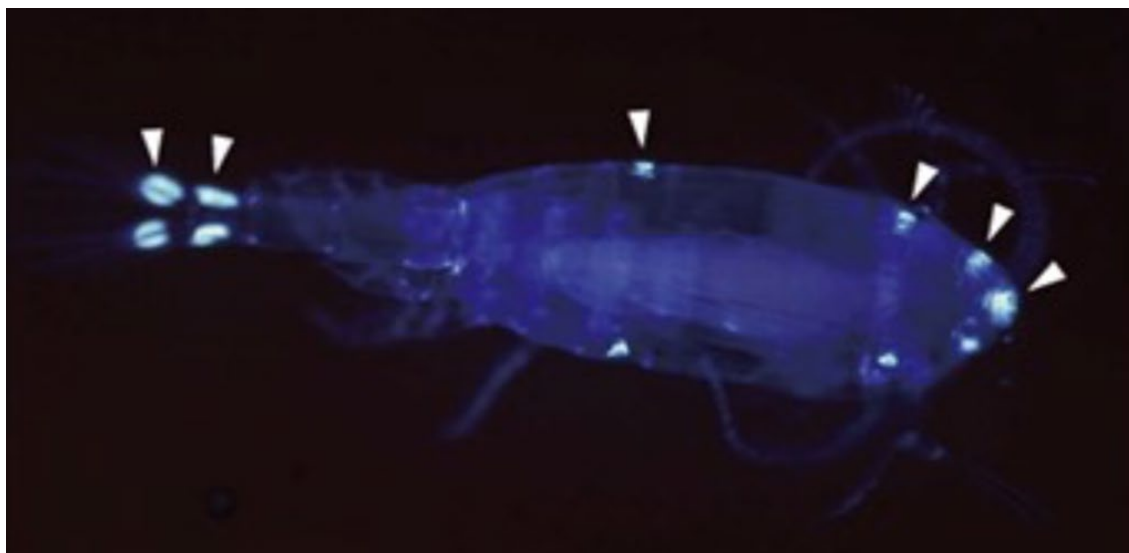


Рисунок 4 — Биолюминесценция у *Metridia pacifica*. Стрелками отмечены светящиеся железы, расположенные на антеннулах, анальном сегменте и каудальных ветвях [34].

Секрет из этих желез выбрасывается в окружающую воду в виде тонкой нити или облака, сопровождающегося мгновенным свечением. Это может создавать иллюзию присутствия особи в другом месте — так называемый "фантом". Такой способ повышает шансы копеподы на спасение от хищника [35].

На сегодняшний день биолюминесценция изучена у нескольких видов копепод, таких как *Metridia longa*, *Metridia pacifica*, *Metridia lucens*, *Gaussia princeps*, *Pleuromamma xiphias* и др. Все эти виды показывают разнообразие в локализации светящихся желез, спектре излучения и генетической организации люцифераз [36]. Так, у *Metridia longa* и *Metridia pacifica* было идентифицировано как минимум четыре гена, кодирующих люциферазы:

- 1) MLuc1 — одна из наиболее ярких изоформ. Демонстрирует высокую активность при физиологическом pH. Обладает умеренной температурной стабильностью. Используется в коммерческих биосенсорах.
- 2) MLuc2 — отличается повышенной устойчивостью к температурным колебаниям. Это делает её удобной для использования в

нестабильных средах. Её излучение — синий свет с максимумом около 480 нм.

- 3) MLuc3 — имеет более узкий диапазон pH-активности, при этом отличается высокой чувствительностью к ионам Ca^{2+} и Mg^{2+} . Данное свойство может быть связано с функциональной специализацией в физиологии копепод.
- 4) MLuc4 — представляет собой наиболее редкую и плохо охарактеризованную форму. Её биологическая роль до конца не ясна, но предполагается, что она может выполнять вспомогательную функцию или быть активной в определённые стадии жизненного цикла [12].

Все четыре изоформы содержат N-концевой сигнальный пептид. Его наличие подтверждает их предназначение для секреции в окружающую среду. Они не демонстрируют гомологии с другими известными люциферазами, кроме представителей семейства *Metridinidae*. Это указывает на их эволюционную уникальность [37].

Биолюминесценция у копепод, в частности у представителей рода *Metridia*, — это сложное явление, включающее строение тела, работу специализированных генов и белков, а также их поведение. Благодаря этому *Metridia* стала удобной модельной группой для изучения того, как биолюминесценция работает на молекулярном, физиологическом и экологическом уровнях. Люциферазы данного рода ярко светятся, устойчивы к изменениям температуры и условиям морской воды, а также различаются по спектру излучения. Эти особенности делают их полезными не только для фундаментальных исследований, но и в разработке биосенсоров и систем визуализации в живых организмах.

1.7 Транскриптомный анализ и поиск дифференциально экспрессируемых генов

RNA-seq (секвенирование РНК) – это основной инструмент молекулярной биологии для изучения экспрессии генов, особенно в организмах без референсного генома. Такой метод позволяет получить данные о транскриптоме организма, в том числе информацию об экспрессируемых генах, уровне их транскрипции, альтернативном сплайсинге и не кодирующих РНК. RNA-seq представляет собой мощный подход в изучении беспозвоночных, так как многие из них не имеют полностью аннотированных геномов. Метод позволяет провести *de novo* сборку транскриптома, а также анализ экспрессии без информации о последовательностях ДНК.

Наиболее часто используемым инструментом для *de novo* сборки транскриптома является Trinity [38]. Программный пакет был разработан в Broad Institute и Hebrew University of Jerusalem. Trinity выполняет сборку в три этапа: Inchworm, Chrysalis и Butterfly [39].

Но, несмотря на свою эффективность, *de novo* транскриптомный анализ имеет ряд сложностей. Во-первых, проблемой может являться фрагментация транскриптов, так как некоторые могут собираться не полностью. Во-вторых, аннотация может приводить к выявлению ложных открытых рамок считывания (ORF), а также включению в сборку примесей рРНК и других нецелевых последовательностей (особенно актуально для морских беспозвоночных, содержащих симбиотические микроорганизмы). В-третьих, избыточность изоформ и паралогов: похожие последовательности затрудняют различение между альтернативным сплайсингом и дубликацией генов.

Для оценки качества и полноты транскриптома, полученного *de novo*, используется инструмент BUSCO (Benchmarking Universal Single-Copy Orthologs) [40]. Он основан на проверке присутствия эволюционно консервативных генов — универсальных одно-копийных ортологов, отобранных из базы данных OrthoDB. BUSCO позволяет определить, какая

доля этих эталонных ортологов представлена в сборке. Высокий процент — указывает на хорошее качество и достаточную глубину сборки, что особенно важно при анализе дифференциальной экспрессии. [41].

После получения данных транскриптомного анализа ключевым этапом является анализ дифференциальной экспрессии генов (DEG — Differential Expression Genes) — один из ключевых этапов в интерпретации транскриптомных данных, в том числе при сравнении образцов из разных условий, тканей или стадий развития. У организмов с биолюминесценцией, такой анализ позволяет выявить гены, которые активны только в светящихся клетках или тканях, и тем самым локализовать молекулярные компоненты, которые могут быть ответственны за свечение.

Современные методы DEG-анализа используют статистические модели для обработки данных, полученных в результате RNA-seq. Наиболее широко применяемыми инструментами являются DESeq2, EdgeR и Limma. Так, DESeq2 позволяет выявлять дифференциально экспрессируемые гены с помощью статистической нормализации и моделей. EdgeR представляет собой альтернативный пакет, также предназначенный для DEG-анализа, но особенно эффективен при наличии выраженных различий в уровнях экспрессии. Или, например, метод Limma с модулем voom — мощный и гибкий инструмент, объединяющий поддержку сложных моделей с высокой точностью при контроле ошибок.

Для устранения нежелательных артефактов, возникающих, например, при использовании разных партий РНК или условий секвенирования, применяются методы коррекции. Одним из них является ComBat, основанный на эмпирическом байесовском подходе [42]. Нормализация данных позволяет устранить различия, которые могут быть связаны с длиной транскриптов и глубиной секвенирования.

Транскриптомный анализ и методы DEG-анализа представляют собой мощные инструменты для выявления молекулярных механизмов. Они

позволяют не только определить набор активных генов, но и установить их возможную роль в биологических процессах.

1.8 FYY-мотив и функциональная аннотация последовательностей

FYY-мотив (трипептид фенилаланин–тирозин–тирозин) привлекает внимание как потенциальный предшественник молекулы целентеразина — люциферина. Этот трипептид рассматривается как структурная основа для образования гетероциклического ядра целентеразина в результате ферментативной циклизации и дальнейших модификаций. Присутствие FYY-мотива было зафиксировано в ряде организмов с биолюминесценцией, включая гребневи́ков (*Ctenophora*), у которых он локализуется на С-конце белка, что может иметь значение для его ферментативной обработки [29].

Интересно то, что гены с FYY-мотивом у биолюминесцентных видов часто кодируют белки, схожие с ферментами семейства изопенициллин-N-синтаз — железозависимыми оксидазами, которые участвуют в формировании гетероциклов [43]. Это предполагает, что FYY-пептиды могут выступать не только как структурные элементы, но и как потенциальные субстраты или регуляторы ферментативной активности.

Поиск функциональных коротких последовательностей, таких как FYY, происходит путем сканирования последовательностей аминокислот, которые были получены на основе транскриптома. Одним из используемых инструментов является TransDecoder [44]. Эта программа, которая умеет извлекать предполагаемые открытые рамки считывания (ORF, Open Reading Frame) и предсказывает кодирующие области. Однако этот метод имеет ограничения: он может пропустить короткие функциональные пептиды, ошибочно классифицировать фрагменты рРНК или фальшивые ORF как белки [45]. При анализе коротких пептидов важно учитывать наличие стоп-кодона после мотива FYY, что повышает вероятность его функциональной значимости.

Следующим этапом после получения транскриптомных данных и предсказания ORF становится функциональная аннотация — попытка предсказать биологические функции полученных белков на основании сходства с уже известными. Для этого наиболее часто используется BLAST-поиск (Basic Local Alignment Search Tool), позволяющий находить гомологичные последовательности в различных референсных базах данных [46].

Результаты BLAST-поиска значительно зависят от выбора базы данных. Базы, такие как NCBI RefSeq и UniProt/Swiss-Prot, содержат высококачественные и курированные записи, что позволяет избежать ложной аннотации. Инструменты, такие как InterPro, объединяют данные из различных доменных баз (Pfam, SMART, PANTHER и др.) и позволяют идентифицировать консервативные белковые домены, сигнальные последовательности, повторы и сайты посттрансляционных модификаций. Интерпретация по доменам особенно полезна для белков без прямых гомологов, часто встречающихся у неклассических моделей.

При аннотации уникальных белков, таких как люциферазы у *Metridia*, могут возникнуть сложности. Так, белок MLuc, несмотря на ярко выраженную люциферазную активность, долгое время не имел достоверных гомологов среди известных ферментов. Первичная последовательность MLuc не имела аналогов в известных доменных базах, что свидетельствует об эволюционной уникальности этого семейства. В таких случаях аннотация требует экспериментального подтверждения — например, экспрессии и функциональных тестов *in vitro*.

Поэтому поиск функциональных мотивов, в нашем случае — FYY, и функциональная аннотация последовательностей являются взаимосвязанными биоинформатическими подходами, позволяющими выявлять новые гены и белки и исследовать их потенциальные биологические роли.

2. МАТЕРИАЛЫ И МЕТОДЫ

2.1 Сбор и подготовка биологического материала

Образцами для РНК-секвенирования послужили планктонные ракообразные *Metridia longa*. Материал был собран в рамках двух арктических экспедиций 2022 и 2023 годов в акваториях Баренцева и Карского морей. Сбор осуществлялся в отличных гидрологических районах: вблизи архипелага Шпицберген и Новой Земли. Это позволило охватить диапазон температурных и биогеохимических условий северных морей. Образцы отбирались в зоне основного распределения *Metridia longa* - в слое 50–200 метров (при помощи планктонных сетей и сетей с фракционированием по размеру).

Ключевым моментом подготовки материала для транскриптомного анализа являлось разделение проб на фракции. Из каждой партии выделялись целые особи *Metridia longa*, а также разделённые вручную фракции с железами свечения и без них.

Разделение проводилось вручную с использованием стереомикроскопа сразу после отлова, в условиях охлаждённого помещения (температура около +4 °С). Каждая особь или фрагмент тщательно промывался стерильным фильтрованным морским раствором для минимизации контаминации.

После отделения, материал незамедлительно помещали в раствор RNeasy (Qiagen) для стабилизации РНК и инкубировали при +4 °С в течение 12–16 часов. Затем образцы замораживались в жидком азоте и далее хранились при температуре –80 °С до транспортировки в лабораторию.

Всего было подготовлено семь образцов РНК: один пул целых особей *Metridia longa*, три фракции особей с железами и три фракции особей без желез. Все образцы были маркированы уникальными идентификационными кодами с указанием координат станции, даты и типа фракции. Подобная схема отбора и хранения обеспечила высокое качество РНК.

2.2 Выделение РНК и секвенирование

Выделение РНК из образцов *Metridia longa* проводилось с применением классического метода фенол-хлороформной экстракции при реагенте TRIzol (Thermo Fisher Scientific). Это позволило обеспечить эффективное лизирование клеток. Процедура выделения включала: гомогенизацию материала, фазовое разделение с хлороформом, осаждение РНК изопропанолом, промывку осадка 75% этанолом. Для повышения выхода и чистоты РНК дополнительно использовалась очистка на колонках RNeasy (Qiagen). Растворение РНК проводилось в DEPC-обработанной воде. Качество и концентрация РНК оценивались спектрофотометрически по соотношению оптической плотности OD260/280 (приемлемые значения составляли от 1.9 до 2.1), а также — электрофореза на агарозном геле.

Для построения библиотек использовались наборы реагентов NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs) с этапом рибосомального деплетирования, так как мРНК-полиА-хвосты могут быть слабо выражены у беспозвоночных. Обработка образцов и подготовка кДНК-библиотек проводились в сотрудничестве с центром высокопроизводительного секвенирования (платформой Core Facility «Биоинформатика» МГУ).

Секвенирование выполнялось на Illumina NovaSeq 6000 по протоколу paired-end с длиной ридов 150 пар оснований (PE150). Каждый образец был отсеквенирован с генерацией в среднем 40–60 миллионов парных прочтений, что обеспечивает достаточную глубину покрытия для дифференциального анализа экспрессии.

Исходные данные были получены в формате FASTQ, с общим объёмом порядка 90 Гб. Первичный контроль качества данных проводился с использованием программы FastQC. Результаты проверки показали высокую общую оценку: значения Q30 превышали 90%, адаптерные последовательности встречались редко и удалялись на этапе препроцессинга.

Эти данные послужили основой для последующей сборки и анализа транскриптома *de novo*.

2.3 Сборка и первичный анализ транскриптома *de novo*

Для сборки транскриптома *de novo* использовалась программа Trinity, одна из наиболее широко применяемых систем для реконструкции транскриптомов без опорного генома. Данный метод сборки транскриптома использует графы де Брёйна для разбиения данных РНК-секвенирования на множество независимых компонентов, каждый из которых соответствует отдельному экспрессируемому гену. С применением параллельных вычислений из этих графов восстанавливаются транскрипты, включая альтернативные сплайсоформы. Метод подходит как для парно-концевых, так и для одиночных ридов [47]. Сборка была проведена на объединённых ридах всех образцов, прошедших контроль качества и фильтрацию, с целью максимально полно представить транскриптомную репрезентацию организма.

Сборка транскриптома включала в себя три модуля Trinity: Inchworm (для генерации уникальных транскриптов), Chrysalis (кластеризация и построение *de Bruijn*-графов) и Butterfly (реконструкция полноразмерных изоформ). В результате работы программы были получены транскрипты, сгруппированные по предполагаемым генам, включая альтернативные сплайс-формы. Средняя длина *contig*'ов составила 922 п.н., общее количество уникальных транскриптов — 131 500, N50 — 1 202 п.н., что является типичным показателем при *de novo* сборке [48].

Оценка качества сборки транскриптома проводилась с помощью *rnaQUAST*. Общая длина сборки составила 19 496 972 пар оснований, было собрано 304 970 транскриптов, из которых 96 315 превышали длину 500 п.н., а 50 697 — 1000 п.н.. Средняя длина транскриптов составила 639 п.н., максимальная длина — 23 074 п.н., значение N50 — 1255 п.н., что свидетельствует о высокой непрерывности сборки.

Сборка транскриптома *de novo* позволила нам получить полноту и разнообразие транскриптов, что необходимо для дальнейшего анализа экспрессии и аннотации белков. Полученные метрики подтверждают пригодность транскрипта для последующих этапов биоинформатического анализа.

2.4 Предсказание ORF, поиск доменных структур и фильтрация по FYY-мотиву

Поиск открытых рамок считывания (ORF) и предсказание потенциально кодирующих последовательностей были выполнены с использованием инструмента TransDecoder. Сперва был выполнен поиск длинных ORF (TransDecoder.LongOrfs). Далее — дополнительная фильтрация кандидатов проводилась с помощью *hmmsearch* (HMMER) против Pfam-базы [49]. Совпадения с доменами белков подтверждали функциональность соответствующих ORF и повышали точность предсказаний. Работа состояла из двух этапов. На первом этапе запускалась команда, в результате которой был получен каталог с файлами ORF-предсказаний. Далее проводился гомологичный поиск в базе Pfam с помощью «*hmmsearch*», чтобы уточнить предсказания. Вторым этапом было использование TransDecoder с учетом найденных ранее доменов.

После предсказания белковых последовательностей была выполнена фильтрация по наличию FYY-мотива. Фильтрация выполнялась при помощи инструментов командной строки (*bash*, *awk*, *grep*). Сначала отбирались последовательности, содержащие трипептид FYY или его вариант YYF. Затем отбирались те, где FYY находился на конце последовательности или конкретно перед стоп-кодоном. Таким образом было получено 85 последовательностей, где FYY находился либо в конце белка, либо перед стоп-кодоном.

Для аннотации потенциальных белков и поиска гомологов был проведён BLAST-поиск с использованием базы данных RefSeq protein, так как она

содержит только проверенные белки, последовательность или свойства которых были подтверждены. Совпадения в базе данных указывали на возможное отношение транскрипта к уже известным семействам белков.

По результатам, был получен набор из 85 кандидатов, обладающих С-концевым FYY-мотивом. Эти транскрипты представляют интерес для дальнейшего анализа.

2.5 Анализ дифференциальной экспрессии (DEG)

Анализ дифференциальной экспрессии (DEG) является одним из самых важных этапов интерпретации транскриптомных данных. Он позволяет выявить гены, различающиеся по уровню экспрессии между сравниваемыми условиями. В данном исследовании DEG-анализ был проведен для сравнения транскриптомов фракций *Metridia longa* с наличием и отсутствием биолюминесцентных желез. Целью анализа являлась идентификация генов, которые вовлечены в биолюминесценцию. В том числе, задачей было поиск кандидатов на роль новых люцифераз, а также белков, которые содержат FYY-мотив, предположительно имеющий связь с синтезом люциферина.

Первым этапом анализа было выравнивание всех ридов, которые прошли контроль качества. Выравнивание проводилось на de novo-собранный транскриптом при помощи алгоритма HISAT2. Индексация транскриптов выполнялась с параметром hisat2-build. Затем, с использованием hisat2 риды индивидуально выравнивались для каждого образца. Как результат были получены BAM-файлы, которые потом были сконвертированы с помощью Samtools, что обеспечило корректную подачу данных на последующие этапы анализа.

На втором этапе была получена количественная информация о транскриптах с использованием утилиты FeatureCounts. Это позволило получить count-матрицу, в которой строки соответствовали транскриптам, а столбцы — образцам. Полученная матрица далее использовалась как

входной формат для анализа на языке программирования – R, на котором и проводилось дальнейшее исследование.

Для проведения статистического анализа был использован пакет DESeq2. В нем был создан объект DESeqDataSet с дизайном модели ~status. В данном случае status обозначает экспериментальную группу (с железами — light, без желез — control). Переменная batch не включалась в модель, так как в данной выборке каждый batch соответствует только одному уровню status, что делает переменные линейно зависимыми. Такой тип зависимости препятствует одновременному включению обеих переменных в модель линейной регрессии, так как приводит к сингулярной матрице дизайна.

Нормализация данных выполнялась встроенным методом DESeq2, дополнительно были рассчитаны преобразования vst (variance stabilizing transformation) и rld (regularized log transformation) для визуализации.

Метод vst стабилизирует дисперсию по всему диапазону значений экспрессии. Это особенно полезно, когда есть большое число низкоэкспрессируемых транскриптов. Rld выполняет логарифмическое преобразование с регуляризацией. Способ "смягчает" вариации для транскриптов с низким покрытием. Он более чувствителен к небольшим различиям, особенно в данных с малым числом образцов.

Выбор данных типов нормализации обоснован тем что: во-первых, в нашем случае отсутствует референсный геном, что исключает ряд других типов нормализации. Во-вторых, у нас малое количество реплик (в данном случае 3 на группу) и это помогает минимизировать шум. В-третьих, эти методы подходят для PCA и кластеризации, так как делают данные более нормальными по распределению и пригодными для линейных моделей.

Таким образом, эти методы обеспечивают более стабильные дисперсии для генов с разным уровнем экспрессии. Это особенно важно при визуализации данных, таких как PCA, кластеризации и тепловых карт.

2.6 Функциональная аннотация белков

Функциональная аннотация предсказанных белков проводилась с использованием разных независимых подходов, обеспечивающих как поиск гомологов, так и оценку структурно-функциональных характеристик белков. Такой комплексный анализ был особенно важен для предварительной интерпретации свойств белков, содержащих FYY-мотив, так как у *Metridia longa* отсутствует аннотированный референсный геном.

На первом этапе был выполнен поиск гомологов с использованием BLASTp. Алгоритм BLASTp (Basic Local Alignment Search Tool for Proteins) — это инструмент, позволяющий находить локальные участки сходства между белковыми последовательностями, что даёт возможность идентифицировать потенциальных гомологов. BLASTp работает по принципу локального выравнивания и способен обнаруживать как близкие, так и отдалённые родственные связи между белками [50]. Для повышения надёжности были использованы две базы данных. Первичная аннотация проводилась по базе uniprotkb_refprotswissprot — высококачественной части UniProt, содержащей проверенные белки с вручную подтверждённой аннотацией [51]. Порог значимости *e-value* был установлен на уровне $1e-5$. На втором этапе использовался BLASTp против базы nr (non-redundant) от NCBI, что позволило расширить спектр поиска за счёт гипотетических и частично аннотированных белков из широкого круга организмов, включая не полностью изученных копепод.

Вторым этапом был применен инструмент InterProScan — интеграционная система для аннотации белковых последовательностей [52]. Программа объединяет в себе предсказательные модели от множества специализированных баз данных, таких как Pfam, SMART, ProSite и другие. С помощью InterProScan был произведен анализ на наличие известных доменных структур, функциональных сайтов и сигнальных пептидов. Данный подход обеспечил комплексную оценку белков по нескольким источникам. В

данном исследовании InterProScan использовался в режиме с полным охватом всех доступных компонентов, что обеспечило максимально подробную аннотацию.

В дополнение к классическим методам был использован HHpred — инструмент, основанный на сравнении скрытых марковских моделей [53]. Он позволяет выявлять отдалённые структурные гомологии даже при низкой последовательностной идентичности (<30%). HHpred дает возможность сопоставить исследуемую последовательность с белками из PDB и других структурных баз и определить вероятность принадлежности к определённым семействам. В нашем случае анализ проводился с использованием базы PDB_mmCIF70 с порогами по умолчанию, а также активной оценкой статистической значимости (e-value) и вероятности предсказания.

Использование этих четырёх подходов позволило максимально охватить как функциональную, так и структурную аннотацию белков, что особенно важно при работе с организмами, не имеющими референсного генома и обширной аннотации, как в случае с *Metridia longa*.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

3.1 Дифференциальная экспрессия генов между светящимися и несветящимися фракциями *Metridia longa*

Одной из ключевых задач настоящей работы являлось изучение молекулярных различий между тканями с биолюминесценцией и без нее. В рамках исследования было проведено сравнение транскриптомных профилей этих двух типов тканей с целью выявления дифференциально экспрессируемых генов, а также поиска потенциальных кандидатов, связанных с свечением у *Metridia longa*.

Для визуализации результатов дифференциальной экспрессии (DEG) были использованы различные методы анализа: МА-график, тепловая карта, PCA-анализ, volcano-плот. Эти подходы позволили оценить структуру различий между группами и выявить приоритетные транскрипты для дальнейшего анализа.

В первую очередь, для визуальной оценки транскриптов, которые показывают значимые изменения экспрессии между условиями, была построена диаграмма разброса МА (Mean Average plot). Результаты показаны на рисунке 5.

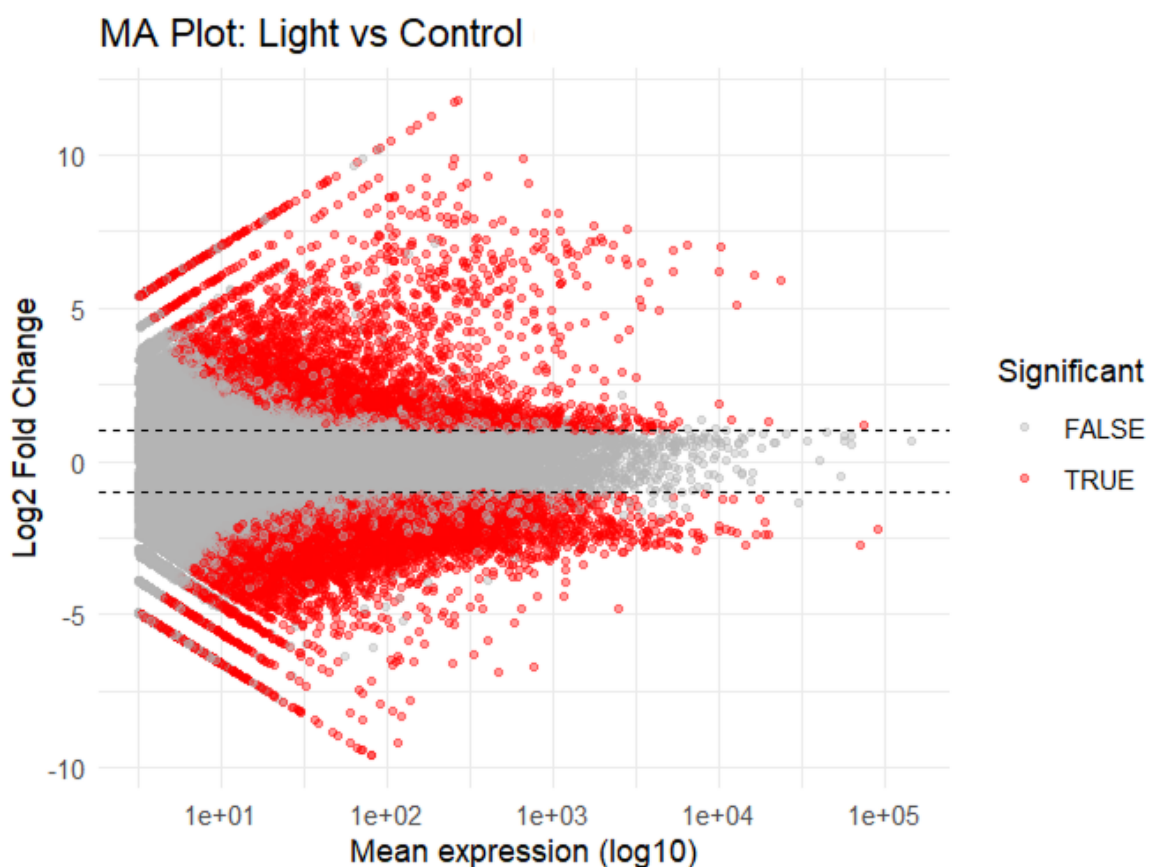


Рисунок 5 — МА-график различий в экспрессии между группами light и control для транскриптов *Metridia longa*. Красным отмечены статистически значимые транскрипты

На графике отражены результаты дифференциальной экспрессии между фракциями *Metridia longa* с биолюминесцентными железами (light) и без них (control). По оси X отложено среднее значение экспрессии (baseMean) в логарифмической шкале (\log_{10}), а по оси Y — логарифм отношения экспрессии между группами (Log2 Fold Change).

Каждая точка на графике соответствует конкретному транскрипту. Красным цветом выделены транскрипты, показавшие статистически значимые изменения экспрессии. Серым, соответственно, транскрипты без значимых изменений. Горизонтальные пунктирные линии обозначают границы биологически значимого изменения экспрессии: $\text{Log}_2\text{FC} = \pm 1$ (соответствует нулевому изменению экспрессии). На графике, точки, удалённые от центральной оси, представляют транскрипты с наиболее выраженной

дифференциальной экспрессией. Большинство транскриптов сосредоточено около нуля, что типично, но также присутствует набор транскриптов с высоким изменением экспрессии.

Дальнейшим этапом анализа была построена тепловая карта (heatmap) для 100 транскриптов с наибольшей вариабельностью экспрессии между образцами. Для построения использовались rlog-нормализованные данные. Применение именно этих данных позволило скорректировать дисперсию и обеспечить сопоставимость уровней экспрессии между генами. Тепловая карта представлена на рисунке 6.

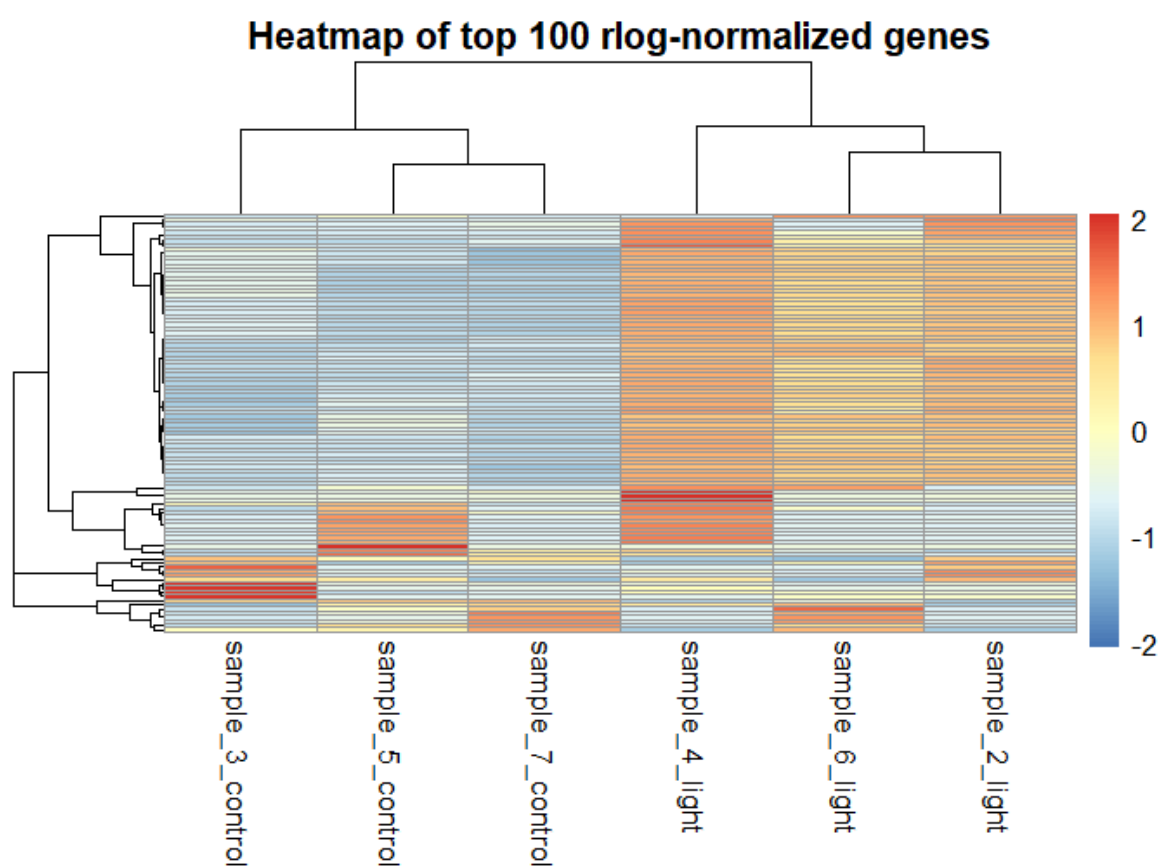


Рисунок 6 — Тепловая карта 100 генов с наибольшей вариабельностью экспрессии в образцах *Metridia longa* (rlog-нормализация)

На тепловой карте показаны изменения экспрессии наиболее вариабельных транскриптов в образцах. Цветовая шкала отражает стандартизованные значения экспрессии. Синие оттенки соответствуют низкой экспрессии, а красные — высокой.

По оси X, на графике, представлены образцы, которые были разделены на две экспериментальные группы. Видимое разделение образцов по группам свидетельствует о наличии характерных различий в экспрессии. Кластеризация образцов и транскриптов проводилась с использованием корреляционной меры сходства, что позволило выявить группы транскриптов с похожим профилем экспрессии между условиями.

Подтверждением различий в экспрессии послужил также PCA-анализ, результаты которого представлены на рисунке 7.

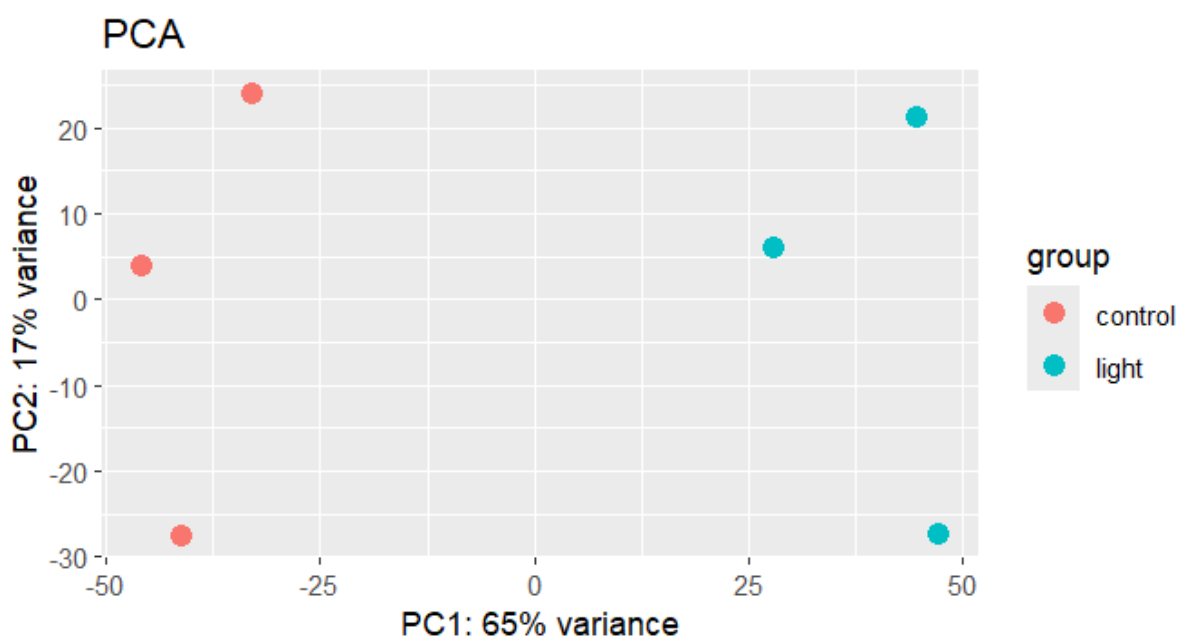


Рисунок 7 — PCA-анализ образцов *Metridia longa* с железами (light, голубой цвет) и без желез (control, розовый цвет)

PCA-график подтверждает наличие различий в экспрессии между группами. По оси X отложена первая главная компонента (PC1), которая объясняет 65% общей изменчивости данных. По оси Y — вторая компонента (PC2), объясняющая 17% вариации. Образцы из разных групп (с железами и

без железа) чётко разделяются друг от друга по оси X. Это указывает на различия в уровне экспрессии между светящимися и несветящимися тканями.

Следующим важным шагом в анализе стало построение volcano-плота для визуализации статистической значимости и амплитуды изменений экспрессии между группами. Так же сразу было учтено поведение транскриптов с FYY-мотивом. Результаты отображены на рисунке 8.

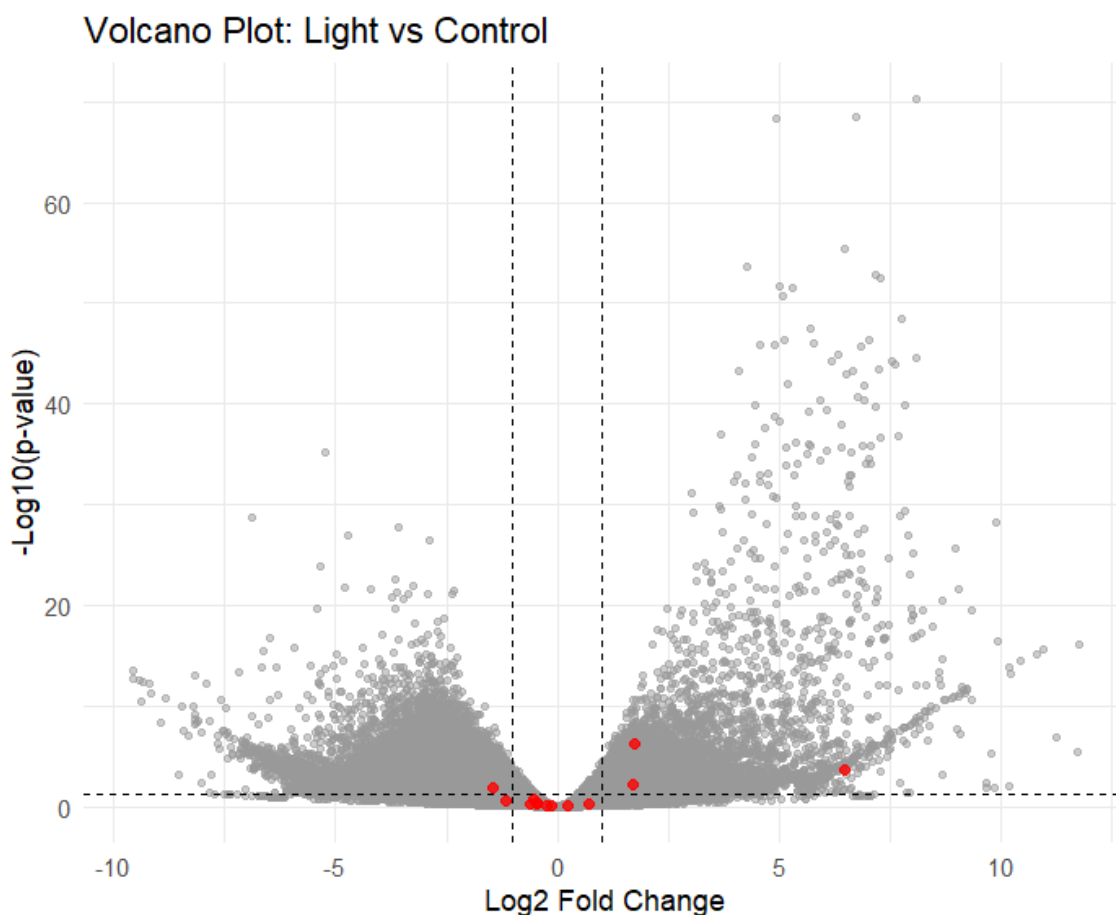


Рисунок 8 — Volcano-диаграмма ($p\text{-value} = 0.05$) различий экспрессии генов между образцами с железами (light) и без желез (control) у *Metridia longa*. Красным выделены транскрипты, содержащие FYY-мотив

На данном графике показано распределение транскриптов по степени изменения экспрессии между исследуемыми группами. По оси абсцисс (X) отложены значения логарифма изменения уровня экспрессии (Log2 Fold Change). Этот показатель отражает то, насколько сильно уровень экспрессии гена отличается между группами. Транскрипты, показывающие повышенную

экспрессивность, отображаются в положительных значениях для группы light (с железами), отрицательных — группы control (без желез). По оси ординат (Y) отображены значения $-\log_{10}(\text{p-value})$. P-value в этом случае показывает статистическую значимость различий: чем выше точка расположена на оси, тем более значимо различие. Каждая точка на графике соответствует отдельному транскрипту. Горизонтальная пунктирная линия на уровне $-\log_{10}(\text{p-value}) = 1.3$ соответствует пороговому значению $\text{p-value} = 0.05$. Ниже этого значения различия считаются статистически незначимыми.

Красными точками были выделены FYY-кандидаты. Анализ их распределения показал, что большинство из них находятся вблизи нулевого значения Log2FoldChange , что указывает на отсутствие значимых изменений экспрессии между группами для основной массы отобранных кандидатов. Однако несколько транскриптов демонстрировали отклонение от центра графика, что позволило выделить их как более перспективные для дальнейшего функционального анализа.

Дальнейший отбор был основан на анализе уровней экспрессии, проведённом с помощью volcano-плота (рисунок 8). На построенном графике были выделены транскрипты с FYY-мотивом. Так же были оценены их характеристики: положение относительно осей, значимость изменений экспрессии (p-value) и амплитуда изменений (Log2 Fold Change).

Большинство FYY-кандидатов оказалось сосредоточено около нулевого значения Log2 Fold Change , что говорит об низко выраженных изменениях экспрессии между группами light (с железами) и control (без желез). Однако среди них были отобраны несколько транскриптов с наиболее заметным отклонением, которые представляли интерес для дальнейшего функционального анализа. Отобранные транскрипты и их характеристики, отображены в таблице 1.

Таблица 1 — Отобранные FYY-кандидаты.

ID транскрипта	Log2 Fold Change	p-value	p-adj
NODE_255237	1.7521	0.0002	0.0414
NODE_267669	6.4817	0.0058	0.0703
NODE_293899	1.7135	0.0120	0.4256

Данная таблица отражает характеристики трёх отобранных транскриптов с FYY-мотивом. Для транскрипта NODE_255237 скорректированное значение p-adj оказалось статистически значимым (0.0414). Транскрипты NODE_267669 и NODE_293899 демонстрируют высокие значения Log2 Fold Change, низкие p-value и их скорректированные значения (p-adj) превышают установленный порог значимости (0.05), что означает отсутствие статистически достоверных изменений экспрессии после поправки на множественные сравнения. Это может говорить о том, что значимость их изменений не велика.

3.2 Функциональная аннотация транскриптов, содержащих FYY-мотив

Следующим этапом была проведена функциональная аннотация транскриптов, содержащих FYY-мотив. Она позволила детально охарактеризовать каждый из отобранных кандидатов. Для анализа применялось несколько подходов. Сначала аминокислотные последовательности были сопоставлены с базой UniProt с помощью BLASTp — для выявления предварительных гомологов среди аннотированных белков. Далее использовался BLASTp (nr) против базы NCBI non-redundant protein sequences для расширенного поиска совпадений, включая неполностью аннотированные белки. Для оценки доменных структур и принадлежности к белковым семействам был применен InterProScan, включающий Pfam и другие профильные базы. Завершающим этапом анализа стал поиск отдалённой

структурной гомологии при помощи HHpred — инструмента, основанного на сравнении скрытых марковских моделей с использованием базы PDB и других структурных библиотек. Комплексное применение этих подходов позволило провести углублённую аннотацию и выделить особенности каждого из трёх отобранных FYY-кандидатов. Характеристики всех трёх транскриптов приведены в таблице 2.

Таблица №2 — Результаты функциональной аннотации отобранных кандидатов.

		NODE_255237	NODE_267669	NODE_293899
Length (aa)		122	104	124
UniProt (BLAST)	Best hit	TCAL_09058 (Tigriopus californicus (Marine copepod))	B9Z55_026724 (Caenorhabditis nigoni)	TCAL_09058 (Tigriopus californicus (Marine copepod))
	Identity	44.9%	27.5%	43.8%
	E-value	1.1e-7	5.9	1.2e-7
BLASTp (nr)	Best hit	CAB4065159.1 unnamed prot. product [Lepeophtheirus salmonis]	low-quality hits	CAB4065159.1 unnamed prot. product [Lepeophtheirus salmonis]
	Identity	45.3%	low-quality hits	41.0%
	E-value	1e-07	low-quality hits	1e-07
Hhpred	Best hit	low-quality hits	PF14651.11 (Lipocalin / cytosolic fatty- acid binding protein family)	low-quality hits
	Probability	low-quality hits	95.8%	low-quality hits
	E-value	low-quality hits	0.063	low-quality hits

Транскрипт NODE_255237 имеет длину 122 аминокислоты и демонстрирует умеренное, но статистически значимое повышение экспрессии в светящихся тканях ($\text{Log2 Fold Change} = 1.7521$, $p\text{-adj} = 0.0414$). По результатам UniProt (BLAST) наилучшим совпадением стал белок TCAL_09058 из *Tigriopus californicus* с идентичностью 44.9% и e-value $3e-05$. BLASTp показал совпадение с белком из *Lepeophtheirus salmonis* (идентичность 41.0%, e-value $1e-07$). Однако эти совпадения не позволили уверенно определить функцию белка. Стоит отметить, что оба вида относятся к классу *Copepoda*, как и *Metridia longa*, что указывает на возможную эволюционную связь и сохранность некоторых белков в пределах группы. Анализ через InterProScan не выявил известных доменных структур, а предсказание структурной гомологии в HHpred показало слабые сигналы, не превышающие порог значимости. Эти данные в совокупности могут свидетельствовать о том, что транскрипт NODE_255237 кодирует ранее неохарактеризованный белок, вероятно, представляет собой белок, специфичный для копепод или же слабо консервативный гомолог известной белковой группы.

Последовательность NODE_267669 демонстрирует самое высокое изменение экспрессии среди отобранных кандидатов ($\text{Log2 Fold Change} = 6.4817$). При этом его $p\text{-adj}$ -значение (0.0703) оказалось выше выбранного порога значимости. Данная последовательность была короче всего — всего 104 аминокислоты. Поиск по UniProt показал низкое сходство с известными белками. Наилучшим хитом стал белок B9Z55_026724 круглого червя *Caenorhabditis nigoni* (идентичность 27.5%, e-value 5.9), что свидетельствует о низкой степени гомологии. При поиске InterProScan и BLASTp значимых совпадений не было выявлено. При этом, HHpred предсказал принадлежность к семейству липокалинов (Lipocalin / cytosolic fatty-acid binding protein) с высокой вероятностью — 95.8% и e-value 0.063. Липокалины представляют

собой большую группу малых внеклеточных белков, обладающих высокой степенью структурной консервативности.

Белки этого семейства способны связывать и транспортировать широкий спектр небольших гидрофобных молекул, включая жирные кислоты, ретиноиды, стероидные гормоны, билины и феромоны. Помимо транспорта, липокалины также участвуют в синтезе простагландинов, обонянии, модуляции иммунного ответа и детоксикации [54]. Их функциональное разнообразие делает липокалины потенциальными участниками в метаболических путях, связанных с биолюминесценцией, в частности, с переносом или трансформацией гидрофобных предшественников целентеразина. Хотя принадлежность транскрипта NODE_267669 к этому семейству требует дальнейшей валидации. Возможное участие подобного белка в метаболизме люциферина или его предшественников не исключено, особенно если учитывать предполагаемую роль липофильных предшественников в биосинтезе целентеразина. Тем не менее, функциональная роль данного кандидата требует дальнейшей валидации и структурных исследований.

Транскрипт с ID номером NODE_293899 по длине и результатам аннотации показал практически идентичные результаты транскрипту NODE_255237. Так, его последовательность составила 124 аминокислоты, а результаты поиска по UniProt и BLASTp показали схожую идентичность (43.8% и 41.0% соответственно). Предсказание гомологии в HHpred, так же, как и у транскрипта NODE_255237 показало низкие сигналы, не подходящие по порогу значимости. Учитывая высокий уровень сходства между этими двумя транскриптами, можно предположить о возможной альтернативной сплайс-форме одного и того же гена. Несмотря на это, оба транскрипта могли бы быть рассмотрены как единая функциональная единица при дальнейшем структурном анализе.

Таким образом, проведенный анализ дифференциальной экспрессии показал наличие ярко выраженных отличий между биолюминесцентными и не светящимися тканями у *Metridia longa*. Это наглядно подтверждается результатами PCA-анализа, MA-плота и кластеризации на тепловой карте, где образцы с светоизлучающими железами отделяются от контрольных. Это свидетельствует о наличии молекулярных механизмов, которые могут быть специфичны для тканей, способных испускать свет.

При этом, на основе volcano-плота с выделенными FYY-кандидатами было показано, что подавляющее большинство таких транскриптов не демонстрируют значимого изменения экспрессии между сравниваемыми условиями. Однако, несколько транскриптов, отклонённых от центра графика, были выбраны для углублённой функциональной аннотации как потенциально вовлечённые в свечение.

Несмотря на то, что основное внимание было уделено транскриптам, содержащим FYY-мотив — предполагаемый предшественник люциферина, — большинство из них не показали значимых различий в экспрессии между группами. Только один из трёх приоритетных кандидатов (NODE_255237) продемонстрировал статистически значимую дифференциальную экспрессию. Другие два транскрипта либо не достигли значимости, либо демонстрировали лишь слабое структурное сходство с известными белками (например, липокалинами).

Эти результаты позволяют предположить, что или роль FYY-мотива в биолюминесценции ограничена, либо отобранные кандидаты не охватывают всю функционально значимую группу генов. Альтернативно, ключевые белки биолюминесценции могут не содержать FYY-мотив вовсе и представлять собой совершенно иные молекулы, задействованные в секреции, окислительно-восстановительных процессах или регуляции экспрессии. В связи с этим целесообразным следующим шагом является проведение структурного анализа отобранных белков (например, с использованием

AlphaFold и DALI), а также расширение поиска кандидатов за рамки FYY-мотива.

3.3 Локализация оксидоредуктазы с FYY-мотивом

Отдельное внимание было уделено транскрипту, предположительно кодирующему оксидоредуктазу и при этом, содержащую FYY-мотив на С-конце. Такой интерес обусловлен несколькими фактами. Во-первых, в исследованиях, проведённых ранее, было высказано предположение о том, что коэлентеразин может формироваться в результате циклизации трипептида FYY с участием ферментативной модификации [6]. Коэлентеразин — это основная молекула, которая используется для биолюминесцентной реакции морских организмов, включая *Metridia longa*. Во-вторых, исследование Francis et al. (2015) показало, что у биолюминесцентных гребневиков кандидаты на роль синтеза люцеферина проявляют сходство с белками из семейства изопенициллин-N-синтаз, относящихся к классу Fe(II)-зависимых оксидоредуктаз [26].

Кандидат, содержащий FYY-мотив на С-конце был предварительно отобран на основе результатов BLASTp-поиска. Все образцы показали высокий процент идентичности, более 95%, что подтверждает их присутствие и консервативность в природных сообществах.

Для оценки изменений уровня его экспрессии между фракциями "light" и "control" был проведён визуальный анализ на volcano plot. На диаграмме этот транскрипт был выделен синей точкой для наглядности (см. рисунок 9).

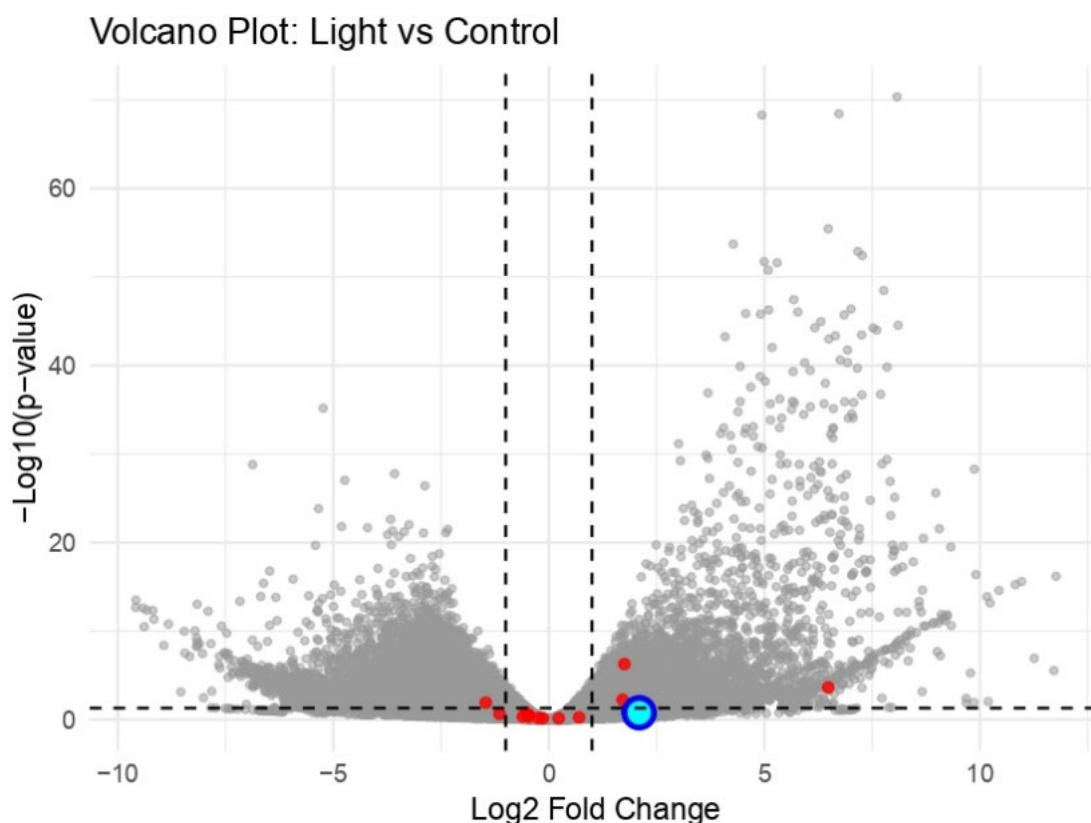


Рисунок 9 — Локализация транскрипта оксидоредуктазы с FYU-мотивом на volcano plot ($p\text{-value} = 0.05$), отображающем различия экспрессии между группами "light" и "control" у *Metridia longa*.

Расположение оксидоредуктазы с FYU оказалось около центральной вертикальной оси графика, в области низких значений по оси Y. Это указывает на то, что разница в уровне экспрессии между светящимися и несветящимися тканями минимальна ($\text{Log}_2 \text{ Fold Change}$ близок к нулю) и статистически незначима (высокое значение $p\text{-value}$).

Исходя из этого, экспрессия предполагаемой оксидоредуктазы с FYU-мотивом не демонстрирует значимых различий между исследуемыми условиями. Результаты могут свидетельствовать о её функции, не связанной с биолюминесцией у *Metridia longa*. Однако отсутствие статистически значимых различий в экспрессии не обязательно указывает на отсутствие функциональной связи данного транскрипта с процессом свечения. Возможна ситуация, при которой экспрессия фермента является конститутивной — то есть стабильной в обеих фракциях, а регуляция свечения осуществляется на

посттранскрипционном или посттрансляционном уровне, либо за счёт доступности субстрата. Кроме того, нельзя исключить, что в момент выделения ткани фермент уже был синтезирован ранее, и его экспрессия прекратилась, но белок всё ещё присутствует и выполняет свою функцию. Такие особенности необходимо учитывать при интерпретации результатов и планировании *in vitro* валидации.

ЗАКЛЮЧЕНИЕ

Биолюминесценция представляет собой сложный и многокомпонентный биологический процесс. Она проявляется в способности организмов самостоятельно испускать свет. Этот процесс может быть вызван разными причинами — чаще всего он связан с химической реакцией между белками (люциферазами) и органическими соединениями (люциферинами). Однако это не единственный механизм, с помощью которого живые организмы могут светиться. Так, например, в некоторых случаях биолюминесценция обусловлена взаимодействием с симбиотическими бактериями, которые живут в организме и сами вырабатывают светоизлучающие вещества. Биолюминесценция играет важную роль в жизни организмов — она используется для привлечения партнёров, отпугивания хищников, маскировки и общения.

Разнообразие биолюминесцентных систем у разных организмов демонстрирует высокую степень конвергентной эволюции. Свечение как форма адаптации сформировалась независимо у многих филогенетических линий — от бактерий до рыб и беспозвоночных. Это подчеркивает её значимость как экологического инструмента: будь то защита, охота или коммуникация.

Копеподы рода *Metridia*, и в частности *Metridia longa*, представляют собой перспективный объект для изучения светопродуцирующих систем благодаря яркому и импульсному характеру свечения. Эти организмы приспособлены к жизни в холодных арктических водах и обладают специализированными световыми железами, секретирующими люциферазу. Несмотря на это, молекулярные механизмы их свечения практически не изучены.

В настоящем исследовании была проведена сборка транскриптома *Metridia longa* de novo, а также выполнен комплексный анализ экспрессии генов в светящихся и контрольных тканях. Особое внимание было уделено

транскриптам, содержащим FYY-мотив — трипептид, ранее предположительно связанный с возможными предшественниками люциферина. На основе анализа экспрессии и аннотации выявлены несколько транскриптов-кандидатов, потенциально вовлечённых в светоизлучающую систему *Metridia longa*. Отдельные из них демонстрировали повышенную экспрессию в светящихся тканях, наличие сигнальных пептидов и принадлежность к семействам белков, вовлечённых в транспорт и метаболизм гидрофобных соединений.

Научная новизна данного исследования заключается в том, что впервые был проведён сравнительный транскриптомный анализ светящихся и несветящихся тканей морского планктонного рачка *Metridia longa* с целью выявления генов, потенциально вовлечённых в биолюминесцентную систему организма. Применение подхода, основанного на поиске FYY-мотива в составе предсказанных аминокислотных последовательностей, позволило сосредоточить внимание на узкой группе транскриптов, которые могут кодировать белки, участвующие в биосинтезе или трансформации молекул люциферина — ключевого компонента в механизме свечения. Такой метод предварительной фильтрации, сочетающий молекулярный маркерный подход и анализ дифференциальной экспрессии, ранее не использовался применительно к данному виду. Полученные результаты демонстрируют перспективность данного подхода для изучения молекулярных механизмов биолюминесценции у немодельных организмов и могут быть использованы как основа для аналогичных исследований у других представителей зоопланктона, особенно тех, для которых пока отсутствуют полноценно охарактеризованные геномные данные.

С практической точки зрения значимость работы заключается в том, что она создаёт основу для дальнейшего изучения молекулярных механизмов биолюминесценции у *Metridia longa* и родственных морских организмов. Хотя и отобранные в ходе анализа транскрипты-кандидаты с FYY-мотивом не могут

быть однозначно отнесены к компонентам светоизлучающей системы, но полученные данные расширяют понимание транскриптомной организации тканей, участвующих в свечении. Также итоги работы демонстрируют, какие белки экспрессируются специфически в светящихся железах. Несмотря на отсутствие прямых признаков участия в свечении, данные транскрипты представляют интерес как возможные элементы, участвующие в транспортировке или метаболизме предшественников люциферина. Таким образом, они позволяют глубже понять транскриптомную организацию светящихся тканей и служат отправной точкой для дальнейших исследований молекулярных механизмов биolumинесценции.

Предложенный подход — комбинация фильтрации по структурным мотивам, дифференциальной экспрессии и функциональной аннотации — может быть в дальнейшем применён для отбора новых, более достоверных кандидатов, включая белки, прямо вовлечённые в биolumинесцентную реакцию. В случае валидации таких белков как люцифераз или компонентов биосинтеза люциферина, они смогут быть использованы в разработке новых биolumинесцентных репортёрных систем, предназначенных для высокочувствительной визуализации *in vivo*, молекулярной диагностики и мониторинга биологических процессов в медицине и биотехнологии.

Несмотря на отсутствие однозначных доказательств включённости отобранных транскриптов с FYY-мотивом в светоизлучающую систему *Metridia longa*, полученные данные открывают ряд перспективных направлений для развития работы. В первую очередь, необходимо провести экспрессию выявленных белков *in vitro* с последующей оценкой их биохимических свойств — в частности, способности катализировать реакции, потенциально связанные с биосинтезом люциферина или световыделением. Это позволит подтвердить или опровергнуть гипотезу об их функциональной роли.

Кроме того, целесообразно осуществить пространственный анализ экспрессии с использованием методов *in situ* гибридизации или RNA-scope, чтобы локализовать отобранные транскрипты в тканях организма и уточнить их связь с железами свечения. Интеграция этих данных с протеомным анализом может дать представление о реальной продукции белков в железах и их взаимодействии.

Отдельное направление — использование AlphaFold3 и других инструментов для предсказания пространственной структуры кандидатов с последующим сравнением с базами ферментативных доменов через DALI. Это может дать подсказки о потенциальной ферментативной функции даже при низкой гомологии на уровне последовательностей.

На уровне методологии перспективным является расширение используемых критериев фильтрации: кроме FYY-мотива, могут быть учтены другие консервативные фрагменты, типичные для известных люцифераз, или структурные признаки, выявленные на основании пространственных моделей. Также возможно применение метагеномных подходов для поиска микробных партнёров, вовлечённых в биосинтез компонентов системы свечения.

Перспективным направлением также остаётся анализ транскриптов с экстремальными значениями Log2 Fold Change и высокой статистической значимостью на volcano plot — даже при отсутствии FYY-мотива. Среди таких генов могут оказаться ранее неописанные участники биолюминесцентной системы, особенно если они обладают сигнальными пептидами или участвуют в метаболизме гидрофобных соединений.

В долгосрочной перспективе результаты данной работы могут быть использованы в рамках более широких исследований, направленных на изучение молекулярных основ биолюминесценции у морских беспозвоночных. Это позволит углубить понимание эволюции светоизлучающих систем и определить потенциальные направления для их

прикладного использования в биомедицинских и биотехнологических задачах.

ВЫВОДЫ

В ходе выполнения дипломного исследования была реализована цель по идентификации транскриптов, потенциально вовлечённых в процессы биолюминесценции у морского планктонного рачка *Metridia longa*. В рамках анализа особое внимание было уделено последовательностям, содержащим FYY-мотив — трипептид, который согласно литературным данным может входить в состав пептидных предшественников люциферина. Работа охватывает полный цикл RNA-seq-анализа: от сборки транскриптома и поиска интересующих мотивов до анализа дифференциальной экспрессии и функциональной аннотации отобранных транскриптов. Полученные результаты позволили сузить круг кандидатов на участие в светоизлучающей системе и сформировать базу для их дальнейшей экспериментальной валидации.

1. По результатам — была успешно проведена сборка транскриптома *Metridia longa* методом de novo, реализованная с использованием программного пакета Trinity. Это позволило получить полный набор транскриптов, репрезентирующих экспрессируемые гены в тканях со светящимися железами и без них. Качество сборки было подтверждено при помощи инструмента rnaQUAST, подтвердившего пригодность полученных данных для последующего анализа экспрессии и аннотации.
2. В рамках работы были отобраны транскрипты, содержащие FYY-мотив, расположенный в С-концевом участке аминокислотной последовательности или перед стоп-кодоном. Это положение мотива рассматривается как потенциально функционально значимое. Всего было выявлено 85 таких транскриптов, что позволило сформировать ограниченную выборку кандидатов для дальнейшего анализа, основанного на экспрессии и функциональных признаках.

3. Анализ дифференциальной экспрессии между светящимися и несветящимися тканями — показал, что большинство FYY-позитивных транскриптов не демонстрируют статистически значимых различий. Однако трое из них отличались от фона и были отобраны для углублённого изучения. Визуализация данных с помощью MA plot, volcano plot, PCA и тепловой карты подтвердила надёжность разделения образцов по условиям и выявила гены, потенциально связанные с функционированием светящихся желез.
4. Была проведена функциональная аннотация трёх транскриптов-кандидатов, отобранных на основании FYY-мотива и уровня экспрессии. Среди них NODE_255237 оказался наиболее перспективным — он демонстрировал статистически значимую экспрессию в светящихся тканях и частичную гомологию с белками других копепод. NODE_293899 оказался близким по всем характеристикам, что позволяет рассматривать его как возможную сплайс-форму того же гена. Транскрипт NODE_267669 был идентифицирован как представитель семейства липокалинов, что может указывать на его участие в переносе гидрофобных молекул — потенциальных предшественников люциферина. При этом роль оксидоредуктазы с FYY-мотивом не подтверждена, так как различия в экспрессии между условиями были статистически незначимыми.

Таким образом, в ходе работы была выполнена комплексная транскриптомная характеристика тканей *Metridia longa*, отличающихся по способности к свечению. Полученные результаты позволили выделить и описать транскрипты-кандидаты, потенциально вовлечённые в биолюминесцентные процессы, хотя их функциональная роль требует дальнейших исследований. Предложенный подход к отбору кандидатов на основе FYY-мотива в сочетании с дифференциальной экспрессией и аннотацией может применяться в будущих проектах, направленных на

расшифровку молекулярной природы свечения у немодельных морских организмов. Работа закладывает фундамент для будущей функциональной валидации отобранных белков и расширяет понимание возможных механизмов биолюминесценции у копепод.

СПИСОК ЛИТЕРАТУРЫ

1. Haddock S.H.D., Moline M.A., Case J.F. Bioluminescence in the sea // Annual Review of Marine Science. – 2010. – Vol. 2. – P. 443–493. – DOI: 10.1146/annurev-marine-120308-081028.
2. Badr, C.E. Bioluminescent imaging: methods and protocols / C.E. Badr // Ch. 1. Bioluminescence imaging: basics and practical limitations. – Springer Science, New York – 2014. – V. 1098. – P. 1–18.
3. Longhurst A.R. The structure and evolution of plankton communities // Progress in Oceanography. – 1985. – Vol. 15, № 1. – P. 1–35. – DOI: 10.1016/0079-6611(85)90036-9.
4. Reid J.W., Williamson C.E. Chapter 21 – Copepoda // Ecology and Classification of North American Freshwater Invertebrates / eds. J.H. Thorp, A.P. Covich. – 3rd ed. – Amsterdam: Academic Press, 2010. – P. 829–899. – DOI: 10.1016/B978-0-12-374855-3.00021-2.
5. Казаченко В.Н. Определитель семейств и родов паразитических копепод (Crustacea: Copepoda) рыб: монография: в 2 ч. Ч. 1. – Владивосток: Дальрыбвтуз, 2001. – 253 с.
6. Sadia N., Ahmed M.K., Khondkar M.M., Rani S., Alam M.J., Karim A.A., Khan M.I. Horizontal and vertical distribution and abundance of zooplankton around the Swatch-of-No-Ground of Northern Bay of Bengal // The Dhaka University Journal of Earth and Environmental Sciences. – 2022. – Vol. 10, № 2. – P. 1–8. – DOI: <https://doi.org/10.3329/dujees.v10i2.57510>.
7. GBIF Secretariat. *Metridia longa* (Lubbock, 1854) in GBIF Backbone Taxonomy. Checklist dataset [Электронный ресурс]. – 2023. – URL: <https://doi.org/10.15468/39omei> (дата обращения 14.04.2025).
8. Markova S.V., Golz S., Frank L.A., Kalthof B., Vysotski E.S. Cloning and expression of cDNA for a luciferase from the marine copepod *Metridia longa*: a novel secreted bioluminescent reporter enzyme // Journal of Biological

- Chemistry. – 2004. – Vol. 279, No. 5. – P. 3212–3217. – DOI: 10.1074/jbc.M311434200.
9. COPEPEDIA. *Metridia longa* (Lubbock, 1854) [Электронный ресурс]. – URL: <https://www.st.nmfs.noaa.gov/nauplius/media/copepedia/taxa/T4000114/> (дата обращения 24.04.2025).
 10. Oba Y., Kato S., Ojika M., Inouye S. Biosynthesis of coelenterazine in the deep-sea copepod *Metridia pacifica* // Biochemical and Biophysical Research Communications. – 2009. – Vol. 390, No. 3. – P. 684–688. – DOI: 10.1016/j.bbrc.2009.10.028.
 11. Евстигнеев П.В., Битюков Э.П. Биолюминесценция морских копепод. – Киев: Наук. думка, 1990. – 144 с.
 12. Markova S.V., Larionova M.D., Burakova L.P., Vysotski E.S. The smallest natural high-active luciferase: cloning and characterization of novel 16.5-kDa luciferase from copepod *Metridia longa* // Biochemical and Biophysical Research Communications. – 2015. – Vol. 457, No. 1. – P. 77–82. – DOI: 10.1016/j.bbrc.2014.12.082.
 13. Delroisse J., Duchatelet L., Flammang P., Mallefet J. Leaving the dark side? Insights into the evolution of luciferases // Frontiers in Marine Science. – 2021. – Vol. 8. – Article 673620. – DOI: 10.3389/fmars.2021.673620.
 14. Hastings J.W. Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems // Journal of Molecular Evolution. – 1983. – Vol. 19. – P. 309–321. – DOI: 10.1007/BF02373036.
 15. Widder E.A. Bioluminescence // In: Archer S.N., Djamgoz M.B.A., Loew E.R., Partridge J.C., Vallerga S. (eds) Adaptive Mechanisms in the Ecology of Vision. – Dordrecht: Springer, 1999. – P. 603–629. – DOI: 10.1007/978-94-017-0619-3_19.
 16. Martini S., Haddock S.H.D. Quantification of bioluminescence from the surface to the deep sea demonstrates its predominance as an ecological trait // Scientific Reports. – 2017. – Vol. 7. – Article 45750. – DOI: 10.1038/srep45750.

17. Widder E.A. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity // *Science*. – 2010. – Vol. 328, No. 5979. – P. 704–708. – DOI: 10.1126/science.1174269.
18. Lau E.S., Oakley T.H. Multi-level convergence of complex traits and the evolution of bioluminescence // *Biological Reviews*. – 2021. – Vol. 96, No. 2. – P. 673–691. – DOI: 10.1111/brv.12672.
19. Shimomura, O. Bioluminescence: chemical principles / O. Shimomura // World Scientific Publishing, Singapore. – 2006.
20. Chalfie M., Tu Y., Euskirchen G., Ward W.W., Prasher D.C. Green fluorescent protein as a marker for gene expression // *Science*. – 1994. – Vol. 263, No. 5148. – P. 802–805. – DOI: 10.1126/science.8303295.
21. Meighen E.A. Molecular biology of bacterial bioluminescence // *Microbiological Reviews*. – 1991. – Vol. 55, No. 1. – P. 123–142. – DOI: 10.1128/mr.55.1.123-142.1991.
22. Syed A.J., Anderson J.C. Applications of bioluminescence in biotechnology and beyond // *Chemical Society Reviews*. – 2021. – Vol. 50. – P. 5668–5705. – DOI: 10.1039/D0CS01492C.
23. Shen L. Safe and sensitive antiviral screening platform based on recombinant human coronavirus OC43 expressing the luciferase reporter gene // *Antimicrobial Agents and Chemotherapy*. – 2016. – Vol. 60, No. 9. – P. 5492–5503. – DOI: 10.1128/AAC.00814-16.
24. Contag C.H., Spilman S.D., Contag P.R., Oshiro M., Eames B., Dennery P., Stevenson D.K., Benaron D.A. Visualizing gene expression in living mammals using a bioluminescent reporter // *Photochemistry and Photobiology*. – 1997. – Vol. 66, No. 4. – P. 523–531. – DOI: 10.1111/j.1751-1097.1997.tb03184.x.
25. Haddock S.H.D. Luminous marine organisms // In: Daunert S., Deo S.K. (eds) *Photoproteins in Bioanalysis*. – Weinheim: Wiley-VCH, 2006. – P. 25–47. – DOI: 10.1002/3527609148.ch2.

26. Inouye S. Firefly luciferase: an adenylate-forming enzyme for multicatalytic functions // Cellular and Molecular Life Sciences. – 2010. – Vol. 67, No. 3. – P. 387–404. – DOI: 10.1007/s00018-009-0170-8.
27. Meighen E.A. Molecular biology of bacterial bioluminescence // Microbiological Reviews. – 1991. – Vol. 55, No. 1. – P. 123–142. – DOI: 10.1128/mr.55.1.123-142.1991.
28. Kendall J.M., Badminton M.N. Aequorea victoria bioluminescence moves into an exciting new era // Trends in Biotechnology. – 1998. – Vol. 16, No. 12. – P. 477–481.
29. Francis W.R., Shaner N.C., Christianson L.M., Powers M.L., Haddock S.H.D. Occurrence of isopenicillin-N-synthase homologs in bioluminescent ctenophores and implications for coelenterazine biosynthesis // PLoS ONE. – 2015. – Vol. 10, No. 6. – e0128742. – DOI: 10.1371/journal.pone.0128742.
30. Tessler M., Gaffney J.P., Crawford J.M., Trautman E., Gujarati N.A., Alatalo P., Pieribone V.A., Gruber D.F. Luciferin production and luciferase transcription in the bioluminescent copepod *Metridia lucens* // PeerJ. – 2018. – Vol. 6. – Article e5506. – DOI: 10.7717/peerj.5506.
31. Маркова С.В., Высоцкий Е.С. Целентеразин-зависимые люциферазы // Биохимия. – 2015. – Т. 80, № 6. – С. 845–866.
32. Herring P.J. Bioluminescence of marine organisms // Nature. – 1977. – Vol. 267. – P. 788–793. – (Ocean Sciences Supplement). – DOI: 10.1038/267788a0.
33. Clarke G.L., Conover R.J., David C.N., Nicol J.A.C. Comparative studies of luminescence in copepods and other pelagic marine animals // Journal of the Marine Biological Association of the United Kingdom. – 1962. – Vol. 42, No. 3. – P. 541–564. – DOI: 10.1017/S0025315400054254.
34. Takenaka Y., Yamaguchi A., Shigeri Y. A light in the dark: ecology, evolution and molecular basis of copepod bioluminescence // Journal of Plankton Research. – 2017. – Vol. 39, No. 3. – P. 369–378. – DOI: 10.1093/plankt/fbx016.

35. Takenaka Y., Yamaguchi A., Shigeri Y. A light in the dark: ecology, evolution and molecular basis of copepod bioluminescence // *Journal of Plankton Research*. – 2017. – Vol. 39, No. 3. – P. 369–378. – DOI: 10.1093/plankt/fbx016.
36. Takenaka Y., Yamaguchi A., Tsuruoka N., Torimura M., Gojobori T., Shigeri Y. Evolution of bioluminescence in marine planktonic copepods // *Molecular Biology and Evolution*. – 2012. – Vol. 29, No. 6. – P. 1669–1681.
37. Takenaka Y., Masuda H., Yamaguchi A., Nishikawa S., Shigeri Y., Yoshida Y., Mizuno H. Two forms of secreted and thermostable luciferases from the marine copepod crustacean *Metridia pacifica* // *Gene*. – 2008. – Vol. 425, No. 1–2. – P. 28–35.
38. Trinity RNA-Seq [Электронный ресурс]: официальный репозиторий и документация. – URL: <https://github.com/trinityrnaseq/trinityrnaseq/wiki> (дата обращения 21.03.2025).
39. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome // *Nature Biotechnology*. – 2011. – Vol. 29, No. 7. – P. 644–652. – DOI: 10.1038/nbt.1883.
40. BUSCO user guide [Электронный ресурс]. – URL: https://busco.ezlab.org/busco_userguide.html (дата обращения: 05.04.2025).
41. 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, No. 19. – P. 3210–3212. – DOI: 10.1093/bioinformatics/btv351.
42. Johnson W.E., Li C., Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods // *Biostatistics*. – 2007. – Vol. 8, No. 1. – P. 118–127. – DOI: 10.1093/biostatistics/kxj037.
43. Rutledge P.J. Isopenicillin N synthase // In: Schofield C.J., Hausinger R.P. (eds.) *2-Oxoglutarate-Dependent Oxygenases*. – Cambridge: The Royal Society of Chemistry, 2015. – P. 414–424. – DOI: 10.1039/9781782621959-00414.

44. TransDecoder [Электронный ресурс]: официальный репозиторий на GitHub. – URL: <https://github.com/TransDecoder/TransDecoder> (дата обращения 15.04.2025).
45. Singh U., Khemka N., Rajkumar M.S., Garg R., Jain M. PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea // *Nucleic Acids Research*. – 2017. – Vol. 45, No. 22. – Article e183. – DOI: 10.1093/nar/gkx866.
46. National Center for Biotechnology Information (NCBI). Entrez Programming Utilities Help [Электронный ресурс]. – URL: <https://www.ncbi.nlm.nih.gov/books/NBK279690/> (дата обращения 26.04.2025).
47. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome // *Nature Biotechnology*. – 2011. – Vol. 29, No. 7. – P. 644–652. – DOI: 10.1038/nbt.1883.
48. Francis W.R., Christianson L.M., Kiko R., Powers M.L., Shaner N.C., Haddock S.H.D. A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly // *BMC Genomics*. – 2013. – Vol. 14. – Article 167. – DOI: 10.1186/1471-2164-14-167.
49. Finn R.D., Coghill P., Eberhardt R.Y., Eddy S.R., Mistry J., et al. The Pfam protein families database: towards a more sustainable future // *Nucleic Acids Research*. – 2016. – Vol. 44, No. D1. – P. D279–D285. – DOI: 10.1093/nar/gkv1344.
50. Pertsemlidis A., Fondon J.W. Having a BLAST with bioinformatics (and avoiding BLASTphemy) // *Genome Biology*. – 2001. – Vol. 2, No. 10. – P. reviews2002.1. – DOI: 10.1186/gb-2001-2-10-reviews2002.
51. UniProt Knowledgebase Manual [Электронный ресурс]. – URL: https://www.uniprot.org/help/uniprotkb_manual (дата обращения 14.05.2025).

52. InterProScan v5. User Guide [Электронный ресурс]. – URL: <https://interproscan-docs.readthedocs.io/en/v5/> (дата обращения 17.05.2025).
53. HHPred: homology detection and structure prediction tool [Электронный ресурс]. – URL: <https://toolkit.tuebingen.mpg.de/tools/hhpred> (дата обращения: 17.05.2025).
54. Flower D.R. The lipocalin protein family: structure and function // Biochemical Journal. – 1996. – Vol. 318, Pt. 1. – P. 1–14. – DOI: 10.1042/bj3180001.