

# Проекты ШМТБ-2022

PR01 Тайны семейства гистидиновых триад (язык: русский)	2
PR02 Вселенная биологических молекул: начало (языки: русский и английский)	3
PR03 Незаменимый инструмент: особенности эволюции клеточных типов гонад позвоночных (языки: русский и английский)	4
PR04 Генетика биомеханики: эволюция генных сетей формирования клеточной стенки растений (языки: русский и английский)	5
PR05 Проверка гипотезы песочных часов на транскриптомых данных эмбриогенеза (языки: русский и английский)	6
PR06 Optimizing Spatiotemporal Models (языки: русский и английский)	7
PR07 Пути неисповедимы: анализ регуляции смены регулонов в ходе эмбрионального развития (языки: русский и английский)	8
PR08 Роль структуры хроматина в регуляции экспрессии у дрожжей (языки: русский и английский)	9
PR09 Роль некодирующих РНК в структуре хроматина (языки: русский и английский)	10
PR10 Исследование экспрессии транспозонов в нервных клетках (языки: русский и английский)	11
PR11 Некодирующие элементы генома в эмбриональном развитии морского ежа (языки: русский и английский)	12
PR12 Show me what you got: making sense of protein language models (язык: английский)	13
PR13 Почему этот? Как нуклеотидный контекст влияет на выбор синонимичного кодона (языки: русский и английский)	14

## PR01 Тайны семейства гистидиновых триад (язык: русский)

**Руководители проекта:** Ольга Бочкарёва, Наталия Драненко

**Сотрудники:** Айгуль Насибуллина, Александр Чистяков, Вера Емельяненко

Белки гистидиновой триады (HIT Proteins) – семейство белков внешней мембраны, играют роль в массе процессов в клетке, у некоторых видов *Streptococcus* они также участвуют в патогенезе. В геномах может встретиться до четырех представителей этого семейства, однако их структура и функции остаются недоизученными.

В рамках школы мы реконструируем и изучим:

1) как возникали и терялись разные гены этого семейства, у каких видов они сохраняются. Это позволит нам их классифицировать.

2) механизмы регуляции экспрессии разных генов. Из литературы мы знаем, что некоторые белки связаны с гомеостазом цинка, а значит предположительно регулируются цинковыми регуляторами. Мы найдем и опишем мотивы сайтов связывания перед разными генами семейства, что позволит сделать гипотезы о различии функций белков и их участии в метаболических процессах.

3) как гены (ко-)локализованы на хромосомах. Мы проверим, формируют ли они опероны/касеты и если да, то с какими генами, опишем их эволюцию.

4) механизмы фазовой вариации (направленного включения и выключения генов), действующие на гены семейства. Несколько лет назад в нашей лаборатории мы биоинформатически предсказали новый механизм фазовой вариации за счет инверсии по повторам внутри генов этого семейства у пневмококков. Теперь мы проверим, у каких еще видов возникла фазовая вариация и различаются ли ее механизмы.

## PR02 Вселенная биологических молекул: начало (языки: русский и английский)

**Руководители проекта:** Пётр Власов

**Сотрудники:** Илья Мазеин

«Ландшафт фитнеса» - ключевое понятие в эволюционной биологии. Этот ландшафт отображает взаимосвязь между генотипом и репродуктивным успехом (фитнесом). В таком представлении, о сходных (по последовательностям составляющих их генов) генотипах говорят, что они «близки» друг к другу, о сильно различающихся, что они «далеки» друг от друга – а по вертикали откладывают фитнес. Это очень удобная теоретическая конструкция - однако, несмотря на годы теории и экспериментов, мы по-прежнему мало знаем о реальном устройстве этого ландшафта, особенно на глобальном масштабе. Также, рассматривая пространство последовательностей и «населяющих» его точек, вполне естественно перейти к наблюдению некоторых областей повышенной и пониженной плотности точек, реально встречающихся в природе – функциональных, «подходящих» последовательностей. Тогда можно оперировать геометрическими и топологическими категориями для изучения этого пространства. До недавнего времени количество доступных (секвенированных) биологических молекул (ДНК, РНК, белки) было недостаточно велико и просто не позволяло вести вышеупомянутый анализ, но теперь аннотированных последовательностей, а значит, и точек в пространстве последовательностей, достаточно велико.

В нашем проекте мы используем подходы из геометрии и топологии для изучения распределения биологических последовательностей (генов, белков) в их «пространстве» и на ландшафте фитнеса на разных - и больших и малых - расстояниях. Мы постараемся оценить глобальные свойства фитнес-ландшафта - а для отдельных семейства генов/белков оценить «индивидуальные» свойства их ландшафтов, в различных эволюционных временных масштабах. Отдельно-интересно проверить гипотезу, что в силу фундаментальных особенностей эволюционного процесса, параметры распределения последовательностей, в глобальном «пространстве» всех их возможных вариантов, остаются постоянными для геновых/белковых семейств на протяжении миллиардов лет эволюции.

## PR03 Незаменимый инструмент: особенности эволюции клеточных типов гонад позвоночных (языки: русский и английский)

**Руководители проекта:** Алексей Дорошков

Развитие тканей и органов животных - красивый танец межклеточных взаимодействий, направленной миграции и сложнейшей координации переключения регуляции работы генов. Огромное разнообразие клеточных типов порождает причудливые решения морфологических и физиологических задач, связанных с адаптацией к условиям среды. Однако, вне зависимости от адаптации, потребность в определении пола и производстве половых клеток остаётся одной из функций высокого приоритета. Мы можем смело утверждать, что яичники и семенники не были утрачены ни в одной из эволюционных ветвей ныне живущих позвоночных животных. Сравнивая семенники у разных видов, мы можем заметить значительные морфологические различия при схожей функции. Эти органы являются отличной моделью для изучения эволюции клеточных типов и межклеточных взаимодействий.

В рамках работы нашей команды планируется провести массовый анализ данных секвенирования транскриптомов одиночных клеток (Single-Cell RNA-Seq) развивающихся семенников для серии видов позвоночных - от рыбок *Danio rerio* до человека *Homo sapiens*, и в результате реконструировать и сопоставить генные сети дифференцировки клеточных типов в этих органах. Это позволит пролить свет на общие принципы эволюции клеточных типов, связанных с репродуктивной функцией у животных.

## PR04 Генетика биомеханики: эволюция генных сетей формирования клеточной стенки растений (языки: русский и английский)

**Руководители проекта:** Алексей Дорошков

События возникновения многоклеточных растений и появление высших растений связаны с расширением возможностей растений к разнообразным адаптациям к внешней среде. Часть из таких адаптаций связана с усложнением клеточной стенки (биохимический состав, вторичные структуры, межклеточные каналы). Также, клеточная стенка у многоклеточных водорослей обеспечивает адгезию между клетками, что является базисом для формирования тканей и органогенеза. Таким образом, этот органоид выполняет множество жизненно важных функций у растений, включая транспортные, регуляторные, адгезивные и защитные. Однако, структурно и функционально клеточные стенки представителей разных таксонов растений в значительной степени различаются (разная толщина, проницаемость и состав метаболитов). Данный проект направлен на межвидовое сопоставление генных сетей, контролирующих метаболизм и развитие клеточных стенок у представителей разных таксономических групп растений. В результате, будут выявлены эволюционно древние компоненты генных систем (общие для разных видов), а также таксон-специфические компоненты. Кроме того, будут использованы доступные данные транскриптомов развития, что позволит охарактеризовать паттерны экспрессии компонент генных сетей в процессе онтогенеза. Участники данного проекта ознакомятся с основой системно-биологического анализа данных: генными онтологиями, методами реконструкции генных сетей, выявлением ортологов, а также методами транскриптомного анализа данных.

## PR05 Проверка гипотезы песочных часов на транскриптомых данных эмбриогенеза (языки: русский и английский)

**Руководитель проекта:** Ирина Жегалова, Дмитрий Крюков

В 19 веке Карл фон Бэр предположил, что в ходе развития эмбрионы разных видов животных обособливаются друг от друга, постепенно приобретая свои уникальные черты. На основе современных данных этот закон был переработан в гипотезу песочных часов. Она предполагает короткую консервативную стадию в развитии животных. На этой стадии все эмбрионы очень друг на друга похожи, а до и после нее количество возможных межвидовых вариаций гораздо шире. Анализ транскриптомов может показать, что эта тенденция описывает не только внешние признаки, но и лежащую в их основе активность генов: в определенный момент активность генов между разными видами «синхронизируется».

В рамках данного проекта предлагается проанализировать транскриптомые данные различных высших (и не только) животных, с целью выявить эту консервативную "узкую" область эмбриогенеза и сопоставить её с соответствующими фенотипическими признаками эмбриогенеза животных.

## PR06 Optimizing Spatiotemporal Models (языки: русский и английский)

**Руководитель проекта:** Лаура Авиньо

**Сотрудники:** Вера Терентьева

Во время эмбриологического развития группа клеток реорганизуется, выделяется и даже проходит через запрограммированную гибель клеток для того, чтобы стать конечностью. Все эти процессы определяются пространственно-временной системой, задающей распределение молекул, клеток и белков в пространстве и то, как эти распределения меняются с течением времени. Как правило, они имеют разные стабильные паттерны и несколько способов их достижения. Это означает, что один и тот же набор уравнений может создавать различные модели в зависимости от значения его параметров. Например, диффузионно-реакционные системы могут описать пятна леопарда и полосы зебры! Кроме того, в настоящее время существует единое мнение в отношении того, что схожая система уравнений, основанная на генетической сети, отвечает за формирование пальцев во время развития конечностей. Тем не менее, эти механизмы всё ещё находятся на грани неизученного, и нам предстоит много работы.

Задачей, требующей дальнейшего изучения, является определение характеристик параметров системы на основе данных изображения. В этом проекте мы поймём, как работает самая простая реакционно-диффузионная система, и смоделируем её во времени и пространстве. После этого мы будем использовать наши данные синтетических изображений для (1) обучения моделей машинного обучения и (2) определения и реализации пользовательского генетического алгоритма для прогнозирования параметров. В конечном счете, мы получим пайплайн, который может быть полезен для изучения более сложных систем, например, для изучения системы, регулирующей развитие конечностей.

Для этого проекта не нужен предыдущий опыт программирования, но имейте в виду, что мы будем программировать довольно много.

## PR07 Пути неисповедимы: анализ регуляции смены регулонов в ходе эмбрионального развития (языки: русский и английский)

**Руководитель проекта:** Сергей Исаев

**Сотрудники:** Даниил Бобровский

Развитие организма — это сложный процесс, который включает в себя согласованное действие множества клеток и работу большого количества генов. Управлять развитием помогает система регулонов — чтобы активировать экспрессию нужного нам набора генов, можно активировать экспрессию всего одного гена (транскрипционного фактора), который в свою очередь будет влиять на остальные (его мишени). Некоторые регулоны работают одновременно, некоторые сменяют друг друга в ходе дифференцировки, и не всегда очень ясна регуляция данного процесса.

В ходе работы на основе датасета из [статьи](#) мы попытаемся понять, есть ли в развитии клеток нервного гребня ситуации, когда один регулон регулирует другой, что приводит к последовательному изменению экспрессий генов при дифференцировке.



## PR08 Роль структуры хроматина в регуляции экспрессии у дрожжей (языки: русский и английский)

**Руководители проекта:** *Кристина Перевощикова*, Александра Галицына

Не секрет, что информация, закодированная в ДНК, во многом определяет то, как будет жить и функционировать конкретная клетка. Относительно недавно стало понятно, что ДНК в ядре эукариотических клеток уложена определенным образом: каждая хромосома занимает отдельную область (их называют хромосомными территориями), в пределах хромосомы часто формируются домены, которые пространственно можно представить себе как “клубочки” ДНК, относительно изолированные пространственно от остальной части молекулы. Эта укладка обеспечивается различными процессами: например, белок когезин умеет делать из ДНК петли, а различные модификации белков гистонов, на которые ДНК намотана как на бусины, в разной степени “притягиваются” друг к другу.

В то же время интересно не только механистическое объяснение того, как ДНК организуется в ядре, но и может ли пространственная укладка хроматина влиять на то, каким путем закодированная в ДНК информация будет реализована, например, на уровень экспрессии генов. Для некоторых высших эукариот (в частности, млекопитающих) уже известна роль хроматиновых доменов в регуляции экспрессии генов. Например, про млекопитающих известно, что гены, находящиеся в одном домене часто экспрессируются совместно. Одно из объяснений этому феномену заключается в особенностях регуляции экспрессии у эукариот, в которой помимо промоторов генов принимают участие энхансеры - участки ДНК служащие площадкой для посадки факторов транскрипции, активирующих экспрессию. Чтобы активировать экспрессию, энхансеры должны физически контактировать с промоторами генов. Энхансеру и промотору, которые находятся в одном домене-клубочке, гораздо проще вступить в контакт, поэтому энхансер преимущественно регулирует экспрессию генов находящихся вместе с ним в одном домене, что приводит к появлению общей для домена регуляторной сети и, как следствие, к совместной экспрессии генов в пределах домена.

Возможно, что появление механизмов поддержания структуры хроматина у эукариот предоставляет им дополнительные возможности для регуляции экспрессии. Однако в настоящее время не понятно, является ли участие структуры хроматина в регуляции экспрессии консервативной особенностью всех эукариот. В частности, непонятно, связаны ли доменная организация хроматина и транскрипция у низших эукариот - дрожжей.

В нашем проекте мы попробуем найти ответ на этот вопрос с помощью анализа данных, полученных на дрожжах, в ходе работы мы решим следующие задачи:

1. Меняются ли границы доменов у дрожжей при стрессе? Скоординировано ли изменение экспрессии генов при стрессе в пределах домена?
2. Похожи ли типы промоторов и наборы транскрипционных факторов, участвующих в регуляции экспрессии у генов, находящихся в одном домене?
3. Будет ли искусственная вставка, встроенная в геном дрожжей вместо разных генов, экспрессироваться по-разному в зависимости от того, в какой домен она была встроена?
4. Будет ли такая вставка, нарушая локальные регуляторные отношения, влиять на экспрессию своих соседей в геноме? Будет ли это влияние сильнее в случае соседей по домену, по сравнению с близко расположенными в геноме генами, расположенными в соседнем домене?

## PR09 Роль некодирующих РНК в структуре хроматина (языки: русский и английский)

**Руководители проекта:** Александра Галицына

Эпигенетическая информация (“дополнительная”, закодированная в негеномной части ядра) - это не только модификации ДНК (метилование, связывание гистоновых меток). Фундаментальным открытием принято считать недавнее открытие, что важнейшим эпигенетическим фактором регуляции является трехмерная организация ДНК: иными словами, то, как разные регионы ДНК друг с другом взаимодействуют.

Как эпигенетика проявляет себя, не секрет: изменяется активность различных генов, меняется их экспрессия, а значит, и набор РНК, имеющих в клетке. Как следствие, меняется набор белков и фенотип клетки. Это - центральная догма молекулярной биологии, постулирующая, что информация, содержащаяся в ДНК, транскрибируется в РНК и транслируется в белок. Однако возможны ли обратные потоки информации в этой системе? Известно, что в эукариотическом ядре есть набор некодирующих РНК, роль которых продолжает исследоваться. Может ли РНК ядра нарушать центральную догму и влиять на ДНК так, чтобы изменять экспрессию генов?

Чтобы попробовать ответить на этот вопрос, мы объединим информацию о нескольких экспериментальных техниках: измерении контактов РНК-ДНК (RedC), ДНК-ДНК (Hi-C), а также нескольких баз данных генов. Мы начнем с разбора классических регионов “из учебника”, например, регуляцию Нох-генов, XIST и HOTTIP. Гипотезы, сформулированные для отдельных генов, мы проверим в масштабе всего генома.

## PR10 Исследование экспрессии транспозонов в нервных клетках (языки: русский и английский)

**Руководитель проекта:** Лариса Окорокова

Транспозоны - это особые элементы генома, которые способны к перемещению и размножению в пределах генома. Так, в геноме человека около 50% ДНК представлена различными мобильными элементами. Мозаичные вставки мобильных элементов создают соматические геномные вариации и обладают значительным потенциалом для создания фенотипического разнообразия, в том числе в мозге. Инсерции транспозонов могут приводить к нарушению кодирующей последовательности генов через нарушение сплайсинга или сдвиг рамки считывания, однако чаще они происходят в некодирующих частях генома и предсказать эффект инсерции в этом случае крайне сложно. Вероятно, что в таких случаях нарушаются транскрипционные программы в клетках, что в свою очередь может привести к развитию заболеваний, например, расстройствам аутистического спектра.

В нашем проекте мы планируем использовать данные РНК-секвенирования органоидов мозга, полученных из индуцибельных плюрипотентных стволовых клеток. Мы определим экспрессию генов и транспозонов в единичных клетках, кластеризуем их и определим клеточные типы. Далее мы сравним контрольные образцы и опытные, обработанные вальпроевой кислотой - противозлептическим препаратом прием которого ассоциирован с развитием симптомов аутизма у лабораторных животных.

## PR11 Некодирующие элементы генома в эмбриональном развитии морского ежа (языки: русский и английский)

**Руководитель проекта:** Николай Панюшев

Морской еж — идеальный модельный объект для эмбриологии. И у него на это сразу пять причин: его личинки прозрачны, ежей легко содержать и ловить, они просто устроены, они эволюционно ближе человеку, чем дрозофила, и их развитие отлично картировано вплоть до судьбы каждой отдельной клетки. Нам интересно проследить роль не кодирующих элементов генома в эмбриогенезе. В ранее проведенном исследовании мы обнаружили не кодирующие РНК, которые специфически экспрессируются в различных частях эмбриона. В этом проекте мы проверим, присутствуют ли найденные РНК в транскриптом морского ежа как самостоятельные транскрипты, или же являются частями каких-то других РНК. Для подтвержденных и уточненных последовательностей построим 2D-структуры и попробуем предложить механизм их работы, а также найдем аналоги в других организмах.

## PR12 Show me what you got: making sense of protein language models (язык: английский)

**Руководитель проекта:** Илья Сенаторов, Ольга Калинина

**Сотрудники:** Анна Тоидзе, Alper Yurtseven, Roman Joeres

Deep Learning - популярная тема практически в каждом разделе науки, который работает с данными. В биологии, особенно при работе с белковыми последовательностями, такие модели часто используются для создания эмбедингов. Эмбединг - процесс превращения входных данных (белковой последовательности) в вектор, в котором похожие белки (из одного семейства) получают близкие векторы, а разные белки (из разных семейств, с разными функциями и структурами) получают далекие векторы.

Вопрос, на который мы хотим получить ответ - понимают ли такие модели, обученные на миллионах последовательностей биохимию белков? Смогут ли они отличить патогенные мутации от доброкачественных? Как сломать такие модели? Как их обмануть? Как их улучшить?

Мы будем искать ответы на эти вопросы двумя способами - биологическим и вычислительным. Мы можем придумать "хитрый" биологический вопрос, на который модель должна дать ответ или мы можем применять современные методы анализа моделей, чтобы разобраться что происходит внутри. Вариантов пытки моделей множество!

TL;DR:

Берем большую белковую модель.

Учимся ее использовать

Учимся ее ломать

?????????

ПРОФИТ

## PR13 Почему этот? Как нуклеотидный контекст влияет на выбор синонимичного кодона (языки: русский и английский)

**Руководитель проекта:** Зоя Червонцева

**Сотрудники:** Ася Менделевич, Евгения Ходжаева, Валентина Бурская

Одним холодным зимним вечером, разглядывая нуклеотидные выравнивания генов из разных видов энтеробактерий, мы с коллегами обнаружили кое-что неожиданное. Аминокислоты в некоторых позициях некоторых генов всегда кодируются строго одним и тем же кодоном. В то время как в других позициях те же аминокислоты могут кодироваться разными синонимичными кодонами.

Почему некоторые позиции так жестко закреплены? Возможно, здесь будет уместна аналогия с тем, как мы выбираем слова в разговоре. Некоторые контексты *требуют* использования одного из синонимов и не допускают других. Синьор Томат и килька в помидоре вам это с легкостью подтвердят.

В нашем проекте мы будем использовать машинное обучение, чтобы понять, как выбор кодона связан с его нуклеотидным контекстом. Кроме того, посмотрим на то, как эволюционируют кодоны “молодых” (получившихся в результате недавней замены) аминокислот: закрепляются ли они сразу или еще долго меняются на синонимичные. Количество и сложность программирования будет зависеть от вашего желания.