



2021
NGS

9-Я ВСЕРОССИЙСКАЯ
НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
ЦЕНТРОВ ГЕНОМНЫХ ИССЛЕДОВАНИЙ
МИРОВОГО УРОВНЯ

ГЕНОМНОЕ СЕКВЕНИРОВАНИЕ И РЕДАКТИРОВАНИЕ

СБОРНИК ТЕЗИСОВ

Министерство здравоохранения Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
“Российский национальный исследовательский
медицинский университет имени Н.И. Пирогова”
(ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России)

«Геномное секвенирование и редактирование»

**СБОРНИК ТЕЗИСОВ
9-й Всероссийской научно-практической
конференции центров геномных исследований
мирового уровня**

Москва
РНИМУ им. Н.И. Пирогова
2021

УДК 577.21
ББК 28.04
Г27

Г27 «Геномное секвенирование и редактирование»
СБОРНИК ТЕЗИСОВ 9-й Всероссийской научно-практической
конференции центров геномных исследований мирового уровня.
Москва: РНИМУ им. Н.И. Пирогова, 2021. — 24 с.

ISBN 978-5-88458-543-0

В сборнике представлены тезисы докладов-участников конференции, посвящённые новейшим методам анализа и модификации генома различных видов организмов, включая человека.

УДК 577.21
ББК 28.04

ISBN 978-5-88458-543-0

© Составители, 2021
© ФГАОУ ВО РНИМУ им. Н.И. Пирогова, 2021

Дифференциальная экспрессия генов сигнальных путей ремоделирования бронхов при тяжелой бронхиальной астме

Е.С. Минина

*Витебский государственный медицинский университет, Витебск, Беларусь
l.mnn1551@gmail.com*

Бронхиальная астма (БА) — это аллергическое заболевание, развитие которого определяет наличие хронического воспаления, гиперреактивности дыхательных путей и структурных изменений в них — ремоделирования. В регуляции гиперплазии гладкомышечных клеток, являющейся одним из звеньев процесса ремоделирования, участвуют сигнальные пути MAPKs, PI3K/AKT, JAK/STAT.

В ходе исследования была выполнена визуализация сигнальных путей (MAPKs, PI3K/AKT, JAK/STAT) на основе данных дифференциальной экспрессии генов. Для анализа транскриптомных данных использовали R-Studio с подключением библиотек pathview, KEGGREST и веб-приложение Phantasus.

Было проанализировано 410 транскриптомов образцов периферической крови, полученных из базы данных Gene Expression Omnibus (GSE69683). В группу исследования были включены некурящие пациенты с тяжелой ($n=246$ — 85 мужчин и 161 женщина) и нетяжелой БА ($n=77$ — 40 мужчин и 37 женщин), здоровые некурящие люди без БА ($n=87$ — 53 мужчины и 34 женщины).

У пациентов с тяжелой БА при сравнении со здоровыми испытуемыми из контрольной группы наблюдалась гиперэкспрессия генов IL1, MYD88, HSP72 сигнального пути MAPKs, TLR2/4, GPCR, PI3K, CREB сигнального пути PI3K/AKT и STATs сигнального пути JAK/STAT (adj. p -value $< 0,05$). При сравнении групп пациентов с нетяжелой БА и здоровых людей без БА была выявлена гиперэкспрессия гена MYD88 сигнального пути MAPKs (adj. p -value $< 0,05$).

При утяжелении течения БА у пациентов происходила активация большего количества митогенных путей и генов, задействованных в процессах ремоделирования. PI3K/AKT и JAK/STAT сигнальные пути давали дополнительный сигнал для полного митогенного ответа, за счет активации системы TLR-рецепторов, G-protein coupled receptors и белков семейства STAT.

Эффективность ТФМ у сибсов с ожирением и СД 1 и СД 2 типа: пилотное исследование

**Е.С. Жгун¹, Е.В. Покровская², Е.И. Олехнович¹, В.А. Веселовский¹,
К.М. Климина¹, И.А.Скляник², Е.А. Шестакова², Е.Н. Ильина¹,
М.В. Шестакова²**

¹ ФГБУ ФНКЦ ФХМ ФМБА России ФГБУ, Москва

² НМИЦ эндокринологии Минздрава России, Москва

Al.androva@gmail.com

Ожирение и связанные с ним метаболические заболевания часто сопровождаются дисбактериозом кишечной микробиоты и характеризуются снижением разнообразия генов. Одним из наиболее эффективных методов коррекции микробиоты кишечника является трансплантация фекальной микробиоты (ТФМ), хорошо зарекомендовавшая себя при лечении инфекций, вызываемых *Clostridium difficile*. Применение ТФМ для коррекции метаболических расстройств перспективно, однако данные о ее использовании ограничены и имеют противоречивые результаты.

В работе двум пациентам (родным братьям) с ожирением 2 степени и различными типами сахарного диабета (СД): старший брат (44 года) с СД 2 типа, младший брат (39 лет) с СД 1 типа была проведена ТФМ.

У обоих пациентов в ходе лечения отмечено снижение массы тела (динамика — 4–5 кг за первый месяц наблюдения, далее — 1–2 кг в месяц). У пациента с СД 2 отмечено снижение выраженности инсулинорезистентности (ИР) (исходный М-индекс 2,42 мг/кг×мин, через 1 год — 3,83 мг/кг×мин), а также улучшение компенсации углеводного обмена. У пациента с СД 1 снижения ИР не и не было отмечено положительной динамики HbA1c в связи с несоблюдением рекомендаций по питанию.

Метагеномное секвенирование образцов кала, собранных до и в течение 6 месяцев после проведения ТФМ проводили на геномном анализаторе HiSeq2500 (Illumina Inc., США). При помощи биоинформатического инструмента MetaPhlAn были получены профили представленности микробных таксонов в исследуемых образцах. На уровне семейств отсутствуют значимые визуальные изменения в таксономическом профиле микробиоты на уровне микробных семейств.

Источник финансирования: НИР «Ремиссия сахарного диабета 2 типа после бариатрической хирургии: роль гормонов желудочно-кишечного тракта (инкретинов), желчных кислот и микробиоты кишечника».

Возможность использования уровня мтДНК в образцах трофэктодермы как дополнительного критерия для предсказания исхода переноса эмбриона после ПГТ-А

**Н.О. Либман, О. Ибитойе, Л.И. Нигматуллина, Н.И. Исаева,
Р.А. Биканов**

*Лаборатория First Genetics Инновационного центра Сколково, Москва
e.miticheva@f-genetics.com*

ПГТ-А (преимплантационное генетическое тестирование на хромосомные аномалии) используется в цикле вспомогательных репродуктивных технологий и позволяет выбрать для переноса эмбрионы с нормальным хромосомным набором.

На успех переноса эмбриона, рекомендованного по результатам ПГТ-А, могут влиять различные факторы: технические (как в эмбриологической, так и в генетической части), клинические — рецептивность эндометрия, наличие сопутствующих гинекологических и экстрагениальных заболеваний, а также качество полученных эмбрионов.

Несмотря на стремительное развитие ПГТ-А, перенос от 35 до 50% зуплоидных по результатам исследования эмбрионов не приводит к наступлению беременности. В связи с этим проводится поиск дополнительных параметров, которые могли бы быть использованы для предсказания успешности переноса эмбриона.

Один из таких параметров — уровень мтДНК (митохондриальной ДНК) в образцах биопсии трофэктодермы. Предполагается, что этот показатель отражает качество энергетического обмена в клетках эмбриона и его жизнеспособность. Однако данные, представленные в публикациях на сегодняшний день, достаточно противоречивы и, как правило, получены на небольших выборках.

В представленной работе проведен ретроспективный анализ уровня мтДНК в образцах биопсии трофэктодермы, полученных для проведения ПГТ-А методом NGS. Для 7720 образцов от эмбрионов, полученных в 2786 циклах ВРТ, проанализирована зависимость уровня мтДНК от возраста пациентов, показаний для проведения ПГТ, а также выявленного кариотипа эмбриона. Для 560 рекомендованных к трансферу образцов проанализирована корреляция уровня мтДНК с исходом переноса эмбриона. В результате работы показано, что использование уровня мтДНК в качестве прогностического фактора успеха эмбриотрансфера нецелесообразно.

Влияние различных факторов на разрешающую способность и качество ПГТ-А

Н.О. Либман, Л.И. Нигматуллина, Н.И. Исаева, Р.А. Биканов

*Лаборатория First Genetics Инновационного центра Сколково, Москва
n.libman@f-genetics.com*

ПГТ-А (преимплантационное генетическое тестирование на хромосомные аномалии) — скрининг эмбрионов, который используется в цикле вспомогательных репродуктивных технологий и выявляет эмбрионы с нормальным набором хромосом. ПГТ-А повышает вероятность и сокращает время до достижения успешной беременности.

В настоящее время основной технологией проведения ПГТ-А является NGS (секвенирование нового поколения). На достоверность полученного результата может влиять множество факторов: качество исходного материала, особенности пробоподготовки, а также биоинформатическая обработка сырых данных.

Определение разрешающей способности и точности метода ПГТ-А — нетривиальная задача, поскольку для образцов нет данных об «истинном кариотипе». Важный источник информации для оценки разрешающей способности ПГТ-А — эмбрионы, полученные от носителей сбалансированных хромосомных транслокаций и унаследовавшие сегментарные нарушения — последствия таких транслокаций. Использование информации о тех регионах, где ожидается появление сегментарных нарушений в связи с родительским кариотипом, позволяет оценить точность метода.

В работе было проведено ПГТ-А методом NGS для 157 образцов биопсии трофэктодермы от 44 семей с известным носительством сбалансированных хромосомных перестроек. Из них 66 образцов показали сегментарные нарушения в ожидаемых регионах. В результате работы продемонстрировано влияние различных факторов на полученный результат, таких как: качество исходного биоматериала, параметры алгоритма обработки данных секвенирования, а также количество прочтений, полученных в результате секвенирования. Разрешающая способность метода составила 5-10 Мб. В то же время, для образцов высокого качества продемонстрирована возможность обнаружения сегментарных нарушений размером от 2,5 Мб.

Разработка протокола экзомного обогащения образцов человека для секвенирования на платформе MGISEQ-2000

**В.А. Белова, А.С. Павлова, Р.Н. Афасижев, М. Коржанова,
А.А. Кривой, В.В. Черанев, Б.А. Никашин, И.А. Булушева,
Д.В. Ребриков, Д.О. Коростин**

*Центр высокоточного редактирования и генетических технологий для биомедицины, ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава РФ, Москва
verusik.belova@gmail.com*

Для платформ секвенирования DNBSeg для экзомного обогащения доступны только коммерческие наборы от MGI Tech, в то время как большинство решений разработано под платформы Illumina. Нашей целью была разработка собственного решения для экзомного обогащения пулов библиотек с использованием зондов Agilent SureSelect All Exon v6 для секвенирования на приборе MGISEQ-2000. Также мы протестировали доступный коммерческий набор от MGI Tech. В этой работе мы продемонстрировали воспроизводимость собственного протокола, называемого RSMU_exome, а также его преимущества в сравнении с набором MGIEasy Exome Capture V4.

Материалы и методы: Мы составили 2 пула из 24 библиотек геномной ДНК (в т.ч. 6 независимых для образца NA12891) — в каждом пуле по 12 (по протоколу от MGI Tech количество библиотек в пуле было уменьшено до 8). Оба пула обогащали 3 подходами — по протоколу RSMU_exome с зондами Agilent v6 и отдельно с зондами MGIEasy v4 и по стандартному протоколу MGIEasy Exome Capture V4. Секвенирование проводили в режиме PE150 на платформе MGISEQ-2000. Итого мы получили набор данных, состоящий из 64 пар fastq файлов для 64 экзотов. Для сравнения протоколов проводили нормализацию данных до 50 М ридов и биоинформатический анализ с помощью программ FastQC, Trimmomatic, BWA, Picard, BCFtools.

Результаты: Нам удалось увеличить количество библиотек ДНК в пуле перед процедурой обогащения до 12, при этом снизить количество циклов финальной ПЦР до 7. Сбалансированность получаемых данных на образец в пуле оказалась лучше для протокола RSMU_exome, чем для MGIEasy (3 образца выпало с on-target covered $\times 10 < 95\%$). После нормализации оценили качество WES библиотек NA12891 для 3 подходов. Для RSMU_exome + Agilent v6 получили значения: dup = 5,8%, off-target = 7,6%, on-target cov $\times 10 = 95,81\%$. Для RSMU_exome + MGIEasy v4: dup = 10%, off-target = 11%, on-target cov $\times 10 = 94,96\%$. Для стандартного MGIEasy Exome Capture V4: dup = 10%, off-target = 19%,

on-target cov x10 = 91,13%. При этом в случае фиксированного сета зондов MGIEasy v4, наш протокол позволяет получить на 2% больше качественных данных для вызова вариантов (англ. variant calling).

Заключение: Мы разработали протокол для экзомного обогащения с увеличенным количество образцов в пуле, совместимый с платформой MGISEQ-2000 и с зондами Agilent v6 (доступен по ссылке <https://doi.org/10.1101/2021.02.02.429394>). Также продемонстрировали, что по качественным и количественным характеристикам наш протокол RSMU_exome превосходит оригинальный протокол MGIEasy Exome Capture V4.

Межклеточная коммуникация, опосредованная компонентами сплайсосомы: роль в формировании химиорезистентности злокачественных опухолей яичника

В.О. Шендер, П.В. Шнайдер, К.С. Ануфриева, Г.П. Арапиди, О.М. Иванова, И.К. Мальянц, В.С. Бойченко, Ж.Ж. Баймуханова, К.М. Климина, М.А. Лагарькова, В.М. Говорун

*Центр высокоточного редактирования и генетических технологий для биомедицины, ФГБУ ФНКЦ ФХМ ФМБА России, Москва
shender_vika@mail.ru*

Рак яичника (РЯ) характеризуется высокой частотой рецидивов (до 80%) и пятилетней выживаемостью не более 35%. Известно, что погибающие опухолевые клетки способны секретировать молекулы, которые стимулируют пролиферацию оставшихся опухолевых клеток, что гарантирует быстрое повторное заселение опухолевой ниши сразу после первоначального ответа на лечение. Таким образом, чрезвычайно важно учитывать роль межклеточной передачи сигналов, изменяющейся во время терапии.

Мы провели всесторонний анализ изменений секретомов клеток РЯ под действием цисплатина и продемонстрировали новый молекулярный механизм приобретения опухолевыми клетками более агрессивного фенотипа. Помимо секретомов *in vitro*, мы проанализировали парные асцитные жидкости от одних и тех же пациентов с РЯ до и после неoadъювантной ХТ. Внеклеточные жидкости после ХТ были обогащены белками регуляторами сплайсинга и стимулировали значительные изменения в транскриптомных профилях реципиентных опухолевых клетках, что сопровождалось повышенной устойчивостью к цисплатину. С помощью флуоресцентных и изотопных меток была подтверждена возможность транспорта белков сплайсосомы в составе внеклеточных везикул от погибающих опухолевых клеток к реципиентным клеткам. Кроме того, искусственное повышение представленности нескольких белков сплайсосомы в опухолевых клетках способствовало приобретению устойчивости к цисплатину. Таким образом, мы предполагаем, что циркулирующие во внеклеточном пространстве факторы сплайсинга могут «подготовить» интактные опухолевые клетки к последующему лекарственному воздействию. Знания об этих механизмах могут быть полезны для разработки новых подходов комбинированной ХТ в сочетании с традиционными препаратами.

Работа поддержана Грантом 075-15-2019-1669 Министерства науки и высшего образования РФ.

Инструмент для дизайна таргетных панелей TOR

Н.С. Пильщикова, С.С. Мозгов, А.Е. Павлов

*Parseq Lab Co. Ltd, Санкт-Петербург
info@parseq.pro*

Таргетное обогащение методом мультиплексной ПЦР позволяет сфокусировать потенциал секвенирования на целевых участках генома и обеспечить специфическую амплификацию при недостаточном количестве матрицы, поэтому является эффективным инструментом для клинических и исследовательских NGS приложений.

Для создания таргетной панели требуется инструмент дизайна ПЦР-праймеров, подходящих для совместной специфической амплификации большого числа участков генома. Кроме того, к решению предъявляются следующие требования:

- (1) возможность работы с разнообразными протоколами пробоподготовки и платформами для секвенирования;
- (2) использование для дизайна праймеров любых базовых/референсных последовательностей ДНК;
- (3) настраиваемые параметры дизайна под разные условия и приложения (нормальная или деградированная ДНК, протяжённые целевые регионы или хотспоты, fusion ампликоны для детекции CNV и RNA fusions, и т.д.).

В свободном доступе отсутствуют полнофункциональные реализации подобных инструментов, поэтому нами был разработан и валидирован собственный инструмент для дизайна таргетных панелей, который использовали при разработке трёх наборов: для секвенирования кодирующей последовательности гена CFTR человека, полной последовательности РНК-вируса SARS-CoV-2, около 500 полиморфных позиций в геноме винограда.

Размер панелей составил от 58 до более чем 500 ампликонов. Было реализовано 2 различных протокола таргетной пробоподготовки. По результатам проведенной валидации *in vitro* получена равномерность покрытия 96 — 98% с долей on target 89 — 99%.

Показано, что разработанный инструмент TOR является перспективным подходом решения широкого спектра задач таргетного обогащения ДНК для проведения высокопроизводительного секвенирования генома в научных и клинических целях.

Kromsatel: программный инструмент для исправления химерных прочтений ампликонов

М.А. Сиколенко, Л.Н. Валентович

*Институт микробиологии НАН, Минск, Беларусь
sikolenko@bio.bsu.by*

Kromsatel — программный инструмент для исправления химерных прочтений ампликонов.

Одним из способов приготовления библиотек ДНК для определения их последовательности является полимеразная цепная реакция со специфическими праймерами. Например, получение ампликонов применяется для приготовления библиотек ДНК на основе РНК матриц, что в данный момент активно используется для секвенирования РНК вирусов, в частности коронавируса SARS-CoV-2 (ARTIC network nCoV-2019 sequencing protocol — <https://artic.network/ncov-2019>).

С помощью ПЦР можно получить подходящего размера фрагменты ДНК, которые интересуют исследователя (прицельное секвенирование). После амплификации, продукты ПЦР очищаются от реакционной смеси, далее к ним добавляют адаптерные, а иногда и индексные последовательности, либо с помощью ДНК-лигазы, либо с помощью ещё одного раунда ПЦР с внешними праймерами. К сожалению, «сырые» данные, получаемые с секвенатора при описанном способе пробоподготовки, могут содержать ненужную информацию (последовательности праймеров, которые использовались для приготовления библиотеки ДНК; контаминантные, контрольные, адаптерные и прочие последовательности) или комбинацию фрагментов ДНК (химер), которая образовалась случайным образом при амплификации ДНК или на стадии лигирования адаптеров.

Как правило, химерные прочтения подлежат удалению из обрабатываемого набора данных. Однако если известна референсная последовательность секвенируемого генома, а также если известна схема отжига праймеров, использованных для амплификации, то химерные прочтения могут быть разделены на фрагменты, каждый из которых содержит последовательность лишь одного секвенируемого ампликона. Такого рода предобработка может значительно повысить качество обрабатываемого набора данных без потери содержащейся в нём информации.

Именно такая процедура была реализована нами в рамках компьютерной программы под названием kromsatel. Программа разработана на языке программирования Python и является кроссплатформенной.

Сама программа, а также её исходный код находятся в открытом доступе по адресу <https://github.com/masikol/kromsatel>.

Идея алгоритма предобработки, реализованного в программе kromsatel, состоит в следующем:

1. На основе набора праймеров, использующихся для амплификации библиотеки ДНК, последовательность референсного генома разбивается на ампликоны в соответствии со схемой отжига праймеров. При этом, помимо последовательностей основных ампликонов, генерируются также и последовательности т. н. минорных ампликонов: ПЦР-продуктов, соответствующих участкам перекрытия между основными ампликонами.

2. Подлежащий предобработке набор прочтений выравнивается на последовательности ампликонов (для этого используется локальное выравнивание). На основании полученных выравниваний прочтения разделяются на фрагменты, которые в пределах прочтения не перекрываются. В результате каждый из полученных фрагментов соответствует лишь одному ампликону.

Таким образом, программа kromsatel помимо разделения химерных прочтений на корректные фрагменты, решает задачу удаления последовательностей адаптеров и (опционально) праймеров, использованных для амплификации. Кроме того, прочтения, попавшие в обрабатываемый набор данных в результате ошибочного считывания индексных последовательностей (т. е. принадлежащие другим образцам), не будучи выровненными на референсный геном, также будут удалены.

Данная программа успешно опробована в институте микробиологии НАН Беларуси для предобработки прочтений ампликонов, получаемых с помощью протокола ARTIC для секвенирования генома вируса SARS CoV 2 на приборах MiSeq (Illumina) и MinION (Oxford Nanopore Technologies).

Полногеномный анализ *Neisseria gonorrhoeae*: установление маркеров адаптивного преимущества и синтонической структуры

И.Д. Кандинов

*Центр высокоточного редактирования и генетических технологий для биомедицины, Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва
ilya9622@gmail.com*

Наиболее распространенными в РФ линиями NG-MAST в 2013-2019 гг. являются G807, G1993, G9476. Для выявления причин, влияющих на формирование современной популяции *N. gonorrhoeae* в России, и изучения маркеров адаптивного преимущества доминирующих генетических линий проведено полногеномное секвенирование 45 изолятов *N. gonorrhoeae* указанных геногрупп на двух платформах: MiniION и MiniSeq. После сборки геномов de novo сравнивали геномы изолятов типа 807 и типа 1407 — потенциально опасного с точки зрения антибиотикорезистентности. Выявлено 144 характерных аллеля для сиквенс-типа 807 и 98 аллелей для типа 1407. Определены белки, кодируемые данными генами, установлена принадлежность таких белков к группам, характеризующим их функцию. Наибольшее количество характерных для NG-MAST 807 белков относится к группе фагов, инсерционных элементов, пилей, адгезинов, белков мутности, факторов патогенности, токсинов, группе белков, отвечающих за взаимодействие с нуклеиновыми кислотами, репликацию, транскрипцию, трансляцию, белков, в метаболизме и транспорте через мембрану. Определено расположение ортологов в геномах изолятов G807. Проведен детальный анализ геномных перестроек у G807 и определено расположение мобильных элементов, включающих инсерционные элементы IS, фаги (NgoФ1 – NgoФ9), элементы Корреи, гонококковый остров, элементы Спенсера-Смита. Выявлено отсутствие прямой связи между геномными перестройками и филогенетической близостью штаммов, поскольку как для филогенетически близких изолятов G807, так и для филогенетически далеких референсных геномов WHO характерно наличие значительного количества инверсий и геномных транслокаций.

Работа выполнена при поддержке гранта РФФИ 17-75-20039 и Соглашения с Министерством науки и высшего образования РФ № 075-15-2019-1660.

Полная последовательность генома *Pseudomonas syringae* B1M B-268

Р.М. Бузиков¹, Т.А. Пилипчук², Л.Н. Валентович², Э.И. Коломиец²,
А.М. Шадрин¹

¹ ФИЦ ПНЦБИ РАН, Пущино;

² Институт Микробиологии НАН, Минск, Беларусь
andrey2010s@gmail.com

Pseudomonas syringae B1M B-268, штамм используемый для тестирования *in vitro* эффективности бактериофагов — основы биопестицида «Мультифаг®», производимого в Беларуси. Для секвенирования генома *P. syringae* B1M B-268 использовали гибридную сборку на основе данных Oxford Nanopore и Illumina. Всего было получено 2 054 402 парноконцевых прочтений Illumina MiSeq длиной 251 нт и 46 590 прочтений Oxford Nanopore со средней длиной 2 137 нт (максимальная длина чтения 122 273 нт). Далее, прочтения, полученные с помощью Oxford Nanopore, были собраны с помощью программы Flye в единичный контиг (покрытие 14x). Затем данные Illumina (покрытие 120x) были использованы для уточнения сборки Flye.

Геном *P. syringae* B1M B-268 (CP068034.2) состоит из единственной кольцевой хромосомы длиной 6 018 586 п.н. (содержание GC 59%). Показатель средней нуклеотидной идентичности (ANI) между типовым штаммом *P. syringae* KCTC 12500T и B1M B-268 составил 95,5%, что подтверждает видовую принадлежность данного штамма. В ходе аннотации генома было идентифицировано 5 165 генов, включая 5 004 белок-кодирующих последовательности, исключая псевдогены, 64 гена тРНК, пять кластеров генов 16S, 23S, 5S рРНК, причем один из них содержит две копии гена 5S рРНК. Каждый кластер генов рРНК содержал тандем генов тРНК изолейцина и аланина между генами 16S и 23S рРНК. Поиск профагов с помощью PHASTER выявил два неполных и один интактный фаг длиной 16,7, 29,7 и 17,3 т.п.н. соответственно. Также в геноме штамма *P. syringae* B1M B-268 были выявлены гены, кодирующие белки нуклеации льда, биосинтез сиринопептина, системы секреции типа III и VI.

Исследование выполнено при финансовой поддержке РФФИ и БРФФИ в рамках научного проекта № 20-516-00015 и № Б20Р-078.

Характеристика изменений, выявленных в геноме *Aspergillus terreus* № 43–16, высокоактивном продуценте ловастатина

А.А. Жгун¹, Г.К. Нураева¹, В.И. Кукушкина^{1,2}

¹ Институт биоинженерии, ФИЦ Биотехнологии РАН, Москва;

² РХТУ им Д.И. Менделеева, Москва
zzhgun@mail.ru

Мицелиальный гриб *Aspergillus terreus* — основной промышленный продуцент холестерина-понижающего препарата ловастатина (ЛОВ). В настоящее время основным методом получения высокоактивных продуцентов вторичных метаболитов в грибах является многораундовый случайный мутагенез и селекция по целевому признаку. В результате такой процедуры у штамма *A. terreus* № 43-16 (*A. terreus* HY) продукцию ЛОВ увеличили в 200–300 раз. Для изучения молекулярных основ, приведших к суперпродукции ЛОВ, в данной работе провели секвенирование этого штамма на платформе Illumina HiSeq 2500 («Illumina» США). Покрывание генома составило 150 раз. Полученные данные сравнивали с ранее аннотированными последовательностями для генома штамма дикого типа *A. terreus* NIH2624 (GenBank: ASM14961v1, RefSeq: NZ_AAJN000000000.1). Выявили 2871 различий относящихся к группе HIGH (сдвиг рамки считывания, нонсенс мутации, нарушения сплайсинга, потери стартового или стоп кодонов). Основные изменения оказались связанными с: i) биосинтезом ацетил-КоА (строительном блоком при биосинтезе ЛОВ) и ацетилированием; ii) биосинтезом холестерина (нарушения в генах цитохрома P450 51 и дельта(14)-стерол редуктазы); iii) глобальной регуляцией (делеции в генах *rasC*, регулятора *pH*, *creA*, регулятора углерода и в гене негативного регулятора митоза); iv) биосинтезом клеточной стенки (нарушение в гене хитин-синтазы); v) биосинтезом полиаминов (делеция в гене аргиназы); vi) разгрузкой потребления S-аденозилметионина, необходимого для биосинтеза ЛОВ (делеция в гене дифтинсинтазы). Полученные данные позволили всесторонне охарактеризовать высокоактивный продуцент ЛОВ и объяснить ряд экспериментальных феноменов, например, потерю способности трансформировать андростендион.

Работа частично поддержана грантом РФФИ 19-04-01173.

Сравнительный анализ геномов штаммов *Fusarium oxysporum* f. sp. *lini*, обладающих различной вирулентностью

Е.М. Дворянинова^{1,2}, Р.О. Новаковский¹, Е.Н. Пушкова¹,
Т.А. Рожмина³, Л.П. Кудрявцева³, Н.В. Мельникова¹, А.А. Дмитриев¹

¹ Институт молекулярной биологии им. В.А. Энгельгардта РАН,
Москва; ² Московский физико-технический институт, Долгопрудный;

³ Федеральный научный центр лубяных культур, Торжок
dvorianinova.em@phystech.edu

Fusarium oxysporum f. sp. *lini* (Fol) является одной из патогенных форм гриба *Fusarium oxysporum*, характеризующегося значительной геномной вариабельностью. Fol вызывает фузариозное увядание льна и насчитывает большое количество штаммов различной вирулентности. На сегодняшний день полногеномные исследования проведены только для единичных высоковирулентных штаммов Fol. В ходе данной работы на платформах Oxford Nanopore Technologies и Illumina секвенировано 6 штаммов Fol низкой (456, 482), средней (476, 525) и высокой (39, 483) вирулентности. С использованием гибридного сборщика MaSuRCA получены геномы секвенированных штаммов на основе прочтений с двух платформ. Длины сборок 6 штаммов Fol варьировали от 53,4 до 69,3 млн.п.н., их N50 — от 1,0 до 2,5 млн.п.н., а их полнота — от 96,5 до 99,7% согласно BUSCO. Также получены полные митохондриальные геномы, которые обладали высокой гомологией друг к другу, но варьировали в длине — ~39 т.п.н. для штаммов 39 и 525, ~46 т.п.н. для штаммов 456, 476, 482 и 483. В сборках 6 штаммов Fol проведен поиск генов вирулентности SIX (secreted in xylem) с помощью выравнивания частичных копий SIX1, SIX7, SIX10, SIX12, SIX13, соответствующих forma specialis *lini*, на полученные сборки. Обнаружена кластеризация SIX7, SIX12, а также SIX7, SIX10, SIX12, SIX13 в сборках штаммов 39, 456, 476 и 483. Геномные сборки штаммов 39, 456, 476, 483 и 525 содержали от 6 до 12 последовательностей гомологичных SIX, а в сборке штамма 482 их обнаружено не было. Полученные данные способствуют накоплению знаний об организации геномов штаммов Fol и необходимы для развития методов борьбы с фузариозным увяданием льна.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 19-34-90055.

Профилирование микробиома для диагностики биопоражения на портрете Е.С. Мациевой работы К. Сомова из Государственной Третьяковской галереи

М.П. Потапов^{1,2}, Д.А. Авданина¹, К.М. Климкина³,
В.А. Веселовский³, Д.Е. Федоров³, Н.П. Симоненко⁴, Е.С. Жгун³,
К.В. Шумихин⁵, А.А. Жгун¹

¹ Институт биоинженерии, ФИЦ Биотехнологии РАН, Москва; ² ПуцГЕНИ, Пущино; ³ ФГБУ ФНКЦ физико-химической медицины ФМБА России, Москва,

⁴ ИОХН РАН, Москва; ⁵ Государственная Третьяковская галерея, Москва
zzhgun@mail.ru

Современные методы, используемые при реставрации произведений масляной живописи на холсте, позволяют достаточно результативно бороться с микробиологическим поражением. Однако эффективность проведения как превентивных, так и экстренных обработок может значительно повыситься благодаря знанию о микробиоме экспонатов того или иного музея и характерных изменениях в его составе при биопоражении. В 2018 в верхней зоне на портрете Е.С. Мациевой работы Константина Сомова (1933 г., холст/масло, Государственная Третьяковская галерея) обнаружили следы предполагаемого микробиологического поражения. Методом ИК-Фурье спектроскопии в этой области выявили появление нескольких характерных для микроорганизмов пиков, соответствующих зонам поглощения амид I, II и полисахаридов. Методом SEM показали отличия в структуре образцов из опытной и контрольных зон. Для окончательной диагностики из зоны предполагаемого биопоражения и контрольной области отобрали пробы, аликвоты выселили на серию стандартных микробиологических сред и провели метагеномное секвенирование исходных проб и полученных после посева культур. При этом на платформе MISeq Illumina определили последовательности после амплификации с метагеномных ДНК фрагментов рДНК бактерий (V3/V4) и грибов (ITS2). Полученные данные (номера доступа GenBank: SRX7729103–SRX7729110, SRX7729112–SRX7729116, SRX7729167–SRX7729171) позволили выявить сдвиг микробиома в проблемной области, в частности, определить значительное увеличение представителей *Saccharomyces*, *Tremellomycetes* и *Agaricomycetes*. Данная работа показывает, что профилирование микробиома картин может служить одним из инструментов для превентивной диагностики и оценки рисков биопоражения в случае локальных отклонений от температурно-влажностного режима, соблюдаемых в музеях мира.

16S типирование микроорганизмов на платформе MGI-2000 в исследовании плоскоклеточного интраэпителиального поражения цервикального канала

И.А. Булушева, А.О. Шмитко, В.Н. Москаленко, Д.О. Коростин, А.А. Кривой

*Центр высокоточного редактирования и генетических технологий для биомедицины, ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России, Москва
bulusheva@phystech.edu*

Работа проводится в рамках проекта исследования плоскоклеточного интраэпителиального поражения (SIL) цервикального канала женской репродуктивной системы. Это состояние характеризуется высокой вероятностью прогрессии до стадии онкологии. Одной из задач исследования является определение бактериального состава слизистой оболочки цервикального канала. Для этого производится выскабливание эпителия, выделение ДНК из находящихся в образце клеток человека и клеток прокариотических организмов. Типирование микроорганизмов опирается на определение последовательностей V4-региона 16s-гена рибосомальной РНК прокариот методом высокопроизводительного секвенирования (NGS) на инструменте MGISEQ-2000. Приготовление библиотек производится ПЦП с общеиспользуемыми праймерами 515F и 806R, содержащими последовательности адаптеров для платформы MGI.

Нашей группой был создан единый биоинформатический пайплайн, включающий в себя: фильтрацию сырых данных по последовательностям праймеров; удаление последовательностей праймеров; фильтрацию по качеству данных секвенирования, кластеризацию данных — определение консенсусных последовательностей V4 региона; фильтрацию химерных фрагментов амликона; фильтрацию по длине консенсус последовательностей на основе ожидаемой длины ампликона V4 региона 16s гена; определение таксономической принадлежности вплоть до уровня род при сравнении консенсус последовательностей с базой данных RDP одноименным ПО. Также нами было внедрено дополнительное таксономическое определение базой данных rRNA_typestrains/16S_ribosomal_RNA алгоритмом BLAST, для вторичной проверки результатов RDP-типирования, а также для определения видовой принадлежности организмов, где это возможно (более 20%).

Для первых образцов слизистой оболочки цервикального канала были успешно определены таксономические составы микроорганизмов

вплоть до их представленности 0.1% от всех последовательностей V4 16s. Были выделены образцы, принадлежащие к типу I здоровой микрофлоры ЖРС с долей бактерий рода *Lactobacillus* более 80%. Были выделены образцы с более высоким таксономическим разнообразием и большой представленностью условно-патогенной микрофлоры, напр. бактерий рода *Gardnerella*. Был определен предел работы метода с точки зрения кол-ва копий 16s генов, входящих в исследуемый образец, который составил 150 000 — 200 000. Было показано, что в случае меньшего кол-ва копий 16s необходимо дополнительное сравнение образца с отрицательным контролем. Дополнительная валидация метода была произведена на образцах из слюны. В настоящий момент пайплайн служит для клинического типирования образцов в вышеописанном проекте.

Геномный каскад промоторов бактериофага T7: участие вызванной суперспиральностью дестабилизации дуплекса днк (SIDD)

М.А. Орлов

*Институт биофизики клетки РАН, Пущино
orlovmikhailanat@gmail.com*

В данной работе рассмотрено значение вызванной суперспиральностью дестабилизации дуплекса ДНК (Stress-induced Duplex Destabilization, SIDD) промоторов бактериофага T7 в регуляции экспрессии его геномного каскада. Ранее для этого физического параметра ДНК показано участие в функционировании ряда областей геномной регуляции, включая промоторы T7. Действительно, исключительно короткие (~20 нуклеотидов) последовательности нативных промоторов T7 не могут объяснить высокоспецифичное настраиваемое (изменяется со сменой фаз жизненного цикла) связывание этих промоторов T7-РНК-полимеразой. Они слишком близки по последовательности — в ряде случаев и вовсе совпадающей. Естественно предполагать дополнительное участие другого типа кодирования такого молекулярного узнавания — физическими свойствами ДНК. Такие свойства напрямую определяют взаимодействия биомолекул и косвенно кодируются первичной структурой ДНК (что затрудняет их однозначное отнесение к фенотипу либо генотипу). Здесь мы применили расчет с последовательным изменением параметров (температура, ионная сила и суперспиральной плотности) профиля SIDD для всего генома бактериофага T7. Показано, что промоторные области более ранних областей геномного каскада более легкоплавки, что может служить для регуляции экспрессии генов и дифференциального узнавания разных промоторов.

Экспрессия генов льна, вовлеченных в биосинтез жирных кислот

Л.В. Повхова^{1,2}, Р.О. Новаковский¹, Т.А. Рожмина^{1,3}, Е.Н. Пушкова¹,
Г.С. Краснов¹, А.А. Дмитриев¹, Н.В. Мельникова¹

¹ Институт молекулярной биологии им. В.А. Энгельгардта РАН, Москва;

² Московский физико-технический институт, Долгопрудный;

³ Федеральный научный центр лубяных культур, Торжок
povhova.lv@phystech.edu

Семена льна (*Linum usitatissimum* L.) богаты ненасыщенными жирными кислотами, включая линоленовую (LIN), линолеовую (LIO) и олеиновую (OLE), роль которых в профилактике сердечно-сосудистых, онкологических и других заболеваний широко изучена. Льняное масло используется в фармацевтике, при производстве пищевых добавок, кормов для животных, композитных и полимерных материалов. Сфера использования льняных семян определяется в первую очередь жирнокислотным составом масла. Сорты льна значительно различаются по составу масла, главным образом содержанием LIN, LIO и OLE. Мы изучили роль экспрессии генов, вовлеченных в биосинтез жирных кислот, в определении жирнокислотного состава льняного масла. Для семи сортов льна с различным содержанием LIN, LIO и OLE выделена РНК из коробочек на стадии жёлто-зелёной спелости, проведено секвенирование транскриптомов на платформе Illumina, и для каждого генотипа получено в среднем 10 тысяч прочтений длиной 81 нуклеотид. Для оценки экспрессии генов прочтения картировали на референсный геном *L. usitatissimum* (GenBank: GCA_000224295.2) с использованием STAR, затем с помощью BEDTools определили количество выравненных последовательностей и провели анализ экспрессии с применением пакета edgeR. Обнаружили различия в экспрессии генов, вовлеченных в биосинтез жирных кислот, включая FAD2A, FAD2B, FAD3A и FAD3B, для сортов с различным содержанием LIN, LIO и OLE в масле. В некоторых сортах с низким содержанием LIN была снижена экспрессия генов FAD, однако это наблюдалось не для всех низколиноленовых генотипов. Полученные результаты необходимы для установления роли экспрессии генов в определении жирнокислотного состава льняного масла.

Работа выполнена при финансовой поддержке гранта РФФ № 21-16-00111.

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК

Научное издание

**«Геномное секвенирование и редактирование»
СБОРНИК ТЕЗИСОВ
9-й Всероссийской научно-практической
конференции центров геномных исследований
мирового уровня**

Подписано в печать 17.05.2021
Формат 60×90 1/16. Усл. печ. л. 1,5. Тираж 200 экз. Заказ № 26-21.

Отпечатано ФГАОУ ВО РНИМУ им. Н.И. Пирогова
Минздрава России, 117997, Москва, ул. Островитянова, 1.
www.rsmu.ru

ISBN 978-5-88458-543-0



9 785884 585430 >