На правах рукописи

ПОПЕНКО Анна Сергеевна

БИОИНФОРМАЦИОННОЕ ИССЛЕДОВАНИЕ ТАКСОНОМИЧЕСКОГО СОСТАВА МИКРОБИОТЫ КИШЕЧНИКА ЧЕЛОВЕКА

03.01.09 – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва – 2014

Работа выполнена в Федеральном государственном бюджетном учреждении науки «Научно-исследовательский институт физико-химической медицины Федерального медико-биологического агентства»

Научный руководитель: доктор биологических наук, профессор, член-корреспондент РАН Говорун Вадим Маркович

Официальные оппоненты: Гельфанд Михаил Сергеевич
доктор биологических наук, профессор, ФГБУН Института проблем передачи информации им. А.А. Харкевича РАН, заместитель директора по научным вопросам
Даниленко Валерий Николаевич
доктор биологических наук, профессор, ФГБУН «Институт общей генетики им. Н.И. Вавилова» РАН, заведующий отделом генетических основ биотехнологии, заведующий лабораторией генетики микроорганизмов

Ведущая организация: Федеральное государственное бюджетное учреждение науки «Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова» РАН

Защита состоится «26» марта 2015 г. в 13 часов на заседании диссертационного совета Д 001.010.01 при Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» по адресу: 119121, Москва, ул. Погодинская, д. 10, стр. 8.

С диссертацией можно ознакомиться в библиотеке и на сайте ИБМХ http//www.ibmc.msk.ru

Автореферат разослан_____2015 г.

Ученый секретарь диссертационного совета,
кандидат химических наук  Карпова Е.А.

ВВЕДЕНИЕ

Актуальность проблемы.

Самыми многочисленными распространенными обитателями нашей планеты являются бактерии. Бактерии, обитающие в одной экологической нише, образуют сложную систему межвидового обобщенного метаболизма, наиболее эффективно используют имеющиеся ресурсы, как то запасы питательных веществ, кислород, свет. Такие бактериальные конгломераты именуются микробиотой. Термин микробиота или микробиом впервые был введен Джошуа Ледербергом и формально определяется как совокупность микроорганизмов, их генетического материала и взаимоотношений внутри экологической ниши (Lederberg J., McCray A., 1995). Изучение микробных сообществ имеет фундаментальное значение: исследования общих и частных взаимосвязей внутри микробиоты, способов поддержания гомеостаза, механизмов ответа на раздражители внешней среды значительно расширят наши познания в области экологии и молекулярной биологии микробных сообществ. Микробиота человека интересна и с медицинской точки зрения. Наиболее многочисленной и разнообразной является микробиота кишечника человека. Так за последние 10 лет выпущено свыше 1500 статей по метагеномике человека, половина из которых – только за 2011 - 2013 года.

Несомненно, микробиом человека, и микробиом кишечника в особенности, непосредственно влияет на организм хозяина. Бактерии кишечника способны переваривать сложные углеводы и другой субстрат, неусвояемый человеком, при этом производя витамины (Jean G. L. et al., 2013), короткоцепочечные жирные кислоты (КЖК) (Gibson Glenn R., 2004). Достоверно неизвестно, какая доля из потребляемой пищи переваривается бактериями в кишечнике человека, однако гнотобиотическим грызунам – животным, лишенным микробиоты, приходится потреблять на 30% больше пищи для сохранения массы тела (Sears Cynthia L. A., 2005).

Первые исследования в большей степени носили эпидемиологический характер. Ученые пытались выявить, что есть эталон здорового микробиома, как он варьирует в зависимости от географического и социального факторов (Yatsunenko T. et al., 2012; Arumugam M. et al., 2011). Помимо общих популяционных исследований, изучается связь между составом микробиома кишечника человека и различными заболеваниями, такими как рак (Balamurugan R. et al., 2008), атеросклероз (Karlsson F.H. et al., 2012),

диабет второго типа(Wu X. et al., 2010). Исследования в данной области носят больше описательный характер, а механизмы взаимодействия микробиома и организма человека все еще слабо изучены.

Несмотря на довольно большое количество метагеномов кишечника собранных из разных регионов Земли, все еще не был сделан метагеном жителей Российской Федерации. Россия представляет собой уникальный пример совокупности совершенно разных географических, этнографических и социальных факторов. Популяционное исследование российских метагеномных образцов позволило расширить представление о существующих микробных сообществах.

Данная работа посвящена разработке оптимального алгоритма анализа метагеномных данных и исследованию возможностей применения метагеномики в прикладных медицинских задачах.

Цель исследования.

1. Разработать алгоритм анализа метагеномных данных для прикладных медицинских и научных исследований.

Задачи исследования.

1. Создать вычислительную инфраструктуру анализа экспериментов, от экспериментальных данных до определения таксономического состава.
2. Разработать алгоритм статистического анализа метагеномных данных.
3. Выполнить сравнение геномного состава кишечных микробиот людей, проживающих на территории РФ, включая городские и сельские зоны, а также в других странах (Дания, США, Китай, Венесуэла, Малави).
4. Изучить метагеномы клинических образцов с использованием созданного алгоритма анализа и сравнить их с образцами кишечного метагенома здоровых людей.
5. Связать данные метагеномного анализа с рядом биохимических параметров, измеренных у доноров на примере зависимости концентрации короткоцепочечных жирных кислот от таксономического состава микробиоты кишечника.

Личный вклад автора.

1. Разработка программного комплекса по обработке первичных данных метагеномного анализа.

2. Создание архитектуры реляционной базы данных для хранения и частичной обработки метагеномных данных.
3. Автоматизация процесса обработки метагеномных данных.
4. Составление таксономического референсного набора последовательностей для оптимизации процесса идентификации метагеномного состава.
5. Составление оптимальной комбинации статистических методов обсчета метагеномных данных, включая модификацию существующей метрики UniFrac.
6. Создана модель влияния состава микробиоты кишечника на метаболизм короткоцепочечных жирных кислот.

Научная новизна исследования.

1. Впервые был изучен состав метагенома кишечника на примере 96 образцов для жителей различных субъектов Российской Федерации.
2. Был разработан алгоритм обработки метагеномных данных, получаемых с прибора ABI SOLiD4, отличающихся малой длиной прочтений.
3. Был произведен поиск возможных связей между составами метагеномов кишечника здоровых людей и пациентов с различными патологиями (онкологические больные, проходящие химиотерапию, пациенты страдающие от алкоголизма и работники производств с повышенным радиационным фоном).
4. Была создана кинетическая модель метаболизма короткоцепочечных жирных кислот (КЖК), позволяющая прогнозировать их концентрации в зависимости от состава микробиоты на основе данных секвенирования 16S рРНК генов.

Практическая значимость.

В ходе исследования был разработан универсальный и полуавтоматизированный алгоритм обработки метагеномных данных, получаемых с использованием любой платформы высокопроизводительного секвенирования. Этот алгоритм успешно применяется в текущих медицинских проектах по изучению таксономических изменений состава микробиоты при различных патологических состояниях. Полное описание состава микробиома когорты здоровых жителей Российской Федерации позволило расширить знания о нормофлоре человека, а так же предоставило контрольный набор данных для последующих исследований микробиоты в России. Дополнительно, была создана предварительная модель метаболизма КЖК, которая в

будущем может быть использована для предсказания концентраций КЖК в зависимости от состава микробиоты кишечника человека.

Апробация работы.

Результаты данного исследования были представлены на научных конференциях, в частности на 5 российских (МССМВ-2011, 54ая научная конференция МФТИ, BGRS/SB, «Постгеномные методы анализа в биологии, лабораторной и клинической медицине», «Инновационные технологии в медицине XXI века») и 2 европейских (IHMC, Париж 2011; ECCB'12 Базель 2012).

Публикации.

По теме диссертации опубликовано 9 работ, из которых 5 статей в рецензируемых научных изданиях и 4 публикации в материалах российских и международных конференций.

Структура и объем диссертации.

Диссертационная работа состоит из 4 глав, выводов и списка литературы, содержащего 149 ссылок, и четырех приложений. Работа изложена на 140 страницах, содержит 19 рисунков, 9 таблиц и 4 приложения.

МЕТОДЫ

Экспериментальные методы

Забор образцов кала

Образец кала забирали у здоровых людей в возрасте 36±18 лет на основе информированного согласия. За три месяца до первого забора и в течение последующего полугода эти люди не принимали антибиотиков и нестероидных противовоспалительных средств. Забор кала осуществляли в индивидуальный пластиковый контейнер, избегая попадания в образцы мочи и туалетной бумаги. Образец весом 10-20 г подвергали немедленной заморозке и хранили при -20°C, либо использовали для выделения ДНК непосредственно после забора. Забор образцов производился медицинским персоналом. Дальнейшая экспериментальная часть проводилась сотрудниками геномного центра НИИ ФХМ.

Выделение ДНК из кала

К замороженной навеске образца кала (150 мг) добавляли кремниево-циркониевые бусины (BioSpec Products, США) диаметром 0,1 мм (300мг) и 0,5 мм (100

мг), а затем 1200 мкл теплого лизирующего буфера (50мМ Tris-HCl, pH 8,0, 500 мМ NaCl, 50 мМ EDTA, 4% SDS), перемешивали на вортексе до однородного состояния и гомогенизировали с помощью MiniBeadBeater (BioSpec Products, США) в течение 3 мин. Полученный лизат инкубировали при 70°С в течение 15 мин, после чего образцы центрифугировали 20 мин при 22000 об./мин. Надосадочную жидкость отбирали в новые пробирки объемом 2 мл и ставили в лёд (4°С). К осадку повторно добавляли 1200 мкл лизирующего буфера и повторяли процесс гомогенизации. Надосадочные жидкости объединяли, добавляли 2 объема 96% этанола и 1/10 объема 3М ацетата натрия. Инкубировали при -20°С не менее часа. После этого образцы центрифугировали при 14000 об/мин 20 мин. Сформировавшийся осадок дважды промывали 1000 мкл 80% этанола, сушили на воздухе и растворяли в деионизованной воде. Осадок был ресуспензирован и растворен в 400 мкл лизирующего буфера. После дополнительного центрифугирования в течение 15 мин при 22000 об/мин, надосадочная жидкость была отобрана в новую пробирку объемом 2 мл и после добавления 1 мкл раствора РНКазы А (5 мг/мл) инкубирована при 37°С в течение часа. Качество полученной ДНК оценивали путем электрофореза 5 мкл очищенной ДНК на 1% агарозном геле.

Концентрацию ДНК в растворе определяли с помощью флуориметра Qubit® (Invitrogen, США) с использованием наборов Quant-iT™ dsDNA Broad-Range Assay Kit и Quant-iT™ dsDNA High Sensitivity Assay Kit (Invitrogen, США), согласно рекомендациям производителя.

Методы секвенирования

Подготовку shotgun-библиотек и их секвенирование с использованием генетических анализаторов SOLiD 4 (Life Technology, США), Ion Torrent PGM (Life Technology, США), GS FLX+ (Roche, США), HiSeq 2000 (Illumina), SOLiD 5500 W (Life Technologies, США) осуществляли согласно рекомендациям производителя (подробнее в диссертационной работе).

Дополнительные метагеномные данные

Для сравнительного анализа образцов российской кишечной микробиоты были использованы общедоступные метагеномные данные: 85 наборов ридов из Дании (Qin et al., 2010), США (Nelson et al., 2010), Венесуэлы, Малави (Yatsunenko et al., 2012) и Китая (Qin et al., 2012).

Биоинформационные методы

Предобработка ридов

Полученные в ходе секвенирования риды в цветовом формате прошли стандартную фильтрацию по качеству. Риды со средним значением баллов качества QV<15 были исключены из анализа. С целью минимизации ошибок секвенирования, оставшиеся риды прошли фильтрацию программой SAET (SOLiD Accuracy Enhancement Tool). Далее риды подверглись редактированию по качеству: все позиции начиная с 5' были удалены вплоть до первой высококачественной позиции (QV >= 30). Все риды, чья длина после фильтраций стала меньше 30 нуклеотидов, были удалены. Высококачественные риды были картированы на геном человека версии hg18 с использованием программы bowtie, те риды, которые не были картированы, были подвергнуты дальнейшему анализу.

Определение таксономического состава

Таксономический состав метагеномных образцов определяли в результате картирования нуклеотидных прочтений на неизбыточный референсный каталог из репрезентативных геномов микроорганизмов, встречающихся в кишечнике человека, с использованием программного пакета bowtie. В качестве источников для составления каталога геномов использовались база проекта Human Microbiome Project (http://www.hmpdacc.org) и NCBI (http://www.ncbi.nlm.nih.gov), а также другие общедоступные источники. Геномы были выравнены друг против друга программой MUMMER 3.0. Те геномы, которые не были похожи ни на один другой более чем на 80%, были включены в каталог. Оставшиеся геномы были кластеризованы по порогу сходства 80% на 80% длины. Из каждого кластера схожих геномов в ручном режиме было выбрано по одной репрезентативной последовательности, которые затем также были добавлены в каталог. Всего в каталог вошли 353 генома, как полностью собранных, так и в виде набора контигов.

В ходе выравнивания набора ридов на референс была получена информация из файлов формата BAM о двух видах покрытия с использованием программного пакета BEDtools: суммарная длина ридов, картировавшихся на референс (покрытие в глубину) и суммарная длина позиций референса, оказавшейся покрытой (покрытие в ширину). Покрытие в ширину было использовано, как порог детектирования референса в метагеноме: должно быть покрыто ридами по крайней мере 1% длины генома.

Покрытие каждого генома было нормализовано на общую длину картировавшихся ридов и его длину (1) для получения относительной представленности.

$$\text{Относительная представленность генома} = 1E12\times\left(\frac{\sum \text{длин картировавшихся ридов}/\text{длина генома}}{\text{общая длина картировавшихся ридов образца}}\right) \quad (1)$$

Относительная представленность бактериальных и архивных родов суммировалась по формуле 2.

$$\text{Относительная представленность рода} = 1E12\times\left(\frac{\sum_{\text{по геномам}}\left(\sum \text{длин картировавшихся ридов}/\text{длина генома}\right)}{\text{общая длина картировавшихся ридов образца}}\right) \quad (2)$$

Принадлежность каждого вида к тому или иному роду определялась с использованием классификации по базе данных последовательностей 16S рРНК генов RDP.

Реализация алгоритма поточной обработки данных

Алгоритм поточной обработки метагеномных данных представлен на Рисунке 1. Он реализован в виде конвейера, получающим на вход наборы ридов. При этом информация об образцах заносится в специально созданную реляционную систему управления базами данных (СУБД) на базе Oracle 11.2. Риды копируются на сервер, где проходят фильтрацию по качеству и производится их картирование на референсные наборы. Информация о покрытиях заносится в СУБД, при выгрузке из нее покрытия нормализуются и создаются файлы с векторами относительной представленности референсных признаков (геномов) по образцам. Эти вектора затем используются для статистического анализа в программной среде R 3.1.0 Конвейер реализован на языках программирования bash, Python 3.2, Perl 5, C++, PL/SQL.

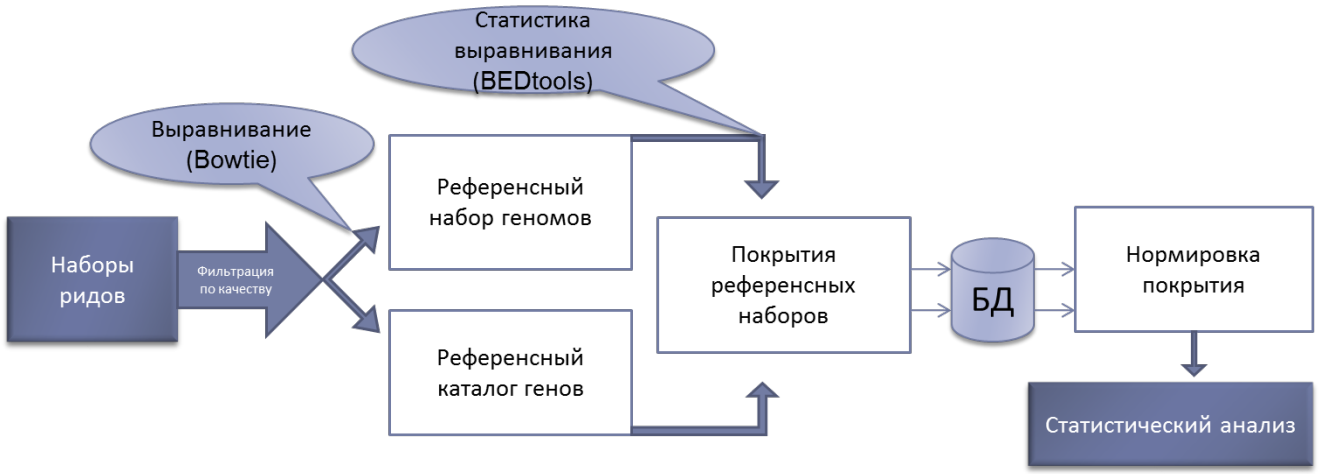

Рисунок 1. Алгоритм поточной обработки метагеномных данных.

Статистический анализ и визуализация

Статистический анализ проводился на языке программирования R 3.1.0. Расстояние между образцами измерялось 7 метриками: на основе корреляции Спирмена, Евклидово расстояние, расстояние Манхэттен, расстояние Канберра, расстояние Брэй-Кертис, расхождение Дженсона-Шэннона и модифицированный UniFrac, применимый для полногеномных данных (процедура описана в диссертационной работе).

Кластеризация образцов проводилась двумя методами: метод k-средних (функция pam пакета cluster языка R) с использованием индекса Калинского-Горабача, и методом, основанном на бутстрэп методе (функция pvclust), заключающимся в том, чтобы из имеющейся выборки образцов сделать большое количество других выборок из этих же элементов, но в случайном порядке, причем образцы могут повторяться или отсутствовать, но размер выборки остается прежним. Таким образом проверяется устойчивость кластеров.

Поиск признаков, по которым различаются те или иные группы образцов проводился тестом Манна-Уитни. Поправка на множественные сравнения производилась с помощью FDR (англ. False Discovery Rate). Гипотеза о различии выборок в целом проверялась тестом ANOSIM (англ. Analysis of similarity).

Моделирование производства и тока короткоцепочечных жирных кислот

Модель была построена совместно с К. В. Песковым и Ю. А. Косинским («Новартис Фарма») по данным из работы Schwiertz et al, 2009 о 98 индивидуумах, для которых есть метаданные (возраст, пол, индекс массы тела), известен родовой состав микробиоты по

результатам секвенирования генов 16S рРНК (численность основных кластеров и некоторых семейств), а также измерены концентрации основных КЖК в фекалиях (мМ).

Описание математической модели

В модели объем содержимого толстого кишечника делится на 4 отсека одинакового объема. Приток субстрата с пищей задан в 1-й отсек, откуда он последовательно переносится во 2-й, 3-й и 4-й отсеки. В каждом отсеке происходят аналогичные процессы: субстрат сбраживается до пропионата и ацетата, а из ацетата синтезируется бутират. Пути метаболизма КЖК сведены в модели тремя обобщенными реакциями, скорости которых задаются уравнениями типа Михаэлиса-Ментен (3-5).

Синтез ацетата (Acet) из субстрата (S):

$$V\_Acet\_pro = k\_S\_Acet \times S/(S + Km\_S\_Acet) \quad (3)$$

Синтез пропионата (Prop) из субстрата:

$$V\_Prop\_pro = k\_S\_Prop \times S/(S + Km\_S\_Prop) \quad (4)$$

Синтез бутирата (But) из ацетата:

$$V\_But\_pro = k\_Acet\_But \times Acet/(Acet + Km\_Acet\_But) \quad (5)$$

В каждом отсеке заданы процессы абсорбции КЖК энтероцитами. Остатки субстрата и КЖК из 4-го отсека элиминируются с постоянной скоростью. Все эти процессы описываются системой обыкновенных дифференциальных уравнений, которые и образуют модель.

Значения для одной части параметров модели оценивали из литературных данных, значения для другой части получали методом поиска соответствия к собственным экспериментальным данным. При этом сопоставляли предсказанные моделью концентрации КЖК в 4-м (терминальном) отсеке с соответствующими концентрациями, измеренными в фекалиях.

Для нахождения значений параметров и анализа ковариационных зависимостей была использована программа MONOLIX, использовался популяционный статистический подход.

При поиске параметров, оказывающих влияние на модель (ковариат) были использованы метаданные и представленность родов, придерживаясь следующих критеиев:

1) Были исследованы на предмет влияния на модель только достаточно представленные группы бактерий, т.е. те, чья доля в подавляющем большинстве образцов и составляют не менее 5% от общего числа бактерий. Малочисленные и редко встречающиеся группы вряд ли могут значительно повлиять на метаболизм КЖК в рамках механистической модели;

2) Статистическая значимость (p-value) и изменение объективной функции (англ. objective function, Log-likelihood estimation), рассчитываемые программой MONOLIX, использовались как параметры включения ковариаты в модель.

РЕЗУЛЬТАТЫ

1. Реализация алгоритма поточной обработки метагеномных данных

Используемый алгоритм обработки метагеномных данных подробно описан в Материалах и методах. Его отличительная особенность от других существующих программных продуктов для обсчета метагеномов заключается в возможности обработки коротких (длиной менее 75 нуклеотидов) ридов, в том числе и в цветовом формате, получаемые с приборов SOLiD4. Полностью автоматизированный конвейер работает и используется в исследованиях на сервере НИИ ФХМ. Обработка ридов включает в себя следующие стадии (рисунок 1):

1) Фильтрация по баллам качества.

Прибор ABI SOLiD 4 производит два типа файлов для каждого образца. В первом содержатся риды, закодированные в цветовом формате, во втором – баллы качества, соответствующие каждому нуклеотиду в каждом риде. Формально они отображают вероятность того, что данный цветовой сигнал был истинным, и могут принимать значения от 0 до 40. В анализ проходят риды, со средним баллом качества выше 15. Помимо фильтрации по среднему значению качества, используется алгоритм коррекции ридов SAET, который определяет по баллам качества низкокачественные 3` концы и обрезает их.

2) Фильтрация по геному человека.

В ходе этого этапа риды картируются на геном человека, некартировавшиеся дальше проходят в анализ. Некоторый процент ДНК человека всегда присутствует в метагеномных образцах ввиду процедур их получения и очищения.

3) Картирование на каталог геномов организмов, найденных в кишечнике человека.

Данный каталог включает набор геномов кишечных бактерий из базы данных HMP, который был дополнен геномами организмов, определенными как значимые для микрофлоры кишечника по данным метагеномных исследований. Этот каталог, по процедуре, описанной в Материалах и методах был преобразован в неизбыточный, содержит 353 генома.

4) Получение значений представленностей микроорганизмов.

По результатам картирования ридов на референсный набор геномов собирается статистика о проценте покрытия геномов (ширина покрытия) и общем количестве картировавшихся на геном нуклеотидов (глубина покрытия). Глубина покрытия нормируется на общее количество нуклеотидов, картировавшихся на весь референсный набор, и длину генома. Затем проводится суммирование нормированной глубины покрытия по родам. Полученные значения относительной представленности, организованные в вектора представленности по образцам, используются в дальнейшем анализе.

5) Визуализация и статистический анализ.

Анализ векторов представленности, в отличии от предыдущих этапов, не автоматизирован, т.к. каждая исследовательская задача требует индивидуального подхода. Он включает в себя визуализации, к примеру, тепловые карты (англ. heatplot), графики многомерного шкалирования (MDS, от англ. Multidimensional Scaling).

Для построения графиков MDS, а также для других видов анализа, необходимо измерить расстояние между образцами по векторам представленностей, которые представляют собой многомерные данные. Наиболее используемыми в метагеномике методами подсчета расстояний являются метрика UniFrac, содержащая в себе информацию о филогенетическом расстоянии между референсными геномами, и часто используемое в экологии расстояние Брэя-Кертиса.

В случае выявления различий между двумя выборками используется тест Манна-Уитни, поскольку метагеномные данные не обладают нормальным распределением.

2. Исследование российских метагеномных образцов от здоровых доноров

Исследуемые образцы были собраны по процедуре, описанной в Материалах и методах, от доноров без патологий желудочно-кишечного тракта, не принимавших антибиотики в течение полугода, в возрасте от 18 лет. Доноры являлись обитателями различных регионов России, это жители сельской местности (46) и крупных городов (50), а именно 4 из 10 наиболее заселенных мегаполисов России (Санкт-Петербург, Саратов, Ростов-на-Дону, Новосибирск). Сельские метагеномы были собраны в деревнях Татарстана, Омской области, Тывы и Хакассии. По этой выборке была поставлена задача выявить вид нормофлоры кишечника взрослых жителей Российской Федерации из различных географических и культурных сред.

Межнациональное сравнение с образцами из других стран, показало, что метагеномные образцы из России, так же, как и в остальном мире, содержат в себе представителей двух отделов – Bacteroidetes и Firmicutes. Однако их соотношение значительно различается, что приводит к разбросу данных на рисунке 2, на котором виден плавное изменение состава от американских образцов, богатых бактериями отдела Bacteroidetes, к российским, в которых преобладают представители отдела Firmicutes. Дополнительно был проведен анализ адекватности сравнения результатов секвенирования с различных платформ и сравнение кластеризаций по различным метрикам и расстояниям (подробнее см. диссертационную работу).

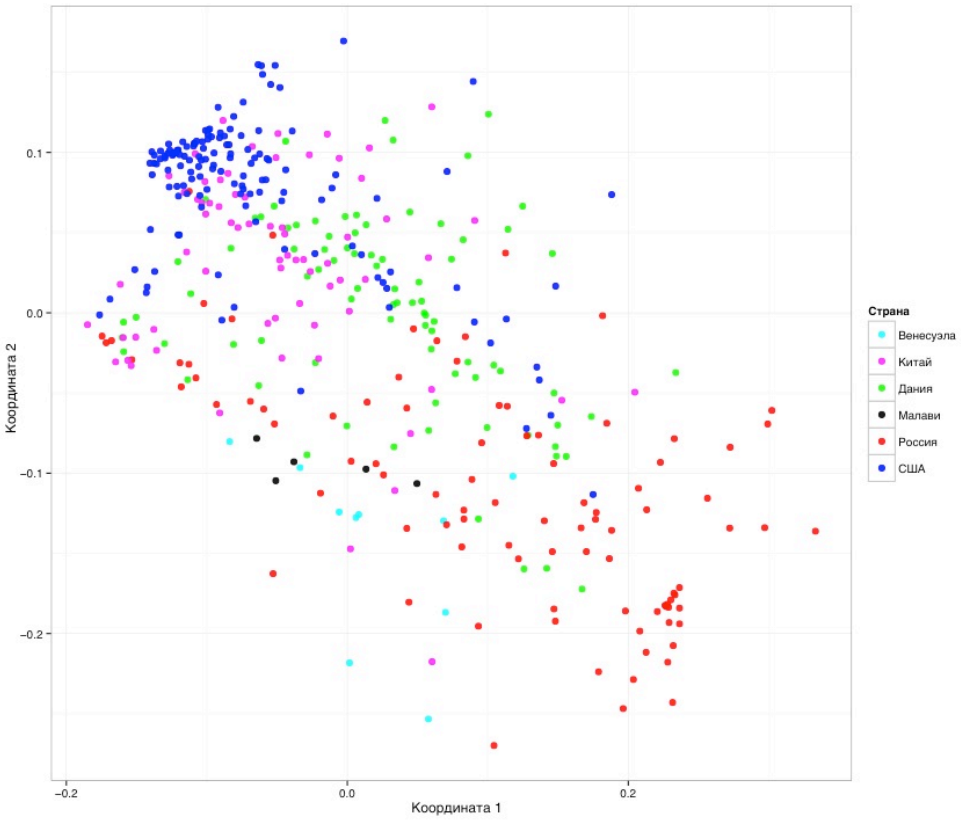

Рисунок 2. График MDS мировых образцов.

Методом pvclust (см. Материалы и методы) были выделены подгруппы образцов в российской метагеномной выборке, отличающиеся от всех прочих по составу (рисунок 3).

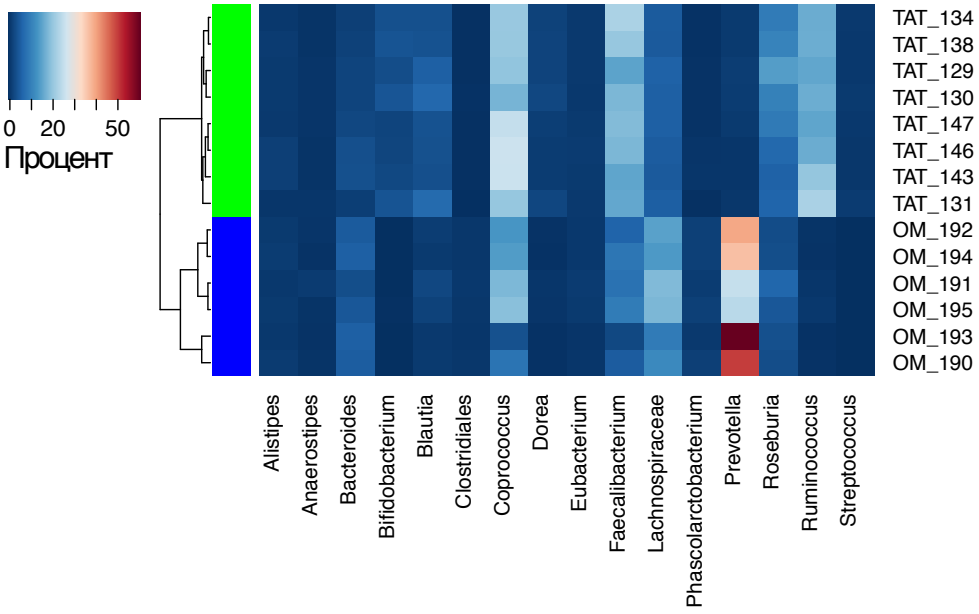

Рисунок 3. Тепловая карта наиболее представленных родов в найденных подгруппах.

3. Исследование клинических метагеномных образцов

Используя полученные данные по нормофлоре кишечника российской популяции в качестве контроля, был исследован ряд образцов от пациентов, в том числе метагеномы кишечника людей страдающих от алкогольной зависимости, принимающих антибиотики, онкологических больных с диагнозом нейробластома и гепатобластома, проходящих курс химиотерапии и работников производства с повышенным радиационным фоном.

Метагеном больных алкоголизмом было обнаружено повышенное содержание бактерий рода *Escherichia* и наличием ряда патогенов, в частности *Salmonella* и *Klebsiella*. Выборка образцов от работников на вредном производстве характеризовалась повышенным содержанием оппортунистических патогенов. Это может быть первым сигналом изменений в иммунной системе доноров и как следствие – изменения иммунной толерантности к этим микроорганизмом. Метагеном кишечника онкологических больных, проходящих химиотерапию, отличается низким бактериальным разнообразием и аномально высоким количеством ДНК человека, по всей видимости, вызванным повышенной эксфолиацией кишечного эпителия в следствие мукозита. Нарушение микробиотного баланса и иммунитета привел к появлению в значимых количествах патогенных организмов, в том числе дрожжей. Как следствие, метагеномы онкологических больных значительно отличались от контрольной выборки (рисунок 4). Некоторые из доноров принимали антибиотики, однако вне зависимости от наличия или отсутствия антибиотикотерапии, значительные отклонения наблюдались во всех образцах.

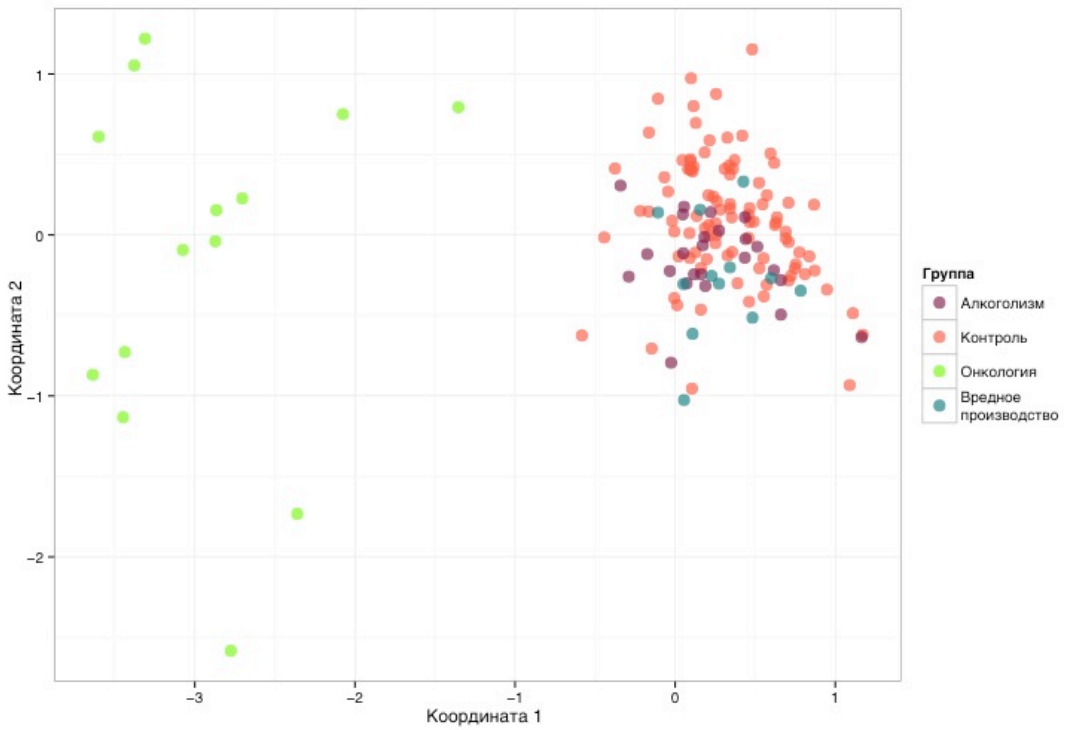

Рисунок 4. График MDS метагеномных образцов из клинических исследований и контрольной выборки. Цветом отмечена принадлежность к группам.

4. Создание кинетической модели производства и метаболизма короткоцепочечных жирных кислот по данным секвенирования генов 16S рРНК.

В данной работе была создана механистическая модель метаболизма КЖК и их дальнейшей реализации в организме хозяина, на основании экспериментальных данных полученных ранее (Schwiertz et al, 2009), включающих секвенирование генов 16S рРНК и измерение концентраций бутирата, пропионата и ацетата в 98 образцах кала.

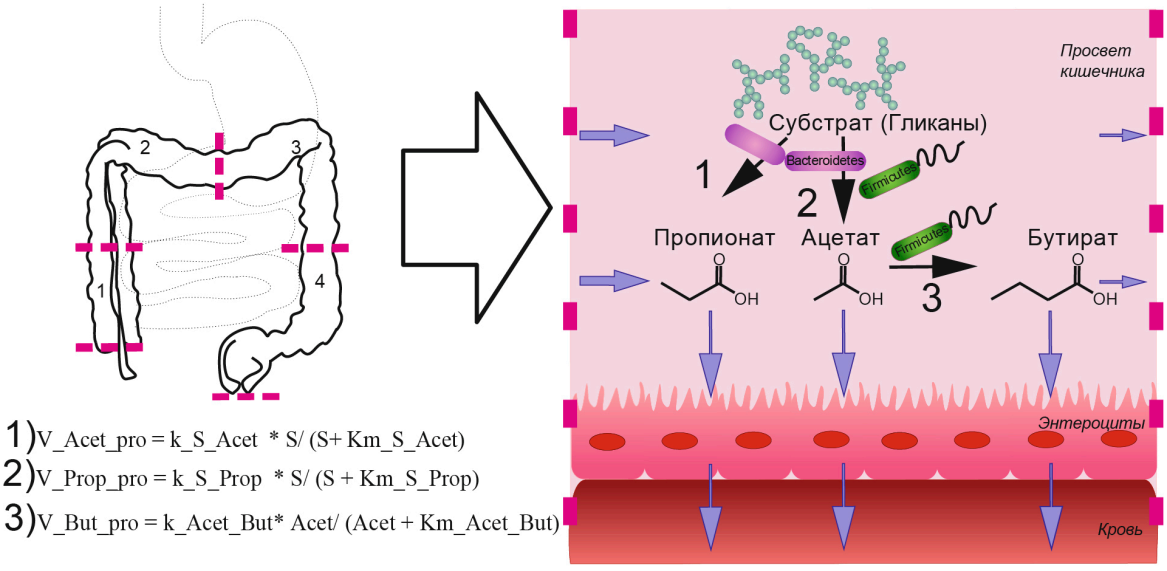

Рисунок 5. Механистическая модель производства и тока КЖК в кишечнике человека.

В модели пути метаболизма КЖК сведены тремя обобщенными реакциями, скорости которых задаются уравнениями типа Михаэлиса-Ментен (рисунок 5). Каждое уравнение включает ряд констант, часть из которых была оценена из литературных данных, значения для другой части были получены путем сопоставления предсказанных выходных концентраций КЖК к экспериментально полученных в исследовании Scwiertz et al. (2009).

После определения значений констант, был произведен поиск ковариат – переменных, влияющих на модель. В числе исследуемых ковариат были представленности бактериальных родов, определенных по секвенированию 16S РНК генов, возраст, ИМТ доноров, pH образцов. Всего было найдено 5 зависимостей. Было выявлено, что бактерии родов *Faecalibacterium* и *Coccoides* определяют изменение константы скорости при образовании бутирата из ацетата. Представители рода *Coccoides* также влияют на увеличение концентрации ацетата при расщеплении субстрата. Отмечено отрицательное влияние представителей родов *Ruminococcus* и *Bifidobacterium* на константу скорости расщепления субстрата до пропионата. Для всех этих процессов существует экспериментальное подтверждение, описанное в литературе (Louis P. et al. 2009, Sartor R.B. 2008, Liu C. et al. 2008, Belenguer A. et al., 2007).

ЗАКЛЮЧЕНИЕ

Результатом данной работы является разработанный программный комплекс для проведения таксономического анализа метагеномных данных. Программа была опробована на первых 96 полногеномно секвенированных метагеномных образцах от жителей Российской Федерации. Было проведено сравнение полученных метагеномов в общемировом контексте, выявлены таксономические особенности метагеномов кишечника жителей удаленных районов.

Описанные 96 образцов также были использованы в качестве контроля при исследовании трех клинических групп: онкологических больных, проходящих курс химиотерапии, лиц, страдающих от алкоголизма, и людей, работающих на производстве с повышенным радиационным фоном. У онкологических больных также наблюдались серьезные изменения состава микробиоты и низкая концентрация бактериальной ДНК, минимальное бактериальное разнообразие. Больные алкоголизмом обладают микробиотой, схожей со здоровой, но есть отличия, включая появления патогенных

организмов. Данный метод показал себя перспективным диагностическим инструментом.

Была построена первая механистическая модель производства короткоцепочечных жирных кислот микробиотой с последующим потреблением их организмом хозяина и выведением наружу. В отличии от существовавших ранее моделей, она применима не только к модельным организмам с заданным составом микробиоты, но и к реальным метагеномным данным от существующих людей. Аналогичный метод может быть использован для поиска взаимосвязей ряда биохимических параметров с составом микробиоты человека, что может способствовать более глубокому пониманию механизмов взаимодействия бактерий с организмом хозяина.

ВЫВОДЫ

1. Была создана вычислительная аналитическая структура, которая включает в себя стадии первичной обработки метагеномных данных, полученных с использованием различных платформ секвенирования (ABI SOLiD, Illumina, 454 GS FLEX+, Ion Torrent), и их статистический анализ.
2. Впервые разработан алгоритм, позволяющий анализировать данные, полученные с использованием платформы SOLiD4.
3. В ходе сравнения 96 образцов из Российского метагеномного проекта и образцов жителей других стран была получена группа, пригодная для реализации в качестве контроля состава микробиоты в других биомедицинских исследованиях. Выявлены новые структуры микробных сообществ.
4. Алгоритм анализа метагеномных данных был опробован на трех вариантах патологий человека. Для онкологических больных было обнаружено значительное состава микробиоты в сторону увеличения количества патогенных микроорганизмов. Для групп больных алкоголизмов и лиц, работающих на предприятиях с повышенным радиационным фоном существенных изменений состава микробиоты по отношению к контролю обнаружено не было. Однако, для лиц страдающих алкоголизмом было отмечено превалирование в составе микробиоты представителей отдела Proteobacteria. Работники производства с повышенным радиационным фоном отличались повышенным содержанием в составе микробиоты условно патогенных микроорганизмов, в норме не вызывающих заболевания, что вероятно связанно с снижением иммунитета.
5. Была создана кинетическая модель метаболизма короткоцепочечных жирных кислот (КЖК), позволяющая прогнозировать их концентрации в зависимости от состава микробиоты на основе данных секвенирования 16S рРНК генов. Созданный метод может быть использован прогнозирования механизмов взаимодействия микробиоты с организмом хозяина.

ПУБЛИКАЦИИ ПО ТЕМЕ РАБОТЫ

Статьи в научных журналах

1. Tyakht AV, Popenko AS, Belenikin MS, Altukhov IA, Pavlenko AV, Kostryukova ES, Selezneva OV, Larin AK, Karpova IY, Alexeev DG. MALINA: a web service for visual analytics of human gut microbiota whole-genome metagenomic reads. // Source Code Biol Med. 2012. Vol. 7, №1. P. 13.
2. Д.Г. Алексеев, Е.С. Кострюкова, А.С. Попенко, И.В. Русаловский, А.В. Тяхт Потенциальные возможности использования распределенных вычислительных систем при решении концептуальных проблем построения информационных комплексов обработки данных высокопроизводительного геномного секвенирования и глубокого протеомного профилирования. // Информатизация и связь. 2012. Т. 8. С. 10-14.
3. Tyakht AV, Kostryukova ES, Popenko AS, Belenikin MS, Pavlenko AV, Larin AK, Karpova IY, Selezneva OV, Semashko TA, Ospanova EA, Babenko VV, Maev IV, Cheremushkin SV, Kucheryavyy YA, Shcherbakov PL, Grinevich VB, Efimov OI, Sas EI, Abdulkhakov RA, Abdulkhakov SR, Lyalyukova EA, Livzan MA, Vlassov VV, Sagdeev RZ,Tsukanov VV, Osipenko MF, Kozlova IV, Tkachev AV, Sergienko VI, Alexeev DG, Govorun VM. Human gut microbiota community structures in urban and rural populations in Russia. // Nat Commun. 2013. Vol. 4. eP. 2469.
4. Tyakht AV, Alexeev DG, Popenko AS, Govorun VM. Rural and urban microbiota: To be or not to be? // Gut Microbes. 2014. Vol. 5, №3. P. 351-356.
5. С.В. Федосенко, Л.М. Огородова, В.М. Говорун, М.А. Карнаушкина, И.В. Салтыкова, Д.Г. Алексеев, Е.С. Кострюкова, А.В. Тяхт, А.С. Попенко. Анализ таксономического состава кишечной микробиоты больных хронической обструктивной болезнью легких. // Уральский медицинский журнал. 2014. Т. 6, №120. С. 168-173.

Материалы конференций

6. Попенко А.С. Программный комплекс для биоинформатического анализа метагеномных данных. // 54ая научная конференция МФТИ “Проблемы

фундаментальных и прикладных, естественных и технических наук в современном информационном обществе”. Москва. 2011. С. 140.

7. Popenko A.S. MALINA - a Web-service for human gut microbiota whole-genome metagenomic reads analysis. // IHMC. Париж. 2012. P №94.
8. Alexeev D.G., Tyakht A.V., Popenko A.S., Belenikin M.S., Altukhov I.A., Pavlenko A.V., Kostryukova E.S., Selezneva O.V., Larin A.K., Karpova I.Y., Govorun V.M. Deep metagenomics and metaproteomics of human gut: dramas and delights. // BGRS\SB. Новосибирск. 2012. С. 33.
9. Попенко А.С., Тяхт А.В., Алексеев Д.Г . Особенности биоинформационного анализа данных полногеномного секвенирования микробных сообществ. // Postgenome. Казань. 2012. С. 209.