

Федеральное государственное бюджетное учреждение науки «Научно-исследовательский институт физико-химической медицины Федерального медико-биологического агентства»

На правах рукописи

ПОПЕНКО Анна Сергеевна

**БИОИНФОРМАЦИОННОЕ ИССЛЕДОВАНИЕ ТАКСОНОМИЧЕСКОГО
СОСТАВА МИКРОБИОТЫ КИШЕЧНИКА ЧЕЛОВЕКА**

03.01.09 – математическая биология, биоинформатика

ДИССЕРТАЦИЯ

на соискание ученой степени кандидата биологических наук

Научный руководитель
д.б.н., проф., чл.-корр. РАН
Говорун Вадим Маркович

Москва – 2014

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
Глава 1. МЕТАГЕНОМИКА ЧЕЛОВЕКА: ИСТОКИ, РАЗВИТИЕ, ПЕРСПЕКТИВЫ	10
1.1 Начало метагеномных исследований.....	10
1.2 Микробиота кишечника человека: состав, функции и значение	13
1.3 Разнообразие здоровой микробиоты кишечника	17
1.4 Микробиота кишечника человека при различных патологиях.....	21
1.4.1 Воспалительные заболевания и рак кишечника	21
1.4.2 Патогенные микроорганизмы и инновационные методы лечения.....	22
1.4.3 Метаболические патологии и роль микробиоты	23
1.4.4 Последствия приема антибиотиков для микробиоты кишечника	25
1.4.5 Генные маркеры патологических состояний	26
1.5 Методика проведения метагеномных исследований	27
1.6 Актуальность метагеномных исследований в Российской Федерации	32
Глава 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ	33
2.1 Забор образцов кала.....	33
2.2 Выделение ДНК из кала.....	33
2.3 Методы секвенирования.....	34
2.4 Дополнительные метагеномные данные.....	35
2.5 Предобработка ридов	35
2.6 Определение таксономического состава	36
2.7 Реализация алгоритма поточной обработки данных.....	37
2.8 Статистический анализ и визуализация	38
2.9 Моделирование производства и тока короткоцепочечных жирных кислот	40
Глава 3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЕ МЕТАГЕНОМНЫХ ОБРАЗЦОВ: РАЗРАБОТКА МЕТОДОЛОГИИ И АНАЛИЗ	44
3.1 Реализация алгоритма поточной обработки метагеномных данных	44
3.2 Исследование российских метагеномных образцов от здоровых доноров	47
3.3 Исследование метагеномных образцов при патологиях	64
3.3.1 Метагеном пациентов, больных онкологическими заболеваниями	65
3.3.2 Метагеном кишечника людей, страдающих от алкогольной зависимости	68
3.3.3 Метагеном кишечника работников производства с повышенным радиационным фоном	73
3.4 Создание кинетической модели производства и метаболизма короткоцепочечных жирных кислот по данным секвенирования генов 16S рРНК.....	77
Глава 4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ	80
4.1 Создание программного комплекса по обработке метагеномных данных	80
4.1.1 Фильтрация ридов	80
4.1.2 Картирование на референсный каталог геномов	81
4.1.3 Подсчет покрытия референсных последовательностей	82
4.1.4 Статистический анализ	83
4.1.5 Реализация программного комплекса	86
4.2 Анализ образцов из Российского метагеномного проекта.....	86
4.3 Видовое разнообразие метагеномов при различных диагнозах.....	90
4.4 Моделирование производства короткоцепочечных жирных кислот	93
4.5 Заключение.....	94
ВЫВОДЫ.....	96

СПИСОК ЛИТЕРАТУРЫ.....	97
СПИСОК ИЛЛЮСТРАЦИЙ.....	111
ПРИЛОЖЕНИЕ 1	114
ПРИЛОЖЕНИЕ 2	123
ПРИЛОЖЕНИЕ 3	131
ПРИЛОЖЕНИЕ 4	138

ВВЕДЕНИЕ

Наш мир представляет собой невероятно сложную экологическую сеть, работа каждого элемента которой выверена с точностью швейцарских часов. Но при этом вся эта система является единым живым целым, адекватно реагирующим на внешние условия и эволюционирующим. Каждый горячий источник, каждая лужа, моря, океаны, почва – все это части всемирной экологической сети. Их внутреннее устройство не проще экосистемы всей планеты. Несмотря на разнообразие условий, будь то земля, вода или кишечник животного, все эти экосистемы объединяет одно – наличие бактерий. На них приходится от половины до 90% всей биомассы нашей планеты¹.

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 ниши². 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. Микробиота человека интересна и с медицинской точки зрения. Наиболее многочисленной и разнообразной является микробиота кишечника человека.

Первым шагом в изучении микробиомов является определение их видового состава. Классические биологические подходы, такие как бактериальный посев или выделение отдельных клонов, весьма затруднительно использовать для этой цели ввиду большого количества видов, составляющих отдельный микробиом, и

невозможности культивировать до 99% бактерий³. Поэтому, действительно широкое распространение микробиомных исследований стало возможным около 10 лет назад с появлением высокопроизводительных секвенаторов нового поколения, которые позволяют за короткие сроки массово секвенировать совокупный геном микробиомов – метагеном. Так настоящим первым прорывом в области метагеномики стало исследование микробиома Саргассова моря, в ходе которого было секвенировано рекордное на тот момент количество ДНК – 1,045 миллиарда нуклеотидов⁴. Но новые экспериментальные методы исследования требуют новых подходов к обработки данных. Анализ столь огромного и разнородного материала стал своего рода вызовом для биоинформатики. Задача определения количественного и качественного бактериального, а также генного, состава по смеси коротких последовательностей ДНК все еще остается сложной. В данной работе описаны универсальный алгоритм метагеномных исследований и результаты первого популяционного исследования микробиома кишечника жителей Российской Федерации.

Актуальность проблемы.

Метагеномика начиналась с исследований микробиомов окружающей среды, как то микробиомов почв⁵, морей⁴, горячих источников⁶, но последние несколько лет особое внимание уделяется исследованиям микробиома человека. Так за последние 10 лет выпущено свыше 1500 статей по метагеномике человека, половина из которых – только за 2011 - 2013 года. Такое лавинообразное появление метагеномных работ вызвано организацией крупных консорциумов: европейский MetaHIT, специализирующийся только на метагеноме кишечника человека, и американский HMP, целью которого является изучение всего микробиома человека. Первые работы этих сообществ^{7,8} стали фундаментом для дальнейших исследований во многих странах мира. Первые исследования носили в большей степени эпидемиологический характер. Ученые пытались выявить, что есть эталон здорового микробиома, как он варьирует в зависимости от географического и социального факторов^{9,10}.

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 витамины¹¹, короткоцепочечные жирные кислоты (КЖК)¹².

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 тела¹³.

Помимо общих популяционных исследований, изучается связь между составом микробиома кишечника человека и различными заболеваниями: рак¹⁴, атеросклероз¹⁵, диабет второго типа¹⁶. 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^{17,9,16}, 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. В процессе исследования был создан алгоритм поточной обработки данных, который позволяет анализировать метагеномы, полученные с приборов ABI SOLiD4 и SOLiD5500, чего не существовало ранее.

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.

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.
4. Была создана кинетическая модель метаболизма короткоцепочечных жирных кислот, 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 (MCCMB-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 (ИНМС, Париж 2011; ECCB'12 Базель 2012).

Публикации.

По результатам работы были написано 5 статей в рецензируемых научных журналах и 4 тезисов в материалах конференций:

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, Kucheryavyu 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. P. 3.
 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

1. 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.
2. Popenko A.S. MALINA - a Web-service for human gut microbiota whole-genome metagenomic reads analysis. // IHMC. Париж. 2012. P №94.
3. Alexeev D.G., Tyakht A.V., Popenko A.S., Belenikin M.S., Altukhov I.A., Pavlenko A.V., Kostyukova 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.
4. Попенко А.С., Тяхт А.В., Алексеев Д.Г. . Особенности биоинформационного анализа данных полногеномного секвенирования микробных сообществ. // Postgenome. Казань. 2012. С. 209.

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, 4 приложения.

Глава 1. МЕТАГЕНОМИКА ЧЕЛОВЕКА: ИСТОКИ, РАЗВИТИЕ, ПЕРСПЕКТИВЫ

1.1 Начало метагеномных исследований

Термин «метагеномика» был впервые введен в 1998 году¹⁸ для обозначения практики анализа совокупной генетической информации из определённой среды. Впервые идея выделения ДНК из образца среды была предложена еще в 1985 году Норманом Пэйсом, а в 1991 году он с коллегами опубликовал работу о выделении и идентификации последовательностей гена 16S рРНК из образца воды Тихого океана¹⁹. Следующей исторической вехой в развитии метагеномики стала двухгодичная Глобальная Океаническая Экспедиция по сбору метагеномных образцов (GOS) в 2003 году, учрежденная Крэйгом Вентером. Состав собранных образцов был определен по результатам секвенирования генов 16S рРНК, и в одном только Саргассовом море было найдено свыше 2000 новых видов бактерий⁴.

Действительно широкое распространение метагеномика получила только в первом десятилетии 21 века благодаря появлению высокопроизводительных приборов для секвенирования. Этот технологический прорыв позволил секвенировать из метагеномных образцов не только последовательности генов 16S рРНК, но и всю тотальную ДНК, включая все гены и некодирующие участки (так называемое шотган-секвенирование, англ. shotgun sequencing).

Вследствие этого, стало возможным получать информацию о таксономическом составе микробиоты, о ее функциональном потенциале – через определение относительной представленности генов, кодирующих те или иные белки, в частности, гены ферментов метаболических путей.

Первая работа с использованием шотган-метагеномного подхода к выделению ДНК и секвенированием на приборах нового поколения была проведена в 2005 году Стефаном Шустером²⁰. С тех пор количество работ по метагеномике растет экспоненциально. Вначале основными объектами

исследования были различные сообщества из окружающей среды, но со временем исследователи все больше стали интересоваться микробными сообществами, живущими и взаимодействующими с высшими организмами. Первая работа по метагеному кишечника человека была опубликована американскими учеными в 2006 году²¹, в ней был проведен функциональный анализ двух наборов метагеномных прочтений, полученных из образцов кала. Вслед за ней вышла статья группы ученых из Японии¹⁰ с описанием уже 13 новых образцов. В обоих исследованиях применялось полногеномное шотган секвенирование с последующей *de novo* сборкой контигов и предсказанием открытых рамок считывания, а таксономическая принадлежность определялась путем поиска ближайшего сходства этих контигов с базой геномных последовательностей NCBI.

Фундаментальным прорывом в развитии метагеномики человека можно считать создание двух крупнейших консорциумов по изучению микробиоты человека: MetaHIT (Metagenome of Human Intestinal Tract) в Европе и HMP (Human Microbiome Project) в США. Ученые из MetaHIT совместно с BGI (Beijing Genomics Institute, Китай) установили каталог из 3,3 млн. преобладающих бактериальных генов в метагеноме кишечника человека¹⁶. В это же время, исследователями из США был опубликован список секвенированных геномов бактерий и архей, найденных в микробиоте человека¹⁷. Эти две работы стали отправными пунктами для дальнейших исследований (Рисунок 1). В современных исследованиях микробиоты методом шотган-секвенирования первым шагом в обработке метагеномных данных (вслед за отсеиванием низкокачественных данных) является выравнивание коротких нуклеотидных последовательностей, полученных в результате секвенирования – ридов (англ. Reads), на референсные последовательности генов и геномов, с целью охарактеризовать метагеном качественно и количественно – как с функциональной, так и с таксономической точки зрения. Поиск по ридам против полных мировых баз занимает огромное количество ресурсов и времени, а сборка полученных от секвенирования коротких ридов длиной менее 100 п.н. мало эффективна.

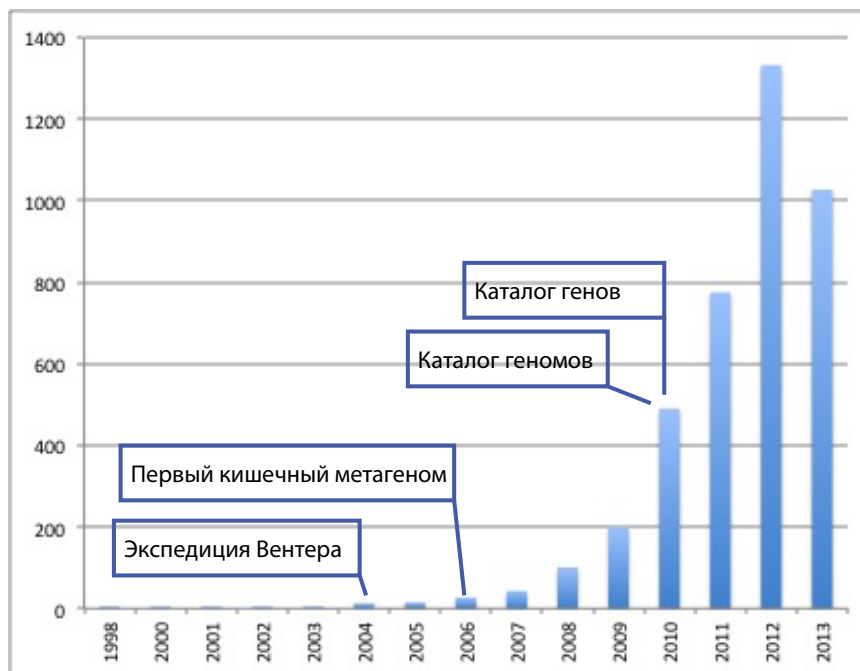


Рисунок 1. Гистограмма количества статей с упоминанием слова «metagenome» согласно статистике PubMed.

Следующие годы ознаменовались широкомасштабными исследованиями микробиоты кишечника населения различных уголков планеты и их сравнение в контексте географических и культурных различий^{17,22,7,8,23}. Параллельно были исследованы различия в микробиоте между здоровыми людьми и пациентами с кишечными заболеваниями, такими как болезнь Крона²⁴, болезнь воспаленного кишечника²⁵, рак толстой кишки²⁶. Помимо кишечных заболеваний изучалась взаимосвязь метагенома и таких системных заболеваний, как метаболический синдром²⁷, диабет²⁸, атеросклероз¹⁵. Исследователи находили маркеры заболеваний как на таксономическом, так и на функциональном уровне, выявляя специфические метаболические пути и гены, чья относительная представленность повышена или понижена у пациентов по сравнению с группой контроля.

К настоящему времени накоплены данные, дающие представление о составе нормофлоре кишечника человека и изучены отклонения от нее, их причины и влияния на организм человека, разрабатываются новые, практические способы применения знаний о микробиоте, такие как трансплантация кала²⁹.

1.2 Микробиота кишечника человека: состав, функции и значение

Микробиота кишечника человека, как показано в метагеномных исследованиях, преимущественно состоит из бактерий, относящихся к двум отделам – Bacteroidetes и Firmicutes, составляющих значительную часть всей микрофлоры. В меньших количествах присутствуют представители отделов Proteobacteria, Actinobacteria, Fusobacteria, Verrucomicrobia и Cyanobacteria³⁰. Также, в метагеноме идентифицируют археи, в основном представленные родом *Methanobrevibacter*. Совокупность микробов кишечника составляет суперколонию из порядка 10^{14} клеток из более чем 800 видов, превышая количество клеток человека в 10 раз, и веся порядка 1-2 кг³¹.

Как только стали появляться первые метагеномные исследования, была осуществлена попытка выявить устойчивые типы кишечной микробиоты, по аналогии с группами крови. В 2011 году вышла работа европейских ученых из консорциума MetaHIT⁹, в которой было показано существование трех типов микробных сообществ, названные энтеротипами. Эти группы характеризовались преобладанием определенных бактериальных родов: центральным в первом энтеротипе был род *Bacteroides*, во втором - *Prevotella*, в третьем - несколько представителей отдела Firmicutes, включая рода *Ruminococcus* и *Faecalibacterium*. Разделение образцов на энтеротипы никак не коррелировало ни с национальной принадлежностью, ни с возрастом или полом. Наличие энтеротипов было подтверждено еще в нескольких исследованиях на других больших группах^{16,32}, но их число варьировало: третий энтеротип иногда не удавалось идентифицировать. В связи с этим, в научном сообществе возникла критика этой теории. Так, выдвигается мнение, что микробиота не поддается категоризации, а правильнее говорить о «непрерывном градиенте состава»³³. Вне зависимости от подхода к классификации, таксономический состав микробиоты кишечника человека значительно варьирует на индивидуальном уровне, при этом он играет важную роль в жизнедеятельности организма человека, образуя, фигурально выражаясь, отдельный орган, выполняющий жизненно-важные функции.

Основная функция микробиоты заключается в деградации непереваренных компонентов пищи. Непереваренный белок утилизируется микроорганизмами в дистальных отделах, с образованием аммиака, фенольных и миндальных соединений³⁴. В день около 20-60 грамм углеводов достигают кишечника избегнув переваривания ферментами человека: наш геном содержит в себе потенциал для расщепления лишь небольшого количества видов гликанов – крахмала, лактозы и сахара, в то время как некоторые микроорганизмы в кишечнике могут расщеплять десятки видов гликанов³⁵. Микробиота ферментирует непереваренные углеводы, преимущественно устойчивые формы крахмала и составные части растительной клеточной стенки, производя короткоцепочечные жирные кислоты и газы, в том числе водород и метан. Крахмал делится на несколько типов по механизму своей устойчивости: или он состоит из полимеров клеточной стенки растений (тип 1), либо обладает гранулярной структурой (тип 2), либо становится менее растворимым после нагревания с последующим охлаждением (тип 3), или имеет дополнительные химические связи (тип 4). Было показано, что добровольцы, употреблявшие пищу с большим содержанием устойчивого крахмала, имели повышенный уровень *Ruminococcus bromii* в метагеноме³⁶.

Представители рода *Bacteroides* способны расщеплять ксиланы, маннаны, галактоманнаны. При этом расщепление происходит на клеточной стенке бактерий, а не во внешнем пространстве. Были показаны сложные пищевые связи между *B. thetaiotaomicron* и *Bifidobacterium spp*, обладающими разными возможностями утилизации фруктанов разной длины³⁷. Бактерия *Bifidobacterium longum subsp infantis* способна утилизировать олигосахариды из человеческого молока³⁸. Целлюлозу могут расщеплять бактерии родов *Ruminococcus* – с образованием метана, и *Bacteroides* – не образуя метан³⁹. Некоторое количество микроорганизмов способно переваривать муцин – основной гликопротеид мукозного слоя кишечника, например *Akkermansia muciniphila*, составляющая около 3% по доле численности в кишечной микробиоте взрослых людей⁴⁰.

Необходимо отметить, что во всех случаях речь идет об определении состава метабенома кишечника по результатам секвенирования ДНК из образцов кала, что примерно соответствует слепку микробиоты всего кишечника. Тем не менее, существуют данные, что состав пристеночной микробиоты, имеющей непосредственный доступ к мукозному слою, и состав микробиоты просвета кишечника значительно различается⁴¹. Однако и те и другие обитатели кишечника достаточно тесно взаимодействуют друг с другом, поэтому говоря о роли микробиоты в метаболизме человека правильно будет рассматривать полную ее совокупность.

Таким образом, микробиота представляется сложным симбиотическим ферментером углеводов, производящим короткоцепочечные жирные кислоты (КЖК), преимущественно, ацетат, пропионат и бутират, которые в свою очередь являются основными энергетическими эквивалентами, получаемые из пищи¹². В дальнейшем, большая часть ацетата и пропионата поступает в кровь, посредством которой ацетат преимущественно разносится по всему организму, а пропионат оседает в печени⁴². Но помимо энергетической функции, КЖК так же выполняют и сигнальную, способны влиять на иммунитет человека^{43,44}, а бутират препятствует канцерогенезу⁴⁵. Соотношение получаемых от микробиоты КЖК предположительно зависит от ее состава, в частности от соотношения бактерий отделов *Bacteroidetes* и *Firmicutes*, у которых значительно различаются профили утилизации поли- и олигосахаридов. Ранее был описан механизм кооперации между бактериями *Bacteroides thetaiotamicron* и *Eubacterium rectale*⁴⁶. Вкратце, *B. thetaiotamicron* способен расщеплять ряд полисахаридов из пищи и организма хозяина до моносахаридов, а так же продуцирует пропионат и ацетат. Ацетат затем потребляется бактерией *E. rectale*, которая из него синтезирует бутират. Однако надо иметь ввиду, что несмотря на то, что бактерии отдела *Firmicutes* являются основными продуцентами бутирата, другие представители могут расщеплять субстрат и синтезировать ацетат.

На сегодняшний день существует несколько работ, посвященных моделированию метаболизма КЖК в кишечнике. Одно из первых таких

исследований было посвящено математическому описанию утилизации лактата бактериями *Eubacterium hallii* и *Anaerostipes coli* SS2/1 и синтеза ими бутирата⁴⁷. Кинетические параметры модели были оценены по экспериментальным данным, а именно по изменению концентраций субстрата и продуктов в совместно выращенных колониях этих бактерий. Позже появилась работа, также описывающая метаболизм КЖК, но опирающаяся на экспериментальные данные, полученные путем подсаживания заданной микробиоты гнотобиотическим мышам⁴⁸. В обоих случаях был создан точный математический аппарат, описывающий процесс синтеза КЖК, но применим он только к тем экспериментальным моделям, на которых он был построен, т.к. в этом процессе задействовано значительно большее количество бактериальных видов, соответственно и параметры модели будут иными.

Помимо КЖК, микробиота кишечника способна синтезировать витамины: в метагеномном исследовании микробиомов жителей городов США и сельских регионов¹⁷ были найдены метаболические пути синтеза кобаламина, фолиевой кислоты, биотина и тиамина.

Второй существенной функцией микробиоты кишечника является ее взаимодействие с иммунной системой либо напрямую, либо через кишечный эпителий: его барьерные функции играют важную роль в достижении баланса с комменсальной микробиотой⁴⁹. Их общение происходит через образ-распознающие рецепторы (англ. Toll-like receptors, TLR) – белки, присутствующие на поверхности клеток иммунной системы и способные узнавать специфичные молекулярные патогенов: в тонком кишечнике, такие факторы, как пептидогликан-узнающие белки (PGRP) и дефенсины, индуцируются комменсалами в ответ на стимуляцию этих рецепторов. Эффекты иммуномодулирования кишечной микробиотой необходимы для иммунотолерантности, достигаются через распространение регуляторных Т-клеток и иммуносупрессоров цитокинов⁵⁰: в толстом кишечнике бактерии *Bacteroides fragilis* и *B. breve* индуцируют через TLR2 клетки Treg⁵¹. Бактериальный капсульный полисахарид (Polysaccharide A из *Bacteroides fragilis*)

способен корректировать дисбаланс Т-хелперов в гнотобиотических мышах⁵². Изучение микробиоты людей, страдающих аллергией, показало пониженное содержание бактерий рода *Bifidobacteria*, и увеличенное – *B. fragilis*, *Staphylococcus aureus* и *E. coli*⁵³.

Комменсальная микробиота предотвращает от инфицирования энтеропатогенами путем конкуренции за субстрат и продуцируя бактерицидные факторы⁵⁴. Так, штамм *Bacillus thuringiensis*, изолированный из образца кала, продуцирует антимикробный пептид (turicine) с узким спектром действия против *Clostridium difficile*⁵⁵.

Комменсальная микробиота продуцирует метаболиты, защищающие от воздействия вирулентных факторов. Например, ацетат производимый *Bifidobacterium* ингибирует перенос токсинов энтерогеморрагической кишечной палочки (ЕНЕС) из кишечника в кровь⁵⁶, а бутират, производимый бактериями отдела Firmicutes, является противовоспалительным фактором⁵⁷. Помимо этого, микробиота кишечника влияет на усвоение лекарств, на их фармакинетику и спектр побочных действий⁵⁸.

Микробиота участвует в развитии центральной нервной системы (ЦНС), в частности влияя на системы, связанные с ответом на раздражение⁵⁹. Бактерии могут воздействовать на вегетативную нервную систему, иннервацию кишечника, путем производства гормоноподобных соединений. Благодаря этому, они могут изменять его перистальтику⁶⁰.

1.3 Разнообразие здоровой микробиоты кишечника

Как ранее было указано, микробиота кишечника человека состоит преимущественно из представителей двух отделов – Bacteroidetes и Firmicutes. Однако, в этих пределах микробные сообщества достаточно вариабельны среди жителей нашей планеты.

Перемены в культурном укладе, образе жизни, режиме питания, рост численности населения, развитие промышленности и сельского хозяйства влияют

на микробиом. Из-за распространения унифицированной промышленной диеты происходит глобализация микробиоты⁶¹. В связи с логистикой всемирного рынка еды возрастает потребность в консервировании пищи, добавлении антибиотиков, термической обработке. Учитывая эти факторы, представляет интерес изучение микробиоты как можно более широкого диапазона сообществ планеты - до того, как индустриализация нанесла свой отпечаток на их микробиоту, снизив разнообразие.

Недавние исследования затронули тему микробиоты у населения стран, где преобладающий образ жизни отличается от т.н. “западного”. Так, в работе Де Филиппо и коллег²² были изучены различия в метагеномах европейских и африканских детей. Они выявили, что в микробиоте кишечника жителей Африки преобладают бактерии рода *Prevotella*, обладающие способностью расщеплять ксилан, что связывают с необходимостью получения большего количества энергии из сложных углеводов, составляющих основной рацион. В микробиоте европейских же детей больше содержание представителей рода *Bacteroides*, ассоциированных с так называемой «западной диетой» – богатой животными жирами и белком³². Было также показано, что уменьшение потребления углеводов ведет к уменьшению количества бутирата и бутират-продуцирующих бактерий отдела Firmicutes, таких как *Roseburia* и *Eubacterium*⁶².

Были изучены микробные сообщества жителей развивающихся стран, таких как Малави (Африка) и Венесуэла (Южная Америка) и проведено сравнение с микробиотой кишечника жителей мегаполисов США на функциональном уровне¹⁷. Интересно, что различия в метаболическом геномном профиле между сельскими и городскими жителями напоминают различия между хищниками и травоядными. В микробиоте горожан повышена представленность генов разложения глутамина, в микробиоте жителей удаленных регионов – ген глутамат синтазы, что предположительно отражает различие в уровне потребления животного белка. В микробиоте жителей США превалировали и другие ферменты расщепления аминокислот: аспартата, пролина, орнитина и лизина, а также катаболизм простых сахаров и сахарозаменителей, гликанов мукозы, метаболизм

ксенобиотиков и метаболизм желчных кислот. Уровень гена альфа-амилазы, участвующей в расщеплении крахмала, был повышен в сельской группе, отражая значительное потребление кукурузы.

Было проведено исследование микробиоты коренных обществ охотников-собирателей, не употребляющих покупную пищу и ведущих исконный образ жизни – народа Хадза, Африка⁶³. На уровне бактериальных родов по сравнению с микробиотой жителей Италии были выявлено повышенное количество представителей *Prevotella*, *Eubacterium*, *Oscillibacter*, *Butyricicoccus*, *Sporobacter*, *Succinivibrio* и *Treponema*, и пониженное – *Bifidobacterium*, *Bacteroides*, *Blautia*, *Dorea*, *Roseburia*, *Faecalibacterium* и *Ruminococcus*.

Помимо географических различий и разного образа жизни, на микробиоту кишечника человека влияет возраст. Изначально мы получаем нашу микробиоту, проходя родовые пути и питаясь материнским молоком⁶⁴, или даже раньше, в перинатальном развитии⁶⁵. Отмечается, что в первые месяцы жизни, микробиота неустойчива⁶⁶, в ней преобладают бифидобактерии, причем их количество коррелирует с количеством у матери и зависит от способа кормления (грудное или искусственное вскармливание)⁶⁷, а их количество падает на протяжении всей жизни⁶⁸. Уже примерно к 3-6 годам микрофлора кишечника стабилизируется¹⁷ и приобретает характер состава как у взрослых людей. Кроме того, были найдены отличия и на уровне метаболических путей между микробиотой младенцев и взрослых¹⁷. Примером изменений, связанных с возрастом может быть метаболизм кобаламина и фолиевой кислоты. Микробиота младенцев обогащена генами биосинтеза фолиевой кислоты, в то время как в микробиоте взрослых выше представленность генов ее разложения и разложения тетрагидрофолата. Биосинтез кобаламина при этом с возрастом увеличивается, также, как и биосинтез биотина и тиамин. Низкая представленность генов биосинтеза кобаламина в микробиоте младенцев коррелирует с низкой представленностью *Bacteroidetes*, *Firmicutes*, и *Archaea*, зато биосинтетический путь кобаламина в них хорошо представлен. *Lactobacillus*, который доминирует в детской микробиоте,

лишен этих генов, но содержит гены биосинтеза и метаболизма фолиевой кислоты.

Теория генетической предрасположенности к той или иной микробиоте не нашла подтверждения. Так, микробиота однояйцевых близнецов не более похожа, чем микробиота других членов семьи⁶⁹. Однако, в рамках одной семьи, микробиота достаточно схожа⁷⁰.

Ближе к преклонному возрасту микробиота кишечника вновь претерпевает изменения, в частности снижается разнообразие микробиоты, повышается доля *Proteobacteria*⁷¹. Как гласит закон Эшби, адаптированный для экологии, чем экосистема разнообразнее, тем она устойчивее, а следовательно, меньше подвержена внешним вредоносным влияниям.

Было выявлено отличие в микробиоте пожилых людей, проживающих в стационаре и в семье⁷¹. На уровне отделов, у первых был выше уровень *Bacteroidetes*, у вторых – *Firmicutes*, а также неклассифицированные бактерии. Исследование микробиоты людей, перешедших вековой рубеж⁶² выявило различие на функциональном уровне. Было найдено, что с возрастом в микробиоте кишечника увеличивается представленность генов, отвечающих за метаболизм триптофана. Это коррелирует с понижением уровня триптофана в сыворотке крови старцев. Видимо, увеличение потребления триптофана микробиотой, делает его менее доступным для хозяина. Также было показано пониженное содержание триптофана в сыворотке крови у пациентов с воспалительным заболеванием кишечника. Кроме того, снижение уровня триптофана ассоциировано со слабоумием, а диета с высоким содержанием триптофана редуцирует бляшки в мышцах, больных болезнью Альцгеймера. Было найдено, что с возрастом уменьшается количество генов, участвующих в синтезе КЖК и общее снижение сахаролитического потенциала.

В течение нескольких лет мировому научному сообществу удалось охарактеризовать метагеном кишечника самых различных географических и социологических групп (Таблица 1). Выявлены отличия между жителями

сельских районов и горожанами. Также охарактеризованы глобальные изменения микробиоты от младенчества до глубокой старости.

Таблица 1.

Результаты сравнительных метагеномных исследований

Исследование	De Filippo et al, 2010	Yatsunenko et al, 2012	Schnorr et al, 2014	Yatsunenko et al, 2012	Claesson et al, 2012	Rampelli, 2013
Первая группа	Европейские дети	США	Италия	Взрослые	Пожилые (стационар)	Старцы (более 100 лет)
Вторая группа	Африканские дети	Малави, Венесуэла	народ Хадза, Африка	Младенцы	Пожилые (семья)	Пожилые
Преобладание в первой группе	Bacteroides	Разложение глутамина и других аминокислот	Bifidobacterium, Bacteroides, Blautia, Dorea, Roseburia, Faecalibacterium и Ruminococcus	Разложение фолиевой кислоты и тетрагидро фолата, биосинтез кобаламина	Bacteroidetes	Метаболизм триптофана
Преобладание во второй группе	Prevotella	Глутамат-синтазы, амилаза	Prevotella, Eubacterium, Oscillibacter, Butyricicoccus, Sporobacter, Succinivibrio и Treponema	Биосинтез фолиевой кислоты	Firmicutes	

1.4 Микробиота кишечника человека при различных патологиях

После того, как были проанализированы сотни кишечных метагеномных образцов, сложилось понимание того, какой состав микробиоты можно назвать здоровым, что есть нормофлора. Следующий шаг состоял в выявлении различий между микробиотой здоровых людей и пациентов, как с кишечными заболеваниями, так и с системными болезнями, затрагивающими весь организм в целом.

1.4.1 Воспалительные заболевания и рак кишечника

Воспалительные заболевания кишечника (ВЗК, англ. Inflammatory Bowel Disease, IBD) это группа патологических состояний кишечника, включающая в

себя болезнь Крона и язвенный колит⁷². Логичным предположением стало, что микробиота кишечника может влиять на течение болезни. Действительно, часто наблюдается дисбиоз¹⁴. Состав микробиоты пациентов с ВЗК характеризовался сниженным количеством бактерий Firmicutes, в частности таких продуцентов бутирата, как *Clostridium* IXa и IV группы и *Bifidobacterium*. При этом у пациентов с ВЗК было увеличено количество сульфат-редукторов и *E. coli*²⁵. Повышение заболеваемости ВЗК и другими аутоиммунными заболеваниями в развитых странах связывают с т.н. гигиенической гипотезой», согласно которой чрезмерное использование антибиотиков и средств гигиены вносит разлад во взаимодействие иммунной системы с «необходимыми» бактериями⁷³.

Воспалительные процессы в мукозном слое повышают риск развития рака кишечника. Как было сказано ранее, бутират является противовоспалительным фактором, но в условиях воспаления количество продуцентов бутирата снижено. Пациенты с колоректальным раком также демонстрируют пониженный уровень продуцентов бутирата *Eubacterium rectale* и *Faecalibacterium prausnitzii* и увеличенный уровень *Enterococcus faecalis*. Супероксид-продуцирующая бактерия *E. faecalis* вызывает повреждения ДНК в эпителии, повышая риск развития рака⁷⁴. На этих наблюдениях основана гипотеза, что сдвиг в составе кишечной бактериальной популяции может быть важным фактором, приводящим к раку²⁶. В значительных количествах в тканях колоректальной опухоли была обнаружена бактерия *Fusobacterium nucleatum* – патоген ротовой полости, обычно вызывающий периодонтит⁷⁵.

1.4.2 Патогенные микроорганизмы и инновационные методы лечения

Обитатели кишечника сами по себе могут стать возбудителями заболеваний. Так клостридиальная инфекция, вызываемая бактерией *Clostridium difficile*, является тяжелым кишечным заболеванием, почти не поддающееся лечению антибиотиками. Даже в случае облегчения симптомов, велика вероятность рецидива. Метагеномные исследования позволили научно обосновать

эффективность такой нестандартной методики лечения, как трансплантация фекальной микробиоты (англ. fecal mass transplantation, FMT)⁷⁶. Суть метода заключается в пересадке кишечной микробиоты от здорового с целью вытеснения патогена. Пока этот метод находится в экспериментальной стадии разработки и стандартизации, но уже показано, что эффективность втрое превышает эффективность антибиотиков⁷⁷. Впрочем, идея не нова, Илья Мечников еще более века назад предложил использовать полезные микроорганизмы для замещения вредоносных⁷⁸.

1.4.3 Метаболические патологии и роль микробиоты

Микробиота кишечника затрагивает функционирование организма хозяина целиком, его метаболизм. Комплексные заболевания, такие как ожирение, аутизм, диабет и аллергия, были ассоциированы с дисбалансом в составе кишечной микробиоты.

Изменения в составе микробиоты могут обуславливать дополнительный «сбор энергии» из клетчатки⁷⁹, в частности имеет значение соотношение бактерий из отделов *Bacteroidetes* и *Firmicutes*. Последние являются основными продуцентами бутирата – главного источника энергии для энтероцитов, и могут быть задействованы в развитии ожирения.

Было проведено исследование отличий кишечной микробиоты людей, больных диабетом второго типа: у них преобладали бактерии рода *Bacteroides* и было меньше бактерий рода *Prevotella* по сравнению с контрольной выборкой. Представленность *Bifidobacterium spp* и *Bacteroides vulgatus* была ниже в опытной группе, а *Clostridium leptum* – выше, однако статистически достоверного результата получено не было²⁸. Так же есть работы по выявлению генных маркеров в метагеноме больных диабетом^{16,80}. Примечательно, что список этих маркеров различается в европейской и азиатской популяции, что говорит о решающей роли образа жизни.

Поскольку микробиота кишечника влияет не только на метаболизм организма хозяина, но и на его иммунитет, была исследована возможная связь

между ней и заболеваниями с аутоиммунными компонентами, например, атеросклерозом. Атеросклероз развивается вследствие накопления холестерина и макрофагов в артериальной стенке. На мышинных моделях было показано, что кишечная микробиота метаболизирует пищевой фосфатидилхолин в триметиламин, который способствует развитию атеросклероза и воспаления. В атеросклеротических бляшках была обнаружена бактериальная ДНК, по своему филогенетическому составу похожая на состав кишечной микробиоты, а количество бактериальной ДНК в атеросклеротических бляшках коррелировало с тяжестью воспаления⁸¹. На функциональном уровне, гены метаболизма пептидогликанов были перепредставлены у пациентов, в то время как у контрольной здоровой группы был повышен уровень генов, задействованных в синтезе противовоспалительных агентов (бутират) и антиоксидантов¹⁵.

Определенный интерес представляет микробиота людей со значительными нарушениями метаболизма. В частности, интерес представляет микробиота больных алкоголизмом. Часто у них развивается дисбиоз, при этом уменьшается представленность бактерий отдела *Bacteroidetes* и увеличивается доля представителей отдела *Proteobacteria*, включая *E. coli*, являющейся в больших количествах маркером воспаления. Это коррелирует с высоким уровнем содержания эндотоксина. В здоровой микробиоте, в отличие от микробиоты алкоголиков, наблюдается большая связность и разнообразие, и как следствие, большая возможность для адаптации. Микробиота же алкоголиков целиком направлена на то, чтобы бороться с последствиями потребления алкоголя⁸².

Другие нарушения питания и метаболизма также сказываются на микробиоте. Квашиоркор – вид тяжелой дистрофии, развивающийся при недостатке белковой пищи и незаменимых аминокислот в частности, а также микроэлементов. В моделях на мышах было показано преобладание водород-потребляющей, сульфит редуцирующей бактерии *Bilophila wadsworthia*, ранее ассоциированной с воспалительными заболеваниями кишечника в человеке и индуцирующей Т-хелперы. Также преобладает *Clostridium innocuum*. При изменении диеты, состав значительно менялся, увеличивался уровень

Bifidobacteria (*B. longum*, *B. bifidum*), *Lactobacilli* (*L. reuteri* и *L. gasseri*) и *Ruminococcus*. При этом уменьшался уровень *Bacteroidales* (*B. uniformis*, *Parabacteroides distasonis*)⁸³.

1.4.4 Последствия приема антибиотиков для микробиоты кишечника

Не только заболевания (Таблица 2), но и лечение может сказываться на микробиоте. Урон от приема антибиотиков весьма значителен для микрофлоры кишечника, при этом восстановление занимает длительный срок⁸⁴. Было показано, что краткосрочное применение антибиотиков может привести к селективному преимуществу популяции резистентных бактерий в кишечнике человека, которое будет сохраняться годами^{85,86,87}. После начальной селекции членов комменсальной микробиоты, обладающих генами резистентности к антибиотикам, потенциально эти гены могут быть перенесены в геномы патогенов путем горизонтального переноса. Известен случай переноса плазмиды с геном бета-лактамазы от резистентного штамма *E. coli* к изначально чувствительному штамму в микробиоте ребенка при лечении ампициллином⁸⁹. Некоторые бактерии являются лишь транзитными обитателями кишки, например, поступающими в ЖКТ с пищевыми продуктами. Однако они также могут нести в своих геномах детерминанты устойчивости и передавать их комменсальной микробиоте по мере прохождения^{89,90}. Использование субтерапевтических доз антибиотиков в сельском хозяйстве в качестве стимуляторов роста у скота и в целях профилактики заболеваний с учетом сохранения антибиотиков в мясе приводит к тому, что человек также получает субтерапевтические дозы антибиотиков, таким образом, происходит рост уровня генов резистентности в микробиоте человека. Это вносит свой вклад в мировой резервуар резистентности, который также влияет через оппортунистические патогены и на клиническую сферу^{91,92}. Микроорганизмы, в норме не приносящие вреда, но несущие гены резистентности к антибиотикам, становясь в определенных

условиях вредоносными (например, при ослаблении иммунитета), уже будут сложнее подвергаться лечению.

Таблица 2.

Отражение патологий на состав кишечной микробиоты

Патология	Понижено	Повышено
ВЗК	Firmicutes (<i>Clostridium</i> IXa и IV группы и <i>Bifidobacterium</i>)	Сульфат-редукторы (<i>Bilophila wadsworthia</i>), <i>E. coli</i>
Колоректальны м рак	<i>Eubacterium rectale</i> и <i>Faecalibacterium prausnitzii</i>	<i>Enterococcus faecalis</i> , <i>Fusobacterium nucleatum</i>
Диабет 2 типа	<i>Prevotella</i> , <i>Bifidobacterium spp</i> и <i>Bacteroides vulgatus</i>	<i>Bacteroides</i> , <i>Clostridium leptum</i> (?)
Атеросклероз	синтез бутирата и антиоксидантов	Метаболизм пептидогликанов
Алкоголизм	Bacteroidetes	Proteobacteria (<i>E. coli</i>)
Квашиоркор	<i>Bifidobacteria</i> (<i>B. longum</i> , <i>B. bifidum</i>), <i>Lactobacilli</i> (<i>L. reuteri</i> и <i>L. gasseri</i>) и <i>Ruminococcus</i>	<i>Bilophila wadsworthia</i> , <i>Clostridium innocuum</i>

После курса лечения антибиотиками пациентам назначаются про- и пребиотические препараты. Пребиотики – селективно ферментируемые компоненты, которые способствуют определенным изменениям в составе и/или активности желудочно-кишечной микрофлоры, приводящим к улучшению здоровья организма. Современные пребиотики - это в основном плохо усваиваемые самим человеком углеводы, являющиеся природными компонентами еды. В их число входят ксилоолигосахариды, галакто-олигосахариды и фруктаны, включая инулин и фруктоолигосахариды³⁶. Пребиотики также используются в связке с пробиотиками. В исследовании группы пожилых людей, которые принимали пробиотическую смесь *Bifidobacterium bifidum* и *B. lactis*, а также инулин с фруктоолигосахаридами в качестве пребиотиков, было обнаружено увеличение выживаемости принимаемых бифидобактерий⁹³.

1.4.5 Генные маркеры патологических состояний

При различных патологических состояниях происходит уменьшение количества видов и родов микроорганизмов и, как следствие, количества генов. На выборке из образцов микробиоты кишечника от 292 человек было показано бимодальное распределение количества генов в их метагеномах. Пики

распределения приходились примерно на значения 380 тысяч и 640 тысяч генов. Соответственно, метагеномы были разбиты на две группы - с низким количеством генов (low gene count, LGC) и с высоким количеством (high gene count, HGC). Среди первых была статистически достоверно повышена доля пациентов с воспалительными заболеваниями кишечника, а также пожилых пациентов с воспалением и больных ожирением. Между этими группами значительно различалась представленность 46 родов. *Bacteroides*, *Parabacteroides*, *Ruminococcus*, *Campylobacter*, *Dialister*, *Porphyromonas*, *Staphylococcus* и *Anaerostipes* доминировали в LGC группе, а 36 родов, включая *Faecalibacterium*, *Bifidobacterium*, *Lactobacillus*, *Butyrivibrio*, *Alistipes*, *Akkermansia*, *Coprococcus* и *Methanobrevibacter* были значительно ассоциированы с HGC группой. Такие виды, как *Faecalibacterium prausnitzii*, которые являются продуцентами бутирата и редко встречаются при воспалениях, были ассоциированы с HGC, а провоспалительные виды (*Bacteroides* и *R. gnavus*) – с LGC⁹⁴. Принадлежность к группе LGC или HGC ассоциирована с наличием или отсутствием ВЗК соответственно⁹⁵. Результаты показывают, что такая характеристика, как число генов в метагеноме, может со временем стать диагностическим инструментом для детекции кишечных воспалений и метаболического синдрома.

1.5 Методика проведения метагеномных исследований

Исследование микробиоты является сложной экспериментальной и математической задачей. Изначально, состав микробиоты оценивался с помощью бактериального посева, и до сих пор это является основной диагностической методикой в медицине. Но у этого метода есть ряд недостатков. Микрофлора кишечника состоит из факультативно аэробных и строго анаэробных микроорганизмов, чья культивация весьма затруднительна. При высевании теряется количественная оценка микробиоты. Поэтому, исследователи начали изучать микробиоту при помощи секвенирования. Здесь есть два основных подхода. Первый, и наиболее применяемый ввиду своей доступности,

относительной дешевизны и методической простоты, является секвенирование 16S рРНК генов – основного фактора классификации бактерий и архей. Таким образом, удастся оценить филогенетическое разнообразие микробиоты, в том числе выявить новые микроорганизмы. Количественная оценка также возможна при использовании этого метода, с поправкой на копийность 16S рРНК генов в геномах. Однако, в этом случае теряется информация о функциональном составе микробиоты: нельзя детектировать метаболические пути и гены, можно лишь строить предположения, исходя из известных последовательностей геномов найденных бактерий. Для решения этой проблемы применяют куда более дорогостоящий и сложный с точки зрения обсчета и эксперимента подход – полногеномное секвенирование (Whole Genome Sequencing, WGS). При этом, выделяется тотальная ДНК из метагеномных образцов и секвенируется на высокопроизводительных приборах (Illumina, ABI SOLiD). На выходе с секвенаторов получают гигабайты данных с информацией о короткоцепочечных (50-100 нуклеотидов, в зависимости от прибора) нуклеотидных последовательностях, именуемые риды. Помимо стандартной фильтрации по качеству, риды проходят проверку на наличие загрязнения человеческим геномом путем картирования на него.

Состав метагенома, как геномный, так и генный, определяется путем картирования ридов на референсные наборы последовательностей. Если первые исследователи были вынуждены картировать риды на всю мировую базу данных нуклеотидных последовательностей из NCBI, то теперь, благодаря усилиям консорциумов MetaHIT и HMP существуют референсные наборы геномов кишечных бактерий⁷⁰ и избыточный каталог преобладающих 3,5 млн. бактериальных генов⁷. Бывают также специфические референсные наборы, в зависимости от целей исследования, например, используют базу данных генов устойчивости к антибиотикам ARDB⁹⁶. Само картирование может проводиться различными способами в зависимости от длины ридов. При длине от 100 нуклеотидов, возможна *de novo* сборка кантиков из ридов. Далее полученные более длинные последовательности либо выравниваются на набор референсных

геномов, либо в них производится поиск открытых рамок считывания, и уже они выравниваются на каталог генов⁹⁷.

Однако, при меньшей длине, например 50 нуклеотидов, получаемой при секвенировании на приборе ABI SOLiD 4, сборка контигов представляется едва ли возможной, слишком высока вероятность получения «химер» – контигов, собранных из оригинально разных геномов. Поэтому в этом случае риды сразу картируют на референс. Для этого существуют специальные программы, наиболее распространенная – bowtie⁹⁸.

Степень представленности той или иной референсной последовательности в метагеноме высчитывается в основном как отношение количества выровненных нуклеотидов на длину референсной последовательности с поправкой на общую длину картированных ридов⁹. Иногда высчитывается гипотетическое значение представленности тех геномов или генов, которых нет в референсном наборе, но присутствуют в метагеноме⁹. В случае функционального анализа также используют raw counts – количество ридов, картировавшихся на референсную последовательность, без каких-либо нормировок или поправок. Эти значения используются в анализе, схожим с транскриптомным анализом на микрочипах, для этого существуют готовые библиотеки для статистического языка программирования R, например пакет metagenomeSeq⁹⁹.

Метагеномные данные, поступающие в статистическую обработку, от WGS или секвенирования генов 16S рРНК, имеют вид набора векторов представленности, где по одной оси расположены образцы, по другой – признаки, гены или геномы, а значения – их представленность в данном образце, их процентное содержание, также по этим данным можно оценить состав всей выборки, альфа- и бета-разнообразие¹⁰⁰.

Для сравнения образцов между собой необходимо ввести меру сходства, т.к. рассматривать каждый признак в отдельности крайне непродуктивно. Для этого считают расстояния между образцами, основываясь на векторах представленности. Есть несколько наиболее используемых метрик в метагеномике. Одна из первых используемых было расстояние Дженсона-

Шэннона¹⁰¹, позволяющее измерить сходство между двумя распределениями случайных величин. Впоследствии более используемым стало расстояние по Брэй-Кертису¹⁰², часто встречающееся в экологических исследованиях. Метагеномной инновацией стала метрика UniFrac, используемая для изучения метагеномов по 16S рРНК генам¹⁰³. Ее особенность заключается в том, что она учитывает филогенетические расстояния между бактериями. Данная метрика включена в популярный сервис анализа данных по 16S рРНК генам QIIME¹⁰⁴.

Имея матрицу расстояний, становится возможным представить изначально многомерные данные в двухмерном пространстве, с помощью MDS (англ. Multidimensional Scaling)¹⁰⁵. Еще один способ визуально оценить метагеномные образцы это метод главных компонент (англ. Principal Component Analysis, PCA). С помощью этого метода можно выявить, по каким признаком наблюдается наибольший разброс образцов, а также оценить результаты кластеризации⁹.

Часто необходимо выявить различающиеся подгруппы образцов. Для этого использует алгоритмы кластеризации. Кластеризации могут быть иерархическими, подразумевающие построение дерева образцов по расстояниям между ними, и неиерархические. К последним относится наиболее востребованный в метагеномике метод k-средних. Его суть в том, что он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров (1).

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (1)$$

k — число кластеров, S_i — полученные кластеры,
 $i = 1, 2, \dots, k$ и μ_i — центры масс векторов $x_j \in S_i$.

Он использовался для выявления энтеротипов: задается количество кластеров, проводится кластеризация, ее качество оценивается индексом Калински-Гарабача, который является отношением суммы квадратов расстояний между образцами из

разных кластеров к сумме квадратов расстояний между образцами внутри кластеров⁹. Это итеративный алгоритм, воспроизводимый до тех пор, пока индекс качества не достигнет своего максимума. Недостаток этого метода заключается в том, что всегда будет хотя бы два кластера, даже если они практически неразличимы. В качестве независимой оценки качества кластеризации используют среднюю силуэтную ширину.

В случае исследования различий между метагеномами здоровых людей и пациентов, проводят дискриминантный анализ. Какой-либо устоявшейся схемы анализа не существует. Часто используют линейные регрессионные модели¹⁰⁶, дисперсионный анализ (ANOVA) или алгоритм Random Forest¹⁰⁷, или иные комбинации статистических методов.

Не существует универсального протокола обработки метагеномных данных, каждое следующее исследование добавляет модификации. Но в целом, алгоритм почти не меняется (Рисунок 2).



Рисунок 2. Схема анализа метагеномных данных от полногеномного секвенирования.

1.6 Актуальность метагеномных исследований в Российской Федерации

К настоящему времени исследование метагенома кишечника человека является одной из центральных тем биомедицинских исследований. Исследовано несколько тысяч метагеномных образцов человека по всему миру методами секвенирования 16S РНК генов и полногеномного секвенирования. Накоплено много данных, позволяющих охарактеризовать нормофлору, найдены некоторые зависимости состава микробиоты от образа жизни, возраста и т.д. Начинается период клинических исследований, изучение патологий желудочно-кишечных и аутоиммунных заболеваний, обусловленных микробиотой. Кроме того, исследуются механизмы взаимосвязи организма хозяина и кишечных микробных сообществ. Методы исследования также значительно продвинулись вперед, но какого-либо единого цельного протокола исследования пока невыбранное.

Тем не менее, до проведения настоящего исследования оставался неохваченным значительный многокультурный регион – Российская Федерация. Особенность изучения микробиоты жителей нашей страны заключается в чрезвычайном разнообразии образов жизни. Здесь есть как мегаполисы мирового уровня, возможные источники урбанистической микробиоты, так и удаленные регионы, в которых люди продолжают соблюдать традиционный уклад. Изучение российского метагенома, как на таксономическом, так и на функциональном уровне, потенциально может значительно расширить рамки определения здоровой микрофлоры, выявить новые зависимости состава от образа жизни.

Глава 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1 Забор образцов кала

Образец кала забирали у здоровых людей в возрасте 36 ± 18 лет (Приложение 2) 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.

2.2 Выделение ДНК из кала

К замороженной навеске образца кала (150 мг) добавляли кремнево-циркониевые бусины (BioSpec Products, США) диаметром 0,1 мм (300мг) и 0,5 мм (100мг), а затем 1200 мкл теплого лизирующего буфера (50мМ Tris-HCl, pH 8,0, 500мМ NaCl, 50мМ 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°C 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°C). 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°C не менее часа. После этого образцы центрифугировали при 14000 об/мин 20 мин. Сформировавшийся осадок дважды промывали 1000 мкл 80%

этанолом, 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°C 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.

2.3 Методы секвенирования

Подготовку shotgun-библиотек и их секвенирование с использованием генетического анализатора SOLiD 4 (Life Technology, США) осуществляли согласно рекомендациям производителя с использованием наборов SOLiD™ Fragment Library Construction Kit, SOLiD™ Fragment Library Barcoding Module 1 – 16, SOLiD™ EZ Bead™ E80 System Consumables, SOLiD™ ToP Sequencing Kit. Фрагментные библиотеки были созданы из 5 мкг тотальной ДНК для каждого образца с баркодами. Были получены риды по флагу F3 длиной 50 нуклеотидов.

Подготовку шотган-библиотек и их секвенирование с использованием генетического анализатора Ion Torrent PGM (Life Technology, США) осуществляли согласно рекомендациям производителя с использованием наборов Ion Xpress™ Plus Fragment Library Kit, Ion OneTouch™ Template Kit, Ion Sequencing Kit, Ion OneTouch™ 200 Template Kit, Ion Sequencing 200 Kit, Ion 318 Chip Kit.

Подготовку shotgun-библиотек и их секвенирование с использованием генетического анализатора GS FLX+ (Roche, США) осуществляли согласно рекомендациям производителя с использованием наборов GS Rapid Library Prep

Kit, GS Titanium SV emPCR Kit (Lib-L) v2, GS Titanium LV emPCR Kit (Lib-L) v2 и GS FLX Titanium Sequencing Kit XL+.

Подготовку shotgun-библиотек и их секвенирование с использованием генетического анализатора HiSeq 2000 (Illumina) осуществляли согласно рекомендациям производителя с использованием наборов TruSeq DNA sample prep kit v.2, TruSeq PE Cluster Kit v3-cBot-HS TruSeq SBS Kit v3-HS с длиной чтения 101 п. н. с каждого конца фрагмента. Демультимплексирование проводилось с помощью программы CASAVA v. 1.8.2.

Подготовка фрагментной библиотеки ДНК и полногеномное секвенирование на платформе SOLiD 5500 W (Life Technologies, Foster City, CA, USA) были произведены в соответствии с инструкциями от производителя с применением следующих наборов: 5500 SOLiD™ Fragment Library Core Kit, 5500 SOLiD™ Fragment Library Barcode Adaptors 1-16, 5500 W Conversion Primers Kit, 5500 W FlowChip V2, 5500 W FlowChip Prep Pack, 5500 W Template Amplification Kit v2, 5500 W FWD1 SP Kit, Double, 5500 W FWD2 SP Kit, Double, 5500 W FWD SR Kit, Double, 5500 W FWD Ligase Kit, Double, 5500 W Run Cycle Buffer Kit, 5500 W FWD Buffer, Double, 5500 W Buffer D. Выходная длина ридов составила 75 п.н.

2.4 Дополнительные метагеномные данные

Для сравнительного анализа образцов российской кишечной микробиоты были использованы общедоступные метагеномные данные: 85 наборов ридов из Дании⁷, 137⁸ из США, 10 и 5 из Венесуэлы и Малави соответственно¹⁷, 69¹⁶ из Китая (Приложение 1).

2.5 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 были прошли фильтрацию программой SAET (SOLiD Accuracy Enhancement Tool), с ранее описанными параметрами¹⁰⁸. Далее риды подверглись редактированию по качеству: все позиции начиная с 5' были удалены вплоть до первой высококачественной позиции ($QV \geq 30$). Все риды, чья длина после фильтраций стала меньше 30 нуклеотидов, были удалены.

Высококачественные риды были картированы на геном человека версии hg18 с использованием программы bowtie⁹⁸. Каждый рид был картирован без допущения вставки или делеции, с максимальным количеством несовпадений 3. В том случае, если рид мог картироваться в 2 и более мест, референс выбирался случайно и равновероятно. Не картировавшиеся риды поступили в дальнейший анализ.

2.6 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 bowtie⁹⁸. 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 3.0¹⁰⁹. 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 (Приложение 2).

В ходе выравнивания набора ридов на референс была получена информация из файлов формата BAM о двух видах покрытия с использованием программного пакета BEDtools¹¹⁰: суммарная длина ридов, картировавшихся на референс

(покрытие в глубину) и суммарная длина позиций референса, оказавшаяся покрытой (покрытие в ширину). Покрытие в ширину было использовано, как порог детектирования референса в метагеноме: должно быть покрыто ридом по крайней мере 1% длины генома., чтобы признать его наличие. Покрытие каждого генома было нормализовано на общую длину картировавшихся ридов и его длину (2) для получения относительной представленности.

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

Относительная представленность бактериальных и архейных родов суммировалась по формуле 3.

$$\text{Относительная представленность рода} = 1E12 \times \left(\frac{\sum_{\text{по геномам}} (\sum \text{длин картировавшихся ридов} / \text{длина генома})}{\text{общая длина картировавшихся ридов образца}} \right) \quad (3)$$

Принадлежность каждого вида к тому или иному роду определялась с использованием классификации по базе данных последовательностей 16S рРНК генов RDP¹¹¹.

2.7 Реализация алгоритма поточной обработки данных

Алгоритм поточной обработки метагеномных данных представлен на Рисунке 3. Он реализован в виде конвейера, получающего 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 происходит картирование на референсные наборы. Информация о покрытиях заносится в СУБД, при выгрузке из нее покрытия нормализуются и на выходе получают файлы с векторами представленностей 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.

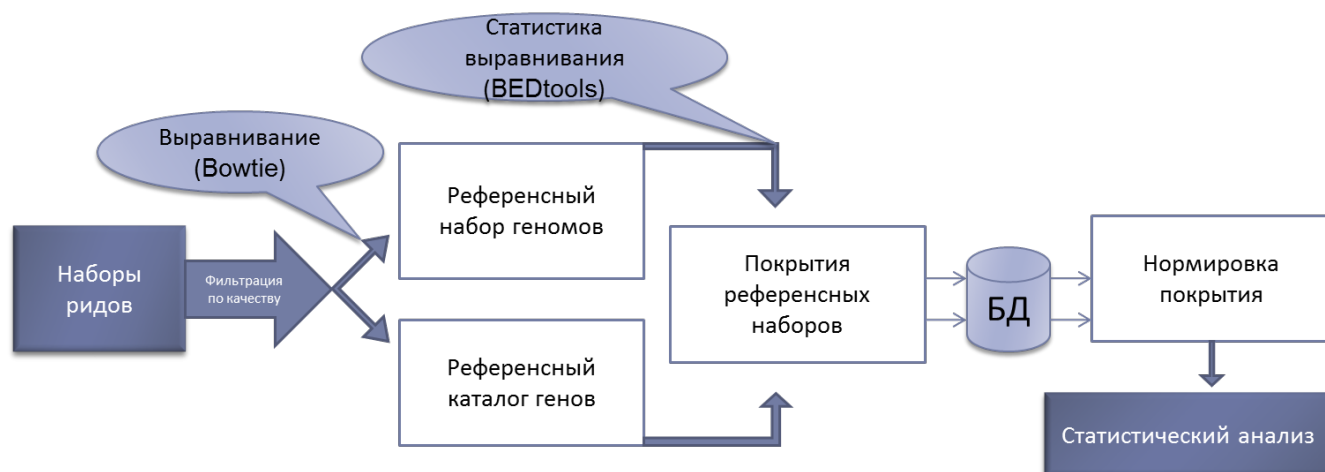


Рисунок 3. Алгоритм поточной обработки метагеномных данных.

2.8 Статистический анализ и визуализация

Статистический анализ проводился на языке программирования R 3.1.0¹¹². 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. Для построения матрицы расстояний по первым 4 метрикам были использованы стандартные команды языка R, для расстояния Брэя-Кертиса была использована команда `bcdist` из пакета `ecodist`¹¹³. Расстояние Дженсена-Шэннона было посчитано специально написанным скриптом на R.

Изначально UniFrac был создан для анализа метагеномов по 16S рРНК генам, поэтому была проведена его адаптация для полногеномных данных. Были найдены последовательности 16S рРНК микроорганизмов, чьи геномы были использованы в референсном наборе, либо их ближайших родственников в пределах одного рода. Последовательности генов 16S рРНК были выравнены программой множественного выравнивания MUSCLE¹¹⁴ и по полученному выравниванию было построено филогенетическое дерево в формате `newick` алгоритмом UPGMA¹¹⁵. Каждой паре образец-геном (используемому для UniFrac) был приписан вес, как относительная представленность этого генома в данном образце. Геномы с нулевой представленностью по всем образцам были удалены.

Далее были использованы скрипты из программного пакета QIIME¹⁰⁴, а именно `convert_biom.py` для преобразования данных о представленности в формат `biom` и `beta_diversity.py` для получения собственно взвешенных расстояний UniFrac по созданной `biom` таблице и полученному филогенетическому дереву референсных геномов.

Кластеризация образцов проводилась двумя методами. Первая кластеризация была сделана по методу определения энтеротипов⁹, использующий метод k-медоед. На вход подавалась таблица расстояний между образцами, посчитанная по метрике Дженсена-Шэннона. Для кластеризации использовалась функция `ram` пакета `cluster`¹¹⁶ языка R. Оптимальное число кластеров определялось путем максимизации индекса Калинского-Гарабача.

Качество полученной кластеризации было оценено значением средней силуэтной величины (функция `silhouette` из пакета `cluster`)¹¹⁷, предсказательной силой (англ. *predictive strength*)¹¹⁸. Дополнительно, было проведено сравнение средней силуэтной величины полученных кластеризаций со случайным разбиением образцов на такое же количество кластеров, повторенное 1000 раз.

Второй тип кластеризации происходил с использованием функции `pvclust`¹¹⁹ и основывался на бутстрэп методе, заключающимся в том, чтобы из имеющейся выборки образцов сделать большое количество других выборок из этих же элементов, но в случайном порядке, причём 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*)¹²⁰.

При построении дерева всех мировых образцов был использован веб-сервис iTOL¹²¹.

2.9 Моделирование производства и тока короткоцепочечных жирных кислот

1. Данные.

Модель была построена по данным о 98 индивидуумах, для которых есть метаданные (возраст, пол, индекс массы тела), известен родовой состав микробиоты по результатам секвенирования генов 16S РНК (численность основных кластеров и некоторых семейств), а также измерены концентрации основных КЖК в фекалиях (мМ)¹²².

2. Описание математической модели

Модель была построена и обчислена совместно с К.В. Песковым и Ю.А. Косинским («Новартис Фарма»). 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-Ментена (4-6).

Синтез ацетата (Acet) из субстрата (S):

$$V_Acet_pro = k_S_Acet \times S / (S + Km_S_Acet) \quad (4)$$

Синтез пропионата (Prop) из субстрата:

$$V_Prop_pro = k_S_Prop \times S / (S + Km_S_Prop) \quad (5)$$

Синтез бутирата (But) из ацетата:

$$V_But_pro = k_Acet_But \times Acet / (Acet + Km_Acet_But) \quad (6)$$

В модели на образование каждого моля ацетата или пропионата расходуется один моль субстрата, на образование моля бутирата тратится два моля ацетата.

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 (см. Приложение 3).

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

3. Оценка физиологических интервалов значений параметров

1) Процессы однонаправленного переноса КЖК и субстрата между отсеками кишечника с постоянной константой скорости, обозначены в модели k_{gut_out} . Из физиологических оценок времени (colonic transition time, СТТ) пребывания субстрата в толстом кишечнике человека 30-40 часов, $k_{gut_out} \sim 0,08$ 1/ч для случая четырёх-частной модели. По некоторым оценкам¹²³ СТТ значительно варьирует в популяции даже у здоровых – от 12 до 80 часов.

2) Из анализа литературы¹²⁴ было заключено, что наиболее вероятная система транспорта КЖК – бикарбонат-зависимый антипортер. Из опубликованных данных эксперимента *in vitro*¹²⁵ были оценены значения K_m для транспорта основных КЖК. Для ацетата и пропионата $K_m = 15,0$ мМ, а для бутирата $K_m = 21,3$ мМ. Измерения общей концентрации КЖК в содержимом кишечника $\sim 100 - 150$ мМ позволяют заключить, что система транспорта КЖК энтероцитов близка к насыщению.

3) Физиологическая оценка нормального уровня абсорбции КЖК колоноцитами 400-600 мМ/день¹²⁶. Это соответствует 24 мМ/ч или абсорбции КЖК 6 мМ/ч в каждом из 4х отсеков модели. В модели было зафиксировано значение $k_{transp} = 8,3$ мМ/ч для транспортной системы энтероцитов в каждом отсеке. Иначе говоря, в модели суммарная абсорбции КЖК не превышает 33,2 мМ/ч (будет стремиться к этому значению при высокой концентрации КЖК). При физиологических концентрациях КЖК их суммарный поток через транспортер составит 6 - 7 мМ/ч.

4) Известно, что большая часть субстрата поступает из пищи, а меньшая – вырабатывается энтероцитами (муцин). Исходя из того, что большая часть субстрата должна метаболизироваться, в модели предполагается постоянный приток субстрата в 1ый отсек, равный 40 мМ/ч (параметр k_{nutr_in1}) и эндогенный приток, равный 1 мМ/ч в каждый из 4х отсеков. Значение k_{nutr_in1} может значительно варьировать из-за различий в питании.

5) Объем 0.2 л каждого из четырех отсеков кишечника был выбран из соображений баланса притока субстрата и количества КЖК на выходе. Экскреция КЖК с фекалиями составляет 10-30 мМ/день¹²⁷, в данной модели эта величина составит 29 мМ/день, что укладывается в физиологический интервал.

Для нахождения значений параметров и анализа ковариационных зависимостей была использована программа MONOLIX¹²⁸.

4. Выбор статистической модели в MONOLIX

Был использован популяционный статистический подход для определения значений параметров, т.к. такие физиологические параметры, как скорость транзита КЖК между отсеками и скорость потребления субстрата может значительно варьировать между индивидуумами. В этом подходе значения параметров модели могут учитываться как случайные величины, распределенные по некоторому закону, например, лог-нормально. В этом случае результатом будет не только средне-популяционное значение, но и дисперсия его распределения (ω). Если для такого параметра была включена интер-индивидуальная вариабельность, то фиксировалось среднее значение, а значение дисперсии распределения (ω) оптимизировалось во время процедуры. Кроме того, в популяционном подходе есть возможность учесть зависимость индивидуальных значений параметров от значений ковариат – некоторых численных характеристик индивидуума.

1) При исследовании интер-индивидуальной вариабельности для всех параметров предполагалось лог-нормальное распределение. Для индивидуума i значение некоторого параметра можно представить как:

$$P_i = P \times \exp(\eta_i), \quad (7)$$

где η_i случайное значение из $N(0; \omega)$; P – оценка среднего популяционного значения параметра; ω – оценка дисперсии.

2) Для измерений концентраций КЖК принята пропорциональная модель остаточной ошибки. Для индивидуума i измерение концентрации КЖК j (1 - ацетат; 2- пропионат; 3- бутират) можно представить в виде:

$$C_{ij} = f_j(P_i) \cdot (1 + b_j \times e_{ij}), \quad (8)$$

где e_{ij} случайное значение из $N(0; 1)$; $f_j(P_i)$ – предсказание концентрации КЖК j в модели с использованием вектора индивидуальных значений параметров P_{iu} .

6. Поиск ковариационных зависимостей

В качестве ковариат пробовали различные метаданные и представленность родов. При выборе ковариат были использованы следующие критерии:

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.

Глава 3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЕ МЕТАГЕНОМНЫХ ОБРАЗЦОВ: РАЗРАБОТКА МЕТОДОЛОГИИ И АНАЛИЗ

3.1 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 SOLiD. Полностью автоматизированный конвейер работает и используется в исследованиях на сервере ФГБУН НИИ ФХМ ФМБА России. Обработка ридов включает в себя следующие стадии:

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. Поскольку прибор не проводит никакой префильтрации, необходим этап отсеивания низкокачественных ридов: в анализ проходят риды, со средним баллом качества выше 15. Таким образом отсеиваются риды не только с точечными ошибками секвенирования, но и риды, полученные с бусин низкого разрешения и поликлональные риды, которые возникают в ходе амплификации на одной бусине нескольких разных последовательностей¹²⁹.

Помимо фильтрации по среднему значению качества, используется алгоритм коррекции ридов SAET, который определяет низкокачественные 3` концы и обрезает их. Это связано с особенностями секвенатора.

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. Эта стадия также позволяет выявить образцы с аномально высоким содержанием человеческой ДНК, что может служить косвенным признаком воспалительных процессов или чрезмерного отшелушивания кишечного эпителия.

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

Данный каталог включает набор геномов кишечных бактерий из базы данных HMP⁷⁰, состоит из как полных геномов, так и наборов контигов. Этот набор был расширен геномами организмов, найденными значимыми для микрофлоры кишечника по данным метагеномных исследований. В частности, были добавлены бактериальные патогены рода *Streptococcus*, частый обитатель больниц *Staphylococcus aureus*, патогенные эукариоты родов *Candida* и *Blastocystis* и т.д. Добавление возможных патогенов позволило использовать данный референсный каталог для исследования различных клинических образцов.

Этот каталог, по процедуре, описанной в Материалах и методах был преобразован в неизбыточный. Это позволило избежать разбавления значения представленности схожих геномов. Полный список 353 бактерий, вошедших в референсный набор, перечислен в Приложении 3.

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). Учитываются только геномы, покрытые ридами метагеномного образца более чем на 1% своей длины. Глубина покрытия нормируется на ряд параметров: общее количество нуклеотидов, картировавшихся на весь референсный набор и длину генома. Также проводится суммирование нормированной глубины

покрытия по родам. Полученные значения, именуемые векторами представленности по образцам, приводятся к процентному содержанию микроорганизмов в образце и используются в дальнейшем анализе.

5) Визуализация и статистический анализ.

Анализ векторов представленности, в отличие от предыдущих этапов, не автоматизирован, т.к. каждая исследовательская задача требует индивидуального подхода. Однако, он включает в себя универсальный базовый анализ, как то оценка представленностей патогенов и организмов-маркеров, оценка альфа-разнообразия, подсчет расстояний между образцами и их визуализация. Наиболее используемыми методами визуализации являются тепловые карты (англ. heatmap), в которых числовые значения представлены в виде цветового градиента, по одной оси расположены образцы, по другой – признаки, например геномы. Сортировка признаков и образцов происходит по их иерархической кластеризации. Другим распространенным методом представления многомерных векторов представленностей является многомерное шкалирование (MDS, от англ. Multidimensional Scaling)¹⁰⁵. В его основе лежит уменьшение размерности до заданного числа, к примеру, до двух, что позволяет визуализировать данные.

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 UniFrac¹⁰³, содержащая в себе информацию о филогенетическом расстоянии между референсными геномами, и часто используемое в экологии расстояние Брэя-Кертиса¹⁰².

В случае выявления различий между двумя выборками используется тест Манна-Уитни, поскольку метагеномные данные не обладают нормальным распределением¹³⁰.

3.2 Исследование российских метагеномных образцов от здоровых доноров

Забор кишечных метагеномных образцов (кала) происходил в медицинских учреждениях, согласно протоколу, описанному в Материалах и методах. Вкратце, критериями включения доноров были отсутствие патологий желудочно-кишечного тракта, отсутствие приема антибиотиков в течение полугода, в возрасте от 18 лет. Средний возраст доноров составил 36 ± 18 , пола были равнопредставлены (48 образцов от женщин и 48 от мужчин). Доноры также 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 Хакасии. Таким образом, по этой выборке была поставлена задача выявить вид нормофлоры кишечника взрослых жителей Российской Федерации из различных географических и культурных сред.

В ходе секвенирования на приборе ABI SOLiD 4 96 образцов кишечного метагенома Российской популяции было получено 2.7 ± 1.1 Гб ридов, длиной 50 нуклеотидов для каждого образца. Была проведена *de novo* сборка контигов по образцам с целью выявления принципиально новых микроорганизмов в значимых количествах (см. Материалы и методы), не присутствующих в референсе, однако, таковые не были найдены. Риды прошли предобработку, был определен таксономический и функциональный состав (см. Материалы и методы).

Дополнительно были использованы общедоступные метагеномные данные образцов от здоровых доноров из Китая (69), Дании (85), США (137), Малави (5) и Венесуэлы (10). Для этих данных доступна информация о донорах, такая как индекс массы тела (ИМТ), возраст (кроме выборки образцов из США). Средние значения этих параметров, предположительно влияющих на состав микробиоты, сопоставимы с аналогичными значениями для российских образцов (Таблица 3).

Таблица 3.

Статистика данных по мировым образцам

	Россия	Дания	Малави	Венесуэла	Китай	США
ИМТ						
медиана	22,63694	25,47	20,85	20,7	23,81	-
среднее	23,86±5,3	27,81±5,84	21,48±1,93	20,77±4,4	23,03±3,4	-
минимум	16,01562	18,59	20	14,1	15,57	-
максимум	36,13946	40,21	24,2	26,1	31,4	-
Возраст						
медиана	31,5	59	25	16,5	52	-
среднее	35,98±18,03	56,67±7,53	25,2±1,5	20,1±15,08	48,03±11,46	-
минимум	14	42	23	5	21	-
максимум	85	69	27	53	69	-
% картирования на человеческий геном						
медиана	0,235	0,05	0,05	0,03	0	0
среднее	1,17±6,78	0,35±1,51	0,06±0,04	0,031±0,009	0	0
минимум	0	0,01	0,02	0,02	0	0
максимум	66,27	13,71	0,12	0,05	0	0
% картирования на каталог бактериальных геномов						
медиана	23,475	30,26	34,68	44,515	43,63	52,92
среднее	24,14±7,5	31,52±10,15	35,86±2,82	44,79±11,41	44,9±11,2	51,97±15,08
минимум	1,17	12,52	34,22	24,58	26,23	18,6
максимум	48,6	56,59	40,89	62,02	70,34	84,95

Дополнительные метагеномные наборы прошли ту же процедуру обработки, что и российские метагеномы. Проценты картирования как на человеческий геном, так и на каталог бактериальных геномов оказались приблизительно равными между выборками (Таблица 3). Первое говорит о сопоставимом качестве выделения метагенома, а второе является еще одним подтверждением отсутствия высокопредставленных новых таксонов и достаточном качестве секвенирования.

Иностранные метагеномы были секвенированы на платформе Illumina, поэтому для проверки адекватности сравнения дополнительно 5 российских метагеномных образцов, также были секвенированы на приборе Illumina HiSeq 2000. Эти образцы были специально выбраны максимально различными по таксономическому составу, а именно с доминирующими родами *Prevotella*, *Bacteroides*, и преобладающими отделами Firmicutes, Actinobacteria. Средняя корреляция Спирмена для этих 5 пар образцов составила $0,8\pm0,09$, что говорит о возможности сравнения метагеномных данных, полученных с разных платформ секвенирования. Дополнительно была проверена адекватность сравнения данных, полученных с прибора IonTorrent. Один из образцов был секвенирован трижды – на приборе SOLiD, и дважды на IonTorrent с различной длиной ридов. (Рисунок 4). Средняя корреляция Спирмена составила $0,89\pm0,03$.

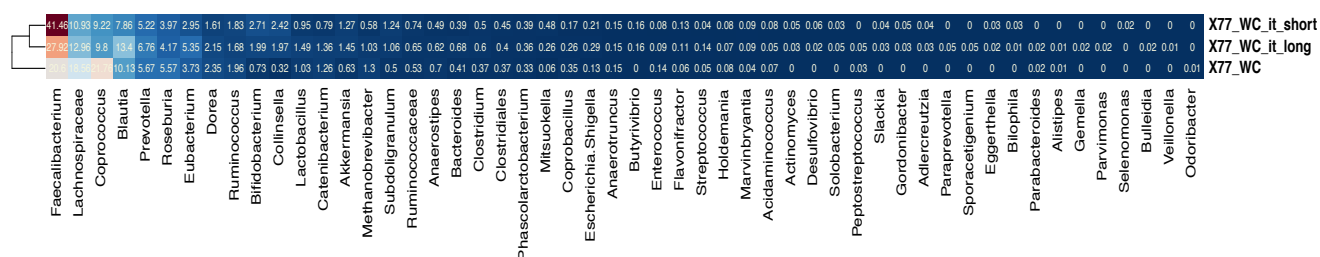


Рисунок 4. Сравнение процентов представленности бактериальных родов в образце, секвенированном на приборе ABI SOLiD 4 (X77_WC) и дважды – на IonTorrent (со средней длиной ридов 120 ± 15 п.о. (X77_WC_short) и 229 ± 58 п.о. (X77_WC_long)).

Аналогичная проверка была проведена и для одного образца, секвенированного на приборе SOLiD, и на 454 GS FLX+. Корреляция Спирмена для двух векторов представленности родов составила 0,89.

Помимо сравнения технологий секвенирования, была проведена проверка качества процедуры обработки метагеномных данных. Поскольку алгоритм обработки аналогичен, используемому в работе⁷, то было проверено сходство векторов представленности 85 датских образцов, полученных в ходе обработки разработанной процедурой и оригинальных векторов. Их корреляция Спирмена по 98 общим родам составила $0,75\pm0,033$, несмотря на различающиеся

референсные каталоги и используемое программное обеспечение для выравнивания. Дополнительно было исследовано сходство векторов представленности с результатами обработки программой MetaPhlAn¹³¹, использующей иной подход: вместо картирования на каталог геномов, происходит картирование на каталог таксономически-специфических генов-маркеров. Для этого российские метагеномные риды были конвертированы в нуклеотидный формат и классифицированы по таксономии с использованием программ. Относительная представленность 57 общих родов, используемых в качестве референса в обеих процедурах, была достаточно хорошо коррелировала (корреляция Спирмена $0,835 \pm 0,04$). Разногласия в этих двух процедурах возникают при выявлении представленности таксономически неоднозначных родов. Например, MetaPhlAn относит *Bacteroides pectinophilus* к семейству Bacteroidaceae, *Eubacterium rectale* к роду Eubacterium, в то время как используемой в данной работе метод относит эти геномы к семейству Lachnospiraceae и роду *Roseburia* соответственно, основываясь на классификации по последовательности 16S рРНК генов.

В целом, метагеномные образцы из России, так же, как и в остальном мире, содержат в себе представителей двух отделов – Bacteroidetes и Firmicutes (Рисунок 5). Однако их соотношение значительно различается.

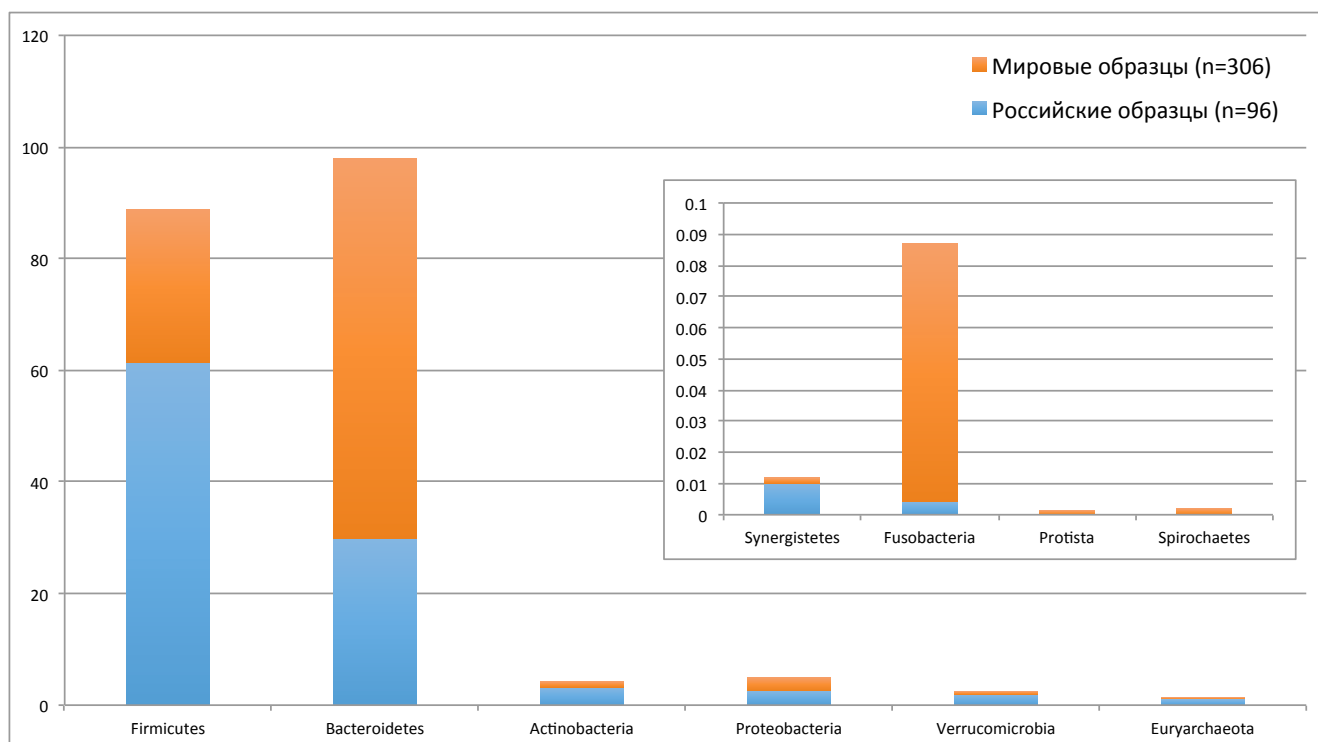


Рисунок 5. Гистограмма среднего процента представленности отделов в российских образцах и в остальных.

Из 124 представленных в референсном каталоге родов, ненулевое покрытие всеми образцами получили 105, а 80 – российскими. Наиболее представленными оказались рода *Bacteroides*, *Prevotella*, отдел *Lachnospiraceae*, *Faecalibacterium*, *Alistipes*, *Coprococcus*, *Parabacteroides* и *Roseburia*. Все вместе они составляют 80% от общего покрытия. Их относительная представленность по географическим группам в логарифмическом масштабе представлена на Рисунке 6.

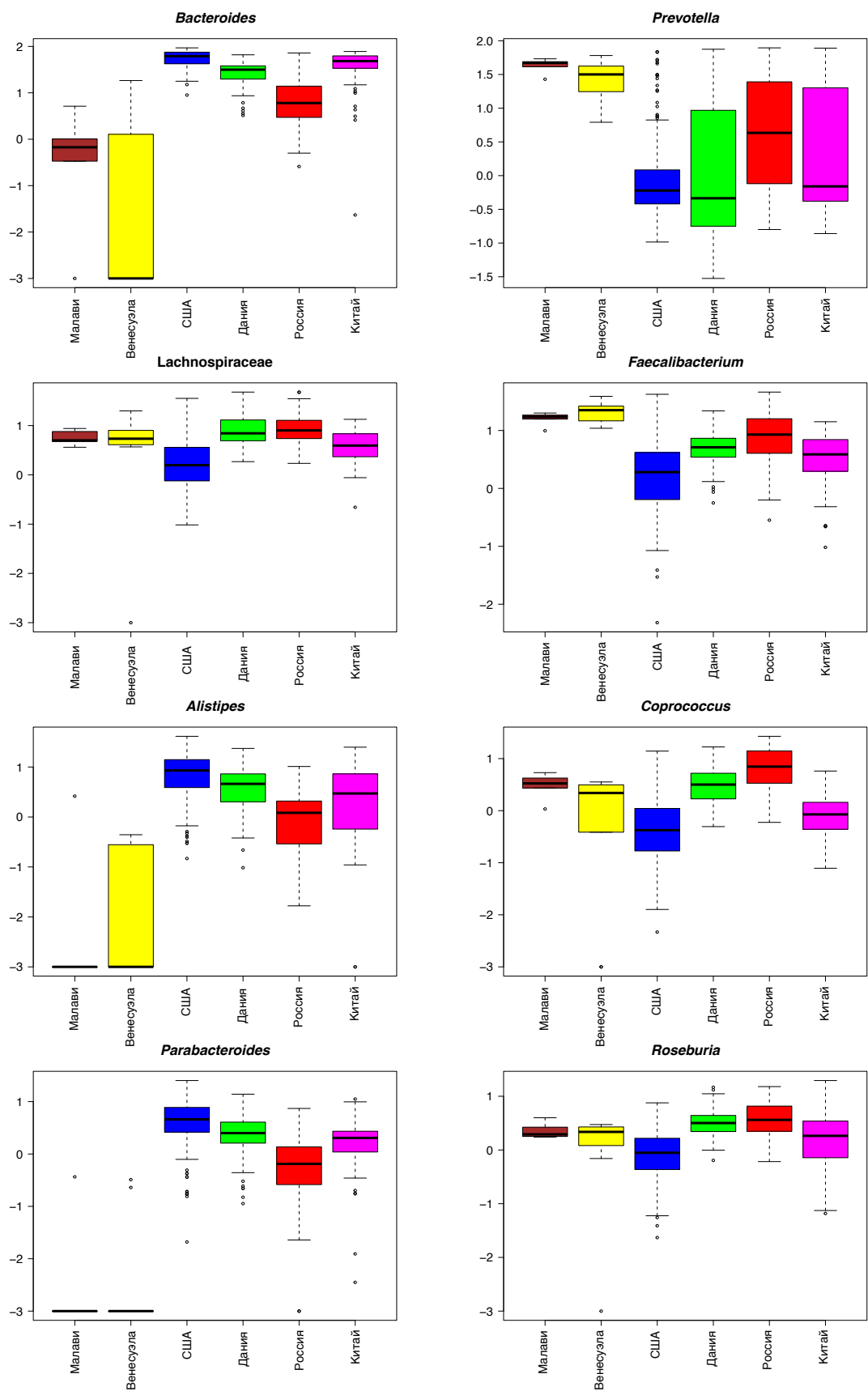


Рисунок 6. Относительная представленность бактериальных родов, составляющих 80% общего покрытия, по географическим группам.

Российские метагеномы представлены городскими и сельскими образцами примерно в равных количествах. При этом, обе эти группы сопоставимы по остальным признакам. Так средний ИМТ в городских и сельских образцах составляет $24,08 \pm 4,1$ и $23,6 \pm 6,5$ соответственно, а возраст – $36,04 \pm 17,3$ и $35,91 \pm 19$. Статистика Манна-Уитни говорит об отсутствии достоверных различий по этим параметрам (p-value равно 0,13 и 0,54 соответственно). Таким образом, говоря о различиях между этими группами, имеется ввиду различия, связанные именно с образом жизни и рационом. На Рисунке 7 изображена тепловая карта представленности прокариотических родов в российских образцах, иерархически кластеризованных (см. Материалы и методы).

Рисунок 7. Тепловая карта представленности прокариотических родов в российских образцах. Синим цветом слева обозначены сельские образцы, красным – городские.

Однозначного распределения на сельскую и городскую группу нет. Статистический тест ANOSIM выявил слабые различия (0,07814, p-value = 0,002). Так же, не наблюдается группировки российских образцов по географическому признаку, за исключением образцов из Татарстана (Рисунок 8).

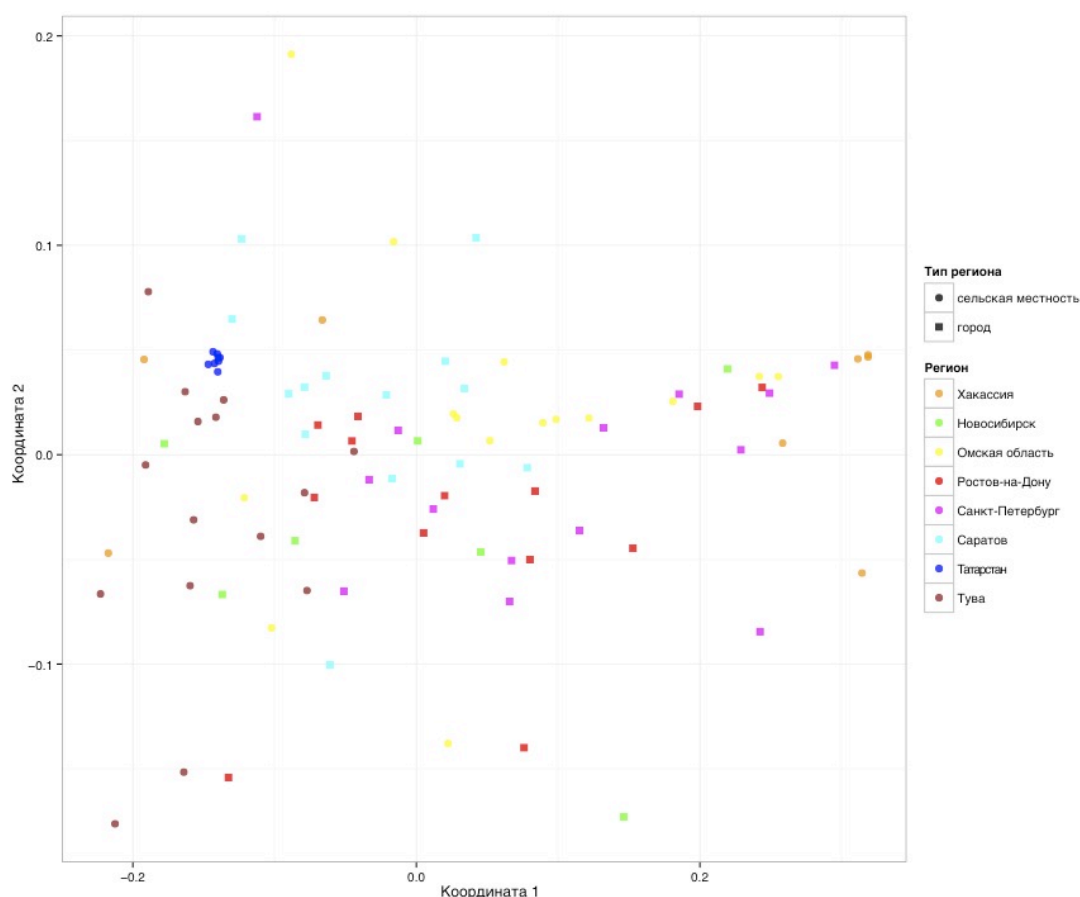


Рисунок 8. График MDS Российских образцов.

Были проанализированы тройки наиболее представленных родов в образцах и сравнены с теми, что встречаются в других мировых образцах. Так было выявлено 46 российских образцов с уникальными тройками высокопредставленных родов (Таблица 4).

Уникальные доминантные тройки родов в российских образцах

Образец	1	2	3
Kh_236	<i>Blautia</i>	<i>Prevotella</i>	<i>Anaerostipes</i>
Kh_240	<i>Blautia</i>	<i>Lachnospiraceae</i>	<i>Escherichia/Shigella</i>
Kh_242	<i>Lactobacillus</i>	<i>Blautia</i>	<i>Lachnospiraceae</i>
Kh_250	<i>Prevotella</i>	<i>Streptococcus</i>	<i>Bacteroides</i>
NOV_282	<i>Ruminococcus</i>	<i>Lachnospiraceae</i>	<i>Coprococcus</i>
NOV_283	<i>Escherichia/Shigella</i>	<i>Methanobrevibacter</i>	<i>Akkermansia</i>
NOV_284	<i>Blautia</i>	<i>Coprococcus</i>	<i>Lachnospiraceae</i>
NOV_285	<i>Blautia</i>	<i>Bacteroides</i>	<i>Eubacterium</i>
OM_178	<i>Faecalibacterium</i>	<i>Lachnospiraceae</i>	<i>Coprococcus</i>
OM_181	<i>Prevotella</i>	<i>Faecalibacterium</i>	<i>Coprococcus</i>
OM_182	<i>Phascolarctobacterium</i>	<i>Alistipes</i>	<i>Prevotella</i>
OM_184	<i>Prevotella</i>	<i>Roseburia</i>	<i>Lachnospiraceae</i>
OM_185	<i>Escherichia/Shigella</i>	<i>Faecalibacterium</i>	<i>Coprococcus</i>
OMS_251	<i>Faecalibacterium</i>	<i>Prevotella</i>	<i>Lactobacillus</i>
RND_304	<i>Ruminococcus</i>	<i>Blautia</i>	<i>Bacteroides</i>
RND_306	<i>Coprococcus</i>	<i>Faecalibacterium</i>	<i>Prevotella</i>
SAR_262	<i>Faecalibacterium</i>	<i>Enterococcus</i>	<i>Lachnospiraceae</i>
SAR_268	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Lachnospiraceae</i>
SAR_274	<i>Akkermansia</i>	<i>Methanobrevibacter</i>	<i>Lachnospiraceae</i>
SAR_275	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Roseburia</i>
Spb_100_40P	<i>Prevotella</i>	<i>Bacteroides</i>	<i>Blautia</i>
Spb_105_45P	<i>Bacteroides</i>	<i>Ruminococcus</i>	<i>Coprococcus</i>
Spb_66_6P	<i>Prevotella</i>	<i>Bacteroides</i>	<i>Phascolarctobacterium</i>
Spb_73_13P	<i>Streptococcus</i>	<i>Dialister</i>	<i>Escherichia/Shigella</i>
Spb_74_14P	<i>Bacteroides</i>	<i>Coprococcus</i>	<i>Ruminococcus</i>
Spb_76_16P	<i>Bacteroides</i>	<i>Escherichia/Shigella</i>	<i>Blautia</i>
TAT_129	<i>Coprococcus</i>	<i>Ruminococcus</i>	<i>Faecalibacterium</i>
TAT_130	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Ruminococcus</i>
TAT_131	<i>Ruminococcus</i>	<i>Coprococcus</i>	<i>Faecalibacterium</i>
TAT_134	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Ruminococcus</i>
TAT_138	<i>Coprococcus</i>	<i>Faecalibacterium</i>	<i>Ruminococcus</i>
TAT_143	<i>Coprococcus</i>	<i>Ruminococcus</i>	<i>Faecalibacterium</i>
TAT_146	<i>Coprococcus</i>	<i>Faecalibacterium</i>	<i>Ruminococcus</i>
TAT_147	<i>Coprococcus</i>	<i>Faecalibacterium</i>	<i>Ruminococcus</i>
TYV_202	<i>Blautia</i>	<i>Akkermansia</i>	<i>Ruminococcus</i>
TYV_203	<i>Bifidobacterium</i>	<i>Ruminococcus</i>	<i>Lachnospiraceae</i>
TYV_205	<i>Blautia</i>	<i>Catenibacterium</i>	<i>Dorea</i>
TYV_206	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Blautia</i>
TYV_207	<i>Faecalibacterium</i>	<i>Lachnospiraceae</i>	<i>Roseburia</i>
TYV_209	<i>Bifidobacterium</i>	<i>Blautia</i>	<i>Catenibacterium</i>
TYV_210	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Ruminococcus</i>

TYV_211	<i>Coprococcus</i>	<i>Akkermansia</i>	<i>Lachnospiraceae</i>
TYV_212	<i>Bifidobacterium</i>	<i>Ruminococcus</i>	<i>Lachnospiraceae</i>
TYV_214	<i>Faecalibacterium</i>	<i>Coprococcus</i>	<i>Bifidobacterium</i>
TYV_215	<i>Faecalibacterium</i>	<i>Bifidobacterium</i>	<i>Blautia</i>
TYV_219	<i>Ruminococcus</i>	<i>Prevotella</i>	<i>Blautia</i>

Интересно заметить, что отношение сельских образцов к городским в этой выборки составило 1,88. Уникальность этих метагеномов также подтверждается, тем что среднее расстояние по метрике UniFrac между ними и остальным миром значительно больше, чем между остальными российскими и мировыми образцами (тест Манн-Уитни, $p\text{-value} = 8,086e-10$).

При представлении таксономических векторов представленности всех мировых образцов в двухмерном пространстве с помощью MDS (Рисунок 9) можно заметить плавное изменение состава образцов, от американских к российским, что приблизительно соответствует уменьшению представленности рода *Bacteroides* и увеличению представленности родов из отдела Firmicutes.

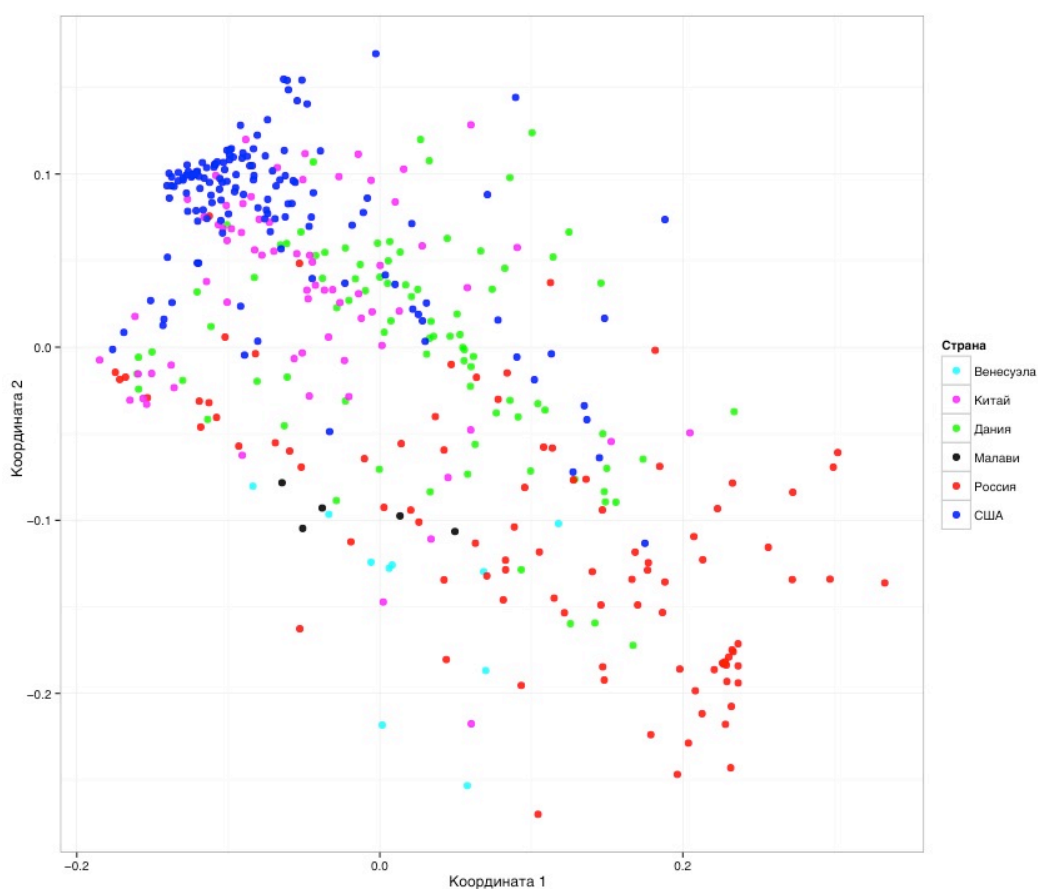


Рисунок 9. График MDS мировых образцов.

Используя тест Манн-Уитни, был найден список различающихся родов между российскими и другими национально-географическими группами образцов (см. Таблица 5). Достоверного различия между российскими и западными образцами по представленности рода *Bacteroides* выявлено не было. Тем не менее, количество образцов, в которых преобладают не представители родов *Bacteroides* и *Prevotella* в российской выборке примерно в 2,3 раза больше.

Таблица 5.

Результаты теста Манна-Уитни: списки различающихся бактериальных родов

		- медиана представленности выше в российских образцах			
Российские против Венесуэльских		Российские против Датских		Российские против Малавийских	
Род	P-value	Род	P-value	Род	P-value
<i>Bacteroides</i>	0,0013	<i>Ruminococcus</i>	0,0421	<i>Bacteroides</i>	0,0147
<i>Prevotella</i>	0,0064	<i>Bifidobacterium</i>	0,0129	<i>Prevotella</i>	0,0273
<i>Faecalibacterium</i>	0,0012	<i>Phascolarctobacterium</i>	0,0036	<i>Alistipes</i>	0,0308
<i>Alistipes</i>	0,0002	<i>Clostridiales</i>	0,0071	<i>Parabacteroides</i>	0,0119
<i>Coprococcus</i>	0,0006	<i>Clostridium</i>	0,0069	<i>Bifidobacterium</i>	0,0049
<i>Roseburia</i>	0,0087	<i>Sutterella</i>	0,0421	<i>Dialister</i>	0,0308
<i>Ruminococcus</i>	0,0087	<i>Paraprevotella</i>	0,0095	<i>Barnesiella</i>	0,0308
<i>Eubacterium</i>	0,0092	<i>Megamonas</i>	0,0071	<i>Akkermansia</i>	0,0147
<i>Escherichia, Shigella</i>	0,0045	<i>Collinsella</i>	0,0071	<i>Odoribacter</i>	0,0415
<i>Akkermansia</i>	0,0002	<i>Flavonifractor</i>	0,0037	<i>Paraprevotella</i>	0,0147
<i>Anaerostipes</i>	0,0101			<i>Tannerella</i>	0,0308
<i>Dorea</i>	0,0152	Российские против Китайских		<i>Coprobacillus</i>	0,0049
<i>Phascolarctobacterium</i>	0,0006	Род	P-value	<i>Anaerotruncus</i>	0,0049
<i>Subdoligranulum</i>	0,0165	<i>Prevotella</i>	0,0095	<i>Acidaminococcus</i>	0,0119
<i>Sutterella</i>	0,0069	<i>Alistipes</i>	0,0002		
<i>Paraprevotella</i>	0,0002	<i>Ruminococcus</i>	0,0006	Российские против США	
<i>Methanobrevibacter</i>	0,0087	<i>Escherichia, Shigella</i>	0,0159	Род	P-value
<i>Collinsella</i>	0,0008	<i>Anaerostipes</i>	0,0009	<i>Akkermansia</i>	0,0003
<i>Tannerella</i>	0,0009	<i>Streptococcus</i>	0,0003	<i>Megamonas</i>	0,0153
<i>Lactobacillus</i>	0,0242	<i>Clostridium</i>	0,0409		
<i>Parasutterella</i>	0,0310	<i>Parasutterella</i>	0,0111		
<i>Coprobacillus</i>	0,0072	<i>Coprobacillus</i>	0,0006		
<i>Anaerotruncus</i>	0,0027	<i>Veillonella</i>	0,0006		

Для проверки гипотезы о статистически достоверном различии российских образцов от других географических и национальных групп, был применен тест ANOSIM (см. Материалы и методы), на вход которому были поданы модифицированные взвешенные расстояния по метрике UniFrac (см. Материалы и методы). В итоге были найдены достоверные различия с группами образцов из США, Дании и Китая (Таблица 6). Вероятно, различия с африканской и южноамериканской группами не были выявлены из-за малого размера выборки.

Таблица 6.

**Результаты статистического теста ANOSIM при сравнении российской
выборки образцов с другими**

ANOSIM России против	Статистика	p-value
США	0,7282	1,00E-04
Китай	0,4897	1,00E-04
Дания	0,2804	1,00E-04
Малави	0,1424	0,083392
Венесуэла	-0,00984	0,50345

Иерархическая кластеризация мировых образцов, построенная по расстоянию метрики UniFrac также показывает разделение образцов по национально-географическому признаку (Рисунок 10).

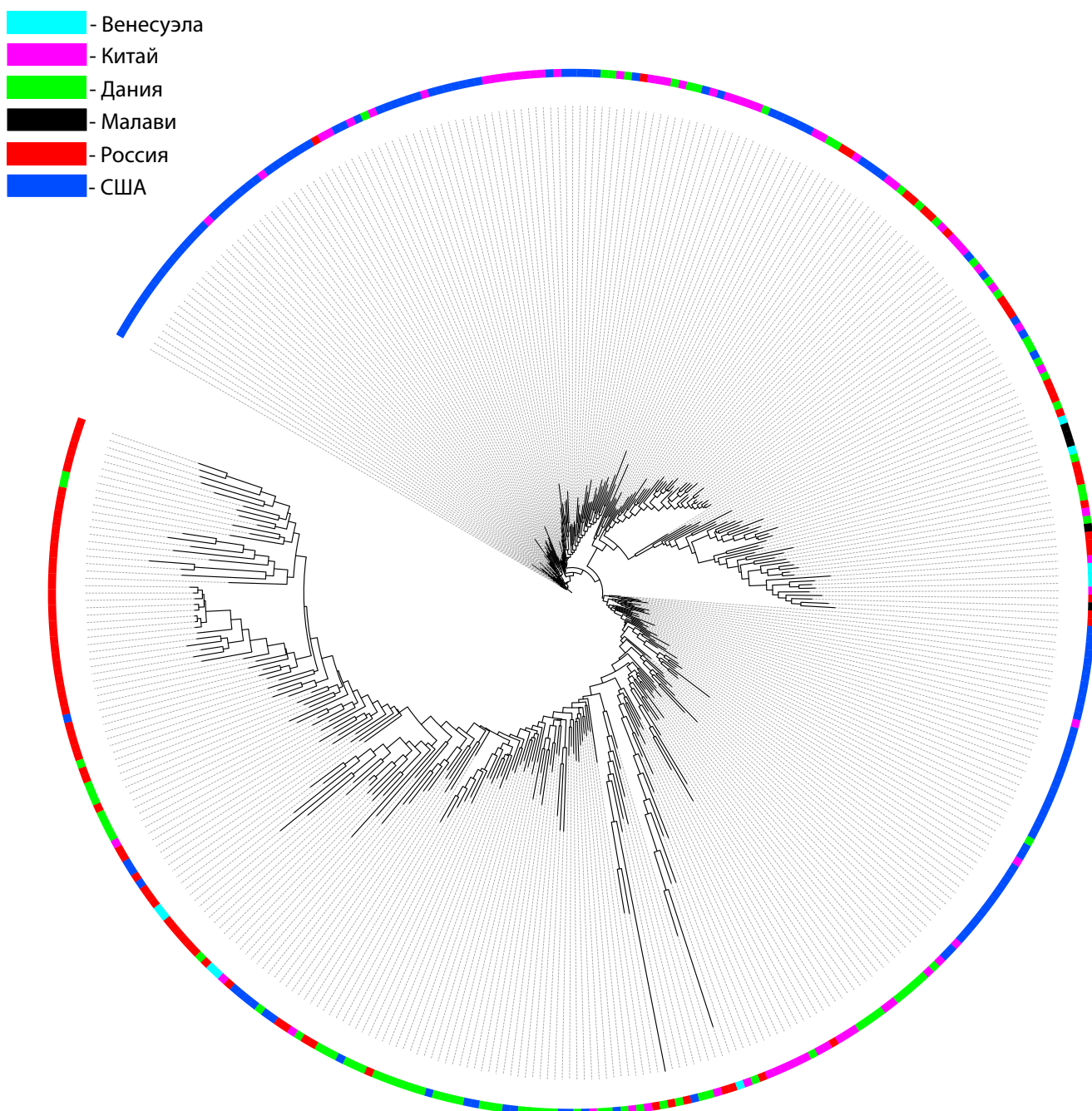


Рисунок 10. Дерево мировых метагеномных образцов.

С использованием данных, полученных нами, была выполнена процедура кластеризации, аналогичная той, которая использовалась для выявления энтеротипов (см. Материалы и методы). Было использовано несколько самых распространенных в биостатистике метрик – Евклидова, Манхэттен, Канберра, расстояние по корреляции Спирмена, Брэй-Кертис, Дженсен-Шэннон и UniFrac. По каждой метрике были посчитаны расстояния между всеми мировыми образцами и кластеризованы алгоритмом k-средних (Приложение 4), как все

образцы, так и отдельно российские образцы. Оптимальное количество кластеров определялось максимизацией индекса Калинского-Гарабача, а качество кластеризации определялось средней силуэтной величиной и предсказательной силой (см. Материалы и методы). На Рисунке 11 приведен график этих параметров в зависимости от используемой метрики и числа кластеров. Так, из него можно заключить, что наилучшее разделение на кластеры происходит при использовании метрики UniFrac, а итоговое количество кластеров равно двум.

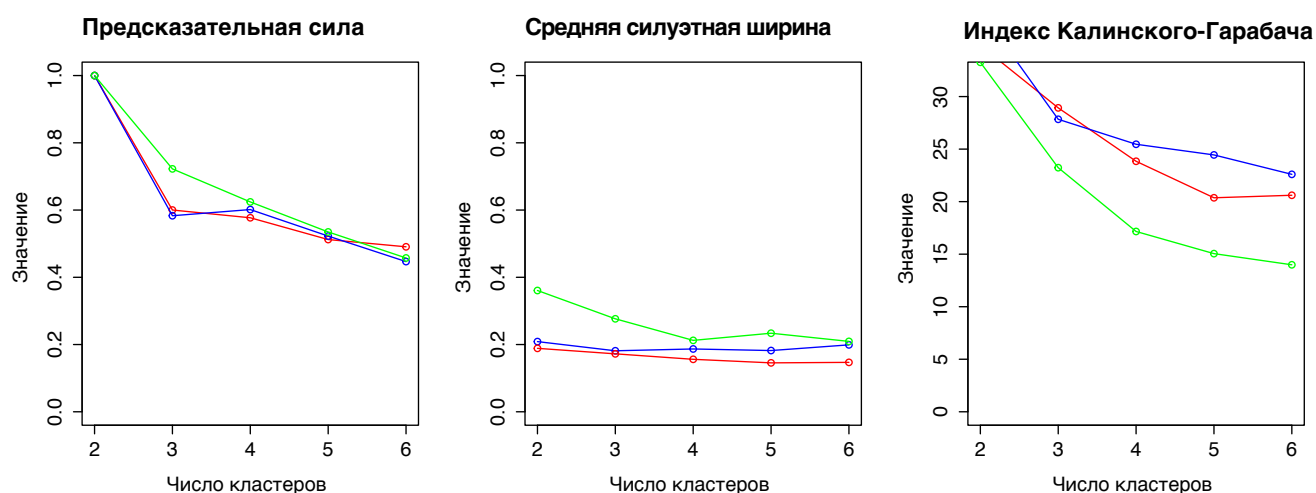


Рисунок 11. Графики изменения параметров оценки кластеризации российских образцов по основным используемым в метагеномике метрикам – расстояния Дженсен-Шеннона (красная линия), Брэй-Кертис (синяя линия), UniFrac (зеленая линия).

Тем не менее, предсказательная сила ниже принятого порога в 0,8, так же, как и средняя силуэтная величина. Однако, при сравнении со случайным распределением 96 российских образцов на 2 кластера (см. Материалы и методы) средняя силуэтная величина кластеризации по метрикам Дженсена-Шэннона и Брэй-Кертиса оказывается примерно в 2 раза меньше – 0,19 и 0,21 против 0,085 и 0,09 соответственно. При кластеризации всех мировых образцов по метрике UniFrac также получено два кластера, причем 10 российских образцов сменили кластер. Преобладающим родом первого кластера является *Prevotella*, а второй

характеризуется высоким содержанием нескольких родов отдела Firmicutes (Таблица 7).

Таблица 7.

Манн-Уитни тест для представленности родов в двух кластерах из кластеризации российских образцов по метрике UniFrac

	- медиана представленности выше в первом кластере
Род	p-value
<i>Bacteroides</i>	0,0056
<i>Lachnospiraceae</i>	0,0019
<i>Faecalibacterium</i>	0,0007
<i>Coprococcus</i>	0,0019
<i>Parabacteroides</i>	0,0041
<i>Bifidobacterium</i>	0,0001
<i>Anaerostipes</i>	0,0036
<i>Odoribacter</i>	0,0262
<i>Ruminococcaceae</i>	0,0002
<i>Sutterella</i>	0,0134
<i>Methanobrevibacter</i>	0,0188
<i>Collinsella</i>	0,0001
<i>Coprobacillus</i>	0,0022
<i>Anaerotruncus</i>	0,0002

Сельские и городские метагеномы в российских кластерах оказались примерно одинаково распределены: 51,5% и 53,5% городских образцов попали в первый кластер и во второй кластер соответственно. При совместной кластеризации с мировыми образцами, картина меняется: в одном кластере оказываются образцы с преобладающей долей родов *Prevotella*, *Bifidobacterium* и различными представителями отдела Firmicutes, а во втором – метагеномы с преобладанием рода *Bacteroides*.

Кластеризация по k-средним согласно своему алгоритму всегда распределит все поданные на вход данные хотя бы по двум кластерам. Однако этот подход неверный, в том случае, если необходимо определить, существуют ли кластеры в принципе или выявить схожие подгруппы образцов. Поэтому была использована иная процедура нахождения подгрупп схожих данных, реализованная на программном языке R в функции `pvclust` одноименной библиотеки (см.

Материалы и методы). Так было найдено 2 подгруппы: одна состоит из образцов от одной семьи из Омской области (6 образцов), вторая из татарских метабеномов (8 образцов) (Рисунок 12).

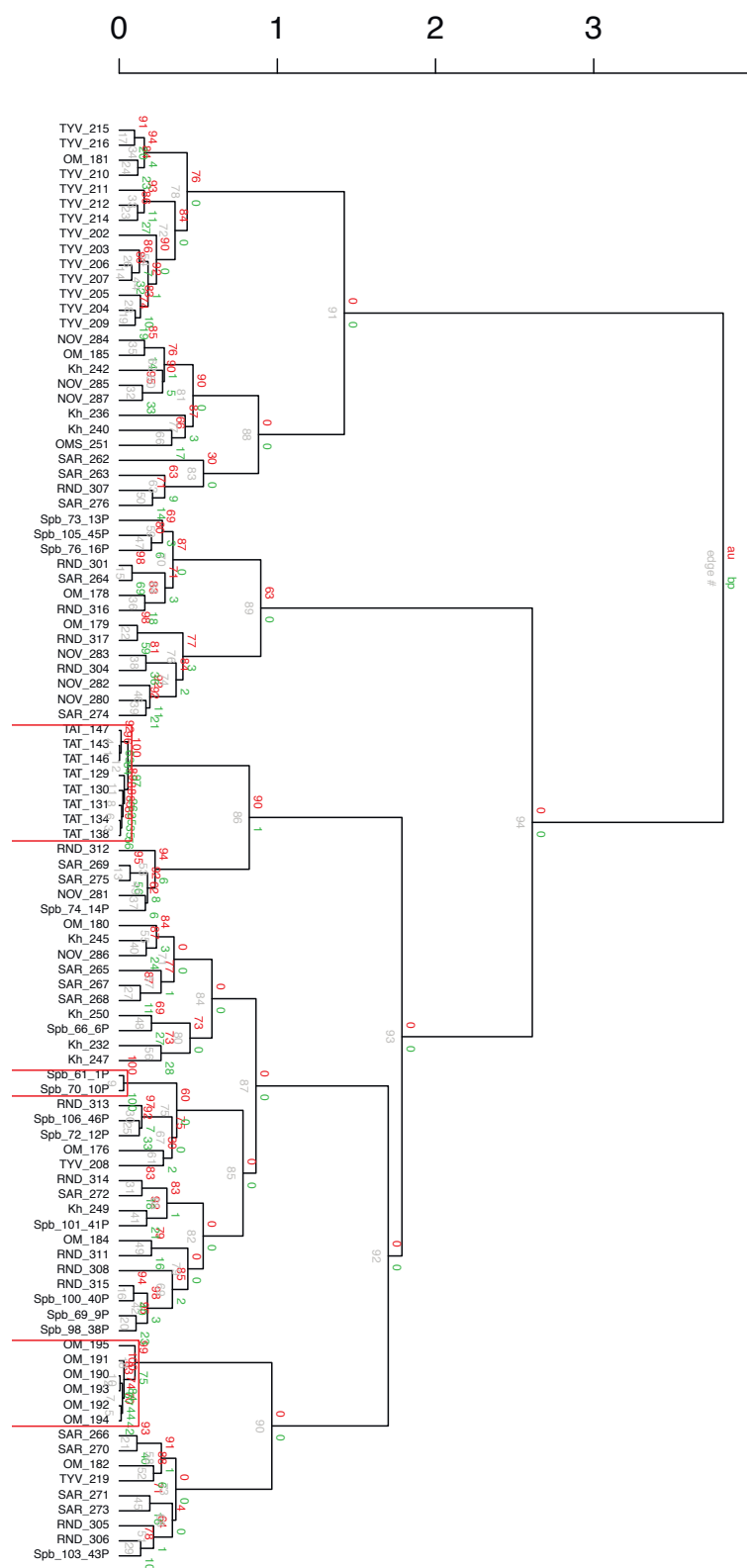


Рисунок 12. Результат иерархической кластеризации российских образцов. Красным выделены найденные схожие подгруппы образцов.

Омский кластер по своему таксономическому составу характеризуется высокой представленностью родов *Prevotella* (40,13%±14,19), *Lachnospiraceae* (15,0%±3,54), *Coprococcus* (13,47%±5,91) *Faecalibacterium* (7,32%±2,66). Вместе, эти рода составляют 80% покрытия в омской группе. В группе образцов из Татарстана преобладают рода *Coprococcus* (22,69%±3,3), *Faecalibacterium* (18,6%±2,6), *Ruminococcus* (18,11%±2,5) и *Roseburia* (9,77%±2,9). В целом, это две достаточно разные группы образцов (Рисунок 13) так омские образцы можно условно отнести ко второму энтеротипу (характеризуемый высок представленностью рода *Prevotella*). Татарстанские же образцы полностью вошли в список образцов, с уникальным тройками высокопредставленных родов.

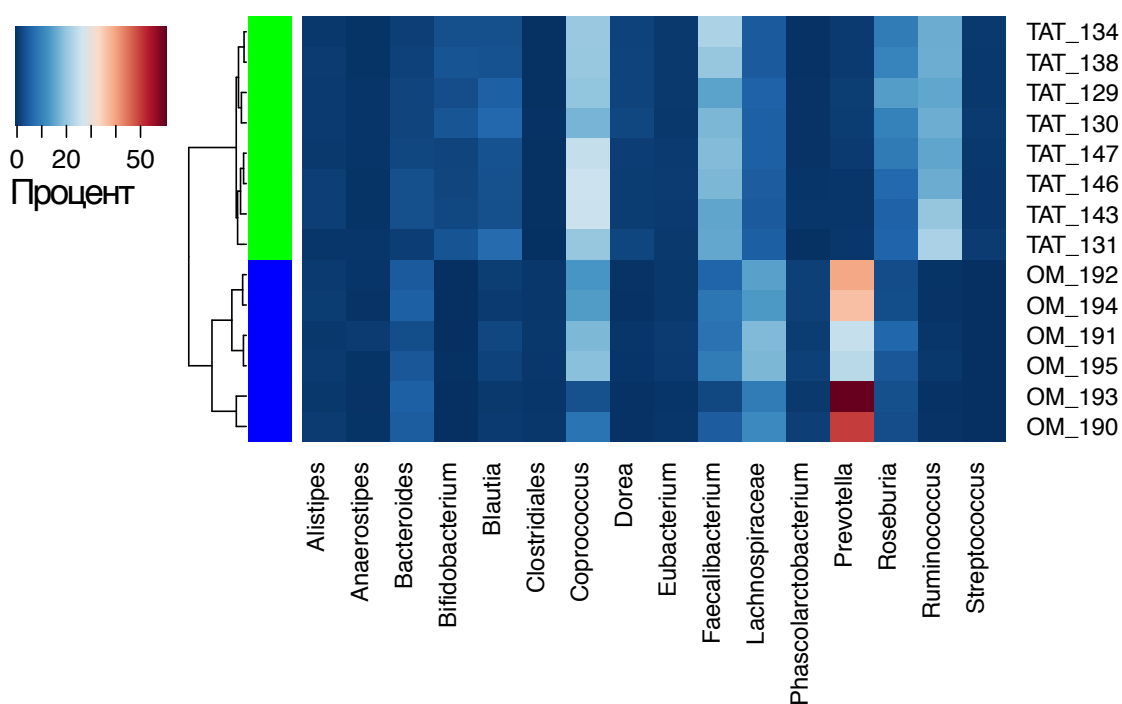


Рисунок 13. Тепловая карта наиболее представленных родов в найденных подгруппах.

Для омской и татарстанской подгрупп средняя корреляция Спирмена таксономических векторов представленности составила $0,97 \pm 0,03$ и $0,93 \pm 0,03$ соответственно, в то время как аналогичная величина для всех российских образцов составляет $0,79 \pm 0,07$. Однако среднее расстояние UniFrac для омской подгруппы оказалось достаточно велико и составило $0,044 \pm 0,03$, а для всех

руссииских образцов – $0,096\pm0,04$. Татарстанская подгруппа значительно более однородна, среднее расстояние UniFrac в ней составило $0,0087\pm0,003$.

В результате данного этапа исследования было расширено общее представление о нормофлоре кишечника человека, впервые был выявлен состав метабенома кишечника жителей РФ.

3.3 Исследование метабеномных образцов при патологиях

Каждое новое метабеномное исследование следует рассматривать в контексте предыдущих. Здесь было проведено сравнение руссииских метабеномных образцов и образцов из других стран мира. Проведенный сравнительный анализ позволил выявить ряд интересных черт здоровой микробиоты, как то преобладание представителей отдела Firmicutes и большее разнообразие. Полученные знания имеют не только фундаментальное значение, но и практическое применение. Теперь, когда есть представление о нормофлоре кишечника человека, в том числе и жителей РФ, появилась возможность проводить клинические исследования: изучать микробиоту при кишечных патологиях, потенциально травмирующих курсах лечения (прием антибиотиков, химиотерапия), иммунологических заболеваниях. Используя полученные данные, был изучен ряд метабеномных образцов из нескольких клинических исследований, в том числе метабеномы 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. Вышеописанные 96 руссииских образцов были использованы в качестве контроля в этих исследованиях.

3.3.1 Метагеном пациентов, больных онкологическими заболеваниями

Известно что, курс противораковой химиотерапии сопровождается нарушением баланса микробиоты кишечника¹³². Это в свою очередь препятствует лечению и уменьшает вероятность успешного исхода, пациенты страдают от диареи и других кишечных расстройств. С целью изучения влияния химиотерапии на метагеном больных онкологическими заболеваниями (гепатобластома, нейробластома) с течением времени, были собраны образцы кала (15 шт.) от 4 пациентов, проходивших лечение в ФНКЦ ДГОИ им. Д. Рогачева. Всего было собрано 3-4 образца на человека в течение нескольких недель прохождения курса лечения. Метагеном были секвенированы на приборе SOLiD 4 (Life Technologies) по стандартному протоколу (см. Материалы и методы). Всего было получено в среднем $1,5 \pm 0,4$ млрд. п.н. на образец. Полученные риды, длиной 50 пн., были обработаны стандартным образом (см. Материалы и методы).

Уже в ходе фильтрации ридов выявился необычайно высокий процент картирования на человеческий геном, который составил в среднем $29,9 \pm 23,7\%$, а максимум – $76,6\%$ (образец K418), в то время как среднее этой величины для контрольной выборки составила $1,17 \pm 6,78$ при аналогичном методе выделения ДНК. Предположительно, такое необычайно высокое содержание ДНК человека связано с развивающимся мукозитом – воспалительным поражением слизистой желудочно-кишечного тракта. В результате отшелушивающийся эпителий кишечника в больших количествах попадает в метагеномные образцы, значительно снижая процент бактериальной ДНК в нем: средний процент картирования на бактериальные геномы составил $22,9 \pm 17,3\%$, что примерно соответствует значению для здоровых образцов, однако разброс значительно больше, так минимальный процент картирования – $0,8\%$, для того же образца K418.

С целью подтверждения гипотезы о падении разнообразия микробиоты кишечника человека в ходе химиотерапии, был посчитан индекс альфа-

разнообразия. В среднем для клинических образцов он составил $1,47 \pm 0,62$, в то время как для всех доступных мировых метагеномов – $3,19 \pm 0,54$. Действительно, индексы разнообразия достоверно различаются между контрольной и клинической выборками (тест Манна-Уитни, $p\text{-value} = 1,176 \times 10^{-9}$). Это доказывает, что разнообразие микробиоты у больных онкологией и проходящих химиотерапию снижается примерно в два раза.

Наиболее представленными в образцах оказались рода *Ruminococcus*, *Klebsiella*, дрожжи *Candida*, *Enterococcus*, *Streptococcus*, *Pseudomonas*, *Parabacteroides* и *Clostridium*, составляя 95% общего совместного покрытия (Рисунок 14).

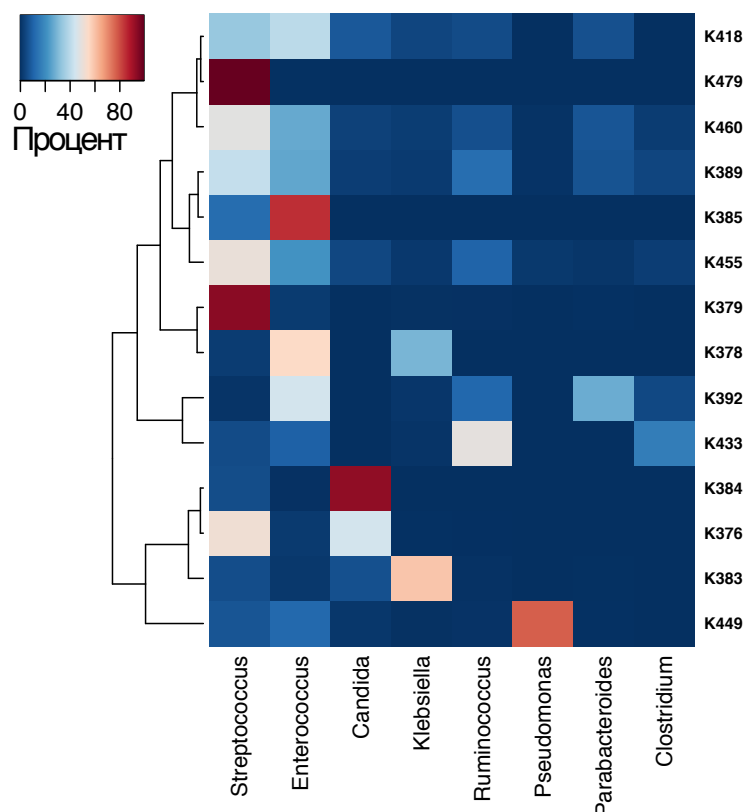


Рисунок 14. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах кала пациентов онкологии.

Как ранее было сказано, было проанализировано по 3-4 образца от каждого пациента. Для того чтобы оценить степень изменения метагенома кишечника во времени, был построен график MDS (Рисунок 15).

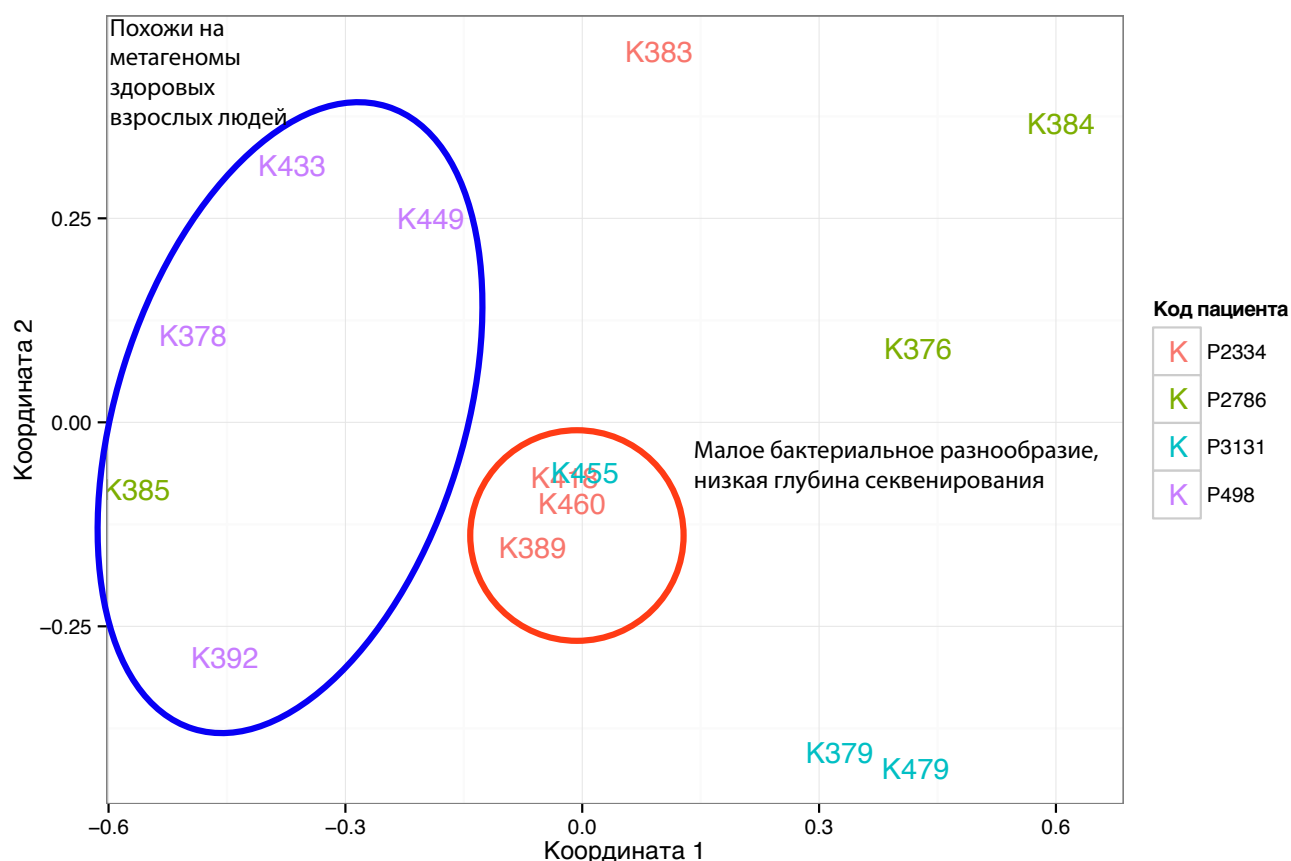


Рисунок 15. График MDS метагеномных образцов онкологических больных. Цветом отмечена принадлежность к тому или иному пациенту.

Из графика видно, что в целом образцы от одного пациента имеют тенденцию группироваться, однако проследить однонаправленного изменения во времени не удастся. Также стоит отметить группу образцов в центре графика, для которой характерно особенно низкое бактериальное разнообразие и низкая глубина секвенирования.

В целом, метагеном кишечника онкологических больных, проходящих антибиотикотерапию отличается низким бактериальным разнообразием и аномально высоким количеством ДНК человека, по всей видимости, вызванным мукозитом. Нарушение микробиотного баланса и иммунитета приводит к появлению в значимых количествах патогенных организмов, в том числе дрожжей. Все это значительно утяжеляет состояние пациентов.

3.3.2 Метагеном кишечника людей, страдающих от алкогольной зависимости

Микробиота кишечника, модулируя метаболизм человека, вынуждена и сама подстраиваться под образ жизни своего хозяина. Так, например, известно, что при алкоголизме повышается уровень бактерий *Escherichia* и снижается разнообразие⁸². В рамках данной работы была исследована когорта из 24 метагеномных образцов (предоставлены Клиническим центром Университетской клинической больницы №2 ГБОУ ВПО Первый МГМУ имени И.М. Сеченова), чьими донорами были люди, страдающие от алкоголизма на протяжении $17,8 \pm 10,6$ лет, но преимущественно без дополнительных диагнозов.

Из образцов была выделена тотальная ДНК и секвенирована на приборе SOLiD 5500 (см. Материалы и методы), затем результаты были обчислены стандартным образом (см. Материалы и методы). Проценты картирования на геном человека и на референсный каталог геномов были в пределах нормы (средние значения $0,36 \pm 0,3$ и $29,75 \pm 7,5\%$ соответственно), индекс разнообразия был даже несколько выше среднего значения по контрольным образцам – $3,5 \pm 0,6$. Основную часть микробиоты формируют рода *Bacteroides* и *Prevotella* (Рисунок 16). Кардинальных отличий от контрольных образцов, как это было в случае с образцами от онкологических больных, не наблюдалось. Однако, уровень бактерий рода *Escherichia* действительно повышен в некоторых образцах: у A25 и A24 он составил 11,7% и 8,8% процентов соответственно. Также стоит отметить образец A17, единственный, течение болезни у которого дошло до запойной стадии – у него доля бактерий рода *Prevotella* составляла 77%. Также у ряда образцов встречались патогенные организмы, в частности представители рода *Klebsiella* детектировались в 10 образцах.

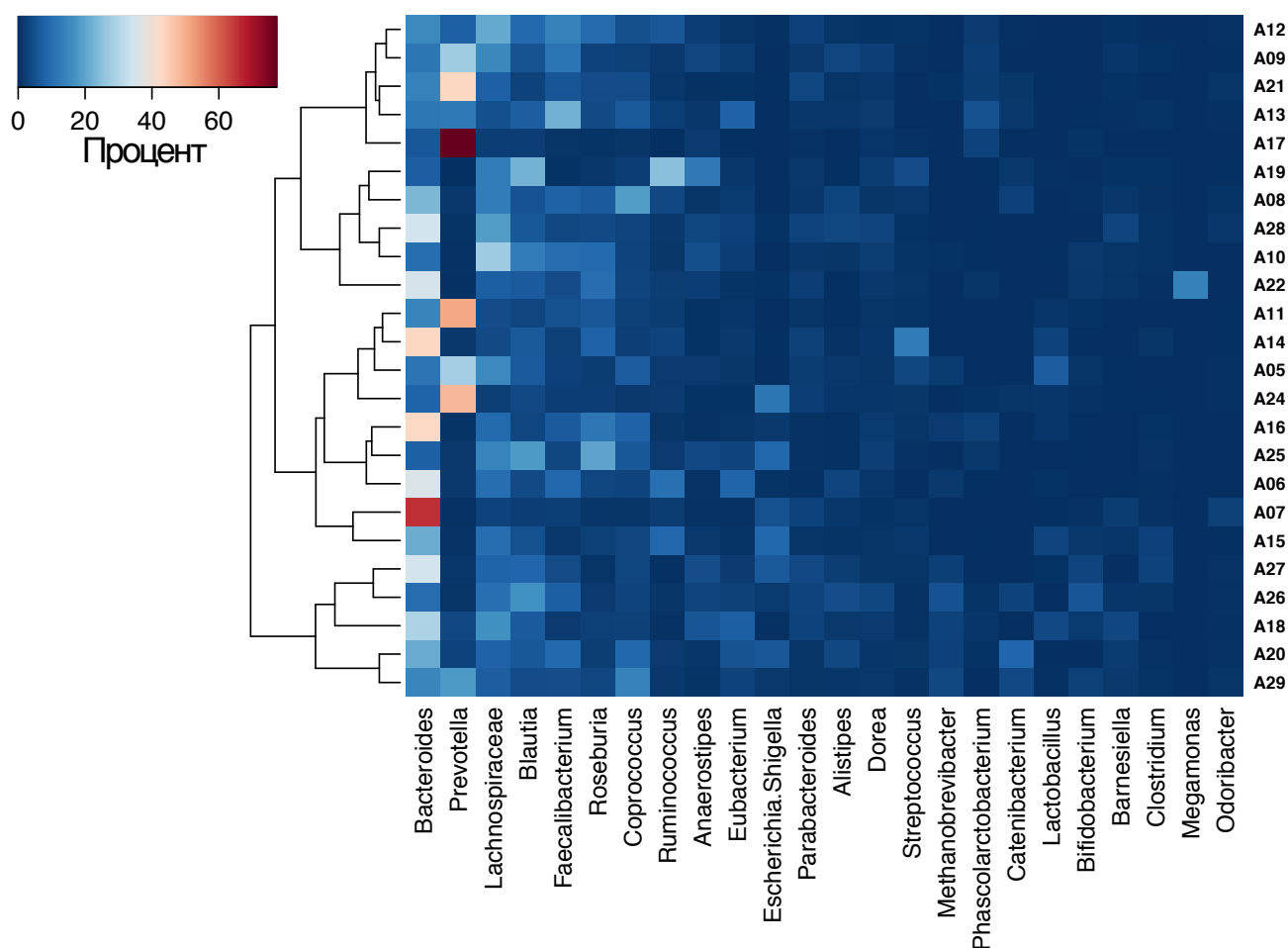


Рисунок 16. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах кала страдающих от алкоголизма.

Статистический тест Манна-Уитни выявил ряд различий по представленности бактериальных родов и отдельных организмов между контрольными и исследуемыми образцами (Таблица 8).

Таблица 8

**Результаты теста Манна-Уитни при сравнении контрольных образцов
и выборки метабеномов больных алгоголизмом**

А) По родовому составу

Род	p-value (FDR)	Средняя представленность в опытной выборке	Средняя представленность в контрольной выборке
<i>Actinomyces</i>	0,037	0	0,002
<i>Akkermansia</i>	5,81e-05	0,15	1,92
<i>Anaerostipes</i>	0,018	2,47	1,36
<i>Anaerotruncus</i>	0,0136	0,12	0,19
<i>Bacteroides</i>	9,097e-05	22,21	10,63
<i>Barnesiella</i>	0,015	0,9	0,4
<i>Coproccoccus</i>	0,017	4,96	9,03
<i>Faecalibacterium</i>	0,018	6,07	10,86
<i>Finegoldia</i>	0,048	0,0008	0
<i>Gordonibacter</i>	0,042	0	0,003
<i>Holdemania</i>	0,02	0,08	0,12
<i>Lactobacillus</i>	0,036	0,96	0,896
<i>Marvinbryantia</i>	0,019	0,002	0,015
<i>Oxalobacter</i>	0,015	0,01	0
<i>Parabacteroides</i>	0,004	1,83	1,13
<i>Rothia</i>	0,019	0,0009	0,015
<i>Ruminococcaceae</i>	0,014	0,42	0,62
<i>Sporacetigenium</i>	0,013	0,26	0,13
<i>Subdoligranulum</i>	0,023	0,36	0,55
<i>Turicibacter</i>	0,026	0,03	0,008
<i>Veillonella</i>	0,02	0,36	0,06

Б) По геномному составу

Организм	p-value (FDR)	Средняя представленность в опытной выборке	Средняя представленность в контрольной выборке
<i>Acidaminococcus fermentans</i> DSM 20731	0,0330	0,0540	0,0330
<i>Acidaminococcus</i> sp D21	0,0300	0,1100	0,0710
<i>Actinomyces odontolyticus</i> ATCC 17982	0,0350	0,0000	0,0020
<i>Akkermansia muciniphila</i> ATCC BAA 835	0,0000	0,1460	1,9190
<i>Alistipes onderdonkii</i> DSM 19147	0,0400	0,2650	0,2140
<i>Anaerostipes caccae</i> DSM 14662	0,0002	0,0150	0,0400
<i>Anaerostipes hadrus</i> DSM 3319	0,0065	1,2920	0,6650
<i>Anaerotruncus colihominis</i> DSM 17241	0,0040	0,1140	0,1940

<i>Bacteroides caccae</i> ATCC 43185	0,0004	0,9410	0,4960
<i>Bacteroides clarus</i> YIT 12056	0,0120	0,1610	0,0340
<i>Bacteroides coprocola</i> DSM 17136	0,0060	0,7530	0,5140
<i>Bacteroides dorei</i> 5 1 36 D4	0,0001	2,3000	1,1970
<i>Bacteroides eggerthii</i> 1 2 48FAA	0,0004	0,5310	0,1970
<i>Bacteroides finegoldii</i> DSM 17565	0,0001	0,3580	0,1820
<i>Bacteroides fluxus</i> YIT 12057	0,0130	0,1180	0,0840
<i>Bacteroides fragilis</i> 3 1 12	0,0040	0,2220	0,2110
<i>Bacteroides intestinalis</i> DSM 17393	0,0063	0,2170	0,1240
<i>Bacteroides oleiciplenus</i> YIT 12058	0,0041	0,1150	0,0630
<i>Bacteroides ovatus</i> 3 8 47FAA	0,0000	0,3730	0,1460
<i>Bacteroides ovatus</i> ATCC 8483	0,0000	0,6820	0,1640
<i>Bacteroides plebeius</i> DSM 17135	0,0002	0,5310	0,4340
<i>Bacteroides</i> sp 3 1 23	0,0042	0,2510	0,1350
<i>Bacteroides</i> sp D2	0,0000	0,3570	0,0931
<i>Bacteroides stercoris</i> ATCC 43183	0,0098	1,1240	0,3420
<i>Bacteroides thetaiotaomicron</i> VPI 5482	0,0041	0,5240	0,3110
<i>Bacteroides uniformis</i> ATCC 8492	0,0029	2,6120	1,4740
<i>Bacteroides vulgatus</i> ATCC 8482	0,0000	7,0530	2,6260
<i>Bacteroides xylanisolvens</i> XB1A	0,0040	0,3930	0,2270
<i>Barnesiella intestinihominis</i> YIT 11860	0,0077	0,8900	0,4020
<i>Bifidobacterium angulatum</i> DSM 20098	0,0105	0,0250	0,1508
<i>Blautia wexlerae</i> AGR2146	0,0450	1,6800	1,6040
<i>Butyrivibrio crossotus</i> DSM 2876	0,0053	0,2470	1,8077
<i>Clostridium bartlettii</i> DSM 16795	0,0063	0,2540	0,1296
<i>Clostridium clostridioforme</i> 2 1 49FAA	0,0490	0,0407	0,0627
<i>Clostridium leptum</i> DSM 753	0,0490	0,0960	0,1698
<i>Clostridium methylpentosum</i> DSM 5476	0,0230	0,0600	0,0357
<i>Clostridium scindens</i> ATCC 35704	0,0063	0,0624	0,1046
<i>Clostridium</i> sp 7 2 43FAA	0,0000	0,0253	0,0012
<i>Clostridium</i> sp HGF2	0,0000	1,2826	0,0491
<i>Clostridium</i> sp SS2 1	0,0000	0,0000	0,6700
<i>Clostridium spiroforme</i> DSM 1552	0,0041	0,0209	0,0054
<i>Clostridium symbiosum</i> WAL 14163	0,0430	0,1005	0,1540
<i>Coprobacillus</i> sp 8 2 54BFAA	0,0080	0,2312	0,0985
<i>Coprococcus comes</i> ATCC 27758	0,0350	1,7665	1,3795
<i>Coprococcus eutactus</i> ATCC 27759	0,0002	0,2598	2,4766
<i>Coprococcus</i> sp ART55 1	0,0095	0,8207	2,3064
<i>Enterococcus faecalis</i> V583	0,0457	0,1055	0,0270
<i>Eubacterium dolichum</i> DSM 3991	0,0458	0,1109	0,1492
<i>Eubacterium hallii</i> DSM 3353	0,0352	1,8280	1,1302
<i>Faecalibacterium cf prausnitzii</i> KLE1255	0,0041	1,2755	3,3510

<i>Faecalibacterium prausnitzii</i> A2 165	0,0133	1,4706	2,7773
<i>Faecalibacterium prausnitzii</i> L2 6	0,0125	1,0421	1,6611
<i>Faecalibacterium prausnitzii</i> SL3 3	0,0236	2,0557	3,0694
<i>Fusobacterium nucleatum</i> subsp <i>nucleatum</i> ATCC 23726	0,0095	0,0013	0,0000
<i>Gordonibacter pamelaee</i> 7 10 1 b	0,0406	0,0000	0,0027
<i>Holdemania filiformis</i> DSM 12042	0,0112	0,0769	0,1193
<i>Klebsiella pneumoniae</i> 342	0,0469	0,0517	0,0642
<i>Lactobacillus delbrueckii</i> subsp <i>bulgaricus</i> ATCC 11842	0,0458	0,0000	0,0055
<i>Lactobacillus fermentum</i> 28 3 CHN	0,0038	0,0258	0,0254
<i>Lactobacillus fermentum</i> IFO 3956	0,0038	0,0414	0,0170
<i>Lactobacillus gasseri</i> ATCC 33323	0,0001	0,0672	0,0467
<i>Lactobacillus salivarius</i> UCC118	0,0001	0,4273	0,0517
<i>Marvinbryantia formatexigens</i> DSM 14469	0,0134	0,0019	0,0150
<i>Oxalobacter formigenes</i> OXCC13	0,0095	0,0104	0,0000
<i>Parabacteroides distasonis</i> ATCC 8503	0,0033	0,6644	0,4509
<i>Parabacteroides johnsonii</i> DSM 18315	0,0047	0,1187	0,0728
<i>Parabacteroides merdae</i> ATCC 43184	0,0069	1,0727	0,6046
<i>Pseudoflavonifractor capillosus</i> ATCC 29799	0,0122	0,0712	0,1190
<i>Roseburia hominis</i> A2 183	0,0211	0,4993	0,7040
<i>Rothia mucilaginosa</i> DY 18	0,0134	0,0008	0,0149
<i>Ruminococcaceae bacterium</i> D16	0,0050	0,2474	0,4174
<i>Ruminococcus gnavus</i> ATCC 29149	0,0006	0,9386	0,1775
<i>Ruminococcus lactaris</i> ATCC 29176	0,0386	0,4515	0,6136
<i>Ruminococcus obeum</i> A2 162	0,0063	0,8845	0,5490
<i>Ruminococcus</i> sp 18P13	0,0000	1,5630	0,1977
<i>Ruminococcus</i> sp 5 1 39BFAA	0,0000	0,0491	2,4269
<i>Ruminococcus</i> sp SR1 5	0,0000	2,6159	1,0201
<i>Ruminococcus torques</i> ATCC 27756	0,0029	1,6779	0,4686
<i>Ruminococcus torques</i> L2 14	0,0227	2,3893	1,5919
<i>Streptococcus anginosus</i> 1 2 62CV	0,0319	0,0569	0,0149
<i>Streptococcus infantarius</i> subsp <i>infantarius</i> ATCC BAA 102	0,0441	0,0088	0,1821
<i>Streptococcus</i> sp 2 1 36FAA	0,0000	0,1385	0,0077
<i>Streptococcus</i> sp C150	0,0378	0,0179	0,0523
<i>Subdoligranulum</i> sp 4 3 54A2FAA	0,0098	0,1913	0,2936
<i>Subdoligranulum variabile</i> DSM 15176	0,0095	0,1532	0,2562
<i>Turicibacter sanguinis</i> PC909	0,0339	0,0279	0,0080
<i>Veillonella atypica</i> ACS 134 V Col7a	0,0069	0,0387	0,0137
<i>Veillonella parvula</i> DSM 2008	0,0236	0,1174	0,0119
<i>Veillonella</i> sp oral taxon 158 str F0412	0,0035	0,1494	0,0089

Из этих списков также видно основное различие между выборками – соотношение представителей отделов Bacteroidetes и Firmicutes. Однако стоит отметить и более детальные различия. Так в образцах от пациентов было в среднем в 6 раз больше представителей рода *Veillonella*, которых часто находят в воспаленных и поврежденных тканях человека¹³³, но при этом являются частью нормофлоры человека, обитают во всем желудочно-кишечном тракте.

3.3.3 Метагеном кишечника работников производства с повышенным радиационным фоном

Выше в данной работе было разобрано влияние образа жизни на микробиоту человека, в частности особенности питания. Тем не менее это может оказаться не единственным определяющим фактором. Вредные условия окружающей среды, например такие, как работа с радиоактивными материалами, гипотетически также может влиять на микробиоту. Так было исследовано 12 образцов от работников закрытого предприятия из города Саров.

Образцы этой группы по своему таксономическому составу принципиально не отличались от контрольной выборки. Средний процент картирования на каталог референсных геномов составил $22,2 \pm 9,9\%$, а на геном человека – $0,35 \pm 0,17\%$. Основную долю состава метагенома формировали рода *Bacteroides* и *Prevotella* (Рисунок 17), но у трех образцов ядро метагенома формировали представители отдела Firmicutes. Таким образом, данная выборка также принципиально не отличается от контрольной.

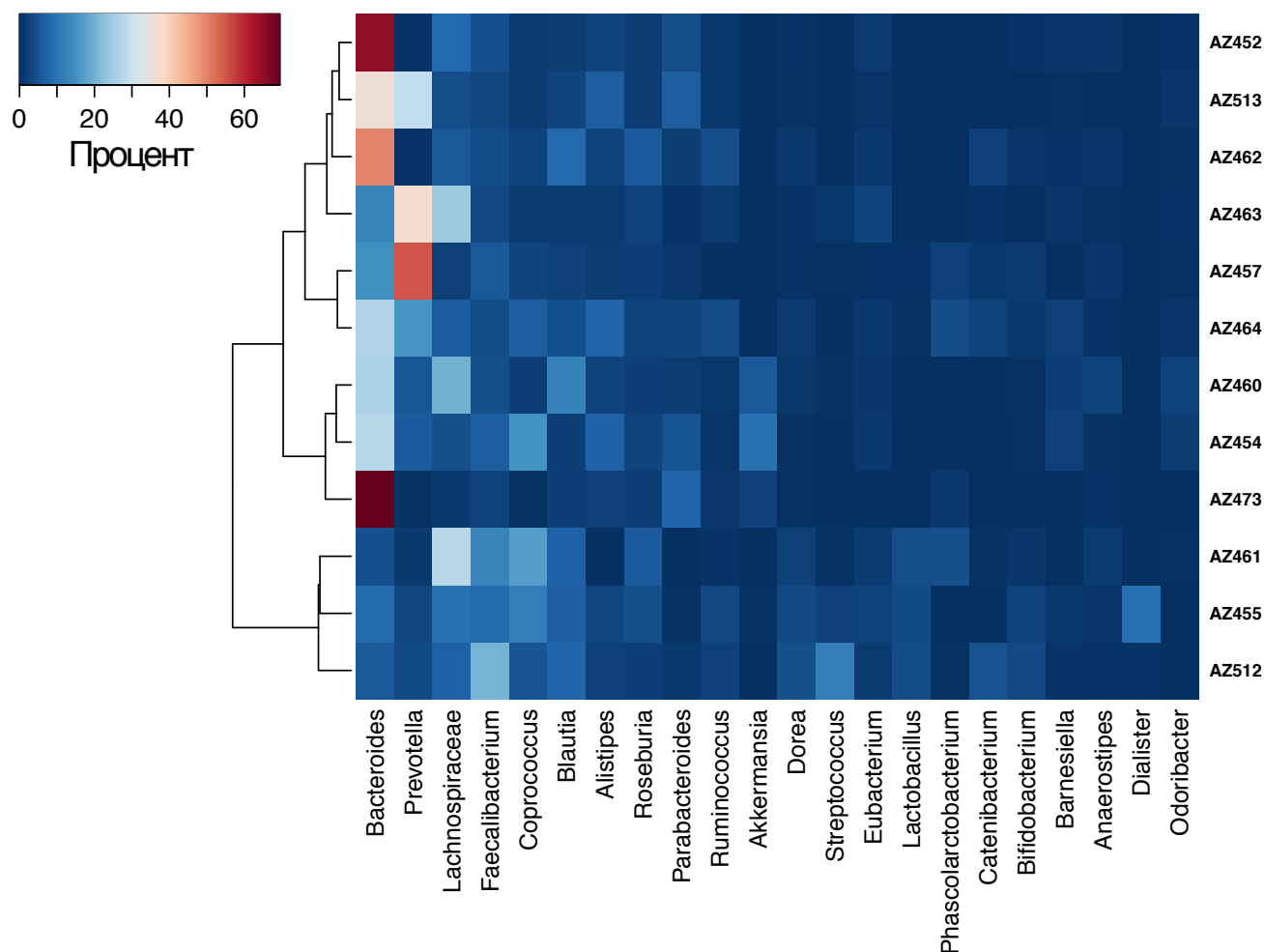


Рисунок 17. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах от работников вредного производства.

Более детально рассмотреть картину позволяет использование теста Манна-Уитни (Таблица 9). Из его результатов видно значительное преобладание вида *Adlercreutzia equolifaciens*, способной перерабатывать изофлавоны, содержащиеся к примеру в сое, в эстроген подобное вещество эквол¹³⁴. Также было повышено содержание бактерий рода *Tannerella*, являющихся оппортунистическими патогенами. Бактерия *Bilophila wadsworthia*¹³⁵, перепредставленная в опытной группе, ассоциирована с аппендицитом, хотя ее патогенность не доказана. У одного образца была найдена бактерия *Tropheryma whipplei*, являющаяся возбудителем болезни Уиппла¹³⁶.

**Результаты теста Манна-Уитни при сравнении контрольных образцов
и выборки метагеномов работников вредного производства**

А) По родовому составу

Род	p-value (FDR)	Средняя представленность в опытной выборке	Средняя представленность в контрольной выборке
<i>Adlercreutzia</i>	0,04	0,02	0,008
<i>Alistipes</i>	0,01	3,46	1,721
<i>Bacteroides</i>	0,008	29,09	10,63
<i>Barnesiella</i>	0,012	1,011	0,4
<i>Bilophila</i>	0,011	0,18	0,049
<i>Dialister</i>	0,038	0,83	0,985
<i>Enterococcus</i>	0,047	0,068	0,35
<i>Escherichia, Shigella</i>	0,038	0,19	2,19
<i>Odoribacter</i>	0,038	0,72	0,33
<i>Oxalobacter</i>	0,011	0,0009	0
<i>Parabacteroides</i>	0,017	2,84	1,13
<i>Tannerella</i>	0,008	0,26	0,11
<i>Tropheryma</i>	0,011	0,0012	0

Б) По геномному составу

Организм	p-value (FDR)	Средняя представленность в опытной выборке	Средняя представленность в контрольной выборке
<i>Adlercreutzia equolifaciens</i> DSM 19450	0,0375	0,0196	0,0078
<i>Alistipes onderdonkii</i> DSM 19147	0,0114	0,3830	0,2137
<i>Alistipes putredinis</i> DSM 17216	0,0149	2,5425	1,2248
<i>Alistipes shahii</i> WAL 8301	0,0369	0,3886	0,2533
<i>Anaerostipes caccae</i> DSM 14662	0,0393	0,0255	0,0398
<i>Bacteroides caccae</i> ATCC 43185	0,0285	0,7534	0,4957
<i>Bacteroides cellulosilyticus</i> DSM 14838	0,0015	0,7488	0,1480
<i>Bacteroides clarus</i> YIT 12056	0,0081	0,7212	0,0335
<i>Bacteroides coprophilus</i> DSM 18228	0,0387	0,5421	0,2674
<i>Bacteroides dorei</i> 5 1 36 D4	0,0209	2,5736	1,1972
<i>Bacteroides eggerthii</i> 1 2 48FAA	0,0052	1,0328	0,1967
<i>Bacteroides finegoldii</i> DSM 17565	0,0001	0,5574	0,1821
<i>Bacteroides fluxus</i> YIT 12057	0,0071	0,1963	0,0843
<i>Bacteroides fragilis</i> 3 1 12	0,0264	0,2304	0,2114
<i>Bacteroides intestinalis</i> DSM 17393	0,0032	0,3097	0,1244
<i>Bacteroides oleiciplenus</i> YIT 12058	0,0046	0,1496	0,0626
<i>Bacteroides ovatus</i> 3 8 47FAA	0,0132	0,5248	0,1461
<i>Bacteroides ovatus</i> ATCC 8483	0,0000	0,9615	0,1640

<i>Bacteroides plebeius</i> DSM 17135	0,0149	1,8500	0,4341
<i>Bacteroides</i> sp D2	0,0042	0,4703	0,0931
<i>Bacteroides stercoris</i> ATCC 43183	0,0445	1,3332	0,3421
<i>Bacteroides thetaiotaomicron</i> VPI 5482	0,0215	0,5747	0,3110
<i>Bacteroides uniformis</i> ATCC 8492	0,0029	4,8189	1,4737
<i>Bacteroides vulgatus</i> ATCC 8482	0,0022	7,6081	2,6261
<i>Bacteroides xylanisolvens</i> XB1A	0,0194	0,6064	0,2272
<i>Barnesiella intestinihominis</i> YIT 11860	0,0132	1,0110	0,4020
<i>Bilophila wadsworthia</i> 3 1 6	0,0110	0,1819	0,0487
<i>Clostridium</i> sp 7 2 43FAA	0,0000	0,0357	0,0012
<i>Clostridium</i> sp D5	0,0149	0,0168	0,0487
<i>Clostridium</i> sp HGF2	0,0000	2,3494	0,0491
<i>Clostridium</i> sp L2 50	0,0424	0,4249	2,0767
<i>Clostridium</i> sp SS2 1	0,0000	0,0000	0,6700
<i>Coprococcus eutactus</i> ATCC 27759	0,0089	0,3135	2,4766
<i>Dialister succinatiphilus</i> YIT 11850	0,0262	0,0473	0,5127
<i>Escherichia coli</i> str K12 substr MG1655	0,0393	0,1023	1,1328
<i>Escherichia</i> sp 1 1 43	0,0422	0,0709	0,8008
<i>Odoribacter splanchnicus</i> DSM 20712	0,0389	0,7219	0,3255
<i>Oxalobacter formigenes</i> OXCC13	0,0114	0,0009	0,0000
<i>Parabacteroides distasonis</i> ATCC 8503	0,0190	1,4290	0,4509
<i>Parabacteroides johnsonii</i> DSM 18315	0,0039	0,2159	0,0728
<i>Parabacteroides merdae</i> ATCC 43184	0,0181	1,1921	0,6046
<i>Roseburia inulinivorans</i> DSM 16841	0,0369	1,1625	2,1032
<i>Ruminococcus obeum</i> ATCC 29174	0,0264	0,3846	0,7949
<i>Ruminococcus</i> sp 18P13	0,0001	0,5213	0,1977
<i>Ruminococcus</i> sp 5 1 39BFAA	0,0000	0,0875	2,4269
<i>Ruminococcus</i> sp SR1 5	0,0342	1,9982	1,0201
<i>Ruminococcus torques</i> L2 14	0,0342	0,8812	1,5919
<i>Streptococcus</i> sp 2 1 36FAA	0,0393	0,0962	0,0077
<i>Streptococcus</i> sp C150	0,0209	0,0086	0,0523
<i>Tannerella</i> sp 6 1 58FAA CT1	0,0046	0,2594	0,1086
<i>Tropheryma whipplei</i> str Twist	0,0114	0,0012	0,0000

Данная выборка метагеномных образцов на первый взгляд не имеет явных отличий от контрольной выборки, но при более детальном рассмотрении становится был обнаружен повышенный уровень оппортунистических патогенов, что может быть связано с изменением иммунного статуса организма хозяина.

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

В главе 1 подробно описано изменение микробиоты в различных патологических состояниях, таких как болезнь Крона, атеросклероз, диабет второго типа. Вероятно, она делает свой вклад в развитие этих состояний, тем не менее механизм этого процесса не известен. Одним из возможных вариантов является влияние на организм хозяина через энергетический метаболизм, а именно путем производства короткоцепочечных жирных кислот – бутирата, пропионата и ацетата, чье количество зависит от состава микробиоты⁴⁶. Возможность предсказания изменений концентраций КЖК по микробному составу в будущем может быть использовано, в диагностических и исследовательских целях.

В данной работе была разработана механистическая модель производства КЖК и их дальнейшего потребления организмом хозяина (Рисунок 18) согласно экспериментальным данным, включающих секвенирование генов 16S рРНК и измерение концентраций бутирата, пропионата и ацетата в 98 образцах кала из более раннего метагеномного исследования¹²².

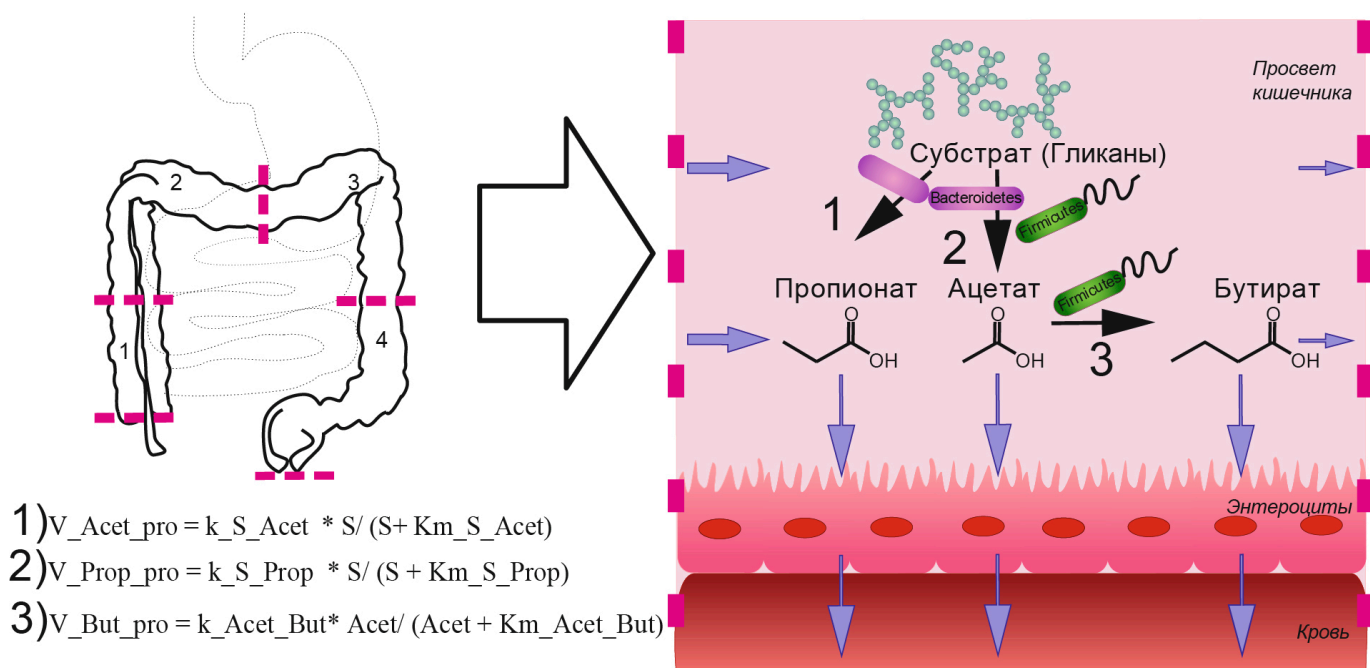


Рисунок 18. Механистическая модель производства и тока КЖК в кишечнике человека.

Алгоритм построения модели подробно рассмотрена (см. Материалы и методы). Вкратце, объем содержимого толстого кишечника делится на 4 отсека одинакового объема. Постоянный приток субстрата с пищей задан в 1ый отсек, откуда он последовательно переноситься во 2ой, 3й и 4й отсеки. В каждом отсеке происходят аналогичные процессы: субстрат сбраживается до пропионата и ацетата, а из ацетата синтезируется бутират. Сложные биохимические пути метаболизма КЖК сведены в модели тремя обобщенными реакциями, скорости которых задаются уравнениями типа Михаэлиса-Ментен (Рисунок 18).

В каждом отсеке заданы процессы абсорбции КЖК энтероцитами. Остатки субстрата и КЖК из 4го отсека элиминируются с постоянной скоростью. Все эти процессы описываются системой обыкновенных дифференциальных уравнений, которые и образуют модель (см. Приложение 3). Экспериментально измеренные концентрации КЖК в фекалиях соответствуют тем, что получаются на выходе из 4го отсека. Константы при уравнениях, описывающих производство КЖК, относятся к параметрам ассоциированным с микробиотой, в то время как

константы скорости транзита по кишечнику, всасывания в энтероциты и притока субстрата – параметры, связанные с особенностями организма хозяина.

Значения для одной части параметров модели были оценены из литературных данных, значения для другой части были получены вычислительным методом. При этом (см. Материалы и методы) были сопоставлены предсказанные моделью концентрации КЖК в 4-м (терминальном) отсеке с соответствующими концентрациями, измеренными в фекалиях.

После определения популяционных 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. Было выявлено, что бактерии рода *Faecalibacterium* определяют изменение константы скорости при образовании бутирата из ацетата. Представители рода *Coccoides* влияют на увеличение концентрации ацетата при расщеплении субстрата. Отмечено отрицательное влияние представителей родов *Ruminococcus* и *Bifidobacterium* на константу скорости расщепления субстрата до пропионата.

Тем не менее, учет этих данных в качестве ковариат обеспечивает около 30% вариабельности концентраций КЖК. Куда более весомый вклад (~ 70%) привносит учет возможной интериндивидуальной вариабельности по значениям параметров.

Итого, была создана механистическая модель производства КЖК в кишечнике человека. Найденные зависимости согласуются с экспериментальными данными.

Глава 4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

4.1 Создание программного комплекса по обработке метагеномных данных

В ходе данной работы был разработан и реализован алгоритм обработки метагеномных данных. Отличительная особенность этого алгоритма заключается в способности обрабатывать данные, полученные с разных платформ секвенирования, в том числе и с приборов ABI SOLiD 4, генерирующие риды в цветовом формате малой длины, что составляет определенную вычислительную сложность.

4.1.1 Фильтрация ридов

Секвенаторы SOLiD 4 и SOLiD 5500 не проводят предобработку ридов, поэтому начальным этапом является фильтрация по среднему значению качества. Это позволяет отсеять риды, содержащие ошибки секвенирования, такие как неправильное определение цвета и поликлональные риды. Кроме того, обрезаются низкокачественные 3' концы ридов алгоритмом SAET. В итоге отсеивается в среднем $18 \pm 3\%$ ридов. Следующий этап фильтрации – картирование на геном человека. Это стандартный этап при обработки метагеномных данных^{8,9,16}, его наличие связано с возможным загрязнением во время получения образцов или во время пробподготовки. Помимо оценки качества экспериментальных процедур, результат этой фильтрации служит первичным маркером возможных патологий. Так, наличие большого количества ДНК человека в метагеномном образце может объясняться воспалительными процессами, такими как мукозит.

4.1.2 Картирование на референсный каталог геномов

Оставшиеся после фильтрации высококачественные риды картируются на референсный каталог геномов. Картирование происходит программой bowtie⁹⁸. Допускается при картировании наличие максимум трех несовпадений, делеции и инсерции невозможны. В существующих метагеномных программных комплексах⁹⁷ часто используется алгоритм выравнивания BLAST¹³⁷, однако он не подходит для картирования коротких ридов и ридов в цветовом формате, в отличие от bowtie. Для картирования более длинных ридов не в цветовом формате, полученных с прибора Ion Torrent, была использована оптимизированная для этих целей программа bowtie2¹³⁸.

Референсный каталог геномов базируется на наборе кишечных бактерий, составленный HMP (Human Metagenome Project)⁷⁰. Это американский проект по изучению всего метагенома человека, в рамках которого запланировано секвенировать около 3000 геномов культивируемых и некультивируемых бактерий и архей. В созданном программном комплексе были использованы геномы из этого проекта, относящиеся к метагеному кишечника. Каталог также был расширен геномами бактерий патогенов и другими, ассоциированными с теми или иными патологическими состояниями согласно литературным данным. Также были добавлены геномы эукариот, а именно несколько представителей рода *Candida* и *Blastocystis hominis*. Это было сделано для более глубокого анализа клинических образцов. Дело в том, что смещение микробного гомеостаза в следствие кишечных заболеваний, приема антибиотиков или химиотерапии предположительно может дать возможность для заселения освободившихся экологических ниш патогенными организмами. Далее полученный каталог геномов был приведен к неизбыточному виду, т.е. все последовательности были выравнены друг против друга и кластеризованы по порогу 80% сходства на 80% длины. В итоговый набор вошли геномы, непохожие ни на один другой и по одному представителю из каждого кластера гомологии. Это было сделано для того, чтобы избежать разбавления представленностей геномов. В том случае, если

в каталоге несколько схожих геномов, ввиду равновероятного покрытия ридом схожих участков, они становятся равнопредставленными, хотя в действительности это может быть не так. Поскольку основной смысл определения бактериального состава сводится к выявлению функционального потенциала всего метагенома, а схожий геном означает чаще всего одинаковый функционал, то целесообразней использовать по одному представителю от кластера гомологии, тогда возможные различия между группами метагеномных образцов будут значительно лучше детектироваться. Такой подход имеет свое отражение в метагеномном анализе по последовательностям 16S рРНК генов. В таком виде анализа оперируют термином OTU (от англ. Operational Taxonomic Unit)¹⁰³ – условная таксономическая единица, соответствующая роду, виду или штамму и выделяемая по гомологии 16S рРНК гена. Ее появление связано с наличием большого количества бактерий, чьи геномы неизвестны. По этой же причине, нельзя назвать используемый каталог конечным. Каждое новое метагеномное исследование выявляет маркеры тех или иных состояний кишечника, секвенируются новые геномы из микробиоты. Увеличение каталога приводит к получению большего количества информации, а следовательно, к расширению исследования, что постепенно нивелирует недостаток описываемого метода, заключающийся в том, что можно детектировать только те микроорганизмы, которые есть в каталоге. В настоящий момент, каталог насчитывает 353 неизбыточных генома.

4.1.3 Подсчет покрытия референсных последовательностей

После картирования происходит подсчет покрытия, которое является числовой оценкой представленности геномов в образце. Существуют два вида покрытия – суммарное количество позиций генома, оказавшихся покрытыми хотя бы одним ридом (ширина покрытия), и суммарная длина всех ридов, картировавшихся на референсную последовательность (глубина покрытия). Ширина покрытия используется в качестве эмпирического порога отсечения:

геном считается хоть сколько-нибудь представленным, если покрыт хотя бы на 1% своей длины. Это сделано для того, чтобы избежать ложно-положительной детекции признаков по случайным причинам. Глубина покрытия нормируется на длину референсной последовательности и общее количество картировавшихся ридов образца. Такая нормировка соответствует преобразованиям, выполненными в других метагеномных исследованиях⁹. Таким образом для каждого образца формируется его вектор представленности признаков, в данном случае – геномов. Значения представленностей приводятся к процентам для более адекватного сравнения образцов между собой. К сожалению, на данный момент не существует способа сравнения абсолютного количества представленности геномов в метагеномных образцах, т.к. глубина секвенирования может значительно варьировать не только между различными приборами, но и в рамках одного запуска секвенатора. Помимо нормированных значений, в анализе используются и значения числа легших на референсную последовательность ридов – каунты⁹⁹. Такой метод позволяет детектировать статистически достоверные различия между группами образцов по низкопредставленным таксонам. Однако подсчет расстояний между образцами по таким данным затруднителен. Например, используемая в данной работе расстояние Брэй-Кертис¹⁰² требует суммы по образцам равной единице.

4.1.4 Статистический анализ

Имея количественную и качественную оценку состава метагеномов, появляется возможность для их анализа. Первая возможная оценка – индекс альфа-разнообразия. В экологии существует понятие «Принцип биоразнообразия», частично приписываемое закону Эшби. Его суть состоит в том, что чем разнообразнее экосистема, тем сложнее в ней сдвинуть равновесие и разрушить. Вероятно, что большое разнообразие является признаком здоровой устойчивой микробиоты. Помимо общей оценки, появляется возможность оценить представленность тех или иных бактерий. Особое значение имеют

различные патогены, такие как *Clostridium difficile*, *Pseudomonas aeruginosa*, *Candida albicans*, *Blastocystis hominis* и другие микроорганизмы, в норме не встречающиеся в метагеноме. Их наличие может быть следствием произошедшего смещения микробного равновесия, приведшее к возможности заселения патогенами. Подобное событие возможно, например, вследствие употребления антибиотиков, из-за которого комменсальная микробиота погибает, а патогены, несущие механизмы резистентности к этим антибиотикам, напротив выживают, занимая освободившиеся роли и активно размножаются. Помимо появления патогенов в значимых количествах, сигналом патологических состояний может стать значительное изменение представленности некоторых бактерий. Так, например *Escherichia coli* в норме составляет примерно 1-10% от общего микробного состава. Однако в случае воспалительных процессов, ее количество может возрасть свыше 80%¹⁶. В дальнейшем вероятно будут выделены организмы-маркеры тех или иных заболеваний и будут выявлены точные границы нормы их представленности. Это станет новым мощным диагностическим инструментом, неинвазивным и достаточно удобным.

Помимо общей оценки видового или родового состава микробиоты, используются различные методы сравнения образцов по всем признакам сразу. Для этого между образцами подсчитывают расстояние. Обычное Евклидово расстояние не подходит для метагеномики, т.к. не выполняется условие ортогональности: в геномах есть консервативные участки, секвенированный с них рид может картироваться в таком случае равновероятно на несколько последовательностей, пропадает условие линейной независимости. В метагеномике есть 2 наиболее используемые расстояния – Брэй-Кертис и UniFrac. Первая метрика BC часто используется в экологии, определяется как наименьшая сумма C_{ij} представлений таксонов, общих для двух сообществ, деленная на сумму их общего количества найденных таксонов S_i и S_j (9).

$$BC_{ij} = \frac{2C_{ij}}{S_i + S_j} \quad (9)$$

Вторая метрика применяется исключительно в метагеномике, и изначально была создана для анализа по 16S рРНК генов. Ее особенность заключается в том, что в ней учитываются не только представленности бактерий, но и филогенетические расстояния между ними. Мной эта метрика была изменена для допустимости использования с результатами полногеномного секвенирования и также использована в анализе (см. Материалы и методы).

Обе метрики показали схожие результаты, в частности при кластеризации по k-средним результат различался у 18 образцов из 96, что довольно немного, учитывая что степень кластеризации метагеномов недостоверна. Это в свою очередь делает спорным теорию о трех энтеротипах⁹ – дискретных кластеров микробиоты, дифференцированных по составу. Изначально предполагалось, что они станут своего рода кишечными аналогами групп крови, однако в последующих исследованиях энтеротипы либо находили лишь частичное подтверждение^{16,32}, либо утверждалось обратное, что кластеров микробиоты не существует, а стоит говорить о непрерывном градиенте сообществ³³.

Одна из главных задач исследования микробиоты это выявление различий между группами образцов, например здоровыми и больными или представителями различных стран. В метагеномике применяется ряд различных статистических методов для обнаружения достоверных различий и этот список расширяется с каждым новым исследованием. Сложность заключается в невозможности использовать параметрические статистические тесты, например t-тест, т.к. метагеномные данные распределены не нормально¹³⁰. Поэтому, он заменяется на непараметрический тест Манна-Уитни (он же U-тест) в качестве фильтрации различающихся признаков. Если необходимо сравнение нескольких групп образцов, используются также вариации от известного метода ANOVA¹⁰⁶ или дисперсионный анализ, например PERMANOVA¹³⁹ или ANOSIM¹²⁰, последний был использован в данной работе для выявления различий между группами образцов по географическому признаку.

4.1.5 Реализация программного комплекса

Алгоритм с базовым набором функций статистического анализа доступен для общественного пользования через веб-сервис MALINA¹⁴⁰. Полная версия алгоритма реализована на сервере НИИ ФХМ и активно используется в текущих метагеномных исследованиях. Программный комплекс состоит из нескольких модулей, написанных на языках программирования bash, Python 3.2, Perl 5. Данные об образцах и результаты картирования хранятся в СУБД Oracle 11.2. Статистический анализ реализован на языке программирования R 3.1. Весь комплекс был создан совместно с А. В. Тяхтом.

4.2 Анализ образцов из Российского метагеномного проекта

Программный комплекс был использован для анализа 96 сельских и городских образцов от здоровых доноров, собранных в рамках Российского метагеномного проекта (www.metagenome.ru). Экспериментальная работа и секвенирование проводилась в геномном центре НИИ ФХМ. Полный анализ образцов проводился совместно с А.В. Тяхтом, состоял из таксономической и функциональной частей. В этой работе рассмотрена таксономическая составляющая анализа.

Анализ включал в себя сравнение с метагеномными образцами из других исследований и его результаты были описаны выше. Была проведена процедура проверки состоятельности сравнения этих образцов, с учетом их получения с разных платформ секвенирования. Так один из российских образцов был дополнительно секвенирован дважды на приборе Ion Torrent и еще 5 образцов на приборе Illumina. Полученные вектора представленностей хорошо коррелировали между одними и теми же образцами, что стало подтверждением приемлемости сравнения образцов, секвенированных на различных приборах. Независимыми методами была подтверждена состоятельность работы алгоритма и адекватность получаемых результатов: некоторые образцы были обработаны этим

программным комплексом и сервисом MetaPhlan, полученные вектора представленностей также хорошо коррелировали.

В ходе исследования было выявлено, что российские метагеномы характеризуются меньшим содержанием представителей рода *Bacteroides*, чем европейские или американские. Было показано разделение российских образцов на 2 группы: в одной группе преобладал род *Prevotella*, в другой доминировало несколько родов из отдела Firmicutes, что приблизительно соответствует 2 из 3 ранее найденных энтеротипов. Еще один энтеротип, с родом *Bacteroides* в качестве доминанты, выявлен не был. Разделение сельских и городских метагеномов между кластерами оказалось примерно одинаковое. Для сравнения, на Западных и Китайских образцах было показано существование 2 или 3 кластеров^{9,32,16}: всегда присутствовал энтеротип с преобладающим родом *Prevotella*, энтеротип с *Bacteroides*, но не всегда удавалось найти энтеротип с доминирующими Firmicutes. Стоит заметить, что метагеномы с преобладанием рода *Bacteroides* ассоциированы с высоким потреблением животного белка и жира и меньшим потреблением клетчатки. Вероятно, рацион российских доноров в меньшей степени состоит из мяса, и в большей – из круп и другой пищи, богатой углеводами.

В выборке российских метагеномных образцов были найдены уникальные микробные сообщества, не встречающиеся в ранее изученных метагеномах. Что характерно, преимущественно такие образцы – от доноров из сельских регионов, что вызывает предположение, что дальнейшее изучение микробиоты кишечника жителей удаленных районов позволит выявить большее разнообразие сообществ. Оригинальность этих сообществ была определена по тройкам преобладающих родов. Такая оценка была выбрана в связи с тем, что около 70% всего покрытия составляют первые три рода, что делает такую классификацию удобной в применении для общей характеристики микробиоты. Хотя безусловно, патоген представленный хотя бы на 1% будет иметь значительное влияние на всю микробиоту. Около 43% российских метагеномов содержали уникальные тройки, не обнаруженные в не-российских образцах. Большая часть содержащихся в них

родов относилась к отделу Firmicutes, но были и представители Bacteroidetes, Verrucomicrobia, Actinobacteria, Proteobacteria, Tenericutes и Archaea.

Из отдела Firmicutes наиболее представленными родами были *Roseburia*, *Coprococcus*, *Faecalibacterium*, *Eubacterium*, *Ruminococcus*, *Blautia*, *Butyrivibrio* и неклассифицированные представители семейства Lachnospiraceae. Так же встречались метагеномы с нетипичными доминантами, к примеру в двух образцах преобладали микроорганизмы *Akkermansia muciniphila* и *Methanobrevibacter smithii*. Каждый из них в отдельности достаточно часто упоминается в исследованиях в той или иной ассоциации. Однако вместе они встречаются редко и ассоциированы со здоровым кишечником, как маркеры микробиоты с большим метаболическим потенциалом⁹⁴. Доминирование архей рода *Methanobrevibacter* вероятно отражает высокий уровень производства водорода бактериями, что также подтверждается отсутствием более эффективных водород-утилизирующих бактерий, таких как *Desulfovibrio* и *Desulfotobacterium*¹⁴¹.

Стоит отметить единичные случаи доминирования бактерий родов *Phascolarctobacterium* и *Lactobacillus* в образцах из Омской области и Хакасии соответственно. Было найдено, что эти же бактерии являются доминантами в микробиоте монголов из сельских районов, и их количество меньше у монголов из городов¹⁴².

Некоторые из обнаруженных оригинальных метагеномов имели в качестве доминант оппортунистические патогены. Так в нескольких образцах преобладала бактерия *Escherichia coli*, которая встречается в больших количествах при воспалительных процессах в кишечнике. В одном из образцов из Санкт-Петербурга было отмечено высокое содержание бактерии *Streptococcus infantarius*, являющейся маркером колоректального рака¹⁴³.

Помимо обычной для метагеномного исследования процедуры кластеризации по k-средним, была применена особая методика, определяющая компактные достоверно схожие подгруппы образцов. Ею были выделены две подгруппы. Первая подгруппа состояла из образцов из Омской области, чьими донорами были представители одной семьи. Ранее было показано, что фактор

совместного проживания значительно влияет на метагеном. Отличительная особенность таксономического состава этой группы заключается в высоком содержании родов *Prevotella*, *Coprococcus* и *Faecalibacterium*, что напоминает состав метагеномов из Малави и Венесуэлы. Вторая выделенная подгруппа состояла из образцов из сельского района Татарстана. Интересно заметить, что все эти образцы относятся к носителям уникальных сообществ. В частности, они на 80% состоят из представителей родов *Roseburia*, *Coprococcus*, *Faecalibacterium* и *Ruminococcus*. Наиболее представленными организмами из этих родов являются *Eubacterium rectale*, *C. eutactus*, *F. prausnitzii* и *R. bromii* соответственно. Такое сообщество может являться примером объединенного метаболизма и приносить пользу организму хозяина. В частности, *R. bromii* и *E. rectale* способны ферментировать устойчивый крахмал второго и третьего типа¹⁴⁴. Высокое содержание представителей родов *Roseburia* и *Coprococcus*, отличало метагеномы здоровых пожилых людей от пожилых пациентов, проживающих в стационаре, а снижение доли представителей рода *Ruminococcus* было ассоциировано с ухудшением состояния здоровья⁶⁸. Метаболомный анализ фекальных вод показал, что эти виды ассоциированы с повышенным уровнем бутирата⁶⁸, который играет важную роль в поддержании гомеостаза кишечника. Кроме того, *E. rectale*, *F. prausnitzii* и *Roseburia spp.* преобладали в образцах контрольной здоровой группы в исследовании метагенома больных диабетом 2го типа¹⁶, а последние 2 бактерии преобладали в контрольной группе по сравнению с метагеномами, больных атеросклерозом¹⁵. Таким образом, это новые выявленные сообщества могут быть примером действительно здоровой микробиоты, являющейся крайне полезной для организма хозяина. Предположительно, такой состав микробиоты поддерживается диетой, богатой натуральными продуктами с преобладанием углеводов, в частности клетчатки и резистентными формами крахмала.

Современный стиль жизни мегаполисов значительно ограничивает нас от взаимодействия с микробным окружением. Антибактериальные средства гигиены, пастеризованная пища, консерванты, все это несомненно влияет на нашу микробиоту, однако степень влияния все еще предстоит оценить. Но уже сейчас

имеются свидетельства того, что потеря некоторой ее составляющей, а также уменьшение взаимодействия с другими микроорганизмами приводит к росту аллергий и метаболических заболеваний¹⁴⁵. В связи с этим возникает вопрос – какую микробиоту следует принять за золотой стандарт? Одно из предположений заключается в поиске здоровой микробиоты среди жителей удаленных регионов, ведущих здоровый образ жизни, питающихся натуральными продуктами. Микробиота таких доноров должна быть достаточно богата ввиду отсутствия потребления антимикробных агентов и разнообразного здорового питания. Подобные сообщества представляют медицинский интерес, как универсальный материал для трансплантации микробиоты, с целью лечения кишечных заболеваний и облегчения болезненных состояний.

4.3 Видовое разнообразие метагеномов при различных диагнозах

Безусловно изучение микробиоты человека имеет фундаментальное значение, но не менее важным является использование полученных знаний на практике. Она тесно связана с нашим организмом, претерпевая изменения вместе с ним, при кардинальном изменении образа жизни или же при болезнях. Следовательно, метагеном может содержать в себе маркеры патологических состояний, даже на ранних стадиях и служить детектором начинающихся изменений в организме. Поэтому следующим этапом после определения нормофлоры кишечника стало проведение медицинских исследований. Описание здоровой микробиоты кишечника населения РФ позволило дополнить знания о нормофлоре, расширить выборку контролей. Эти образцы были использованы в данной работе для выявления специфических черт метагеномов из трех выборок образцов: от онкологических больных, страдающих алкоголизмом и работников на вредном производстве.

Не только болезни, но и методы лечения могут значительно сказываться на здоровье организма. Одним из наиболее тяжелых по своим последствиям видом лечения является химиотерапия. Противоопухолевые лекарственные препараты

блокируют рост и деление клеток, что приводит к целому ряду побочных эффектов, включая снижение иммунитета. В связи с этим, радикально изменяется состав кишечной микробиоты. При сравнении группа образцов от детей с экологическими заболеваниями со здоровой когортой, в ней были выявлены несколько значительно перепредставленных родов, в норме не содержащихся вовсе или же в малых количествах, например *Klebsiella*, дрожжи *Candida*, *Streptococcus*, *Pseudomonas*. Все они являются патогенными микроорганизмами и их появление является отражением тяжелой дисфункции иммунитета. Даже в численном соотношении метагеном онкологических больных отличается от здоровых образцов: глубина секвенирования отличается на порядок, что говорит о значительно меньшей концентрации бактерий в кале. В некоторых образцах содержание ДНК человека превышало 40%, что может быть следствием мукозита, не являющегося редкостью для онкологических больных, проходящих химиотерапию. Помимо общего количества, снижено и бактериальное разнообразие по сравнению с контролем, в свою очередь это может означать неустойчивость микробиоты пациентов к внешним воздействиям, что еще более усугубляет положение.

Микробиота человека является неотъемлемой частью его метаболизма, синтезируя КЖК, витамины, аминокислоты. Нарушения микробиотного гомеостаза ухудшает состояние больных, осложняет процесс и восстановление. Это исследование микробиоты онкологических больных было пилотным проектом, одним из первых исследований медицинского значения, с использованием метагеномного подхода. Уже сейчас существуют методы лечения, основанные на знаниях о микробиоте. Существуют примеры успешного лечения клостридиальной инфекции методом трансплантации кала²⁹. Дальнейшая работа в этом направлении позволит помочь онкологическим больным, компенсировать дисфункцию микробиоты, а следовательно повысить долю успешных исходов.

Исследование образцов микробиоты больных алкоголизмом и сотрудников предприятия из Сарова, работающих с радиоактивными материалами не выявило

кардинальных отличий от контрольной выборки из 96 российских образцов (Рисунок 19).

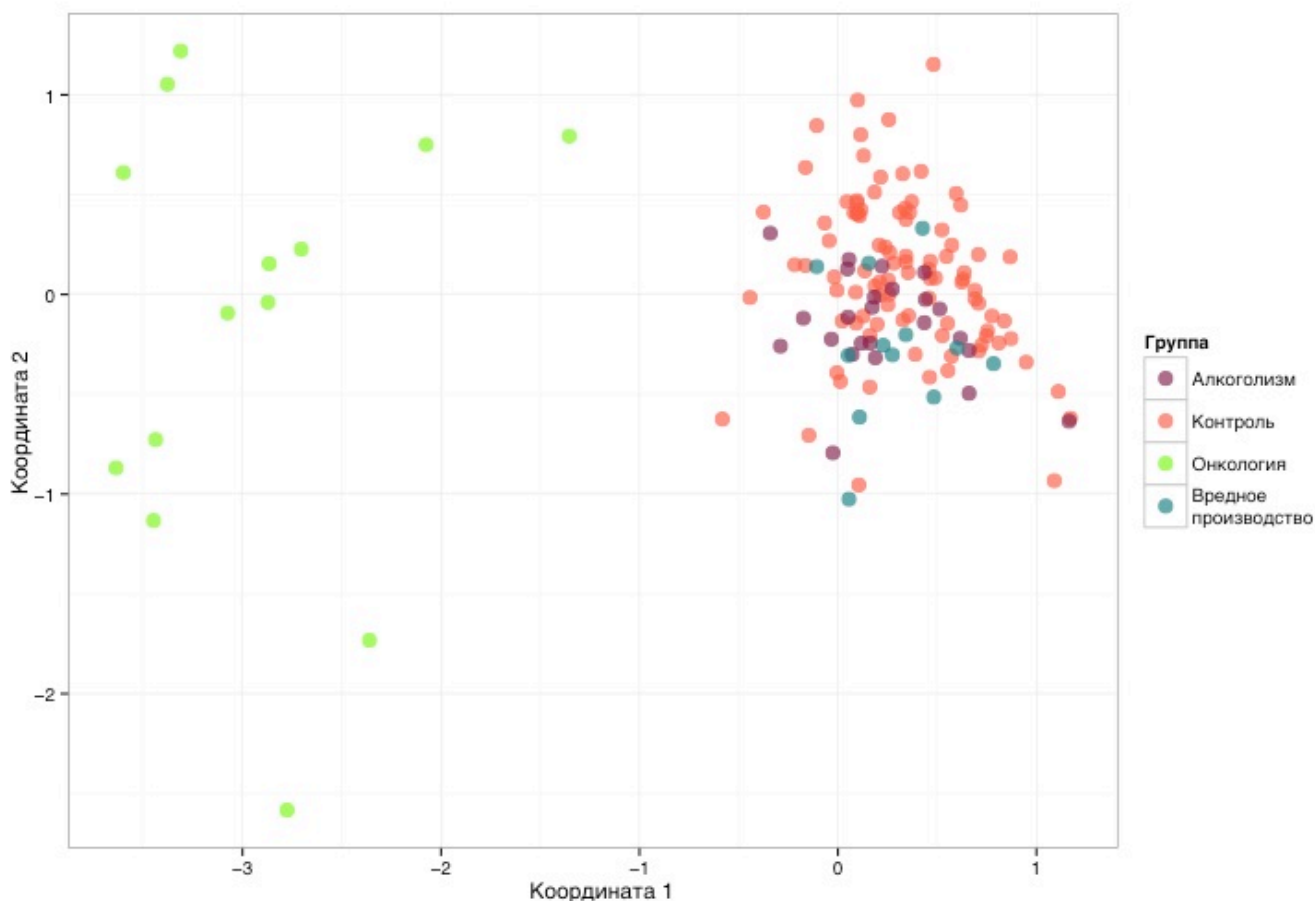


Рисунок 19. График MDS метагеномных образцов из клинических исследований и контрольной выборки. Цветом отмечена принадлежность к группам.

Обе группы можно условно отнести к первому и второму энтеротипу, т.к. доминантными родами в них являются *Bacteroides* и *Prevotella*, хотя у трех образцов из Сарова преобладают рода из отдела Firmicutes, что их относит к условному третьему энтеротипу. Метагеном больных алкоголизмом также отличился повышенным содержанием бактерий рода *Escherichia* и наличием ряда патогенов, в частности *Salmonella* и *Klebsiella*. Тем не менее, при более детальном рассмотрении в обоих случаях есть статистически достоверное повышенное содержание оппортунистических патогенов и бактерий, ассоциированных с

болезненными состояниями. Это может быть первым сигналом изменений в иммунной системе доноров и как следствие – изменения иммунной толерантности к этим микроорганизмам. Хотя на данном этапе картина не выглядит критичной, но при более продолжительном воздействии вредных условий иммунитет может претерпевать куда более сильные изменения, приводящие микробиотному дисбалансу, который наблюдается у онкологических больных.

4.4 Моделирование производства короткоцепочечных жирных кислот

До сих пор метагеномные исследования оставались вещью в себе. Влияние микробиоты кишечника на организм хозяина предполагалось лишь исходя из ее общего метаболического потенциала, т.е. по найденным в метагеноме биохимическим путям. Однако, для подтверждения такого взаимодействия необходима привязка к другим биохимическим параметрам организма. Таким параметром могут служить концентрации короткоцепочечных жирных кислот (КЖК), померенные в кале. Связывая эти значения с составом микробиоты можно получить представление о степени и характере влияния на организм хозяина. До текущего момента было предпринято несколько попыток создания подобной модели^{47,48}, однако ни одну из них не представляется возможным использовать для реальных данных, т.к. создаваемые предсказания работали только для модельных организмов с заданным составом микробиоты.

В этой работе была создана первичная обучаемая модель предсказания концентраций КЖК, поступающих в организм человека, в зависимости от состава метагенома. В ходе исследования были выявлены бактериальные рода, значительно влияющие на результат. Наиболее ожидаемая из найденных зависимостей, это влияние представленности бактерии *Faecalibacterium prausnitzii* на значение константы скорости при реакции образования бутирата из ацетата. У этой бактерии есть фермент бутирил КоА: ацетат КоА трансфераза, позволяющий преобразовывать из ацетата бутират, что экспериментально подтверждает полученные результаты¹⁴⁶. Также представленность этого рода влияет на

производство ацетата из субстрата, что подтверждается экспериментально¹⁴⁷. Также было обнаружено отрицательное влияние представленности бактерий *Ruminococcus* и *Bifidobacterium* на константу скорости преобразования субстрата в пропионат. В случае *Bifidobacterium* это может быть объяснено тем, что в кислой среде они активно размножаются и производят лактат, но из-за низких pH он не перерабатывается в КЖК, в том числе в пропионат¹⁴⁸.

Информация которая была использована – концентрации трех основных КЖК в фекалиях – может быть недостаточной для однозначной идентификации значений параметров модели. Дело в том, что эти концентрации являются суммарным результатом действия многих процессов, происходящих в течение продолжительного времени в разных отделах толстого кишечника. Это вклады в метаболизм КЖК различных групп бактерий, изменение содержания полисахаридного субстрата по ходу толстого кишечника, абсорбция КЖК энтероцитами.

Важно помнить, что измеренные концентрации КЖК в фекалиях не однозначно отражают их производство в кишечнике. Вероятно, что именно по этой причине не выявлена связь между уровнем ацетата и/или пропионата и плотностью семейства *Bacteroidetes*, – основного продуцента ацетата и пропионата из полисахаридов^{149,47}. С одной стороны, тот факт, что в нашей модели-реконструкции не удалось учесть вклад *Bacteroidetes* в метаболизм КЖК в качестве ковариационной зависимости может быть следствием ее несовершенства. С другой стороны, может быть так, что в исследуемой популяции за синтез ацетата и пропионата отвечают в равной мере как представители *Bacteroidetes*, так и *Firmicutes*.

4.5 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, 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 отличие 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 изменение состава микробиоты в сторону увеличения количества патогенных микроорганизмов. Для групп больных алкоголизмом 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 связано 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 для прогнозирования механизмов взаимодействия микробиоты с организмом хозяина.

СПИСОК ЛИТЕРАТУРЫ

1. Добрецов Н.Л. О ранних стадиях зарождения и эволюции. // Вестник ВОГиС. 2005. Т. 9. С. 43-54.
2. Lederberg J., McCray A. Ome Sweet 'Omics—a genealogical treasury of words. // Scientist. 2001. Vol. 15. P. 8.
3. Amann R.I., Ludwig W., Schleifer K.H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. // Microbiol Rev. 1995. Vol. 59. P. 143-169.
4. Venter J.C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. // Science. 2004. Vol. 304, №5667. P. 66-74.
5. Daniel R. The soil metagenome—a rich resource for the discovery of novel natural products. // Curr Opin Biotechnol. 2004. Vol. 15, №3, P. 199-204.
6. Pride D.T., Schoenfeld T. Genome signature analysis of thermal virus metagenomes reveals Archaea and thermophilic signatures. // BMC Genomics. 2008. Vol. 9. P. 420.
7. Qin J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. // Nature. 2010. Vol. 464, №7285. P. 59-65.
8. Nelson K.E. et al. A catalog of reference genomes from the human microbiome. // Science. 2010. Vol. 328, №5981. P. 994-999.
9. Arumugam M. et al. Enterotypes of the human gut microbiome. // Nature. 2011. Vol. 473, №7346. P. 174-180.
10. Kurokawa K. et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. // DNA res. 2007. Vol. 14. P. 169-181.
11. Jean G. L. et al. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. // Current Opinion in Biotechnology. 2013. Vol. 24, №2. P. 160-168.
12. Gibson Glenn R. Fibre and effects on probiotics (the prebiotic concept). // Clinical Nutrition Supplements. 2004. Vol. 1, №2. P. 25–31
13. Sears Cynthia L. A dynamic partnership: Celebrating our gut flora. // Anaerobe. 2005. Vol. 11, №5. P. 247–251.

14. Vipperla K., O'Keefe S. J. The microbiota and its metabolites in colonic mucosal health and cancer risk. // NCP. 2012. Vol. 27. P. 624-635.
15. Karlsson F.H. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. // Nat. comm. 2012. Vol. 3. P. 1245.
16. Qin J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. // Nature. 2012. Vol. 490. P. 55-60.
17. Yatsunenko T. et al. Human gut microbiome viewed across age and geography. // Nature. 2012. Vol. 486. P. 222-227.
18. Handelsman J. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. // Chemistry & biology. 1998. Vol. 5. P. 245-249.
19. Pace N.R., Delong, E.F. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. // Journal of Bacteriology. 1991. Vol. 173, №14. P. 4371–4378.
20. Poinar H.N. et al. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. // Science. 2005. Vol. 311, №5759, P. 392–394.
21. Gill S. R. et al. Metagenomic analysis of the human distal gut microbiome. // Science. 2006. Vol. 312. P. 1355–1359.
22. De Filippo C., Cavalieri D., Di Paola M., Ramazzotti M., Poullet J.B., Massart S., Collini S., Pieraccini G., Lionetti P.. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. // Proc Natl Acad Sci U S A. 2010. Vol. 107, №33. P. 14691-14696
23. Tyakht A. et al. Human gut microbiota community structures in urban and rural populations in Russia. // Nat Commun. 2013. Vol. 4, P. 2469.
24. Hofer U. Microbiome: bacterial imbalance in Crohn's disease. // Nat Rev Microbiol. 2014. Vol. 12, №5. P. 312.
25. Fava F., Danese S. Intestinal microbiota in inflammatory bowel disease: friend of foe? // World J Gastroenterol. 2011. 1Vol. 7, №5. P. 557-566.
26. Balamurugan R., Rajendiran E., George S., Samuel G.V., Ramakrishna B.S. Real-time polymerase chain reaction quantification of specific butyrate-

- producing bacteria, *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. // *J Gastroenterol Hepatol*. 2008. Vol. 23, №8. P. 1298-1303.
27. Furlow B. Gut microbe composition and metabolic syndrome. // *Lancet Diabetes Endocrinol*. 2013. Vol. 1. P. 4-5
 28. Wu X. et al. Molecular characterisation of the faecal microbiota in patients with type II diabetes. // *Current microbiology*. 2010. Vol. 61. P. 69–78.
 29. Borody T. J., Khoruts A. Fecal microbiota transplantation and emerging applications. // *Nature reviews Gastroenterology and Hepatology*. 2012. Vol. 9. P. 88–96.
 30. Eckburg P.B. et al. Diversity of the human intestinal microbial flora. // *Science*. 2005. Vol. 308, №5728. P. 1635-1638.
 31. Ley R.E., Peterson D.A., Gordon J.I.. Ecological and evolutionary forces shaping microbial diversity in the human intestine. // *Cell*. 2006. Vol. 124, №4. P. 837-848.
 32. Wu G.D. et al. Linking long-term dietary patterns with gut microbial enterotypes. // *Science*. 2011. Vol. 334. P. 105-108.
 33. Jeffery I. B., Claesson, M. J., O'Toole P. W., Shanahan F. Categorization of the gut microbiota: enterotypes or gradients? // *Nature Reviews Microbiology*. 2012. Vol. 10, №9. P. 591–592.
 34. Macfarlane S., Macfarlane G.T. Regulation of short-chain fatty acid production. // *Proc Nutr Soc*. 2003. Vol. 62, №1. P. 67-72.
 35. Koropatkin N. M., Cameron E. A., Martens E. C. How glycan metabolism shapes the human gut microbiota. *Nature Reviews. Microbiology*. 2012. Vol. 10, №5. P. 323–335.
 36. Flint H. J., Scott K., Duncan S. H. Microbial degradation of complex carbohydrates in the gut. // *Gut Microbes*. 2012. Vol. 3, №4. P. 289–306.
 37. Falony G., Calmeyn T., Leroy F., De Vuyst L. Coculture fermentations of *Bifidobacterium* species and *Bacteroides thetaiotaomicron* reveal a mechanistic

- insight into the prebiotic effect of inulin-type fructans. // *Appl Environ Microbiol.* 2009. Vol. 75, P. 2312-2319.
38. Barboza M., et al. Glycoprofiling bifidobacterial consumption of galacto-oligosaccharides by mass spectrometry reveals strain-specific, preferential consumption of glycans. // *Appl Environ Microbiol.* 2009. Vol. 75, №23. P. 7319-7325.
 39. Chassard C., Delmas E., Robert C., Bernalier-Donadille A. The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens. // *FEMS Microbiol Ecol.* 2010. Vol. 74. P. 205-213.
 40. Derrien M., Collado M.C., Ben-Amor K., Salminen S., de Vos W.M.. The Mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. // *Appl Environ Microbiol.* 2008. Vol. 74. P. 1646-1648.
 41. Chen W., Liu F., Ling Z., Tong X., Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. // *PLoS One.* 2012. Vol. 7, №6. P. e39743
 42. Cummings J.H., Pomare E.W., Branch W.J., Naylor C.P., Macfarlane G.T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. // *Gut.* 1987. Vol. 28, №10. P. 1221-1227.
 43. Lin H.V., Frassetto A., Kowalik E.J. Jr., Nawrocki A.R., Lu M.M., Kosinski J.R., Hubert J.A., Szeto D., Yao X., Forrest G., Marsh D.J. Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. // *PLoS One.* 2012. Vol. 7, №4. P. e35240.
 44. Brown A.J. et al. The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. // *J Biol Chem.* 2003. Vol. 278, №13. P. 11312-11319
 45. Ventura M., Turrioni F., Canchaya C., Vaughan E.E., O'Toole P.W., van Sinderen D. Microbial diversity in the human intestine and novel insights from metagenomics. // *Front Biosci (Landmark Ed).* 2009. Vol. 14. P. 3214-3221.

46. Mahowald M.A. et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. // *Proc Natl Acad Sci U S A*. 2009. Vol. 106, №14. P. 5859-5864.
47. Muñoz-Tamayo R., Laroche B., Walter E., Doré J., Duncan S.H., Flint H.J., Leclerc M. Kinetic modelling of lactate utilization and butyrate production by key humancolonial bacterial species. // *FEMS Microbiol Ecol*. 2011. Vol. 76, №3. P. 615-624.
48. Shoaie S., Karlsson F., Mardinoglu A., Nookaew I., Bordel S., Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. // *Sci Rep*. 2013. Vol. 3. P. 2532.
49. Wells, J. M., Rossi, O., Meijerink, M., & van Baarlen, P. (2011). Epithelial crosstalk at the microbiota-mucosal interface. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl , 4607–14
50. McLoughlin R.M., Mills K.H. Influence of gastrointestinal commensal bacteria on the immune responses that mediate allergy and asthma. // *J Allergy Clin Immunol*. 2011. Vol. 127, №5. P. 1097-1107.
51. Hiutung C., Sarkis K.M. Innate immune recognition of the microbiota promotes host-microbial symbiosis. // *Nature Immunology*. 2013. Vol. 14. P. 668–675.
52. Mazmanian S.K., Liu C.H., Tzianabos A.O., Kasper D.L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. // *Cell*. 2005. Vol. 122. P. 107–118.
53. Penders J., Stobberingh E.E., van den Brandt P.A., Thijs C. The role of the intestinal microbiota in the development of atopic disorders. // *Allergy*. 2007. Vol. 62, №11. P. 1223-1236.
54. Wardwell L.H., Huttenhower C., Garrett W.S. Current concepts of the intestinal microbiota and the pathogenesis of infection. // *Curr Infect Dis Rep*. 2011. Vol. 13, №1. P. 28-34.
55. Rea M.C. et al. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. // *Proc Natl Acad Sci USA*. 2010. Vol. 107. P. 9352–9357.

56. Fukuda S. et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. // *Nature*. 2011. Vol. 469. P. 543–547.
57. J Segain D.R. de la Bletiere et al. Butyrate inhibits inflammatory responses through NF κ B inhibition: implications for Crohn's disease. // *Gut*. 2000. Vol. 47, №3. P. 397–403.
58. O'Keefe S.J. et al. Products of the colonic microbiota mediate the effects of diet on colon cancer risk. // *J Nutr*. 2009. Vol. 139, №11. P. 2044-2048.
59. Foster J.A., McVey Neufeld K.A. Gut-brain axis: how the microbiome influences anxiety and depression. // *Trends Neurosci*. 2013. Vol. 36, №5. P. 305-312.
60. Forsythe P., Kunze W. Voices from within: gut microbes and the CNS. // *Cellular and molecular life sciences* . 2013. Vol. 70. P. 55–69.
61. Raoult D. The globalization of intestinal microbiota. // *Eur J Clin Microbiol Infect Dis*. 2010. Vol. 29, №9. P. 1049-50.
62. Rampelli S., Candela M., Turroni S., Biagi E., Collino S., Franceschi C., O'Toole P.W., Brigidi P. Functional metagenomic profiling of intestinal microbiome in extreme ageing. // *Aging*. 2013. Vol. 5, №12. P. 902-912.
63. Schnorr S.L. et al. Gut microbiome of the Hadza hunter-gatherers. // *Nat Commun*. 2014. Vol. 5. P. 3654.
64. Ward T.L., Hosid S., Ioshikhes I., Altosaar I. Human milk metagenome: a functional capacity analysis. // *BMC Microbiol*. 2013. Vol. 13. P. 116.
65. Jiménez E., Marín M.L., Martín R., Odriozola J.M., Olivares M., Xaus J, Fernández L.,Rodríguez J.M. Is meconium from healthy newborns actually sterile? // *Res Microbiol*. 2008. Vol. 159, №3. P.187-193.
66. Palmer C., Bik E.M., DiGiulio D.B., Relman D.A., Brown P.O. Development of the human infant intestinal microbiota. // *PLoS Biol*. 2007. Vol. 5, №7. P. e177
67. Grönlund, M., Grześkowiak, Ł. (2011). Influence of mother's intestinal microbiota on gut colonization in the infant. // *Gut Microbes*. 2011. Vol. 2, №4. P. 227-233.

68. Hopkins M.J., Sharp R., Macfarlane, G.T. Age and disease related changes in intestinal bacterial populations assessed by cell culture. 16S rRNA abundance, and community cellular fatty acid profiles. // Gut . 2001. Vol. 48. P. 198–205.
69. Turnbaugh P.J. et al. A core gut microbiome in obese and lean twins. // Nature. 2009. Vol. 457, №7228. P. 480-484.
70. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. // Nature. 2012. Vol. 486, №7402. P. 207-214.
71. Claesson M.J. et al. Gut microbiota composition correlates with diet and health in the elderly. // Nature. 2012. Vol. 488. P. 178–184.
72. Baumgart D.C., Sandborn W.J. Inflammatory bowel disease: clinical aspects and established and evolving therapies. // The Lancet. 2007. Vol. 369, №9573. P. 1641–1657.
73. Strachan D.P. Hay fever, hygiene, and household size. // BMJ. 1989. Vol. 299, №6710. P. 1259-1260.
74. Huycke M.M., Gaskins H.R. Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. // Exp Biol Med. 2004. Vol. 229, №7. P. 586-597.
75. Castellarin M. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. // Genome Res. 2012. Vol. 22, №2. P. 299-306.
76. van Nood E. et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. // New England Journal of Medicine. 2003. Vol. 368, №5. P. 407-415.
77. Metchnikoff E., Mitchell P.C.S. The Prolongation of Life: optimistic studies // William Heinemann. 1907.
78. Sekirov I., Russell S.L., Antunes L.C., Finlay B.B. Gut microbiota in health and disease. // Physiol Rev. 2010. Vol. 90. P. 859–904.
79. Turnbaugh P.J., Ley R.E., Mahowald M.A., Magrini V., Mardis E.R., Gordon J.I. An obesity-associated gut microbiome with increased capacity for energy harvest. // Nature. 2006. Vol. 444. P. 1027-1031.
80. Karlsson F.H., Tremaroli V., Nookaew I., Bergström G., Behre C.J., Fagerberg B., Nielsen J., Bäckhed F. Gut metagenome in European women with normal,

- impaired and diabetic glucose control. // *Nature*. 2013. Vol. 498, №7452. P. 99-103.
81. Oren O. et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. // *Proc. Natl Acad. Sci. USA*. 2011. Vol. 108. P. 4592–4598.
 82. Mutlu E. a et al. Colonic microbiome is altered in alcoholism. // *Am J Physiol Gastrointest Liver Physiol*. 2012. Vol. 302. P. G966–978.
 83. Smith, M.I. et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. // *Science (New York, N.Y.)*. 2013. Vol. 339, №6119. P. 548–554.
 84. Jernberg C., Löfmark S., Edlund C., Jansson J. K. Long-term impacts of antibiotic exposure on the human intestinal microbiota. // *Microbiology*. 2010. Vol. 156, №11. P. 3216–3223.
 85. Jakobsson H.E., Jernberg C., Andersson A.F., Sjölund-Karlsson M., Jansson J.K., Engstrand L. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. // *PLoS One*. 2010. Vol. 5, №3. P. e9836.
 86. Jernberg C., Löfmark S., Edlund C., Jansson J.K. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. // *ISME J*. 2007. Vol. 1, №1. P. 56-66.
 87. Löfmark S., Jernberg C., Jansson J.K., Edlund C. Clindamycin-induced enrichment and long-term persistence of resistant *Bacteroides* spp. and resistance genes. // *J Antimicrob Chemother*. 2006. Vol. 58, №6. P. 1160-1167.
 88. Karami N., Martner A., Enne V.I., Swerkersson S., Adlerberth I., Wold A.E. Transfer of an ampicillin resistance gene between two *Escherichia coli* strains in the bowel microbiota of an infant treated with antibiotics. // *J Antimicrob Chemother*. 2007. Vol. 60, №5. P. 1142-1145.
 89. Andremont A. Commensal Flora May Play Key Role in Spreading Antibiotic Resistance. // *ASM news*. 2003. Vol. 69, №12. P. 601-607.
 90. Salyers A.A., Gupta A., Wang Y. Human intestinal bacteria as reservoirs for antibiotic resistance genes. // *Trends Microbiol*. 2004. Vol. 12, №9. P. 412-416.
 91. Aubry-Damon H. et al. Antimicrobial resistance in commensal flora of pig farmers. // *Emerg Infect Dis*. 2004. Vol. 10, №5. P. 873-879.

92. Heuer H., Smalla K. Horizontal gene transfer between bacteria. // *Environ Biosafety Res.* 2007. Vol. 6, №1-2. P. 3-13.
93. Bartosch S., Woodmansey E.J., Paterson J.C.M., McMurdo M.E.T., Macfarlane G.T. Microbiological effects of consuming a synbiotic containing *Bifidobacterium bifidum*, *Bifidobacterium lactis* and oligofructose in elderly persons, determined by real-time polymerase chain reaction and counting of viable bacteria. // *Clin Infect Dis.* 2005. Vol. 40. P. 28-37.
94. Le Chatelier et al., Richness of human gut microbiome correlates with metabolic markers. // *Nature.* 2013. Vol. 500 №7464. P. 541-546.
95. Lepage P. et al. A metagenomic insight into our gut's microbiome. // *Gut.* 2013. Vol. 62, №1. P. 146–158.
96. Liu B., Pop M. ARDB-Antibiotic Resistance Genes Database. // *Nucleic Acids Res.* 2009. Vol. 37. P. D443-447
97. Arumugam M., Harrington E.D., Foerstner K.U., Raes J., P. Bork. SmashCommunity: A metagenomic annotation and analysis tool. // *Bioinformatics.* 2010. Vol. 26. P. 2977-2978.
98. Langmead B., Trapnell C., Pop M., Salzberg S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. // *Genome Biology.* 2009. Vol. 10. P. R25.
99. Paulson J.N., Stine O.C., Bravo H.C., and Pop M. Differential abundance analysis for microbial marker-gene surveys. // *Nature Methods.* 2013. Vol. 10. P. 1200–1202.
100. Whittaker R.H. Evolution and measurement of species diversity // *Taxon.* 1972. Vol. 2. P. 213—251.
101. Österreicher F., Vajda I. A new class of metric divergences on probability spaces and its statistical applications. // *Ann. Inst. Statist. Math.* 2003. Vol. 55, №3. P. 639–653.
102. Bray J. R., J. T. Curtis. An ordination of upland forest communities of southern Wisconsin. // *Ecological Monographs.* 1957. Vol. 27. P. 325-349.

103. Lozupone C., Hamady M., Knight R.. UniFrac - An Online Tool for Comparing Microbial Community Diversity in a Phylogenetic Context. // BMC Bioinformatics. 2006. Vol. 7. P. 371.
104. Caporaso J.G., et al. QIIME allows analysis of high-throughput community sequencing data. // Nature Methods. 2010. Vol. 7, №5. P. 335-336.
105. Kruskal J. B., Wish M. Multidimensional Scaling, Sage University Paper series on Quantitative Application in the Social Sciences. Beverly Hills and London: Sage Publications. 1978. 07-011.
106. Field A. Analysis of variance (ANOVA). Encyclopedia of measurement and statistics. 2007. P. 33-36.
107. Breiman L. Random Forests. // Machine Learning. 2001. Vol. 45, №1. P. 5–32.
108. Iverson V. et al. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. // Science. 2012. Vol. 335, №6068. P. 587-590.
109. Kurtz S., Phillippy A., Delcher A.L., Smoot M., Shumway M., Antonescu C., Salzberg S.L. Versatile and open software for comparing large genomes. // Genome Biol. 2004. Vol. 5, №2. P. R12.
110. Aaron R., Quinlan I., Hall M. BEDTools: a flexible suite of utilities for comparing genomic features. // Bioinformatics. 2010. Vol. 26, №6. P. 841–842.
111. Wang Q., Garrity G.M., Tiedje J.M., Cole J.R. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. // Appl Environ Microbiol. 2007. Vol. 73, №16. P. 5261-5267.
112. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2010.
113. Goslee, S.C., Urban, D.L. The ecodist package for dissimilarity-based analysis of ecological data. // Journal of Statistical Software. 2007. Vol. 22, №7. P. 1-19.
114. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. // Nucleic Acids Res. 2004. Vol. 32, №5. P.1792-1797.

115. Sokal R., Michener C. A statistical method for evaluating systematic relationships. // University of Kansas Science Bulletin. 1958. Vol. 38. P. 1409–1438.
116. Maechler M., Rousseeuw P., Struyf A., Hubert M., Hornik, K. cluster: Cluster Analysis Basics and Extensions. // R package version 1.15.2. 2014.
117. Rousseeuw P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. // J. Comput. Appl. Math. 1987. Vol. 20. P. 53–65.
118. Tibshirani R., Walther G. Cluster Validation by Prediction Strength. // Journal of Computational and Graphical Statistics. 2005. Vol. 14. P. 511-528.
119. Ryota Suzuki, Hidetoshi Shimodaira Pvcust: an R package for assessing the uncertainty in hierarchical clustering. // Bioinformatics. 2006. Vol. 22, №12. P. 1540-1542.
120. Clarke K.R. Non-parametric multivariate analysis of changes in community structure. // Australian Journal of Ecology. 1993. Vol. 18. P. 117–143.
121. Letunic I., Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. // Bioinformatics. 2007. Vol. 23, №1. P. 127-128.
122. Schwiertz A., Taras D., Schäfer K., Beijer S., Bos N.A., Donus C., Hardt P.D. Microbiota and SCFA in lean and overweight healthy subjects. // Obesity. 2010. Vol. 18, №1. P. 190-195.
123. Wagener S., Shankar K.R., Turnock R.R., Lamont G.L., Baillie C.T. Colonic transit time--what is normal? // J Pediatr Surg. 2004. Vol. 39, №2. P. 166-169.
124. von Engelhardt W., Gros G., Burmester M., Hansen K., Becker G., Rechkemmer G. Functional role of bicarbonate in propionate transport across guinea-pig isolated caecum and proximal colon. // J Physiol. 1994. Vol. 477. P. 365-371.
125. Harig J.M., Soergel K.H., Barry J.A., Ramaswamy K. Transport of propionate by human ileal brush-border membrane vesicles. // Am J Physiol. 1991. Vol. 260, №5. P. G776-782.

126. Bergman E.N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. // *Physiol Rev.* 1990. Vol. 70, №2. P. 567-590.
127. Gijs den Besten et al. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. // *Journal of Lipid Research.* 2013. Vol. 54, №9. P. 2325-2340
128. Chan P.L., Jacqmin P., Lavielle M., McFadyen L., Weatherley B. The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. // *J Pharmacokinet Pharmacodyn.* 2011. Vol. 38, №1. P. 41-61.
129. Sasson A., Michael T.P. Filtering error from SOLiD Output. // *Bioinformatics.* 2010. Vol. 26, №6. P. 849-850.
130. Sanli K., Karlsson F.H., Nookaew I., Nielsen J. FANTOM: Functional and taxonomic analysis of metagenomes. // *BMC Bioinformatics.* 2013. Vol. 14. P. 38.
131. Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. // *Nat Methods.* 2012. Vol. 9, №8. P. 811-814.
132. Ferreira M., Muls A., Dearnaley D., Andreyev J. Microbiota and radiation-induced bowel toxicity: lessons from inflammatory bowel disease for the radiation oncologist. // *The Lancet Oncology.* 2014. Vol. 15, №3. P. e139-147
133. Brook I. Veillonella infections in children. // *J. Clin. Microbiol.* 1996. Vol. 34, №5. P. 1283-1285.
134. Juan J.P., Wang J.H., Liu X. Metabolism of dietary soy isoflavones to equol by human intestinal microflora – implications for health. // *Mol Nutr Food Res.* 2007. Vol. 51, №7. P. 765-781.
135. Finegold S., Summanen P., Hunt Gerardo S., Baron E. Clinical importance of *Bilophila wadsworthia*. // *Eur J Clin Microbiol Infect Dis.* 1992. Vol. 11. P. 1058-1063.

136. Relman D.A., Schmidt T.M., MacDermott R.P., Falkow S. Identification of the uncultured bacillus of Whipple's disease. // *N Engl J Med*. 1992. Vol. 327, №5. P. 293-301.
137. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. // *J Mol Biol*. 1990. Vol. 215. P. 403-410.
138. Langmead B., Salzberg S. Fast gapped-read alignment with Bowtie 2. // *Nature Methods*. 2012. Vol. 9. P. 357-359.
139. Anderson M.J. A new method for non-parametric multivariate analysis of variance. // *Austral Ecology*. 2001. Vol. 26. P. 32–46.
140. Tyakht A.V. et al. 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.
141. Hansen E.E. et al. Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. // *Proc Natl Acad Sci U S A*. 2011. Vol. 108. P. 4599-4606.
142. Zhang J. et al. The diversity of intestinal microbiota of Mongolians living in Inner Mongolia, China. // *Benef Microbes*. 2013. Vol. 4. P. 319-28.
143. Biarc J. et al. Carcinogenic properties of proteins with pro-inflammatory activity from *Streptococcus infantarius* (formerly *S.bovis*). // *Carcinogenesis*. 2004. Vol. 25. P. 1477 14- 84.
144. Leitch E.C., Walker A.W., Duncan S.H., Holtrop G., Flint H.J. Selective colonization of insoluble substrates by human faecal bacteria. // *Environ. Microbiol*. 2007. Vol. 9. P. 667–679.
145. Blaser M.J., Falkow S. What are the consequences of the disappearing human microbiota? // *Nat Rev Microbiol*. 2009. Vol. 7. P. 887-894.
146. Louis P., Flint H. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. // *FEMS Microbiol Lett*. 2009. Vol. 294, №1. P. 1-8.
147. Liu C., Finegold S.M., Song Y., Lawson P.A.. Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*,

- Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. // *Int J Syst Evol Microbiol*. 2008. Vol. 58. P. 1896-1902.
148. Belenguer A., Duncan S.H., Holtrop G., Anderson S.E., Lobley G.E., Flint H.J. Impact of pH on lactate formation and utilization by human fecal microbial communities. // *Appl Environ Microbiol*. 2007. Vol. 73, №20. P. 6526-6533.
149. Miller T.L. The pathway of formation of acetate and succinate from pyruvate by *Bacteroides succinogenes*. // *Arch Microbiol*. 1978. Vol. 117, №2. P. 145-152.

СПИСОК ИЛЛЮСТРАЦИЙ

Рисунок 1. Гистограмма количества статей с упоминанием слова «metagenome» согласно статистике PubMed.

12

Рисунок 2. Схема анализа метагеномных данных от полногеномного секвенирования.

31

Рисунок 3. Алгоритм поточной обработки метагеномных данных.

38

Рисунок 4. Сравнение процентов представленности бактериальных родов в образце, секвенированном на приборе ABI SOLiD 4 (X77_WC) и дважды – на IonTorrent (со средней длиной ридов 120 ± 15 п.о. (X77_WC_short) и 229 ± 58 п.о. (X77_WC_long)).

49

Рисунок 5. Гистограмма среднего процента представленности отделов в российских образцах и в остальных.

51

Рисунок 6. Относительная представленность бактериальных родов, составляющих 80% общего покрытия, по географическим группам.

52

Рисунок 7. Тепловая карта представленности прокариотических родов в российских образцах.

53

Рисунок 8. График MDS Российских образцов.

54

Рисунок 9. График MDS мировых образцов.

56

Рисунок 10. Дерево мировых метагеномных образцов.

59

Рисунок 11. Графики изменения параметров оценки кластеризации российских образцов по основным используемым в метагеномике метрикам – расстояния Дженсен-Шеннона (красная линия), Брэй-Кертис (синяя линия), UniFrac (зеленая линия).

60

Рисунок 12. Результат иерархической кластеризации российских образцов. Красным выделены найденные схожие подгруппы образцов.

62

Рисунок 13. Тепловая карта наиболее представленных родов в найденных подгруппах.

63

Рисунок 14. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах кала пациентов онкологии.

66

Рисунок 15. График MDS метагеномных образцов онкологических больных. Цветом отмечена принадлежность к тому или иному пациенту.

67

Рисунок 16. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах кала страдающих от алкоголизма.

69

Рисунок 17. Тепловая карта наиболее представленных родов (95% суммарного покрытия) в метагеномных образцах от работников вредного производства.

74

Рисунок 18. Механистическая модель производства и тока КЖК в кишечнике человека.

78

Рисунок 19. График MDS метагеномных образцов из клинических исследований и контрольной выборки. Цветом отмечена принадлежность к группам.

92

Таблица 1. Результаты сравнительных метагеномных исследований.	21
Таблица 2. Отражение патологий на состав кишечной микробиоты.	26
Таблица 3. Статистика данных по мировым образцам.	48
Таблица 4. Уникальные доминантные тройки родов в российских образцах.	55
Таблица 5. Результаты теста Манна-Уитни: списки различающихся бактериальных родов.	57
Таблица 6. Результаты статистического теста ANOSIM при сравнении российской выборки образцов с другими.	58
Таблица 7. Манн-Уитни тест для представленности родов в двух кластерах из кластеризации российских образцов по метрике UniFrac.	65
Таблица 8. Результаты теста Манна-Уитни при сравнении контрольных образцов и выборки метагеномов больных алкоголизмом.	70
Таблица 9. Результаты теста Манна-Уитни при сравнении контрольных образцов и выборки метагеномов работников вредного производства.	75

ПРИЛОЖЕНИЕ 1

Образец	Страна	Технология секвенирования	Город\село	Возраст	Пол	ИМТ	%картирования на hg18	%картирования на референс
AFR_4461147	Малави	pyro454	село	26	Ж	20,2	0,05	40,89
AFR_4461154	Малави	pyro454	село	23	Ж	21,5	0,07	34,86
AFR_4461159	Малави	pyro454	село	25	Ж	na	0,12	34,68
AFR_4461162	Малави	pyro454	село	27	Ж	20	0,02	34,22
AFR_4461167	Малави	pyro454	село	25	Ж	24,2	0,04	34,66
AMA_4461122	Венесуэла	pyro454	село	15	Ж	24,7	0,04	60,15
AMA_4461123	Венесуэла	pyro454	село	35	Ж	na	0,03	45
AMA_4461125	Венесуэла	pyro454	село	5	М	17,4	0,02	31,41
AMA_4461126	Венесуэла	pyro454	село	6	М	18,2	0,03	43,98
AMA_4461127	Венесуэла	pyro454	село	9	Ж	14,1	0,02	62,02
AMA_4461128	Венесуэла	pyro454	село	11	Ж	16,5	0,03	44,58
AMA_4461129	Венесуэла	pyro454	село	18	Ж	20,7	0,03	50,81
AMA_4461130	Венесуэла	pyro454	село	29	Ж	24,2	0,03	24,58
AMA_4461131	Венесуэла	pyro454	село	53	М	26,1	0,03	40,94
AMA_4461138	Венесуэла	pyro454	село	20	Ж	25	0,05	44,45
Kh_232	Россия	solid	село	57	Ж	na	0,09	22,34
Kh_236	Россия	solid	село	57	Ж	na	0,79	30,51
Kh_240	Россия	solid	село	43	М	na	0,19	29,69
Kh_242	Россия	solid	село	62	М	na	0,21	20,16
Kh_245	Россия	solid	село	45	М	na	0,06	20,25
Kh_247	Россия	solid	село	48	М	na	0,09	17,37
Kh_249	Россия	solid	село	43	М	na	0,1	44,39
Kh_250	Россия	solid	село	45	Ж	na	0,09	19,53
MH0001	Дания	illumina	город	49	Ж	25,55	0,05	50,16
MH0002	Дания	illumina	город	59	Ж	27,28	0,05	26,32
MH0003	Дания	illumina	город	69	М	33,19	0,01	34,67
MH0004	Дания	illumina	город	59	М	31,18	0,97	28,07
MH0005	Дания	illumina	город	64	М	21,68	0,23	35
MH0006	Дания	illumina	город	59	Ж	22,38	0,4	37,21
MH0007	Дания	illumina	город	69	М	33,6	0,32	30,46
MH0008	Дания	illumina	город	59	М	24,35	0,11	46,85
MH0009	Дания	illumina	город	64	М	29,04	0,02	14,46
MH0010	Дания	illumina	город	64	М	33,27	0,08	56,59
MH0011	Дания	illumina	город	na	Ж	22,31	0,09	20,58
MH0012	Дания	illumina	город	42	Ж	32,1	0,02	18,78
MH0013	Дания	illumina	город	54	М	20,46	1,65	31,29
MH0014	Дания	illumina	город	54	Ж	38,49	0,06	34,7
MH0015	Дания	illumina	город	59	М	25,47	0,05	40,07
MH0016	Дания	illumina	город	49	Ж	30,5	0,01	42,24
MH0017	Дания	illumina	город	64	М	21,81	0,04	32,84
MH0018	Дания	illumina	город	49	М	31,37	2,12	53,67
MH0019	Дания	illumina	город	44	Ж	20,01	0,04	47,21
MH0020	Дания	illumina	город	63	Ж	33,23	0,08	54,14

МН0021	Дания	illumina	город	49	Ж	25,42	0,07	36,86
МН0022	Дания	illumina	город	64	М	24,42	0,17	16,24
МН0023	Дания	illumina	город	69	М	31,74	0,02	31,65
МН0024	Дания	illumina	город	59	Ж	22,72	0,03	28,42
МН0025	Дания	illumina	город	49	Ж	34,2	0,04	25,63
МН0026	Дания	illumina	город	49	Ж	37,32	0,08	26,21
МН0027	Дания	illumina	город	59	Ж	23,07	0,05	43,8
МН0028	Дания	illumina	город	44	Ж	22,7	0,02	29,66
МН0030	Дания	illumina	город	59	М	35,21	0,29	12,52
МН0031	Дания	illumina	город	69	М	22,34	0,29	13,58
МН0032	Дания	illumina	город	69	М	35,28	0,04	33,68
МН0033	Дания	illumina	город	59	Ж	31,95	0,01	30,63
МН0034	Дания	illumina	город	54	М	39,97	0,01	38,04
МН0035	Дания	illumina	город	49	М	22,66	1,15	24,02
МН0036	Дания	illumina	город	64	М	30,74	0,07	30,19
МН0037	Дания	illumina	город	44	М	24,02	0,02	31,94
МН0038	Дания	illumina	город	54	Ж	21,97	0,12	30,26
МН0039	Дания	illumina	город	58	М	23,07	0,04	24,97
МН0040	Дания	illumina	город	67	Ж	20,87	0,07	23,41
МН0041	Дания	illumina	город	59	М	23,17	0,32	40,7
МН0042	Дания	illumina	город	49	М	24,46	0,2	27,49
МН0043	Дания	illumina	город	69	М	23,72	0,35	31,87
МН0044	Дания	illumina	город	64	М	24,48	0,02	39,12
МН0045	Дания	illumina	город	59	М	25,11	0,05	40,9
МН0046	Дания	illumina	город	54	М	23,74	0,02	23,62
МН0047	Дания	illumina	город	69	Ж	30,4	0,21	14,23
МН0048	Дания	illumina	город	54	Ж	19,4	0,19	21,96
МН0049	Дания	illumina	город	44	Ж	35,52	0,09	29,52
МН0050	Дания	illumina	город	49	М	25,08	0,01	26,96
МН0051	Дания	illumina	город	69	Ж	23,15	0,31	25,03
МН0052	Дания	illumina	город	49	Ж	33,18	0,08	28,46
МН0053	Дания	illumina	город	49	Ж	32,7	0,03	14,47
МН0054	Дания	illumina	город	49	М	20,31	0,07	27,83
МН0055	Дания	illumina	город	59	М	30,29	0,02	22,72
МН0056	Дания	illumina	город	54	М	25,35	1,01	23,93
МН0057	Дания	illumina	город	54	Ж	32,98	0,15	28,6
МН0058	Дания	illumina	город	54	Ж	22,04	0,05	36,42
МН0059	Дания	illumina	город	59	М	33,27	0,08	32,36
МН0060	Дания	illumina	город	54	М	23,52	0,02	17,46
МН0061	Дания	illumina	город	69	Ж	30,12	0,88	49,87
МН0062	Дания	illumina	город	49	Ж	37,54	13,71	30,85
МН0063	Дания	illumina	город	59	М	30,23	0,02	17,92
МН0064	Дания	illumina	город	54	Ж	23,18	0,02	33,82
МН0065	Дания	illumina	город	59	М	28,23	0,04	18,35
МН0066	Дания	illumina	город	44	Ж	20,79	0,03	19,06
МН0067	Дания	illumina	город	54	М	21,07	0,01	39,44
МН0068	Дания	illumina	город	54	Ж	28,97	0,03	51,07
МН0069	Дания	illumina	город	59	Ж	36,71	0,04	24,33
МН0070	Дания	illumina	город	49	М	22,69	0,04	29,25
МН0071	Дания	illumina	город	44	Ж	25,37	0,09	24,69

MH0072	Дания	illumina	город	64	Ж	40,21	0,03	49,26
MH0073	Дания	illumina	город	54	М	32,49	0,08	38,85
MH0074	Дания	illumina	город	49	Ж	20,46	0,01	38,6
MH0075	Дания	illumina	город	64	М	30,55	0,01	33,52
MH0076	Дания	illumina	город	69	Ж	34,78	0,34	29
MH0077	Дания	illumina	город	49	Ж	24,92	0,15	27,72
MH0078	Дания	illumina	город	49	Ж	36,9	0,63	21,78
MH0079	Дания	illumina	город	64	Ж	19,97	0,7	23,7
MH0080	Дания	illumina	город	59	Ж	18,59	0,08	44,72
MH0081	Дания	illumina	город	49	Ж	37,95	0,02	28,07
MH0082	Дания	illumina	город	59	Ж	22,56	0,11	32,39
MH0083	Дания	illumina	город	54	Ж	30,59	0,01	28,89
MH0084	Дания	illumina	город	64	М	31,67	0,03	40,11
MH0085	Дания	illumina	город	59	Ж	36,46	0,01	45,11
MH0086	Дания	illumina	город	59	Ж	21,59	0,01	37,69
NOV_280	Россия	solid	город	42	М	21,63	3,22	31,09
NOV_281	Россия	solid	город	47	М	22,6	0,08	36,68
NOV_282	Россия	solid	город	52	М	31,41	0	25,89
NOV_283	Россия	solid	город	43	Ж	23,14	0,41	37,8
NOV_284	Россия	solid	город	41	М	25,08	0,17	19,04
NOV_285	Россия	solid	город	29	М	22,92	0,58	18,63
NOV_286	Россия	solid	город	27	Ж	18,71	0,12	20,75
NOV_287	Россия	solid	город	48	Ж	27,85	0,18	15,22
OM_176	Россия	solid	село	42	Ж	20,06	0,43	18,83
OM_178	Россия	solid	село	25	Ж	na	0,11	24,05
OM_179	Россия	solid	село	62	М	na	0,16	11,74
OM_180	Россия	solid	село	43	Ж	21,64	0,05	18,02
OM_181	Россия	solid	село	51	Ж	23,51	0,14	14,67
OM_182	Россия	solid	село	31	Ж	19,71	0,5	18,74
OM_184	Россия	solid	село	18	М	23,89	0,1	25,49
OM_185	Россия	solid	село	16	М	17,11	0,56	17,6
OM_190	Россия	solid	село	42	Ж	33,6	0,29	27,37
OM_191	Россия	solid	село	43	Ж	35,01	0,25	23,46
OM_192	Россия	solid	село	40	Ж	35,64	0,16	26,35
OM_193	Россия	solid	село	44	Ж	34,72	0,18	27,82
OM_194	Россия	solid	село	41	Ж	35,86	0,25	25,7
OM_195	Россия	solid	село	43	Ж	36,14	0,27	21,56
OMS_251	Россия	solid	село	42	Ж	na	0,46	31,92
RND_301	Россия	solid	город	55	Ж	24,54	0,57	22,42
RND_304	Россия	solid	город	55	Ж	24,09	0,97	23,49
RND_305	Россия	solid	город	35	М	30,42	0,53	27,07
RND_306	Россия	solid	город	65	Ж	30,12	0,23	21,76
RND_307	Россия	solid	город	32	Ж	22,99	0,18	23,36
RND_308	Россия	solid	город	41	М	33,21	0,21	38,91
RND_311	Россия	solid	город	21	М	31,44	0,24	33,84
RND_312	Россия	solid	город	21	М	21,04	0,14	35,04
RND_313	Россия	solid	город	23	Ж	24,02	0,74	24,95
RND_314	Россия	solid	город	23	Ж	21,99	0,1	20,27
RND_315	Россия	solid	город	26	Ж	20,18	0,39	22,68
RND_316	Россия	solid	город	27	М	26,06	0,31	29,96

RND_317	Россия	solid	город	50	Ж	23,39	0,82	24,94
SAR_262	Россия	solid	город	26	М	na	66,27	1,17
SAR_263	Россия	solid	город	63	Ж	na	0,6	27,22
SAR_264	Россия	solid	город	47	Ж	na	0,51	23,74
SAR_265	Россия	solid	город	72	Ж	na	0,41	15,43
SAR_266	Россия	solid	город	30	Ж	na	0,48	18,73
SAR_267	Россия	solid	город	61	Ж	na	0,46	21,61
SAR_268	Россия	solid	город	24	М	na	0,26	19,77
SAR_269	Россия	solid	город	68	Ж	na	0,52	25,09
SAR_270	Россия	solid	город	72	М	na	0,33	23,45
SAR_271	Россия	solid	город	70	М	na	1,41	14,46
SAR_272	Россия	solid	город	22	М	na	0,51	17,27
SAR_273	Россия	solid	город	28	М	na	0,17	14,41
SAR_274	Россия	solid	город	64	М	na	0,46	23,91
SAR_275	Россия	solid	город	31	М	na	0,19	24,21
SAR_276	Россия	solid	город	39	М	na	1,05	23,45
Spb_100_40P	Россия	solid	город	20	М	23,53	0,05	30,18
Spb_101_41P	Россия	solid	город	20	М	20,34	0,09	14,89
Spb_103_43P	Россия	solid	город	19	М	23,55	0,1	21,46
Spb_105_45P	Россия	solid	город	20	М	22,53	0,12	27,35
Spb_106_46P	Россия	solid	город	20	М	22,95	0,52	24,53
Spb_61_1P	Россия	solid	город	20	М	28,08	0,22	31,24
Spb_66_6P	Россия	solid	город	20	М	32,15	0,03	18,25
Spb_69_9P	Россия	solid	город	21	М	20,76	0,49	28,89
Spb_70_10P	Россия	solid	город	21	М	22,86	0,07	27,86
Spb_72_12P	Россия	solid	город	21	М	20,53	0,04	23,18
Spb_73_13P	Россия	solid	город	20	М	17,16	0,11	37,01
Spb_74_14P	Россия	solid	город	20	М	21,97	0,04	27,71
Spb_76_16P	Россия	solid	город	20	М	19,37	0,05	48,6
Spb_98_38P	Россия	solid	город	20	М	20,11	0,07	40,01
SRR413556	Китай	illumina	na	28	М	26,73	0	52,77
SRR413557	Китай	illumina	na	26	Ж	16,87	0	42,17
SRR413558	Китай	illumina	na	27	Ж	23,92	0	36,55
SRR413559	Китай	illumina	na	27	Ж	17,26	0	50,54
SRR413560	Китай	illumina	na	27	М	26,3	0	45,21
SRR413561	Китай	illumina	na	27	Ж	17,4	0	40,37
SRR413562	Китай	illumina	na	29	Ж	17,69	0	49,5
SRR413563	Китай	illumina	na	21	Ж	26,3	0	37,72
SRR413564	Китай	illumina	na	28	Ж	17,29	0	29,95
SRR413565	Китай	illumina	na	27	Ж	18,75	0	37,5
SRR413566	Китай	illumina	na	28	М	26,83	0	52,07
SRR413575	Китай	illumina	na	52	М	23,88	0	52,03
SRR413576	Китай	illumina	na	53	М	24,61	0	31,42
SRR413577	Китай	illumina	na	48	М	25,4	0	44,53
SRR413578	Китай	illumina	na	53	Ж	23,03	0	62,23
SRR413579	Китай	illumina	na	53	Ж	26,67	0	49,51
SRR413580	Китай	illumina	na	60	М	22,34	0	27,52
SRR413581	Китай	illumina	na	69	М	20,76	0	30,88
SRR413582	Китай	illumina	na	61	Ж	25,07	0	30,03
SRR413583	Китай	illumina	na	44	Ж	24,97	0	50,5

SRR413584	Китай	illumina	na	56	Ж	19,23	0	28,97
SRR413585	Китай	illumina	na	54	Ж	19,48	0	50,74
SRR413586	Китай	illumina	na	41	Ж	17,4	0	39,04
SRR413587	Китай	illumina	na	55	Ж	25,39	0	57,32
SRR413588	Китай	illumina	na	50	Ж	21,08	0	36,98
SRR413589	Китай	illumina	na	39	М	15,57	0	38,89
SRR413590	Китай	illumina	na	40	М	31,4	0	33,47
SRR413592	Китай	illumina	na	54	Ж	22,22	0	55,42
SRR413593	Китай	illumina	na	67	Ж	24,34	0	43,36
SRR413594	Китай	illumina	na	67	М	24,02	0	67,18
SRR413595	Китай	illumina	na	38	М	22,68	0	27,07
SRR413596	Китай	illumina	na	45	Ж	21,08	0	42,93
SRR413597	Китай	illumina	na	56	М	23,88	0	48,58
SRR413598	Китай	illumina	na	53	Ж	26,67	0	26,23
SRR413599	Китай	illumina	na	52	Ж	26	0	48,13
SRR413600	Китай	illumina	na	55	М	21,55	0	48,43
SRR413601	Китай	illumina	na	52	Ж	23,81	0	60,29
SRR413602	Китай	illumina	na	47	М	24,68	0	52,75
SRR413603	Китай	illumina	na	55	Ж	25,64	0	66,57
SRR413604	Китай	illumina	na	50	М	24,22	0	61,12
SRR413606	Китай	illumina	na	58	Ж	22,06	0	56,65
SRR413608	Китай	illumina	na	60	Ж	20,2	0	65,46
SRR413609	Китай	illumina	na	46	Ж	29,64	0	35,64
SRR413610	Китай	illumina	na	61	Ж	23,73	0	51,7
SRR413611	Китай	illumina	na	40	М	20,02	0	41,18
SRR413612	Китай	illumina	na	43	Ж	16,65	0	44,97
SRR413613	Китай	illumina	na	60	Ж	21,33	0	35,96
SRR413614	Китай	illumina	na	52	Ж	24,84	0	48,86
SRR413616	Китай	illumina	na	54	М	23,36	0	38,92
SRR413617	Китай	illumina	na	57	М	29,38	0	43,48
SRR413618	Китай	illumina	na	56	Ж	20,81	0	70,34
SRR413619	Китай	illumina	na	53	М	26,77	0	41,15
SRR413620	Китай	illumina	na	59	М	20,45	0	39,18
SRR413621	Китай	illumina	na	57	М	24,62	0	60,64
SRR413622	Китай	illumina	na	50	М	25,6	0	52,36
SRR413623	Китай	illumina	na	62	М	23,43	0	42,53
SRR413624	Китай	illumina	na	50	Ж	18,31	0	56,03
SRR413625	Китай	illumina	na	55	М	23,36	0	28,2
SRR413627	Китай	illumina	na	41	Ж	20,69	0	42,66
SRR413628	Китай	illumina	na	42	М	21,34	0	47,46
SRR413629	Китай	illumina	na	47	Ж	23,83	0	68,06
SRR413630	Китай	illumina	na	49	Ж	23,92	0	27,08
SRR413635	Китай	illumina	na	44	Ж	24,14	0	43,63
SRR413636	Китай	illumina	na	48	Ж	24,06	0	29,46
SRR413637	Китай	illumina	na	55	М	24,61	0	44,28
SRR413638	Китай	illumina	na	47	Ж	21,64	0	40,95
SRR413639	Китай	illumina	na	42	Ж	27,99	0	37,39
SRR413758	Китай	illumina	na	56	Ж	22,04	0	44,49
SRR413768	Китай	illumina	na	56	М	27,64	0	32,63
SRS011061	США	illumina	город	na	na	na	0	38,13

SRS011134	СИА	illumina	город	na	na	na	0	27,15
SRS011239	СИА	illumina	город	na	na	na	0	55,49
SRS011271	СИА	illumina	город	na	na	na	0	37,37
SRS011302	СИА	illumina	город	na	na	na	0	20,41
SRS011405	СИА	illumina	город	na	na	na	0	58,5
SRS011452	СИА	illumina	город	na	na	na	0	59,97
SRS011529	СИА	illumina	город	na	na	na	0	42,62
SRS011586	СИА	illumina	город	na	na	na	0	45,94
SRS012273	СИА	illumina	город	na	na	na	0	45,33
SRS012902	СИА	illumina	город	na	na	na	0	80,88
SRS013158	СИА	illumina	город	na	na	na	0	73,5
SRS013215	СИА	illumina	город	na	na	na	0	84,95
SRS013476	СИА	illumina	город	na	na	na	0	45,34
SRS013521	СИА	illumina	город	na	na	na	0	53,19
SRS013687	СИА	illumina	город	na	na	na	0	39,85
SRS013800	СИА	illumina	город	na	na	na	0	52,91
SRS013951	СИА	illumina	город	na	na	na	0	29,32
SRS014235	СИА	illumina	город	na	na	na	0	39,74
SRS014287	СИА	illumina	город	na	na	na	0	27,77
SRS014313	СИА	illumina	город	na	na	na	0	49,42
SRS014459	СИА	illumina	город	na	na	na	0	49,37
SRS014613	СИА	illumina	город	na	na	na	0	65,29
SRS014683	СИА	illumina	город	na	na	na	0	67,47
SRS014923	СИА	illumina	город	na	na	na	0	27,63
SRS014979	СИА	illumina	город	na	na	na	0	18,6
SRS015065	СИА	illumina	город	na	na	na	0	47,56
SRS015133	СИА	illumina	город	na	na	na	0	29,09
SRS015190	СИА	illumina	город	na	na	na	0	73,79
SRS015217	СИА	illumina	город	na	na	na	0	23,63
SRS015264	СИА	illumina	город	na	na	na	0	62,29
SRS015369	СИА	illumina	город	na	na	na	0	65,62
SRS015578	СИА	illumina	город	na	na	na	0	65,16
SRS015663	СИА	illumina	город	na	na	na	0	41,33
SRS015782	СИА	illumina	город	na	na	na	0	35
SRS015794	СИА	illumina	город	na	na	na	0	60,76
SRS015854	СИА	illumina	город	na	na	na	0	48,33
SRS015960	СИА	illumina	город	na	na	na	0	68,4
SRS016018	СИА	illumina	город	na	na	na	0	66,03
SRS016056	СИА	illumina	город	na	na	na	0	63,1
SRS016095	СИА	illumina	город	na	na	na	0	36,16
SRS016203	СИА	illumina	город	na	na	na	0	50,17
SRS016267	СИА	illumina	город	na	na	na	0	72,48
SRS016335	СИА	illumina	город	na	na	na	0	25,95
SRS016495	СИА	illumina	город	na	na	na	0	19,27
SRS016753	СИА	illumina	город	na	na	na	0	74,87
SRS016954	СИА	illumina	город	na	na	na	0	54,62
SRS016989	СИА	illumina	город	na	na	na	0	55,3
SRS017103	СИА	illumina	город	na	na	na	0	65,05
SRS017191	СИА	illumina	город	na	na	na	0	68,47
SRS017247	СИА	illumina	город	na	na	na	0	74,97

SRS017307	CША	illumina	город	na	na	na	0	45,24
SRS017433	CША	illumina	город	na	na	na	0	52,92
SRS017521	CША	illumina	город	na	na	na	0	64,6
SRS017701	CША	illumina	город	na	na	na	0	56,41
SRS017821	CША	illumina	город	na	na	na	0	46,19
SRS018133	CША	illumina	город	na	na	na	0	53,16
SRS018351	CША	illumina	город	na	na	na	0	53,06
SRS018427	CША	illumina	город	na	na	na	0	55,59
SRS018575	CША	illumina	город	na	na	na	0	59,76
SRS018656	CША	illumina	город	na	na	na	0	27,05
SRS018817	CША	illumina	город	na	na	na	0	67,62
SRS019030	CША	illumina	город	na	na	na	0	62,14
SRS019161	CША	illumina	город	na	na	na	0	52,57
SRS019267	CША	illumina	город	na	na	na	0	67,13
SRS019397	CША	illumina	город	na	na	na	0	63,58
SRS019582	CША	illumina	город	na	na	na	0	39,52
SRS019601	CША	illumina	город	na	na	na	0	68,29
SRS019685	CША	illumina	город	na	na	na	0	41,92
SRS019787	CША	illumina	город	na	na	na	0	71,37
SRS019910	CША	illumina	город	na	na	na	0	59,46
SRS019968	CША	illumina	город	na	na	na	0	63,42
SRS020233	CША	illumina	город	na	na	na	0	44,45
SRS020328	CША	illumina	город	na	na	na	0	36,95
SRS020869	CША	illumina	город	na	na	na	0	39,73
SRS021484	CША	illumina	город	na	na	na	0	56,81
SRS021948	CША	illumina	город	na	na	na	0	38,37
SRS022071	CША	illumina	город	na	na	na	0	45,41
SRS022137	CША	illumina	город	na	na	na	0	63,51
SRS022609	CША	illumina	город	na	na	na	0	32,21
SRS023346	CША	illumina	город	na	na	na	0	72,05
SRS023526	CША	illumina	город	na	na	na	0	39,08
SRS023583	CША	illumina	город	na	na	na	0	64,97
SRS023914	CША	illumina	город	na	na	na	0	27,63
SRS023971	CША	illumina	город	na	na	na	0	73,36
SRS024009	CША	illumina	город	na	na	na	0	59,84
SRS024075	CША	illumina	город	na	na	na	0	55,85
SRS024132	CША	illumina	город	na	na	na	0	46,02
SRS024265	CША	illumina	город	na	na	na	0	29,85
SRS024331	CША	illumina	город	na	na	na	0	64,59
SRS024388	CША	illumina	город	na	na	na	0	50,12
SRS024435	CША	illumina	город	na	na	na	0	40,43
SRS024549	CША	illumina	город	na	na	na	0	81,76
SRS024625	CША	illumina	город	na	na	na	0	54,69
SRS042284	CША	illumina	город	na	na	na	0	65,92
SRS042628	CША	illumina	город	na	na	na	0	24,92
SRS043001	CША	illumina	город	na	na	na	0	71,39
SRS043411	CША	illumina	город	na	na	na	0	66,44
SRS043701	CША	illumina	город	na	na	na	0	71,23
SRS045004	CША	illumina	город	na	na	na	0	48,05
SRS045645	CША	illumina	город	na	na	na	0	60,82

SRS045713	США	illumina	город	na	na	na	0	68,35
SRS047014	США	illumina	город	na	na	na	0	52,36
SRS047044	США	illumina	город	na	na	na	0	45,85
SRS048164	США	illumina	город	na	na	na	0	51,16
SRS048870	США	illumina	город	na	na	na	0	55,05
SRS049712	США	illumina	город	na	na	na	0	38,48
SRS049900	США	illumina	город	na	na	na	0	42,16
SRS049959	США	illumina	город	na	na	na	0	22
SRS049995	США	illumina	город	na	na	na	0	49,02
SRS050299	США	illumina	город	na	na	na	0	71,65
SRS050422	США	illumina	город	na	na	na	0	42,02
SRS050752	США	illumina	город	na	na	na	0	38,32
SRS050925	США	illumina	город	na	na	na	0	50,53
SRS051031	США	illumina	город	na	na	na	0	44,13
SRS051882	США	illumina	город	na	na	na	0	47,81
SRS052027	США	illumina	город	na	na	na	0	57,36
SRS052697	США	illumina	город	na	na	na	0	42,72
SRS053335	США	illumina	город	na	na	na	0	38,28
SRS053398	США	illumina	город	na	na	na	0	51,7
SRS054590	США	illumina	город	na	na	na	0	67,28
SRS054956	США	illumina	город	na	na	na	0	68,66
SRS055982	США	illumina	город	na	na	na	0	61,38
SRS056259	США	illumina	город	na	na	na	0	65,57
SRS056519	США	illumina	город	na	na	na	0	62,06
SRS057717	США	illumina	город	na	na	na	0	66,21
SRS058723	США	illumina	город	na	na	na	0	67
SRS058770	США	illumina	город	na	na	na	0	66,35
SRS062427	США	illumina	город	na	na	na	0	28,75
SRS063040	США	illumina	город	na	na	na	0	29,66
SRS063985	США	illumina	город	na	na	na	0	36,25
SRS064276	США	illumina	город	na	na	na	0	46
SRS064557	США	illumina	город	na	na	na	0	59,57
SRS064645	США	illumina	город	na	na	na	0	56,78
SRS065504	США	illumina	город	na	na	na	0	48,81
SRS077730	США	illumina	город	na	na	na	0	62,15
SRS078176	США	illumina	город	na	na	na	0	37,27
TAT_129	Россия	solid	село	47	М	20,68	0,24	28,72
TAT_130	Россия	solid	село	34	Ж	na	0,2	30,67
TAT_131	Россия	solid	село	85	М	na	0,11	30,2
TAT_134	Россия	solid	село	33	Ж	na	0,12	29,3
TAT_138	Россия	solid	село	77	М	na	0,11	29,35
TAT_143	Россия	solid	село	56	М	na	0,11	29,15
TAT_146	Россия	solid	село	27	М	na	0,07	29,11
TAT_147	Россия	solid	село	71	Ж	28,89	0,12	27,07
TYV_202	Россия	solid	село	14	Ж	22,68	0,26	32,52
TYV_203	Россия	solid	село	14	Ж	18,43	0,47	18,37
TYV_204	Россия	solid	село	15	Ж	20,83	1,99	16,48
TYV_205	Россия	solid	село	15	Ж	17,31	0,17	31,22
TYV_206	Россия	solid	село	15	Ж	20,33	0,28	12,15
TYV_207	Россия	solid	село	17	Ж	20,24	0,51	11,31

TYV_208	Россия	solid	село	17	Ж	20,69	1,39	19,41
TYV_209	Россия	solid	село	14	М	23,18	1,91	18,72
TYV_210	Россия	solid	село	15	Ж	16,02	0,21	18,4
TYV_211	Россия	solid	село	14	Ж	18,29	0,72	13,25
TYV_212	Россия	solid	село	16	Ж	20,45	3,74	20,43
TYV_214	Россия	solid	село	28	Ж	22,04	0,22	17,65
TYV_215	Россия	solid	село	16	Ж	16,39	0,28	20,29
TYV_216	Россия	solid	село	14	М	19,68	0,75	20,07
TYV_219	Россия	solid	село	15	Ж	21,26	7,35	27,16

ПРИЛОЖЕНИЕ 2

Геном	Род	Отдел
Acidaminococcus fermentans DSM 20731	Acidaminococcus	Firmicutes
Acidaminococcus intestini RyC MR95	Acidaminococcus	Firmicutes
Acidaminococcus sp D21	Acidaminococcus	Firmicutes
Acinetobacter baumannii ATCC 17978	Acinetobacter	Proteobacteria
Acinetobacter haemolyticus ATCC 19194	Acinetobacter	Proteobacteria
Acinetobacter junii SH205	Acinetobacter	Proteobacteria
Acinetobacter oleivorans DR1	Acinetobacter	Proteobacteria
Acinetobacter radioresistens SK82	Acinetobacter	Proteobacteria
Acinetobacter sp ADP1	Acinetobacter	Proteobacteria
Acinetobacter sp RUH2624	Acinetobacter	Proteobacteria
Acinetobacter sp SH024	Acinetobacter	Proteobacteria
Actinobacillus succinogenes 130Z	Actinobacillus	Proteobacteria
Actinomyces odontolyticus ATCC 17982	Actinomyces	Actinobacteria
Actinomyces sp oral taxon 848 str F0332 A	Actinomyces	Actinobacteria
Adlercreutzia equolifaciens DSM 19450	Adlercreutzia	Actinobacteria
Aeromonas hydrophila SSU	Aeromonas	Proteobacteria
Aeromonas veronii AMC34	Aeromonas	Proteobacteria
Aggregatibacter segnis ATCC 33393	Aggregatibacter	Proteobacteria
Akkermansia muciniphila ATCC BAA 835	Akkermansia	Verrucomicrobia
Alistipes indistinctus YIT 12060	Alistipes	Bacteroidetes
Alistipes onderdonkii DSM 19147	Alistipes	Bacteroidetes
Alistipes putredinis DSM 17216	Alistipes	Bacteroidetes
Alistipes shahii WAL 8301	Alistipes	Bacteroidetes
Anaerobaculum hydrogeniformans ATCC BAA 1850	Anaerobaculum	Synergistetes
Anaerococcus hydrogenalis DSM 7454	Anaerococcus	Firmicutes
Anaerococcus lactolyticus ATCC 51172	Anaerococcus	Firmicutes
Anaerococcus tetradius ATCC 35098	Anaerococcus	Firmicutes
Anaerococcus vaginalis ATCC 51170	Anaerococcus	Firmicutes
Anaerofustis stercorihominis DSM 17244	Anaerofustis	Firmicutes
Anaerostipes caccae DSM 14662	Anaerostipes	Firmicutes
Anaerostipes hadrus DSM 3319	Anaerostipes	Firmicutes
Anaerotruncus colihominis DSM 17241	Anaerotruncus	Firmicutes
Arcobacter butzleri RM4018	Arcobacter	Proteobacteria
Bacillus alcalophilus ATCC 27647	Bacillus	Firmicutes
Bacillus halodurans C 125	Bacillus	Firmicutes
Bacillus smithii 7 3 47FAA	Bacillus	Firmicutes
Bacillus sp 7 6 55CFAA CT2	Bacillus	Firmicutes
Bacteroides caccae ATCC 43185	Bacteroides	Bacteroidetes
Bacteroides cellulosilyticus DSM 14838	Bacteroides	Bacteroidetes
Bacteroides clarus YIT 12056	Bacteroides	Bacteroidetes
Bacteroides coprocola DSM 17136	Bacteroides	Bacteroidetes
Bacteroides coprophilus DSM 18228	Bacteroides	Bacteroidetes
Bacteroides dorei 5_1_36 D4	Bacteroides	Bacteroidetes

Bacteroides eggerthii 1_2_48FAA	Bacteroides	Bacteroidetes
Bacteroides finegoldii DSM 17565	Bacteroides	Bacteroidetes
Bacteroides fluxus YIT 12057	Bacteroides	Bacteroidetes
Bacteroides fragilis 3_1_12	Bacteroides	Bacteroidetes
Bacteroides fragilis 638R genome	Bacteroides	Bacteroidetes
Bacteroides intestinalis DSM 17393	Bacteroides	Bacteroidetes
Bacteroides oleiciplenus YIT 12058	Bacteroides	Bacteroidetes
Bacteroides ovatus 3_8_47FAA	Bacteroides	Bacteroidetes
Bacteroides ovatus ATCC 8483	Bacteroides	Bacteroidetes
Bacteroides ovatus SD CMC	Bacteroides	Bacteroidetes
Bacteroides pectinophilus ATCC 43243	Bacteroides	Firmicutes
Bacteroides plebeius DSM 17135	Bacteroides	Bacteroidetes
Bacteroides sp 3_1_23	Bacteroides	Bacteroidetes
Bacteroides sp D2	Bacteroides	Bacteroidetes
Bacteroides stercoris ATCC 43183	Bacteroides	Bacteroidetes
Bacteroides thetaiotaomicron VPI 5482	Bacteroides	Bacteroidetes
Bacteroides uniformis ATCC 8492	Bacteroides	Bacteroidetes
Bacteroides vulgatus ATCC 8482	Bacteroides	Bacteroidetes
Bacteroides xylanisolvens XB1A	Bacteroides	Bacteroidetes
Barnesiella intestinihominis YIT 11860	Barnesiella	Bacteroidetes
Bifidobacterium adolescentis ATCC 15703	Bifidobacterium	Actinobacteria
Bifidobacterium angulatum DSM 20098	Bifidobacterium	Actinobacteria
Bifidobacterium animalis subsp lactis V9	Bifidobacterium	Actinobacteria
Bifidobacterium bifidum PRL2010	Bifidobacterium	Actinobacteria
Bifidobacterium breve DSM 20213	Bifidobacterium	Actinobacteria
Bifidobacterium catenulatum DSM 16992	Bifidobacterium	Actinobacteria
Bifidobacterium dentium Bd1	Bifidobacterium	Actinobacteria
Bifidobacterium gallicum DSM 20093	Bifidobacterium	Actinobacteria
Bifidobacterium longum subsp infantis ATCC 15697	Bifidobacterium	Actinobacteria
Bifidobacterium longum subsp longum JDM301	Bifidobacterium	Actinobacteria
Bifidobacterium pseudocatenulatum DSM 20438	Bifidobacterium	Actinobacteria
Bilophila wadsworthia 3_1_6	Bilophila	Proteobacteria
Blastocystis hominis	Blastocystis	Protista
Blautia hansenii DSM 20583	Blautia	Firmicutes
Blautia hydrogenotrophica DSM 10507	Blautia	Firmicutes
Blautia producta ATCC 27340	Blautia	Firmicutes
Blautia wexlerae AGR2146	Blautia	Firmicutes
Brachyspira pilosicoli 95/1000	Brachyspira	Spirochaetes
Bulleidia extructa W1219	Bulleidia	Firmicutes
Butyrivibrio crossotus DSM 2876	Lachnospiraceae	Firmicutes
Butyrivibrio fibrisolvens 16 4	Butyrivibrio	Firmicutes
Campylobacter coli JV20	Campylobacter	Proteobacteria
Campylobacter concisus 13826	Campylobacter	Proteobacteria
Campylobacter curvus 52592	Campylobacter	Proteobacteria
Campylobacter gracilis RM3268 contig00052	Campylobacter	Proteobacteria
Campylobacter hominis ATCC BAA 381	Campylobacter	Proteobacteria
Campylobacter jejuni subsp jejuni NCTC	Campylobacter	Proteobacteria

11168		
Campylobacter upsaliensis JV21	Campylobacter	Proteobacteria
Candida albicans	Candida	Ascomycota
Candida dubliniensis CD36 1	Candida	Ascomycota
Candida glabrata CBS 138 A	Candida	Ascomycota
Candidatus Methanomethylophilus alvus Mx1201	Methanomethylophilus	Euryarchaeota
Capnocytophaga ochracea DSM 7271	Capnocytophaga	Bacteroidetes
Catenibacterium mitsuokai DSM 15897	Catenibacterium	Firmicutes
Citrobacter freundii 4 7 47CAA	Citrobacter	Proteobacteria
Citrobacter koseri ATCC BAA 895	Citrobacter	Proteobacteria
Citrobacter youngae ATCC 29220	Enterobacteriaceae	Proteobacteria
Clavispora lusitaniae ATCC 42720 1	Clavispora	Dikarya
Clostridiales bacterium 1_7_47FAA	Clostridiales	Firmicutes
Clostridiales sp SM4 1	Lachnospiraceae	Firmicutes
Clostridiales sp SS3 4	Clostridiales	Firmicutes
Clostridium asparagiforme DSM 15981	Clostridiales	Firmicutes
Clostridium bartlettii DSM 16795	Sporacetigenium	Firmicutes
Clostridium bolteae ATCC BAA 613	Clostridium	Firmicutes
Clostridium botulinum A2 str Kyoto	Clostridium	Firmicutes
Clostridium butyricum E4 str BoNT E BL5262	Clostridium	Firmicutes
Clostridium celatum DSM 1785	Clostridium	Firmicutes
Clostridium citroniae WAL 17108	Clostridium	Firmicutes
Clostridium clostridioforme 2_1_49FAA	Clostridium	Firmicutes
Clostridium difficile 630	Clostridium	Firmicutes
Clostridium hathewayi DSM 13479	Clostridium	Firmicutes
Clostridium hathewayi WAL 18680	Clostridium	Firmicutes
Clostridium hiranonis DSM 13275	Peptostreptococcus	Firmicutes
Clostridium hylemonae DSM 15053	Clostridium	Firmicutes
Clostridium leptum DSM 753	Ruminococcaceae	Firmicutes
Clostridium methylpentosum DSM 5476	Ruminococcaceae	Firmicutes
Clostridium nexile DSM 1787	Lachnospiraceae	Firmicutes
Clostridium perfringens ATCC 13124	Clostridium	Firmicutes
Clostridium saccharolyticum like K10	Lachnospiraceae	Firmicutes
Clostridium scindens ATCC 35704	Dorea	Firmicutes
Clostridium sp 7_2_43FAA	Clostridium	Firmicutes
Clostridium sp D5	Clostridium	Firmicutes
Clostridium sp HGF2	Clostridium	Firmicutes
Clostridium sp L2 50	Coproccoccus	Firmicutes
Clostridium sp SS2 1	Anaerostipes	Firmicutes
Clostridium spiroforme DSM 1552	Coproccoccus	Firmicutes
Clostridium symbiosum WAL 14163	Clostridium	Firmicutes
Clostridium thermocellum ATCC 27405	Clostridium	Firmicutes
Collinsella aerofaciens ATCC 25986	Collinsella	Actinobacteria
Collinsella intestinalis DSM 13280	Collinsella	Actinobacteria
Collinsella stercoris DSM 13279	Collinsella	Actinobacteria
Collinsella tanakaei YIT 12063	Collinsella	Actinobacteria
Coproccoccus sp 8 2 54BFAA	Coproccoccus	Firmicutes
Coproccoccus sp D6 cont11	Coproccoccus	Firmicutes

Coprococcus catus GD 7	Coprococcus	Firmicutes
Coprococcus comes ATCC 27758	Coprococcus	Firmicutes
Coprococcus eutactus ATCC 27759	Coprococcus	Firmicutes
Coprococcus sp ART55 1	Coprococcus	Firmicutes
Corynebacterium ammoniagenes DSM 20306	Corynebacterium	Actinobacteria
Desulfitobacterium hafniense Y51	Desulfitobacterium	Firmicutes
Desulfovibrio alaskensis G20	Desulfovibrio	Proteobacteria
Desulfovibrio piger ATCC 29098	Desulfovibrio	Proteobacteria
Desulfovibrio sp 6 1 46AFAA	Desulfovibrio	Proteobacteria
Dialister invisus DSM 15470	Dialister	Firmicutes
Dialister microaerophilus UPII 345 E	Dialister	Firmicutes
Dialister succinatiphilus YIT 11850	Dialister	Firmicutes
Dorea longicatena DSM 13814	Dorea	Firmicutes
Dysgonomonas gadei ATCC BAA 286	Dysgonomonas	Bacteroidetes
Dysgonomonas mossii DSM 22836	Dysgonomonas	Bacteroidetes
Edwardsiella tarda ATCC 23685	Edwardsiella	Proteobacteria
Edwardsiella tarda EIB202	Edwardsiella	Proteobacteria
Eggerthella lenta DSM 2243	Eggerthella	Actinobacteria
Enterobacter asburiae LF7a	Enterobacter	Proteobacteria
Enterobacter cancerogenus ATCC 35316	Enterobacter	Proteobacteria
Enterobacter cloacae SCF1	Enterobacter	Proteobacteria
Enterobacter cloacae subsp cloacae ATCC 13047	Enterobacter	Proteobacteria
Enterobacter cloacae subsp cloacae NCTC 9394	Enterobacter	Proteobacteria
Enterobacter sp 638	Enterobacter	Proteobacteria
Enterococcus casseliflavus EC10	Enterococcus	Firmicutes
Enterococcus faecalis V583	Enterococcus	Firmicutes
Enterococcus faecium 1230933	Enterococcus	Firmicutes
Enterococcus saccharolyticus 30 1	Enterococcus	Firmicutes
Enterocytozoon bieneusi H348	Enterocytozoon	Microsporidia
Erysipelotrichaceae bacterium 3_1_53	Eubacterium	Firmicutes
Erysipelotrichaceae bacterium 5_2_54FAA	Eubacterium	Firmicutes
Escherichia albertii TW07627	Escherichia/Shigella	Proteobacteria
Escherichia coli str K12 substr MG1655	Escherichia/Shigella	Proteobacteria
Escherichia fergusonii ATCC 35469	Escherichia/Shigella	Proteobacteria
Escherichia hermannii NBRC 105704	Escherichia/Shigella	Proteobacteria
Escherichia sp 1_1_43	Escherichia/Shigella	Proteobacteria
Eubacterium bifforme DSM 3989	Eubacterium	Firmicutes
Eubacterium cylindroides T2 87	Eubacterium	Firmicutes
Eubacterium dolichum DSM 3991	Eubacterium	Firmicutes
Eubacterium eligens ATCC 27750	Lachnospiraceae	Firmicutes
Eubacterium hallii DSM 3353	Lachnospiraceae	Firmicutes
Eubacterium limosum KIST612	Eubacterium	Firmicutes
Eubacterium siraeum 70 3	Eubacterium	Firmicutes
Eubacterium ventriosum ATCC 27560	Lachnospiraceae	Firmicutes
Faecalibacterium cf prausnitzii KLE1255	Faecalibacterium	Firmicutes
Faecalibacterium prausnitzii A2 165	Faecalibacterium	Firmicutes
Faecalibacterium prausnitzii L2 6	Faecalibacterium	Firmicutes

Faecalibacterium prausnitzii SL3 3	Faecalibacterium	Firmicutes
Finegoldia magna ATCC 29328	Finegoldia	Firmicutes
Flavonifractor plautii ATCC 29863	Flavonifractor	Firmicutes
Fusobacterium gonidiaformans ATCC 25563	Fusobacterium	Fusobacteria
Fusobacterium mortiferum ATCC 9817	Fusobacterium	Fusobacteria
Fusobacterium necrophorum subsp funduliforme 1 1 36S	Fusobacterium	Fusobacteria
Fusobacterium nucleatum subsp nucleatum ATCC 23726	Fusobacterium	Fusobacteria
Fusobacterium sp 2_1_31	Fusobacterium	Fusobacteria
Fusobacterium ulcerans ATCC 49185	Fusobacterium	Fusobacteria
Fusobacterium varium ATCC 27725	Fusobacterium	Fusobacteria
Gardnerella vaginalis ATCC 14019	Gardnerella	Actinobacteria
Gemella moribillum M424	Gemella	Firmicutes
Gordonibacter pamelaiae 7 10 1 b	Gordonibacter	Actinobacteria
Granulicatella adiacens ATCC 49175	Granulicatella	Firmicutes
Granulicatella elegans ATCC 700633	Granulicatella	Firmicutes
Haemophilus parainfluenzae ATCC 33392	Haemophilus	Proteobacteria
Hafnia alvei ATCC 51873	Hafnia	Proteobacteria
Halorubrum lipolyticum DSM 21995	Halorubrum	Euryarchaeota
Helicobacter bilis ATCC 43879	Helicobacter	Proteobacteria
Helicobacter canadensis MIT 98 5491	Helicobacter	Proteobacteria
Helicobacter cinaedi CCUG 18818	Helicobacter	Proteobacteria
Helicobacter pullorum MIT 98 5489	Helicobacter	Proteobacteria
Helicobacter pylori J99	Helicobacter	Proteobacteria
Helicobacter winthamensis ATCC BAA 430	Helicobacter	Proteobacteria
Holdemania filiformis DSM 12042	Holdemania	Firmicutes
Klebsiella pneumoniae 342	Klebsiella	Proteobacteria
Lachnospiraceae bacterium 1_4_56FAA	Lachnospiraceae	Firmicutes
Lachnospiraceae bacterium 2_1_46FAA	Lachnospiraceae	Firmicutes
Lachnospiraceae bacterium 3_1_57FAA_CT1	Lachnospiraceae	Firmicutes
Lachnospiraceae bacterium 9_1_43BFAA	Lachnospiraceae	Firmicutes
Lactobacillus acidophilus NCFM	Lactobacillus	Firmicutes
Lactobacillus amylolyticus DSM 11664	Lactobacillus	Firmicutes
Lactobacillus antri DSM 16041	Lactobacillus	Firmicutes
Lactobacillus brevis ATCC 367	Lactobacillus	Firmicutes
Lactobacillus brevis subsp gravesensis ATCC 27305	Lactobacillus	Firmicutes
Lactobacillus casei ATCC 334	Lactobacillus	Firmicutes
Lactobacillus crispatus ST1	Lactobacillus	Firmicutes
Lactobacillus delbrueckii subsp bulgaricus ATCC 11842	Lactobacillus	Firmicutes
Lactobacillus fermentum 28 3 CHN	Lactobacillus	Firmicutes
Lactobacillus fermentum IFO 3956	Lactobacillus	Firmicutes
Lactobacillus gasseri ATCC 33323	Lactobacillus	Firmicutes
Lactobacillus helveticus DPC 4571	Lactobacillus	Firmicutes
Lactobacillus iners AB 1 NZ ADHG01000001	Lactobacillus	Firmicutes

Lactobacillus jensenii 115321	Lactobacillus	Firmicutes
Lactobacillus johnsonii NCC 533	Lactobacillus	Firmicutes
Lactobacillus oris PB013 T2 3	Lactobacillus	Firmicutes
Lactobacillus plantarum str 13	Lactobacillus	Firmicutes
Lactobacillus reuteri SD2112	Lactobacillus	Firmicutes
Lactobacillus rhamnosus GG	Lactobacillus	Firmicutes
Lactobacillus ruminis ATCC 25644	Lactobacillus	Firmicutes
Lactobacillus sakei strain 23K	Lactobacillus	Firmicutes
Lactobacillus salivarius UCC118	Lactobacillus	Firmicutes
Lactobacillus ultunensis DSM 16047	Lactobacillus	Firmicutes
Lactobacillus vaginalis ATCC 49540	Lactobacillus	Firmicutes
Lactococcus lactis subsp cremoris NZ9000	Lactococcus	Firmicutes
Leminorella grimontii ATCC 33999	Leminorella	Firmicutes
Leuconostoc mesenteroides ATCC 8293	Leuconostoc	Firmicutes
Listeria grayi DSM 20601	Listeria	Firmicutes
Listeria innocua Clip11262	Listeria	Firmicutes
Listeria monocytogenes Finland 1998	Listeria	Firmicutes
Mannheimia succiniciproducens MBEL55E	Basfia	Proteobacteria
Marvinbryantia formatexigens DSM 14469	Marvinbryantia	Firmicutes
Megamonas hypermegale	Megamonas	Firmicutes
Methanobrevibacter smithii ATCC 35061	Methanobrevibacter	Euryarchaeota
Methanomassiliicoccus luminyensis B10	Methanomassiliicoccus	Euryarchaeota
Methanomassiliicoccus sp Mx1 Isoire	Methanomassiliicoccus	Euryarchaeota
Methanosphaera stadtmanae DSM 3091	Methanosphaera	Euryarchaeota
Mitsuokella multacida DSM 20544	Mitsuokella	Firmicutes
Mycoplasma hominis ATCC 23114	Mycoplasma	Firmicutes
Neisseria macacae ATCC 33926	Neisseria	Proteobacteria
Odoribacter laneus YIT 12061	Odoribacter	Bacteroidetes
Odoribacter splanchnicus DSM 20712	Odoribacter	Bacteroidetes
Oxalobacter formigenes HOxBLS	Oxalobacter	Proteobacteria
Oxalobacter formigenes OXCC13	Oxalobacter	Proteobacteria
Paenibacillus sp HGF5	Paenibacillus	Firmicutes
Paenibacillus sp HGF7	Paenibacillus	Firmicutes
Parabacteroides distasonis ATCC 8503	Parabacteroides	Bacteroidetes
Parabacteroides johnsonii DSM 18315	Parabacteroides	Bacteroidetes
Parabacteroides merdae ATCC 43184	Parabacteroides	Bacteroidetes
Paraprevotella xylaniphila YIT 11841	Paraprevotella	Bacteroidetes
Parasutterella excrementihominis YIT 11859	Parasutterella	Proteobacteria
Parvimonas micra ATCC 33270	Parvimonas	Firmicutes
Pediococcus acidilactici DSM 20284	Pediococcus	Firmicutes
Peptoniphilus lacrimalis 315 B	Peptoniphilus	Firmicutes
Peptostreptococcus anaerobius 653 L	Peptostreptococcus	Firmicutes
Peptostreptococcus stomatis DSM 17678	Peptostreptococcus	Firmicutes
Phascolarctobacterium sp YIT 12067	Phascolarctobacterium	Firmicutes
Porphyromonas gingivalis ATCC 33277	Porphyromonas	Bacteroidetes
Porphyromonas uenonis 60 3	Porphyromonas	Bacteroidetes
Prevotella bryantii B14	Prevotella	Bacteroidetes
Prevotella buccae ATCC 33574	Prevotella	Bacteroidetes
Prevotella copri DSM 18205	Prevotella	Bacteroidetes

Prevotella disiens FB035 09AN	Prevotella	Bacteroidetes
Prevotella melaninogenica ATCC 25845	Prevotella	Bacteroidetes
Prevotella salivae DSM 15606	Prevotella	Bacteroidetes
Prevotella stercorea DSM 18206	Prevotella	Bacteroidetes
Propionibacterium sp 5 U 42AFAA	Propionibacterium	Actinobacteria
Proteus mirabilis HI4320	Proteus	Proteobacteria
Proteus penneri ATCC 35198	Proteus	Proteobacteria
Providencia alcalifaciens DSM 30120	Providencia	Proteobacteria
Providencia rettgeri DSM 1131	Providencia	Proteobacteria
Providencia rustigianii DSM 4541	Providencia	Proteobacteria
Providencia stuartii ATCC 25827	Providencia	Proteobacteria
Pseudoflavonifractor capillosus ATCC 29799	Bacteroides	Firmicutes
Pseudomonas aeruginosa PAO1	Pseudomonas	Proteobacteria
Pseudomonas mendocina ymp	Pseudomonas	Proteobacteria
Ralstonia sp 5 2 56FAA	Ralstonia	Proteobacteria
Roseburia hominis A2 183	Roseburia	Firmicutes
Roseburia intestinalis XB6B4	Roseburia	Firmicutes
Roseburia inulinivorans DSM 16841	Roseburia	Firmicutes
Rothia mucilaginosa DY 18	Rothia	Actinobacteria
Ruminococcaceae bacterium D16	Ruminococcaceae	Firmicutes
Ruminococcus bromii L2 63	Ruminococcus	Firmicutes
Ruminococcus gnavus ATCC 29149	Ruminococcus	Firmicutes
Ruminococcus lactaris ATCC 29176	Lachnospiraceae	Firmicutes
Ruminococcus obeum A2 162	Blautia	Firmicutes
Ruminococcus obeum ATCC 29174	Blautia	Firmicutes
Ruminococcus sp 18P13	Ruminococcus	Firmicutes
Ruminococcus sp 5_1_39BFAA	Blautia	Firmicutes
Ruminococcus sp SR1 5	Blautia	Firmicutes
Ruminococcus torques ATCC 27756	Lachnospiraceae	Firmicutes
Ruminococcus torques L2 14	Lachnospiraceae	Firmicutes
Salmonella enterica subsp enterica serovar Typhimurium LT2	Salmonella	Proteobacteria
Selenomonas sputigena ATCC 35185	Selenomonas	Firmicutes
Slackia exigua ATCC 700122	Slackia	Actinobacteria
Slackia piriformis YIT 12062	Slackia	Actinobacteria
Solobacterium moorei F0204Scfld3	Solobacterium	Firmicutes
Staphylococcus aureus subsp aureus TW20	Staphylococcus	Firmicutes
Staphylococcus saprophyticus subsp saprophyticus ATCC 15305 DNA	Staphylococcus	Firmicutes
Staphylococcus warneri L37603 contig00102	Staphylococcus	Firmicutes
Stenotrophomonas maltophilia K279a	Stenotrophomonas	Proteobacteria
Streptococcus agalactiae 2603V/R	Streptococcus	Firmicutes
Streptococcus anginosus 1_2_62CV	Streptococcus	Firmicutes
Streptococcus bovis ATCC 700338	Streptococcus	Firmicutes
Streptococcus cristatus ATCC 51100 ctg1129913971918	Streptococcus	Firmicutes
Streptococcus equinus ATCC 9812	Streptococcus	Firmicutes
Streptococcus infantarius subsp infantarius ATCC BAA 102	Streptococcus	Firmicutes

Streptococcus mutans UA159	Streptococcus	Firmicutes
Streptococcus oralis ATCC 35037	Streptococcus	Firmicutes
Streptococcus parasanguinis ATCC 15912	Streptococcus	Firmicutes
Streptococcus peroris ATCC 700780	Streptococcus	Firmicutes
Streptococcus salivarius SK126	Streptococcus	Firmicutes
Streptococcus sp 2_1_36FAA	Streptococcus	Firmicutes
Streptococcus sp C150	Streptococcus	Firmicutes
Streptococcus thermophilus LMD 9	Streptococcus	Firmicutes
Streptococcus vestibularis ATCC 49124	Streptococcus	Firmicutes
Subdoligranulum sp 4 3 54A2FAA	Subdoligranulum	Firmicutes
Subdoligranulum variabile DSM 15176	Subdoligranulum	Firmicutes
Succinatimonas hippei YIT 12066	Anaerobiospirillum	Proteobacteria
Sutterella parvirubra YIT 11816	Sutterella	Proteobacteria
Sutterella wadsworthensis 2 1 59BFAA	Sutterella	Proteobacteria
Sutterella wadsworthensis 3_1_45B	Sutterella	Proteobacteria
Synergistes sp 3 1 syn1	Synergistes	Synergistetes
Tannerella sp 6 1 58FAA CT1	Tannerella	Bacteroidetes
Treponema denticola ATCC 35405	Treponema	Spirochaetes
Treponema pallidum subsp pallidum SS14	Treponema	Spirochaetes
Treponema vincentii ATCC 35580 contig00031	Treponema	Spirochaetes
Tropheryma whipplei str Twist	Tropheryma	Actinobacteria
Turicibacter sanguinis PC909	Turicibacter	Firmicutes
Veillonella atypica ACS 134 V Col7a	Veillonella	Firmicutes
Veillonella dispar ATCC 17748	Veillonella	Firmicutes
Veillonella parvula DSM 2008	Veillonella	Firmicutes
Veillonella sp 6_1_27	Veillonella	Firmicutes
Veillonella sp oral taxon 158 str F0412	Veillonella	Firmicutes
Vibrio cholerae O1 biovar El Tor str N16961	Vibrio	Proteobacteria
Weissella paramesenteroides ATCC 33313	Weissella	Firmicutes
Wolinella succinogenes DSM 1740	Wolinella	Proteobacteria
Yersinia bercovieri ATCC 43970	Yersinia	Proteobacteria
Yokenella regensburgei ATCC 43003	Yokenella	Proteobacteria

ПРИЛОЖЕНИЕ 3

; =====

\$PROBLEM Model of scfa metabolism in human gut

; =====

; Developed by Yuri Kosinsky

; Ver. 9 from Feb 8, 2014

; =====

\$MODEL

; =====

COMP = (Acet1)

COMP = (But1)

COMP = (Prop1)

COMP = (Nutr1)

COMP = (Acet2)

COMP = (But2)

COMP = (Prop2)

COMP = (Nutr2)

COMP = (Acet3)

COMP = (But3)

COMP = (Prop3)

COMP = (Nutr3)

COMP = (Acet4)

COMP = (But4)

COMP = (Prop4)

COMP = (Nutr4)

; =====

\$PSI k_Nutr_Acet Km_Nutr_Acet k_Nutr_Prop Km_Nutr_Prop k_Acet_But Km_Acet_But k_gut_out k_transp
k_nutr_in1

```

; =====
; =====
$ODE
; =====
; Definition of the solver (STIFF or NONSTIFF) for the integration (we only use NONSTIFF for the moment)
NONSTIFF
; Model parameters
; -----
; L, ltestine content volume:
volume_gut = 0.2
; mM, Estimated from Am. J. Physiol. (1991)
Km_But_transp = 21.3
Km_Acet_transp = 15
Km_Prop_transp = 15
; mmol/h, Nutr from mucin

k_nutr_in = 1
; mmol/h
;k_transp =8.3
; mmol/h
;k_nutr_in1 = 40.0
; 1/h
;k_gut_out = 0.08
; Model variables
; -----
Nutr_conc1 = Nutr1/volume_gut
But_conc1 = But1/volume_gut
Acet_conc1 = Acet1/volume_gut
Prop_conc1 = Prop1/volume_gut

```

```

;

D_transp_1 = 1+Acet_conc1/Km_Acet_transp+Prop_conc1/Km_Prop_transp+But_conc1/Km_But_transp

But_flux1 = k_transp*(But_conc1/Km_But_transp)/D_transp_1

Acet_flux1 = k_transp*(Acet_conc1/Km_Acet_transp)/D_transp_1

Prop_flux1 = k_transp*(Prop_conc1/Km_Prop_transp)/D_transp_1

;

Nutr_conc2 = Nutr2/volume_gut

But_conc2 = But2/volume_gut

Acet_conc2 = Acet2/volume_gut

Prop_conc2 = Prop2/volume_gut

;

D_transp_2 = 1+Acet_conc2/Km_Acet_transp+Prop_conc2/Km_Prop_transp+But_conc2/Km_But_transp

But_flux2 = k_transp*(But_conc2/Km_But_transp)/D_transp_2

Acet_flux2 = k_transp*(Acet_conc2/Km_Acet_transp)/D_transp_2


Prop_flux2 = k_transp*(Prop_conc2/Km_Prop_transp)/D_transp_2

;

Nutr_conc3 = Nutr3/volume_gut

But_conc3 = But3/volume_gut

Acet_conc3 = Acet3/volume_gut

Prop_conc3 = Prop3/volume_gut

;

D_transp_3 = 1+Acet_conc3/Km_Acet_transp+Prop_conc3/Km_Prop_transp+But_conc3/Km_But_transp

But_flux3 = k_transp*(But_conc3/Km_But_transp)/D_transp_3

Acet_flux3 = k_transp*(Acet_conc3/Km_Acet_transp)/D_transp_3

Prop_flux3 = k_transp*(Prop_conc3/Km_Prop_transp)/D_transp_3

;

Nutr_conc4 = Nutr4/volume_gut

But_conc4 = But4/volume_gut

```

```

Acet_conc4 = Acet4/volume_gut
Prop_conc4 = Prop4/volume_gut
;
D_transp_4 = 1+Acet_conc4/Km_Acet_transp+Prop_conc4/Km_Prop_transp+But_conc4/Km_But_transp
But_flux4 = k_transp*(But_conc4/Km_But_transp)/D_transp_4
Acet_flux4 = k_transp*(Acet_conc4/Km_Acet_transp)/D_transp_4
Prop_flux4 = k_transp*(Prop_conc4/Km_Prop_transp)/D_transp_4
;
; Initial time (always zero)
; -----
T0 = 0
; Initial conditions
; -----
Acet1_0 = 10
But1_0 = 3
Prop1_0 = 3
Nutr1_0 = 200
Acet2_0 = 10
But2_0 = 3
Prop2_0 = 3
Nutr2_0 = 200
Acet3_0 = 10
But3_0 = 3
Prop3_0 = 3
Nutr3_0 = 200
Acet4_0 = 10
But4_0 = 3
Prop4_0 = 3
Nutr4_0 = 200

```

; Model reactions

; -----

; in 1st cmt (mmol/h):

$v_Nutr_intake1 = k_nutr_in1 + k_nutr_in$

$v_PolyNutr_Acet1 = k_Nutr_Acet * Nutr_conc1 / (Km_Nutr_Acet + Nutr_conc1)$

$v_PolyNutr_Prop1 = k_Nutr_Prop * Nutr_conc1 / (Km_Nutr_Prop + Nutr_conc1)$

$v_Acet_But1 = k_Acet_But * Acet_conc1 / (Acet_conc1 + Km_Acet_But)$

$v_Acet_transport1 = Acet_flux1$

$v_Prop_transport1 = Prop_flux1$

$v_But_transport1 = But_flux1$

$v_Nutr_trans1 = k_gut_out * Nutr1$

$v_Acet_trans1 = k_gut_out * Acet1$

$v_Prop_trans1 = k_gut_out * Prop1$

$v_But_trans1 = k_gut_out * But1$

; in 2nd cmt (mmol/h):

$v_Nutr_intake2 = k_nutr_in$

$v_PolyNutr_Acet2 = k_Nutr_Acet * Nutr_conc2 / (Km_Nutr_Acet + Nutr_conc2)$

$v_PolyNutr_Prop2 = k_Nutr_Prop * Nutr_conc2 / (Km_Nutr_Prop + Nutr_conc2)$

$v_Acet_But2 = k_Acet_But * Acet_conc2 / (Acet_conc2 + Km_Acet_But)$

$v_Acet_transport2 = Acet_flux2$

$v_Prop_transport2 = Prop_flux2$

$v_But_transport2 = But_flux2$

$v_Nutr_trans2 = k_gut_out * Nutr2$

$v_Acet_trans2 = k_gut_out * Acet2$

$v_Prop_trans2 = k_gut_out * Prop2$

$v_But_trans2 = k_gut_out * But2$

; in 3d cmt (mmol/h):

$v_Nutr_intake3 = k_nutr_in$

$v_PolyNutr_Acet3 = k_Nutr_Acet * Nutr_conc3 / (Km_Nutr_Acet + Nutr_conc3)$

$$v_PolyNutr_Prop3 = k_Nutr_Prop * Nutr_conc3 / (Km_Nutr_Prop + Nutr_conc3)$$

$$v_Acet_But3 = k_Acet_But * Acet_conc3 / (Acet_conc3 + Km_Acet_But)$$

$$v_Acet_transport3 = Acet_flux3$$

$$v_Prop_transport3 = Prop_flux3$$

$$v_But_transport3 = But_flux3$$

$$v_Nutr_trans3 = k_gut_out * Nutr3$$

$$v_Acet_trans3 = k_gut_out * Acet3$$

$$v_Prop_trans3 = k_gut_out * Prop3$$

$$v_But_trans3 = k_gut_out * But3$$

; in 4th cmt (mmol/h):

$$v_Nutr_intake4 = k_nutr_in$$

$$v_PolyNutr_Acet4 = k_Nutr_Acet * Nutr_conc4 / (Km_Nutr_Acet + Nutr_conc4)$$

$$v_PolyNutr_Prop4 = k_Nutr_Prop * Nutr_conc4 / (Km_Nutr_Prop + Nutr_conc4)$$

$$v_Acet_But4 = k_Acet_But * Acet_conc4 / (Acet_conc4 + Km_Acet_But)$$

$$v_Acet_transport4 = Acet_flux4$$

$$v_Prop_transport4 = Prop_flux4$$

$$v_But_transport4 = But_flux4$$

$$v_Nutr_trans4 = k_gut_out * Nutr4$$

$$v_Acet_trans4 = k_gut_out * Acet4$$

$$v_Prop_trans4 = k_gut_out * Prop4$$

$$v_But_trans4 = k_gut_out * But4$$

; Differential equations

; -----

$$DDT_Acet1 = +v_PolyNutr_Acet1 - v_Acet_But1 - v_Acet_But1 - v_Acet_transport1 - v_Acet_trans1$$

$$DDT_But1 = +v_Acet_But1 - v_But_transport1 - v_But_trans1$$

$$DDT_Prop1 = +v_PolyNutr_Prop1 - v_Prop_transport1 - v_Prop_trans1$$

$$DDT_Nutr1 = +v_Nutr_intake1 - v_PolyNutr_Acet1 - v_PolyNutr_Prop1 - v_Nutr_trans1$$

$$DDT_Acet2 = +v_Acet_trans1 + v_PolyNutr_Acet2 - v_Acet_But2 - v_Acet_But2 - v_Acet_transport2 - v_Acet_trans2$$

$$DDT_But2 = +v_But_trans1 + v_Acet_But2 - v_But_transport2 - v_But_trans2$$


```

DDT_Prop2 = +v_Prop_trans1+v_PolyNutr_Prop2-v_Prop_transport2-v_Prop_trans2
DDT_Nutr2 = +v_Nutr_trans1+v_Nutr_intake2-v_PolyNutr_Acet2-v_PolyNutr_Prop2-v_Nutr_trans2
DDT_Acet3 = +v_Acet_trans2+v_PolyNutr_Acet3-v_Acet_But3-v_Acet_But3-v_Acet_transport3-v_Acet_trans3
DDT_But3 = +v_But_trans2+v_Acet_But3-v_But_transport3-v_But_trans3
DDT_Prop3 = +v_Prop_trans2+v_PolyNutr_Prop3-v_Prop_transport3-v_Prop_trans3
DDT_Nutr3 = +v_Nutr_trans2+v_Nutr_intake3-v_PolyNutr_Acet3-v_PolyNutr_Prop3 - v_Nutr_trans3
DDT_Acet4 = +v_Acet_trans3+v_PolyNutr_Acet4-v_Acet_But4-v_Acet_But4-v_Acet_transport4-v_Acet_trans4
DDT_But4 = +v_But_trans3+v_Acet_But4-v_But_transport4-v_But_trans4
DDT_Prop4 = +v_Prop_trans3+v_PolyNutr_Prop4-v_Prop_transport4-v_Prop_trans4
DDT_Nutr4 = +v_Nutr_trans3+v_Nutr_intake4-v_PolyNutr_Acet4-v_PolyNutr_Prop4-v_Nutr_trans4

; =====
$OUTPUT
; =====

; mM Output variable:
OUTPUT1 = Acet_conc4

; mM Output variable:
OUTPUT2 = Prop_conc4

; mM Output variable:
OUTPUT3 = But_conc4

```

ПРИЛОЖЕНИЕ 4

Кластеризация российских образцов по родовому составу						- в мировой кластеризации сменили кластер по метрике UniFrac		
	Средняя силуэтная ширина	0,189	0,12	0,362	0,209	0,077	0,209	0,361
	Число кластеров	2	3	2	2	4	2	2
	Образец	Д-Ш	Спирман	Евклидово	Манхэттен	Канберра	Б-К	UniFrac
1	Kh_232	1	1	1	1	1	1	1
2	Kh_236	1	2	2	1	2	1	2
3	Kh_240	2	2	2	2	2	2	2
4	Kh_242	2	2	2	2	3	2	2
5	Kh_245	1	1	1	1	4	1	1
6	Kh_247	1	1	1	1	4	1	1
7	Kh_249	2	3	2	2	4	2	1
8	Kh_250	1	1	1	1	4	1	1
9	NOV_280	1	2	2	2	1	2	1
10	NOV_281	2	3	2	2	1	2	2
11	NOV_282	2	2	2	2	3	2	2
12	NOV_283	2	2	2	2	1	2	2
13	NOV_284	2	2	2	2	3	2	2
14	NOV_285	2	2	2	2	3	2	2
15	NOV_286	1	2	1	1	3	1	1
16	NOV_287	1	3	2	2	3	2	2
17	OM_176	1	2	2	1	2	1	1
18	OM_178	2	2	2	2	3	2	2
19	OM_179	2	3	2	2	1	2	2
20	OM_180	1	3	1	1	4	1	1
21	OM_181	1	2	2	1	3	1	1
22	OM_182	2	3	2	2	1	2	2
23	OM_184	1	1	1	1	4	1	1
24	OM_185	2	2	2	2	3	2	2
25	OM_190	1	3	1	1	4	1	1
26	OM_191	1	3	2	1	4	1	2
27	OM_192	1	3	1	1	4	1	1
28	OM_193	1	3	1	1	4	1	1
29	OM_194	1	3	1	1	4	1	1
30	OM_195	1	3	2	1	4	1	2
31	OMS_251	2	3	2	2	4	2	2
32	RND_301	1	1	2	1	1	1	2
33	RND_304	2	2	2	2	1	2	2
34	RND_305	1	3	1	1	4	1	1

35	RND_306	2	3	2	2	1	2	2
36	RND_307	2	3	2	2	1	2	2
37	RND_308	1	1	2	1	4	1	1
38	RND_311	1	1	1	1	4	1	1
39	RND_312	2	3	2	2	1	2	2
40	RND_313	1	1	2	2	1	2	2
41	RND_314	2	1	2	2	1	2	2
42	RND_315	2	1	2	2	1	2	1
43	RND_316	2	1	2	2	1	2	2
44	RND_317	2	3	2	2	1	2	1
45	SAR_262	2	2	2	2	4	2	2
46	SAR_263	1	1	2	1	2	1	2
47	SAR_264	1	3	2	1	4	1	2
48	SAR_265	2	1	2	2	1	2	2
49	SAR_266	2	3	2	2	1	2	2
50	SAR_267	1	1	2	1	4	1	2
51	SAR_268	2	3	2	2	3	2	2
52	SAR_269	2	3	2	2	1	2	2
53	SAR_270	2	3	2	2	1	2	2
54	SAR_271	1	3	2	1	3	1	2
55	SAR_272	1	1	2	2	1	2	1
56	SAR_273	2	3	2	2	3	2	2
57	SAR_274	2	2	2	2	2	2	2
58	SAR_275	2	3	2	2	2	2	2
59	SAR_276	1	3	2	1	3	1	2
60	Spb_100_40P	1	1	1	1	4	1	1
61	Spb_101_41P	1	3	2	1	1	1	2
62	Spb_103_43P	1	3	2	2	1	2	1
63	Spb_105_45P	2	1	2	2	1	2	2
64	Spb_106_46P	2	3	2	2	1	2	2
65	Spb_61_1P	1	3	1	1	1	1	1
66	Spb_66_6P	1	1	1	1	4	1	1
67	Spb_69_9P	1	1	1	1	4	1	1
68	Spb_70_10P	1	3	1	1	4	1	1
69	Spb_72_12P	2	3	2	2	1	2	2
70	Spb_73_13P	2	2	2	2	1	2	2
71	Spb_74_14P	2	3	2	2	1	2	2
72	Spb_76_16P	2	1	2	2	1	2	1
73	Spb_98_38P	2	1	2	2	1	2	2
74	TAT_129	2	1	2	2	2	2	2
75	TAT_130	2	1	2	2	2	2	2
76	TAT_131	2	1	2	2	2	2	2
77	TAT_134	2	1	2	2	2	2	2
78	TAT_138	2	1	2	2	2	2	2
79	TAT_143	2	1	2	2	2	2	2

80	TAT_146	2	1	2	2	2	2	2
81	TAT_147	2	1	2	2	2	2	2
82	TYV_202	2	2	2	2	1	2	2
83	TYV_203	2	2	2	2	3	2	2
84	TYV_204	2	2	2	2	3	2	2
85	TYV_205	2	2	2	2	2	2	2
86	TYV_206	2	2	2	2	3	2	2
87	TYV_207	2	2	2	2	3	2	2
88	TYV_208	2	2	2	2	1	2	2
89	TYV_209	2	2	2	2	3	2	2
90	TYV_210	2	2	2	2	3	2	2
91	TYV_211	2	2	2	2	3	2	2
92	TYV_212	2	2	2	2	3	2	2
93	TYV_214	2	2	2	2	3	2	2
94	TYV_215	2	2	2	2	3	2	2
95	TYV_216	2	2	2	2	3	2	2
96	TYV_219	1	3	2	1	2	1	2