

Геномное
секвенирование

|2014
|NGS

СБОРНИК ТЕЗИСОВ

2-ой Всероссийской научно-практической конференции
по геномному секвенированию

15 мая 2014 года
Москва

Конференция поддержана Российским фондом фундаментальных исследований, проект No.14-04-06046

Электронный вариант тезисов: <http://ngsconference.ru/thesis/2014>

ISBN 978-5-600-00383-5

Программный комитет конференции:

Дмитрий Алексеев (НИИ ФХМ)
Михаил Гельфанд (ИППИ РАН)
Сергей Лукьянов (ИБХ РАН)
Всеволод Макеев (ИОГен РАН)
Николай Раввин (Центр "Биоинженерия" РАН)
Денис Ребриков (ДНК-Технология)
Евгений Рогаев (UMASS Medical School)
Николай Янковский (ИОГен РАН)

Организационный комитет конференции:

Сергей Лукьянов (ИБХ РАН)
Денис Ребриков (ДНК-Технология)

РАЗРАБОТКА АЛГОРИТМА ВЫРАВНИВАНИЯ NGS-ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МНОГОМЕРНЫХ ХЭШ-ФУНКЦИЙ

*Михаил В. Иванов, Максим В. Иванов, А.С. Касьянов, В.А. Милейко
Московский Физико-Технический Институт (Государственный университет)
e-mail: immichail@mail.ru*

Одним из наиболее востребованных применений NGS является секвенирование амплификационных библиотек: от отдельных генов до целого экзона. Особенность пробоподготовки с использованием ПЦР обеспечивает важное преимущество на этапе выравнивания данных: возможность однозначного соотнесения концов прочитанных последовательностей с концами референсных ампликонов. Авторами разработан оригинальный двухстадийный алгоритм, реализующий данное преимущество. Предлагаемый алгоритм основан на хэш-функции учета количества всех нуклеотидных дуплетов в фиксированном концевом участке анализируемых последовательностей. Такая хэш-функция непрерывна на пространстве всех возможных последовательностей, что позволяет заменить побуквенное сравнение анализируемых ридов и референсов простым арифметическим сравнением значений хэша. Алгоритм включает предварительное хэширование и индексацию всего набора референсных ампликонов, и две последовательные стадии анализа каждого рида: вычисление значения хэша и поиск кандидатного референсного ампликона на основе сравнения хэшей, а, затем, побуквенное выравнивание с кандидатным референсом. Такой подход значительно повышает общую производительность и понижает порядок сложности алгоритма, поскольку общая ресурсоемкость алгоритма слабо растет при увеличении размера референса, который влияет только на первую стадию анализа, и линейно зависит только от объема полученных данных секвенирования. Результаты первых сравнений производительности разработанного алгоритма с другими, эквивалентными по качеству выравнивания, свидетельствуют о радикальном приросте производительности. Полученные данные позволяют судить о том, что предлагаемый алгоритм после необходимой оптимизации сможет найти самое широкое применение в биоинформатических пайплайнах анализа данных NGS.

COMPUTER ANNOTATION OF BACTERIAL GENES USING PHYLOGENETIC PROFILES

*М.А. Голышев, Е.В. Коротков
Центр "Биоинженерия" РАН
e-mail: bioinf@yandex.ru*

Over the last years a great number of bacterial genomes were sequenced. Now one of the most important challenges of computational genomics is the functional annotation of nucleic acid sequences, and especially of nucleic acid sequences of genes. The current mathematical methods of functional annotation require the high level of homology between two sequences to reliably predict the possible gene function. Therefore, a sufficient share of the sequenced genes from bacterial genomes is not yet annotated. In this study we presented the computational method and the annotation system for predicting biological functions using phylogenetic profiles. The phylogenetic profile of a gene was created by way of searching for similarities between the nucleotide sequence of the gene and 1204 reference genomes, with further estimation of the statistical significance of found similarities. The profiles of the genes with known functions were used for prediction of possible functions and functional groups for the new genes. We conducted the functional annotation for genes from 104 bacterial genomes and compared the functions predicted by our system with the already known functions. For the genes that have already been annotated, the known function matched the function we predicted in 63% of the time, and in 86% of the time the known function was found within the top five predicted functions. Besides, our system increased the share of annotated genes by 19%. The developed system may be used as an alternative or complementary system to the current annotation systems and for creating functional groups of genes. The coincidence with functional groups of KEGG data bank is more than 50%. The demo version of system is: <http://genefunction.ru>.

QUALIMETRIC APPROACH TO MOLECULAR PHYLOGENETICS

*O.V. Gradov
Institute for energy problems of chemical physics RAS
e-mail: o.v.gradov@gmail.com*

This paper proposes a novel qualimetric approach to the analysis of molecular phylogenetic data obtained by sequencing and similar techniques. The above approach is shown to be applicable to the semantides analysis (according to L. Pauling terminology) from the early prebiotic stages including RNA-world models and the primary genetic code carriers up to modern synthetic nucleic acids underlying the alternative abiogenetic models of the genetic code carriers, methods and mathematical approaches for deciphering which are still being developed. The analysis of phenotypic expression of such genetic code carriers is provided from the standpoint of Rosen's optimality theory. The qualimetric value of such analysis is also considered in terms of Azgaldov approach. We also propose a number of qualimetric criteria, which are also the function similarity criteria for semantide sequences according to bijective correspondence or isomorphism based on Hamming distance (as a measure of proximity) and the similarity theory.

СЕКВЕНИРОВАНИЕ ГЕНОМА ПРОБИОТИЧЕСКОГО ШТАММА *LAC-TOBACILLUS CRISPATUS* 2029, ИСПОЛЬЗУЯ ION TORRENT PERSONAL GENOME MACHINE (PGM)

А.В. Карлышев¹, В.М. Абрамов², В.С. Хлебников², И.В. Косарев², В.Г.Мельников², Г.Т.Сухих³

¹*School of Life Sciences, Faculty of Science, Engineering and Computing; Kingston University; Penrhyn Road, Kingston upon Thames, United Kingdom*

²*ОАО «Институт инженерной иммунологии», Любучаны, Московская область*

³*Научный центр акушерства, гинекологии и перинатологии им. В.И.Кулакова
e-mail: a.karlyshev@kingston.ac.uk*

Сравнительный анализ 726 штаммов *L. crispatus* выявил шт. 2029 (VKM В-2727D), обладающий повышенной адгезию к эпителиальным клеткам репродуктивной системы человека и высоким уровнем продукции перекиси водорода. Целью данного исследования являлось выяснение генетических особенностей этого штамма используя геномное секвенирование на базе IonTorrent PGM с применением чипа 314. Библиотека фрагментов ДНК с конструировалась с помощью набора New England Biolabs. Генерирование и обогащение наночастиц для секвенирования проводилось с помощью системы OT2 (OneTouch-system 2). Для сборки и анализа контигов, картирования фрагментов на референсный штамм ST1 (1) использовали CLC Genomics Workbench. Размер генома штамма *L. crispatus* 2029 (2,193,123 nt) и G+C содержание (36.9%) соответствовали опубликованным данным по другим штаммам этой бактерии (2.04-2.46 Mb and 36. 7-37.1% соответственно). Аннотирование генома с помощью RAST (2) позволило выявить гены, ответственные за синтез и транспорт бактериоцина, белков S слоя и адгезина, специфичного к фибронектину. В отличие от штамма *L. crispatus* ST1, ген, кодирующий муцин-специфичный адгезин, отсутствовал. Ген sagB, отвечающий за продукцию стрептолизина S, присутствовал в исследуемом, но не в референсном штамме. Один из генов кодирует лактоцепин с возможной противовоспалительной активностью, связанной с деградацией про-воспалительных хемокинов (3). Эти и другие особенности штамма *L. crispatus* 2029 способствуют лучшему пониманию молекулярных механизмов действия пробиотиков. Геном штамма *L. crispatus* 2029 депонирован в GenBank под номером AVFH00000000 (версия AVFH02000000). 1. Ojala, et al. (2010). J. Bacteriol. 192:3547-3548 2. Aziz, R. K., et al. (2008). BMC Genomics.9:75 3. von Schillde, M., G. et al. (2012).Cell Host & Microbe. 11:387-396.

АКТУАЛЬНЫЕ МЕТОДИЧЕСКИЕ АСПЕКТЫ ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ БАКТЕРИЙ

*О.Л. Воронина, М.С. Кунда, Е.И. Аксенова, А.Н. Семенов, В.Г. Лунин, А.Л. Гинцбург
ФГБУ НИИЭМ им. Н.Ф. Гамалеи Минздрава России
e-mail: olv550@gmail.com*

Полногеномное секвенирование – эффективный инструмент в изучении геномов прокариот, позволяющий получить большой объем информации, как для описания отдельных микроорганизмов, так и для сравнительного анализа. Изучение 13 геномов бактерий, актуальных для медицинской микробиологии, позволило выявить основные трудности на пути работы исследователей. Прежде всего, это проблемы амплификации GC-богатых геномов (GC% - более 65,5), например, представителей родов *Mycobacterium*, *Burkholderia*, *Pseudomonas*. При работе с такими геномами необходимы модификации даже в высокоэффективной эмульсионной ПЦР метода пиросеквенирования: увеличение количеств копий ДНК на шарик (в 4-5 раз), уменьшение объема смеси для амплификации в индивидуальной ячейке. Отдельные области GC-богатых геномов не могут быть амплифицированы и при таких модификациях. Анализ структуры подобных участков ДНК в геноме *M. bovis* показал, что минимальная свободная энергия образования центроидной вторичной структуры составляет -1700.33 kcal/mol, термодинамического множества (с разнообразием 1232.86) -2568.90 kcal/mol. Добавление таких энхансеров, как трегалоза и DMSO в предельных концентрациях не всегда помогает достаточному снижению температуры плавления вторичных структур, что влечет за собой проблемы сборки геномов, особенно участков с повторами. Проблему определения кратности повторения идентичных и сходных участков в геномах бактерий для областей 20 000 п.н. и более помогает разрешить карта рестрикции, полученная посредством комплекса оптического картирования OpGen Argus. Использование карты рестрикции позволило нам доказать 3-х кратное повторение области DU-2 в геноме *M. bovis* BCG-Russia. Для менее протяженных повторов по-прежнему эффективно только секвенирование по Сэнгеру, например, для тес-кассет *S. epidermidis*.

ИЗУЧЕНИЕ ГЕНОМОВ *G. VAGINALIS* РАЗНЫХ ГЕНОТИПОВ МЕТОДАМИ ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ

Е.С. Шубина, Д.О. Коростин, В.В. Муравьева, Д.Ю. Трофимов
Научный центр акушерства, гинекологии и перинатологии им. В.И.Кулакова
e-mail: jekaterina.shubina@gmail.com

Полагают, что *G.vaginalis* является одним из основных микроорганизмов, который встречается при бактериальном вагинозе (БВ). БВ – самая частая причина патологических вагинальных выделений у женщин. Причем, *G.vaginalis* встречается как у больных, так и у здоровых женщин и механизмы развития БВ до сих пор неясны. На основании опубликованных последовательностей генов рибосомальной РНК *G.vaginalis* можно разделить на 3 группы, полагают, что штаммы, относящиеся к разным группам обладают разными свойствами.

Целью данной работы было получить последовательности геномов бактерий относящиеся ко всем трем группам и сравнить присутствующие в них гены, отвечающие за патогенность бактерий.

С помощью специфичных ПЦР праймеров был проверен 51 штамм *G.vaginalis*, для последующего секвенирования отобрано по одному штамму, относящемуся к каждой из групп.

Все образцы были секвенированы на секвенаторе Ion Torrent PGM на химии 100 п.н. с чипами 316v1, библиотеки были приготовлены по протоколу производителя.

Было получено 1 060 502 - 1 599 008 ридов со средней длиной 93-106 п.н. на образец. Сборка в контиги проводилась с помощью программы Mira3. Гены были проаннотированы с помощью ресурса RAST (Rapid Annotation using Subsystem Technology). Сравнение геномов выявило различия в генах отвечающих за цитотоксичность, резистентность и формирование биопленок, что может отвечать за различия в свойствах бактерий разных групп.

МОЛЕКУЛЯРНО-ФИЛОГЕНЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ РОДА ELYMUS L ТРИБЫ TRITICEAE ФЛОРЫ РОССИИ

В.А. Дмитриева, К.С. Добрякова, Н.Н. Носов, А.В. Родионов
Ботанический институт им. В.Л.Комарова РАН
e-mail: valeriya.dm1@gmail.com

Согласно системе, принятой Н.Н.Цвелевым, известные на территории России 53 вида рода *Elymus*, разделены на 4 секции: Sect.1. *Turczaninovia* Tzvel., Sect.2. *Goulardia* Tzvel., Sect.3. *Clinelymopsis* Tzvel., Sect.4. *Elymus*. В геноме видов рода *Elymus* L. выявлены пять разных геномых комбинаций: геном S предположительно, происходит из *Pseudoroegneria*, H от рода *Hordeum*, Y от неизвестного донора, P от *Agropyron* и W от *Australopyrum*. У разных видов рода *Elymus* известно 7 стабильных субгеномных комбинаций: SSHH и SSYY ($2n=4x=28$), SSHHYY, SSYYWW, SSYYPP, SSSSHH и SSSSYU ($2n=6x=42$). Цель представляемого исследования – изучить методами молекулярной филогении генетическое родство видов *Elymus* L., определить их место среди других родов трибы Triticeae флоры России. Для выделения ДНК нами был использован гербарный материал, собранный в Карачаево-Черкесской Респ., Алтайском крае, Респ. Горный Алтай. Молекулярно-филогенетическое исследование проводили используя нуклеотидные последовательности района ITS1–5.8SpPHK–ITS2 ядерного гена 45S рPHK. Мы секвенировали 15 последовательностей рода *Elymus*, 3 последовательности рода *Elytrigia*, и по последовательности одного вида каждого из родов: *Leymus*, *Psathyrostachys* и *Agropyron* (триба Triticeae). Используя полученные нами последовательности и 10 последовательностей видов из базы данных GenBank, мы реконструировали филогенетическое дерево в программе Mega5.2. Род *Elymus* входит в состав клад S и W. Клада P образована ITS последовательностями видов родов *Agropyron*. Отметим, что клада S не является монофилетичной, так как она образована видами *Elytrigia* и *Elymus*. Виды рода *Elymus*, входящие в состав клады S, являются представителями 2 и 4 секции. Мы предполагаем, что род *Elymus* является полифилетичным, что согласуется с его гибридным происхождением.

АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК, ГЕНЕРИРУЕМЫХ В РЕЗУЛЬТАТЕ АПОПТОЗА

*Е.М. Лосева¹, Л.О. Брызгалов², Е.С. Морозкин¹, А.М. Курильщиков¹, И.В. Морозов¹,
Е.В. Антонцева², Е.Ю. Рыкова¹, Т.И. Меркулова², В.В. Власов¹, П.П. Лактионов¹*

¹Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск

²Институт цитологии и генетики СО РАН, Новосибирск

e-mail: losevaem@gmail.com

Факторы, влияющие на процесс формирования пула внеклеточной ДНК (внДНК) могут приводить к различной представленности участков генома в пуле внДНК по сравнению с геномной ДНК родительских клеток. Поскольку основным механизмом генерации внДНК является апоптоз, нами было выполнено исследование состава последовательностей апоптотической и геномной ДНК при помощи секвенирования на платформе SOLiD 3. Апоптотическая ДНК была выделена из культуральной среды пяти линий первичных эндотелиоцитов после индукции в клетках апоптоза, геномная ДНК – из ядер этих же клеток без индукции апоптоза. В результате секвенирования было получено 458 и 412 миллионов прочтений длиной 50 н. для апоптотической и геномной ДНК, соответственно. В результате анализа полученных данных (peak calling, MACS14) было выявлено 764 последовательности ДНК (пиков), перепредставленных в составе апоптотической ДНК. В результате функциональной аннотации пиков было показано, что выявленные пики обогащены районами CpG островков, сайтами гиперчувствительности к ДНКазе I. Верификацию пиков проводили при помощи количественной ПЦР с использованием апоптотической ДНК эндотелиоцитов и циркулирующей ДНК, выделенной из плазмы здоровых доноров. В результате анализа 8 выбранных пиков было показано, что удельная концентрация последовательностей пиков в составе апоптотической ДНК выше в 2-12 раз в составе апоптотической ДНК. При сравнении представленности пиков в составе циркДНК плазмы и геномной ДНК лимфоцитов было обнаружено, что удельная концентрация всех выбранных последовательностей выше в составе циркДНК в 2-61 раз. Работа поддержана грантом РФФИ № 13-04-01460, МКБ №6.1, ФНМ № 23, проектом № 65 фундаментальных исследований, выполняемых СО РАН совместно со сторонними научными организациями на 2012-2014 гг.

ПОЛНОГЕНОМНЫЙ АНАЛИЗ СТРУКТУРЫ ХРОМАТИНА ЧЕЛОВЕКА МЕТОДОМ ДОСТУПНОСТИ ДЛЯ МЕТИЛИРОВАНИЯ МЕТИЛАЗОЙ DAM

*С.С. Буланенкова, О.Б. Филюкова, А.В. Кудрявцева, А.В. Снежкина,
С.Б. Акопов, Л.Г. Николаев, Е.Д. Свердлов*

*Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН
e-mail: sun-lioness@mail.ru*

Геномная ДНК эукариотических клеток со связанными с ней ядерными белками образует хроматин. Подобная структура служит не только для упаковки геномной ДНК в пределах ядра, но и играет важную регуляторную роль, влияя на доступность ДНК для ДНК-связывающих регуляторных факторов. В связи с этим условно выделяют участки открытого, транскрипционно активного и «молчащего» закрытого хроматина. Исследование структуры хроматина является актуальной проблемой современной молекулярной биологии. Нами был предложен подход к анализу структуры хроматина на полногеномном уровне *in vivo*. Основываясь на предположении, что доступность ДНК для метилирования метилазой Dam определяется степенью открытости хроматина, мы экспрессировали ген бактериальной метилазы Dam в клетках человека HEK293. После фрагментации геномной ДНК был произведен отбор фрагментов, подвергшихся Dam-метилированию. Параллельно сходным образом была получена библиотека всех геномных фрагментов независимо от статуса метилирования. Полученные полногеномные библиотеки были проанализированы крупномасштабным секвенированием. В результате была получена карта распределения Dam-метилированных сайтов в геноме человека. Распределение Dam-сайтов не было равномерным. В частности, было обнаружено, что регуляторные области кодирующих белки генов обладают повышенным уровнем Dam-метилирования, и соответственно, повышенной открытостью хроматина. Также наблюдается соответствие в расположении меток открытого хроматина и областей с повышенным уровнем Dam-метилирования. Можно заключить, что доступность для Dam-метилирования является хорошей меткой открытости хроматина в живой клетке, а также методом выявления регуляторных последовательностей, и таким образом, важным новым инструментом для полногеномного анализа хроматина млекопитающих.

ПРОБЛЕМЫ ДИАГНОСТИКИ ОРВИ И СЕКВЕНИРОВАНИЕ НОВОГО ПОКОЛЕНИЯ

А.П. Богоявленский

Институт микробиологии и вирусологии, Алматы, Казахстан

e-mail: anpav_63@mail.ru

Острые респираторные заболевания (ОРЗ) являются наиболее частой патологией в человеческой популяции. Инфекции дыхательных путей представляют серьезную проблему для здравоохранения во всех странах не только из-за частоты и тяжести течения, но и вследствие наносимого ими значительного экономического ущерба, как и отдельным лицам, так и обществу в целом. В среднем каждый взрослый 2-3 раза в год переносит ОРВИ, а ребенок заболевает 8-10 раз. Полиэтиологичность заболеваний, нередкое участие в процессе нескольких возбудителей, нестойкость и строгая типоспецифичность иммунитета являются причиной частых повторных заболеваний, что, в свою очередь, приводит в ряде случаев к формированию рецидивирующих и хронических процессов в бронхах и легких. Общеизвестно, что наиболее значимыми возбудителями ОРВИ у детей являются вирусы гриппа, аденовирусы, респираторно-синцитиальный вирус, вирус парагриппа и коронавирусы. Как правило, причиной ОРВИ становятся коинфекция 2-3 инфекционных агентов, раздельное выделение которых на сегодняшний день практически невозможно из-за неэффективности систем культивирования вирусов. Решение подобных проблем диагностики ОРВИ на сегодняшний день возможно 2 способами, мультиплексной ПЦР, имеющей недостатки, связанные с чувствительностью к отдельным формам коинфекции, и секвенирование нового поколения, позволяющее не только выявить коинфицирующий агент, но и дающее возможность оценки всего виroma индивидуума в целом. Подобная персонализация медицины может привести к совершенно новым системам профилактики и лечения вирусных заболеваний и создать комплекс мероприятий, позволяющий проводить эффективную противоэпидемическую или противоэпизоотическую работу.

ДИАГНОСТИЧЕСКАЯ ВИРУСОЛОГИЯ: ОТ ПЦР К СЕКВЕНИРОВАНИЮ НОВОГО ПОКОЛЕНИЯ

А.П. Богоявленский

Институт микробиологии и вирусологии, Алматы, Казахстан

e-mail: anpav_63@mail.ru

Несмотря на успешные исследования в области диагностики вирусных инфекций, создающих все новые эффективные конструкции качественного и количественного определения вирусов в окружающей среде, лишь немногие адаптированы к рутинным исследованиям, позволяющим их использование для скрининга большого количества клинических образцов. Это обусловлено рядом причин, основным из которых является четкое понимание того, какой метод может быть использован для рутинной диагностики в клинических лабораториях. В качестве примера можно проследить эволюцию диагностических вариантов анализа, основанных на методе цепной полимеразной реакции. Возможность стандартизации анализа и высокая воспроизводимость результатов исследований дали возможность широкого применения данной технологии в клинических лабораториях. Среди методов амплификации в клинической практике на сегодняшний день можно выделить несколько вариантов: обычная или ступенчатая амплификация, по принципу которой построено большинство коммерческих диагностических тест-систем; мультиплексная амплификация, позволяющая выявлять в одной пробе несколько вирусных инфекций, реакция петлевой изотермальной амплификации, используемая в слабо-оснащенных лабораториях или полевых условиях; секвенирование следующего поколения, позволяющее выявить неизвестные вирусы, за счет большого массива данных, получаемых в результате реакции. Опираясь на основные тенденции развития технологии конца XX начала XXI века можно предположить, что в случае существенного снижения стоимости оборудования и набора реагентов, секвенирование следующего поколения является единственным перспективным методом анализа в клинической практике, что в первую очередь обусловлено возможностью изучения инфекционного процесса в целом, без вычленения отдельных возбудителей инфекции.

ONCOFINDER: НОВЫЙ МЕТОД ВЫСОКОПРОИЗВОДИТЕЛЬНОГО АНАЛИЗА ТРАНСКРИПТОМОВ ДЛЯ ИССЛЕДОВАНИЯ АКТИВНОСТИ СИГНАЛЬНЫХ ПУТЕЙ И ПОДБОРА ЭФФЕКТИВНОЙ ТЕРАПИИ

*М.Б. Корзинкин, А.М. Алипер, А.В. Гаража, К. Лежнина, Н. Тереханова,
М. Баранова, Н.М. Борисов, А. Жаворонков, А.А. Буздин
Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН
Федеральный научно-клинический центр детской гематологии,
онкологии и иммунологии им. Д.Рогачева Минздрава России
Pathway Pharmaceuticals
e-mail: bu3din@mail.ru*

Анализ полных транскриптомов связан со сложностью интерпретации получаемых профилей генной экспрессии и с очевидными трудностями в понимании их комплексного влияния на биологические процессы. Для решения этой задачи мы создали новый метод, названный OncoFinder, впервые позволяющий исследовать уровни активации отдельных внутриклеточных сигнальных путей (ВСП) в исследуемом образце исходя из полнотранскриптомных данных. Отличием от конкурирующих подходов является то, что математический аппарат OncoFinder учитывает индивидуальную функциональную роль каждого белка-участника в прохождении сигнала и позволяет дать количественную характеристику уровня активации каждого ВСП. Впервые, метод позволяет нивелировать различия в полнотранскриптомных данных, полученных с использованием разных платформ, например, микрочипирования и широкомасштабного секвенирования. Применение данного подхода позволило нам охарактеризовать ряд новых сигнальных путей в качестве более эффективных маркёров развития некоторых видов рака, чем экспрессия индивидуальных генов. Кроме того, подход позволяет предсказывать успешность терапии для индивидуальных пациентов, а также прогнозировать течение заболевания. Для использования метода OncoFinder биомедицинским сообществом, мы создали одноимённый пакет программного обеспечения (ПО). Некоторые аспекты технологии отражены в недавних публикациях: (Буздин и др., <http://journal.frontiersin.org/Journal/10.3389/fgene.2014.00055/abstract>) и (Жаворонков и др., <http://journal.frontiersin.org/Journal/10.3389/fgene.2014.00049/abstract>). Дополнительная информация о ПО: <http://www.insilicomedicine.com/> и <http://pathwaypharmaceuticals.com/>. Авторы с энтузиазмом участвуют во многих совместных исследовательских программах. Контактный адрес электронной почты: buzdin@ponkc.com.

РАЗРАБОТКА СИСТЕМЫ ПРЕДСКАЗАНИЯ КЛИНИЧЕСКОЙ ЗНАЧИМОСТИ МУТАЦИЙ В ЧЕЛОВЕЧЕСКИХ ГЕНАХ, СВЯЗАННЫХ С РЕДКИМИ ЗАБОЛЕВАНИЯМИ

*Максим Иванов, Камиль Хафизов, Сергей Коваленко
Московский Физико-Технический Институт (Государственный университет)
e-mail: m4merg@gmail.com*

Наряду с бурным ростом количества нуклеотидных последовательностей, активно развиваются биоинформатические инструменты предсказания фенотипической значимости мутаций. Несмотря на то, что уже разработано значительное количество предназначенных для этих целей пакетов, используемые ими методы функционально ограничены, а их прогностическая эффективность зачастую вызывает немало вопросов. При этом данная область исследований развивается скорее экстенсивно, чем интенсивно. Разработанный нами метод, предназначенный для оценки фенотипической значимости мутации, помимо широко распространенных способов анализа физико-химических свойств аминокислот, консервативности сайта мутации, аннотаций последовательностей и геометрических свойств трехмерной структуры, исследует также и изменение свободной энергии белка при мутации, что до сих пор на нашло широкого применения в этой области. Классификация мутаций строится исходя из рассчитанных параметров с помощью эмпирических правил. Обучение алгоритма проводилось на выборке из 1521 мутаций. На тестовой выборке (1019 патогенных и 498 нейтральных) была показана прогностическая способность метода: точность составила 80,6%, корреляционный коэффициент Мэтью - 0.61, площадь под ROC-кривой - 0.878 (что достоверно выше соответствующих параметров Polyphen-2 и SIFT). Помимо прогностической эффективности бинарной классификации наблюдалась корреляция классифицирующей функции и степени патогенности мутации (коэффициент линейной корреляции — 0.5). Как следствие, разработанный нами метод способен поднять планку эффективности классификаторов мутаций по фенотипической значимости, что может облегчить такие задачи, как сужение области поиска мутаций, связанных с генетическими заболеваниями и интерпретацию мутаций неизвестной фенотипической значимости.

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ СЕКВЕНИРОВАНИЯ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ В ДИАГНОСТИКЕ ПЕРВИЧНЫХ КАНАЛОПАТИЙ И КАРДИОМИОПАТИЙ

*О.В. Глазова, Е.В. Залязьминская
Российский научный центр хирургии им. Б.В.Петровского РАМН
e-mail: olga_glazova_inbox.ru*

Как известно, первичные каналопатии и кардиомиопатии опосредованы нарушением структуры целого спектра генов, что затрудняет проведение диагностики классическими методами ПЦР и секвенирования по Сенгеру. Напротив, технологии секвенирования следующего поколения (NGS) перспективны для анализа большого количества генов. Эти технологии позволяют снизить стоимость и длительность диагностики благодаря высокой диагностической производительности. Экспертный консенсус по генетическому тестированию для каналопатий и кардиомиопатий (HRS/EHRA Expert Consensus Statement on the State of Genetic Testing for the Channelopathies and Cardiomyopathies) рекомендует проведение ДНК-диагностики более 5-10 генов как для кардиомиопатий, так и для каналопатий. Классический метод секвенирования по Сенгеру в таком случае занимает более чем 6 месяцев, что уменьшает ценность результатов диагностики для пациента. Нами были смоделированы две панели праймеров для мультиплексной амплификации генов, которые наиболее часто детерминируют развитие соответствующих патологий (10 генов для HCM/LVNC и 11 генов для каналопатии) с использованием сервиса Ampliseq Designer (LifeTechnologies). Каждая панель позволяет одновременно анализировать 15 пациентов на полупроводниковом чипе 314* поколения, либо 150 человек на чипе 316* поколения. Мы проанализировали две пилотные группы из 15 пациентов на каждую панель. Продолжительность анализа составила 4 недели. Конечная стоимость диагностики в пересчете на одного пациента оказалась в 10 раз ниже стоимости подобной диагностики, проведенной методом прямого секвенирования по Сенгеру. Полученные результаты свидетельствуют о том, что технологии NGS могут быть и должны внедряться в клиническую диагностику в качестве рутинных методов.

МУТАЦИЯ ГЕНА COL4A5 C.3098G>A ОБНАРУЖЕНА У ПАЦИЕНТОВ С ТЯЖЕЛОЙ ФОРМОЙ СИНДРОМА АЛЬПОРТА

*Л.И. Шагам, Д.В. Шенцева, А.В. Полякова, О.С. Грознова, В.С. Сухоруков,
Н.М. Карахан, Н.Е. Конькова, В.В. Длин
Научно-исследовательский клинический институт педиатрии
РНИМУ им. Н.И.Пирогова
e-mail: levshagam@gmail.com*

Синдром Альпорта (СА) - наследственный нефрит, характеризующийся гематурией, протеинурией и прогрессирующей почечной недостаточностью, часто сопровождаемый нейросенсорной тугоухостью и глазной патологией. Заболевание может быть обусловлено мутациями в генах альфа-субъединиц коллагена IV типа - COL4A3, COL4A4 и COL4A5. Патологические изменения в последнем гене наиболее распространены среди больных (~85% случаев). Всего описано более тысячи мутаций и полиморфизмов гена, среди которых значительную часть патогенных составляют замены глицина, входящего в состав повторяющихся фрагментов Gly-Xaa-Yaa в альфа-спиральном домене белка. Мы проводили исследуемым полное секвенирование кодирующих экзонов и смежных 5-нуклеотидных участков интронов генов COL4A3, COL4A4 и COL4A5 на полупроводниковом секвенаторе Ion PGM, амплификацию участков осуществляли методом мультиплексной ПЦР по технологии Ion AmpliSeq™. Всего было покрыто прочтениями 98% целевых участков. Исследование четырех родственников с семейной формой СА выявило у каждого из них мутацию гена COL4A5 c.3098G>A, p.G1033D в гетерозиготном состоянии на фоне отсутствия других потенциально патогенных мутаций. У всех исследованных пациентов отмечено раннее начало гематурии, протеинурии (до 1 года), нейросенсорная тугоухость, а также раннее развитие хронической почечной недостаточности. Патологий зрения не обнаружено. Выявленная мутация относится к классу вышеописанных глициновых замен, причем ранее была показана патогенность замен соседних глицинов в позициях 1030 и 1036.

НАРУШЕНИЯ ГЕНА ДИСТРОФИНА У ПАЦИЕНТОВ С МЫШЕЧНОЙ ДИСТРОФИЕЙ ДЮШЕННА/БЕККЕРА В РОССИЙСКИХ ПОПУЛЯЦИЯХ

*В.В. Ильинский, Д.В. Ребриков
Институт общей генетики им. Н.И. Вавилова РАН
e-mail: vilinsky@ya.ru*

Среди двигательных нарушений различного генеза особое место занимает прогрессирующая мышечная дистрофия Дюшенна/Беккера (МДД/Б), которая является одним из наиболее тяжелых наследственных болезней человека. Первые признаки тяжелых форм заболевания обычно начинают проявляться в возрасте 6-8 лет, дистрофия быстро прогрессирует, что при отсутствии купирующей терапии к десяти годам ведет к полному обездвиживанию. В то же время, ряд форм данного заболевания, обозначаемых как мышечная дистрофия Беккера, проявляются лишь в возрасте 15-25 лет и не приводят к столь тяжким нарушениям.

К настоящему моменту достигнут большой прогресс в изучении природы данного заболевания, идентифицирован ген и белок, ответственный за развитие миодистрофии, налажена система ДНК-диагностики. Установлено, что заболевание могут вызывать делеции, дубликации и однонуклеотидные полиморфизмы в экзонах, а также некоторые нарушения в интронах и регуляторной области. Частота встречаемости данных нарушений отлична в различных мировых популяциях, что приводит к необходимости создания различающихся диагностических тестов. Целью данной работы была оценка частоты встречаемости нарушений в гене *dmd* в российских популяциях.

В ходе работы был проведен поиск делеций, дубликаций и однонуклеотидных мутаций в гене дистрофина у 159 пациентов с МДД/Б. В результате анализа было установлено, что делеции встречаются в 44% случаев, дубликации – в 4,5% случаев, однонуклеотидные мутации – в 51,5% случаев. Данные результаты хорошо соотносятся с распространенностью нарушений гена *dmd* в европейских популяциях по данным литературных источников.

Кроме этого, были определены «горячие участки» гена *dmd*, частота мутаций в которых больше, чем в других участках. Данные результаты позволяют по-новому взглянуть на разработку методов молекулярно-генетической диагностики пациентов с МДД/Б.

Работа будет продолжена с целью поиска и установления взаимосвязи локализации мутаций и их типов с клиническими проявлениями МББ/Б.

Работа была поддержана целевым грантом ООО «Генотек».

АНАЛИЗ ТАРГЕТНЫХ УЧАСТКОВ ГЕНОМА С ИСПОЛЬЗОВАНИЕМ СЕКВЕНИРОВАНИЯ НОВОГО ПОКОЛЕНИЯ ПРИ ЗЛОКАЧЕСТВЕННЫХ НОВООБРАЗОВАНИЯХ ЖКТ

*Т.В. Наседкина, И.С. Абрамов, Ч.А. Джумакова, Л.Н. Любченко
Институт молекулярной биологии имени В.А.Энгельгардта РАН
Российский онкологический научный центр им. Н.Н.Блохина РАМН
e-mail: tanased06@rambler.ru*

Около 20-25% случаев колоректального рака (КРР) и около 10% случаев рака желудка имеют семейную историю. Среди семейных форм КРР выделяют несколько наследственных синдромов. Семейный аденоматозный полипоз ассоциирован с мутациями в гене APC, наследственный неполипозный колоректальный рак (синдром Линча) вызывается мутациями в генах системы репарации неспаренных нуклеотидов (mismatch repair, MMR), МУН-ассоциированный полипоз связан с мутациями в гене MUTYH. При наследственной предрасположенности к раку желудка имеют значение мутации в гене CDH1. Для анализа участков генома, ассоциированных с наследственными злокачественными новообразованиями желудочно-кишечного тракта (ЖКТ), использовали отбор последовательностей с помощью библиотеки зондов NimbleGen Sequence Capture (Жидкий чип) или мультиплексную амплификацию кодирующих участков целевых генов (MLH1, MSH2, MSH6, PMS2) (Multiplicom) с последующим секвенированием на платформе GS Junior (454/Roche). Библиотека зондов NimbleGen Sequence Capture включала 2.1 миллиона олигонуклеотидов длиной 50-105 н.о., комплементарных кодирующим участкам генов MLH1, PMS1, PMS2, MSH2, TP53, MSH6, CDH1, APC, BMPR1A, CHEK2, EPCAM, MYT1H, PTEN, SMAD4, STK11. Всего проанализировано 20 пациентов. Выявлено более 20 редких вариантов и патогенных аллелей в генах MUTYH, MSH2, MSH6, APC, PMS2, PTEN и CHEK2. Обсуждаются возможности использования результатов секвенирования в диагностике злокачественных новообразований ЖКТ.

ПОИСК ЭПИГЕНЕТИЧЕСКИХ МАРКЕРОВ РАКА ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ В ЦИРКУЛИРУЮЩЕЙ ДНК ПЛАЗМЫ КРОВИ

*А.А. Бондарь¹, Е.С. Морозкин¹, А.М. Курильщиков¹, И.А. Запороженко¹,
М.Р. Кабилов¹, А.Е. Тупикин¹, М.М. Зарипов², В.Е. Войццицкий², Е.Д. Чикова¹,
В.В. Власов¹, П.П. Лактионов¹*

*¹Институт химической биологии и фундаментальной медицины СО РАН,
г. Новосибирск*

*²Новосибирский областной онкологический диспансер, г. Новосибирск
e-mail: a_bondar@niboch.nsc.ru*

Было проведено сравнение паттернов метилирования фрагментов генов RNF219, KIAA1539, ZC3H4 и GSTP1, потенциальных маркеров рака предстательной железы, в составе циркулирующей ДНК (цирДНК) плазмы 5 здоровых доноров (ЗД) и 5 больных раком предстательной железы (РПЖ) при помощи секвенирования платформе MiSeq. После выделения цирДНК проводили бисульфитную конверсию, амплифицировали целевые фрагменты генов и секвенировали. В результате секвенирования были получены данные с покрытием для каждого продукта ПЦР от 10 до 105 тысяч прочтений. Для анализа данных секвенирования был использован стандартный подход, основанный на анализе диагностической значимости статуса метилирования индивидуальных CpG сайтов в общем пуле циркДНК от каждого донора, и разработанный нами, подход (Representation of Individual Types of Molecules (RITM) analysis), основанный на подсчете количества молекул, имеющих идентичный профиль метилирования и оценки их доли в общем пуле секвенированных молекул для каждого донора и мишени. При помощи RITM анализа было показано, что в пуле циркДНК GSTP1 больных раком предстательной железы встречаются уникальные паттерны метилирования, которые отсутствуют в крови здоровых доноров. Для генов RNF219, ZC3H4 и KIAA1539 не было обнаружено паттернов достоверно отличающихся у больных и здоровых, однако в пуле фрагментов генов RNF219 и KIAA1539 обнаружены паттерны метилирования, достоверно отличающие доноров по возрасту. Работа поддержана грантом ФНМ № 18; грантом РФФИ № 13-04-01460; именной стипендией Президента Российской Федерации для молодых ученых и аспирантов, осуществляющих перспективные научные исследования и разработки по приоритетным направлениям модернизации российской экономики (Конкурс 2013-2015 годов, № СП-4710.2013.4).

НЕИНВАЗИВНАЯ ПРЕНАТАЛЬНАЯ ДИАГНОСТИКА (NIPT) НА БАЗЕ ПЛАТФОРМЫ ION PROTON

Д.О. Коростин, Е.С. Шубина, Н.А. Каретникова, Д.В. Ребриков, Д.Ю. Трофимов
Научный центр акушерства, гинекологии и перинатологии им. В.И.Кулакова
e-mail: d.korostin@gmail.com

В конце 90-х гг XX века было доказано наличие небольшой, около 5-10%, фракции фетальной ДНК, свободно циркулирующей в кровяном русле матери вне клеток. На основе использования платформ для высокопроизводительного секвенирования, появившихся в начале XXI века, были разработаны подходы к диагностике наличия анеуплоидий у плода по внеклеточной ДНК в крови матери. Серия научных работ подтвердила высокую чувствительность и специфичность методики, что в итоге привело к внедрению коммерческого теста в США и Китае в клиническую практику.

Ранее NIPT осуществлялась на наиболее производительных платформах типа SOLiD 5500 и Illumina HiSeq, так как алгоритмы обработки результатов, основанные на оценке доли представленности каждой хромосомы в массиве данных секвенирования, успешно работают с большим количеством прочтений, измеряемым миллионами ридов на образец. Недавно на рынке появилась платформа Ion Proton, основанная на технологии полупроводникового секвенирования и обладающая в первом приближении необходимым уровнем производительности для применения в NIPT.

Нами был разработан алгоритм обработки данных секвенирования, протокол подготовки образцов под платформу Ion Proton, а также проведено исследование первых 50 образцов (36 нормальных, 8 с трисомией по 21 хромосоме, 4 с трисомией по 18 хромосоме, 2 с кариотипом 45X) внеклеточной ДНК (10-14 неделя беременности) на выборке с известным клиническим диагнозом плода, поставленным инвазивными методами.

Образцы были секвенированы с использованием химии v2 и v3 и чипа Ion Proton P1v2. На одном чипе мультиплексировали 5 образцов. В среднем для каждого образца было получено 6,5 млн. ридов. По нашим оценкам достаточным для анализа является 5 млн. ридов на образец, что выполнялось для 28 нормальных образцов, 7 с трисомией по 21 хромосоме, 3 с трисомией по 18 хромосоме и 2 с кариотипом 45X, все эти образцы удалось диагностировать правильно.

Мы можем утверждать, что Ion Proton по своим характеристикам удовлетворяет требованиям к диагностике методом NIPT и может быть использовано для этих целей.

NGS В СТОМАТОЛОГИИ: БАКТЕРИАЛЬНЫЕ КОМПЛЕКСЫ ПАРОДОНТА В НОРМЕ И ПРИ АГРЕССИВНОМ ПАРОДОНТИТЕ

А.В. Шибалева¹, А.Б. Шевелев², О.А. Зорина³, О.А. Борискина³, Е.С. Шубина⁴,
О.В. Осипенкова⁵, Д.В. Ребриков⁴

¹Институт биохимической физики им. Н.М. Эмануэля РАН

²Институт полиомиелита им. М.П. Чумакова РАМН

³Центральный научно-исследовательский институт стоматологии и
черепно-лицевой хирургии МЗ РФ

⁴ЗАО «НПФ ДНК-Технология»

⁵ООО «Иннотех-21»

e-mail: drebrikov@gmail.com

Причина роста заболеваемости агрессивным пародонтитом (АП) в последние годы остается неизвестной, предполагается участие в этом микроорганизмов. Цель работы: с помощью NGS-секвенирования выявить различия в составе бактерий, населяющих пародонт у здоровых лиц и пациентов с АП.

В исследовании участвовало 5 здоровых доноров и 5 пациентов с АП. Банки рДНК были получены с помощью праймеров к гипервариабельной области V4 гена 16S. Они анализировались на приборе Illumina MiSeq. Для 10 образцов было определено 1017669 последовательностей: от 69566 до 139385 на образец. Анализ последовательностей проводился с помощью пакета QIIME и базы NCBI GenBank. Полученные последовательности были сгруппированы в таксономические единицы ранга рода и вида. В результате было получено около 350 таксономических группировок (OUT), представленность которых в выборках исследовалась методом параметрического статистического анализа.

Было выявлено доминирование у больных АП 8 родов бактерий: Porphyromonas, Treponema, Synergistes, Tannerella, Filifactor, Ruminococcus, Parvimonas, Mycoplasma, из которых только Porphyromonas, Treponema и Tannerella традиционно рассматриваются как пародонтопатогены. У лиц со здоровым пародонтом выявлено доминирование рода Veillonella.

Существенное различие между выборками показали виды рода Prevotella. Доминирование вида *P. intermedia* ассоциировано с АП, а комплекса *P. nigrescens*, *P. oris*, *P. veroralis* - со здоровым пародонтом. Опасный пародонтопатоген *Aggregatibacter actinomycetemcomitans* не был выявлен ни у одного из пациентов, что не позволяет считать его причиной роста заболеваемости АП в г. Москва.

Выявлены комплексы бактерий, ассоциированных со здоровым пародонтом. В состав основного комплекса входят виды рода Veillonella и непатогенные виды Prevotella. В качестве альтернативного «защитного комплекса», способного замещать главный комплекс, может рассматриваться группа видов *Streptococcus sanguinis*, *Kingella oralis*, *Granulicatella paradiacens* и *Bergeyella sp.*

Подписанно в печать 13.05.2014 г.

Заказ №1074 Тираж: 400 экз.

Печать цифровая.

Типография «Премиум Принт»
Москва, ул. Миклухо-Маклая 8/3

8 (499) 739-56-97

www.premium-print.ru