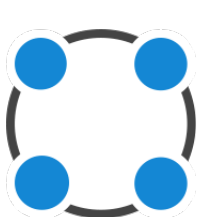


Институт биоинформатики
Московский государственный университет
Санкт-Петербургский Академический университет РАН



Третья летняя школа по биоинформатике

Москва, 20-25 июля 2015

Тезисы докладов

УДК 004.8
ББК 28.0



EMC²

Третья летняя школа по биоинформатике (Москва, 20-25 июля 2015). Тезисы докладов. — Санкт-Петербург: Свое издательство, 2015. — 30 с.

ISBN 978-5-4386-0855-4

Содержание

Секвенирование генома штамма <i>Bacillus ginsengihumi</i> M2.11, продуцирующего фитазу	5
Characterization of <i>C6orf97</i> , a Novel Breast Cancer Susceptibility Candidate Gene	7
<i>De novo</i> сборка транскриптома лиственницы сибирской (<i>Larix sibirica</i> Ledeb.)	8
Сборка и сравнительный анализ хлоропластных геномов хвойных рода <i>Larix</i>	9
Кластеризации данных генной экспрессии при раке молочной железы	10
Молекулярные механизмы защиты опухолевых клеток от противоракового белка TRAIL	12
MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data	14
Apache Taverna: scientific workflow management system	16
Изучение генетических основ социального поведения на основе анализа генов селекционированных линий лис	17
Исследование микробиоты полости рта у детей при терапии кариеса жевательной группы зубов	18
Совершенствование метода микроядерного тестирования для измерения генетической токсичности и исследования геномной стабильности	20

Исследование ассоциации полиморфных вариантов генов цитокинов и факторов роста с бесплодием в супружеских парах	22
Изучение генетического разнообразия земляники садовой (<i>Fragaria ananassa</i>) с использованием молекулярных маркеров SSR-типа	24
Изучение полиморфизма генов устойчивости класса NBS-LRR пшеницы с использованием NGS	25
Анализ лизосомных болезней накопления, связанных с мутациями	26
The method of analysis inherited interspersed and satellite repeats using genome sequencing data and application on genome sequencing data of DNA samples patients with schizophrenia family's trio	27
Компьютерный анализ и обработка экспрессионных данных	28
Использование технологии RNA-seq для анализа некодирующих РНК гена <i>SGMS1</i> человека	29

Секвенирование генома штамма *Bacillus ginsengihumi* M2.11, продуцирующего фитазу

Д.С. Баранова, А.А. Тойменцева, М.Р. Шарипова

Казанский (Приволжский) федеральный университет, ул. Кремлевская,
18, 420000, Казань, Россия

Контактная информация: DashulyaBaranova@mail.ru, 89656088327

Bacillus ginsengihumi – грамположительная бактерия, характеризующаяся множественной чувствительностью к антибиотикам, отсутствием протеолитической активности и наличием фитазной активности. Фитазы способны гидролизовать фитат, высвобождая неорганический фосфат и делая его доступным для усваивания животными. Поэтому фитазы штамма *B. ginsengihumi* могут найти практическое применение в биотехнологии, в качестве кормовых добавок. О физиологических особенностях данного микроорганизма, как и о его геноме известно очень мало. Задачей настоящей работы являлось полногеномное секвенирование штамма *B. ginsengihumi* M2.11, для более полного изучения особенностей генома бактерии.

Штамм *Bacillus ginsengihumi* M2.11 был выделен с территории Тепличного комбината «Майский» Республики Татарстан. Полногеномное секвенирование штамма *B. ginsengihumi* M2.11 было выполнено на платформе GS Junior (454 пиросеквенирование). *De novo* сборка и анализ контигов проводили с использованием ассемблера SPAdes 3.5.0. Статистику изучали с помощью программы QUAST 2.3.. Секвенированный геном состоит из 144 контигов длиной >200 п.о.. Общий размер генома составляет 3 722 429 п.о., значение $N50 = 58040$ п.о., $L50 = 23$ п.о.. Среднее содержание GC в геноме равно 36.14%, что говорит о слабой консервативности генома бактерии. BLAST анализ последовательности 16s рНК показал высокую идентичность (97%) с известными штаммами *Bacillus ginsengihumi*. Автоматическая аннотация с помощью Pipeline (PGAAP) показала наличие 3082 кодирующих последовательностей (CDS), 34 рНК, 88 тРНК и 400 рамок считывания генов. Система аннотаций

RAST предсказала 449 субсистем. В геноме были найдены гены устойчивости к фторхинолонам и фосфомицину. Несмотря на способность штамма продуцировать фитазы, ген фитазы найден не был. Однако, геном штамма содержит гены, участвующие в ассимиляции фосфора: гены ответственные за секрецию щелочных фосфатаз, полифосфат киназ и пирофосфатаз. Полное секвенирование генома данной бактерии позволит предсказать физиологические характеристики штамма и использовать эти данные для создания наиболее благоприятных условий для культивирования бактерии.

Characterization of *C6orf97*, a Novel Breast Cancer Susceptibility Candidate Gene

N.M. Baulina¹, O.G. Kulakova¹, X. Chen²

¹ Pirogov Russian National Research Medical University
Ostrovityanova st., 1, Moscow, Russia

² Fox-Chase Cancer Center
333 Cottman ave, Philadelphia, Pennsylvania, USA
e-mail: tati.90@mail.ru, phone +7 (985) 245-06-45

Background: Breast cancer (BCa) is the most common cancer affecting women, and most familial aggregation of BCa remains unexplained. There is mounting evidence suggesting that there may be additional, but less prevalent, BCa susceptibility genes; however, identification of these genes remains elusive.

Hypothesis: Genome-wide association studies (GWAS) have been used to identify genetic factors that account for breast tumorigenesis, however, this method has some limitations. Differential allele specific expression (DASE) is a common phenomenon that has been shown to contribute to phenotypic variability in human and pathogenesis of cancer. Overlying the resulting DASE maps with GWASes helps to pinpoint *cis*-acting regulatory variants that are specifically active in cancer tissue and cause allelic imbalance. This strategy allowed to maximize the yield of true positive loci from BCa GWASes, including novel BCa candidate locus of *C6orf97* gene. As nothing was known about this gene and its' encoded protein CCDC170 the aim of our study was to functionally characterize this gene.

Results: Our results showed that gene *C6orf97* inhibits BCa cell proliferation and migration. CCDC170 demonstrates a unique cellular localization in nuclear envelope; moreover, disruption of C-terminal part of CCDC170 leads to the loss of the nuclear envelope localization. CCDC170 was identified to interact with NonO protein, which involved in many nuclear processes including transcription, RNA processing, and retention of hyperedited RNAs. Our findings demonstrate that *C6orf97* is a novel BCa suppressor gene, involved in post-transcriptional regulation of gene expression.

De novo сборка транскриптома лиственницы сибирской (*Larix sibirica* Ledeb.)

В.В.Бирюков

Сибирский федеральный университет
Российская Федерация, 660041, г. Красноярск, пр. Свободный, 79
e-mail: vladbir2010@gmail.com

В связи с быстрым развитием секвенирования нового поколения и биоинформатических инструментов в течение последних нескольких лет, стали реальными полногеномный и полнотранскриптомный анализы для немодельных генетических систем, например, хвойных растений, которые отличаются от модельных организмов медленным ростом, высокой фенотипической изменчивостью и огромным размером генома, который богат повторяющимися последовательностями.

В ходе работы, на основе данных секвенирования мРНК, была получена первичная сборка транскриптома лиственницы сибирской (*Larix sibirica* Ledeb.) длиной 26493048 п.н.о. Было произведено картирование ридов обратно на сборку, а также были определены основные показатели сборки: общее число контигов – 43686, длина максимального контига – 8512 п.н.о. и N50 – 878 п.н.о. В ходе работы был проведен сравнительный анализ полученной сборки с транскриптомами других хвойных[1], который показал достаточно высокую степень сходства собранных контигов и контигов предыдущих сборок.

Сборка транскриптома – важный этап в изучении регуляции экспрессии генов и регуляторных механизмов клетки. Изучение хвойных видов играет большую роль, так как они оказывают огромное влияние на экологию и климат, занимают лидирующее место в поглощении углерода и предотвращении последствий глобального потепления климата.

Работа выполнена в рамках проекта «Геномные исследования основных бореальных лесобразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации», финансируемого Правительством РФ (договор №14.Y26.31.0004).

1. J.L.Wegrzyn, J.M.Lee, B.R.Tearse, D.B.Neale. (2008) TreeGenes: A Forest Tree Genome Database, International Journal of Plant Genomics, 2008.

Сборка и сравнительный анализ хлоропластных геномов хвойных рода *Larix*

Е. И. Бондар

Сибирский федеральный университет
660041, г. Красноярск, пр. Свободный, д. 79 (ИФБиБТ), Красноярск,
Россия, -mail: bone-post@yandex.ru, 89135249300

На сегодняшний день в базе данных NCBI Genbank опубликовано более 100 последовательностей геномов хлоропластов для семейства сосновых. Однако большинство из них относятся к роду *Pinus*. Для рода *Larix* хлоропластные геномы известны только для европейской и западной лиственниц.

Целью данного исследования является сборка и аннотирование хлоропластного генома лиственницы сибирской (*Larix sibirica* Ledeb., 1833), а также поиск однонуклеотидных полиморфизмов.

Для сборки хлоропластного генома использовались данные полногеномного секвенирования *L. sibirica*, полученные в Лаборатории лесной геномики СФУ под руководством проф. К. В. Крутовского. Сборка производилась с использованием программ картирования Bowtie2 и геномного ассемблера SPAdes. Аннотирование генома проводилось при помощи сервиса Rapid Annotation using Subsystem Technology (RAST). Для поиска SNP между тремя деревьями одного вида, произрастающими в разных местностях, использовались программы Bowtie2 и Ugene (Call Variants with SAMtools).

Длина хлоропластного генома *L. sibirica* составила 122561 bp и близка к 122474 bp у близкородственной *Larix decidua* Mill.. В результате аннотирования и дальнейшего сравнения полученных данных с аннотациями хлоропластных геномов близкородственных видов *L. decidua* и *L. occidentalis*, был выявлен 121 кодирующий участок, из которых 34 соответствуют генам тРНК и 87 CDS. Для трех деревьев найдено 13 позиций с SNP, 2 из них находятся в кодирующих участках генома.

Работа выполнена в рамках проекта «Геномные исследования основных бореальных лесообразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации», финансируемого Правительством РФ (договор № 14.Y26.31.0004).

Кластеризации данных геной экспрессии при раке молочной железы

Д. Вальтер¹

Под руководством: В. Степанова², И. Колесниченко

1. Национальный исследовательский университет «Высшая школа экономики», Факультет Компьютерных наук.

Россия, Москва, Кочновский проезд, 3, 125319

2. Сколковский институт науки и технологий.

Россия, Московская область, Одинцовский район, дер. Сколково, ул.

Новая, 100, 143025

3. Компания «Яндекс»

Россия, Москва, ул. Льва Толстого, 16, 119021

e-mail: dasha-walter@yandex.ru

В последнее время активно развиваются способы оценки степени тяжести раковых опухолей по данным геной экспрессии. Анализ данных геной экспрессии – это потенциальный способ диагностики и прогнозирования рака. Внедрение технологий анализа и классификации данных геной экспрессии в медицину сделает возможным диагностировать болезнь и наблюдать за ее развитием за кратчайшее время и минимизировать использование инвазивных методов взятия анализов у пациентов.

В нашем проекте были поставлены следующие задачи:

Реализовать алгоритм кластеризации методом k-means и найти разбиение пациентов на кластеры в зависимости от экспрессии генов.

Для осуществления эффективной кластеризации подобрать метрику, способы нормализации данных и определить эффективный способ подбора количества кластеров разбиения.

Сравнить полученное разбиение на кластеры с разбиением по морфологическим характеристикам. За биологические разбиения взяты разбиения по наличию или отсутствию метастаз/рецидива у пациента.

В результате сравнения определить, каковы параметры кластеризации, наиболее близкой к морфологическим наблюдениям.

Для каждого кластера выбрать наборы генов, отличающихся наибольшими показателями экспрессии, провести enrichment analysis, определяющий в каком биологическом пути участвуют данные наборы генов.

Решить задачу определения набора дифференциально экспрессированных генов, необходимого для эффективной кластеризации раковых транскриптов.

Визуализация полученных результатов. Создание интерфейса для кластеризации раковых транскриптов.

Используемые технологии:

- БД GEO
- Язык программирования C++11
- Enrichment Analysis

Молекулярные механизмы защиты опухолевых клеток от противоракового белка TRAIL

Долгих Н.В., Чеканов А.В., Акатов В.С.

*Институт теоретической и экспериментальной биофизики РАН,
Пуцзино
Пуцинский государственный естественно-научный институт, Пуцзино
n.v.dolgikh@gmail.com*

Со времени описания белка TRAIL, как наиболее селективного природного участника иммунной системы, способного вызывать гибель исключительно опухолевых клеток, остается вопрос каким образом опухоль выживает и прогрессирует, несмотря на присутствие TRAIL продуцирующих клеток и растворимого TRAIL в опухолевом окружении? Чувствительность опухолевых клеток к TRAIL-индуцированному апоптозу является ключевым фактором эффективности противоопухолевой терапии на основе рекомбинантного белка TRAIL. Однако механизмы чувствительности и резистентности опухолевых клеток к TRAIL-индуцированному апоптозу до конца не выяснены. Поэтому целью нашей работы является выяснение механизмов резистентности опухолевых клеток к действию TRAIL.

Ранее нами было показано, что при увеличении концентрации клеток и достижения монослоя (конфлюентной культуры) клетки линии A431 становятся резистентными к действию белка TRAIL, в то время как в редкой культуре действие TRAIL вызывает гибель этих клеток. Для того чтобы исследовать каким образом опухолевые клетки избегают гибели, вызванной TRAIL, мы получили стабильные TRAIL-резистентные клетки (сублиния A431-R) и TRAIL-чувствительные клетки (сублиния A431-S). В отличие от клеток A431 дикого типа полученные клетки A431-R резистентны в редкой культуре, а A431-S – чувствительны как в редкой, так и в конфлюентной культуре.

Часто резистентность опухолевых клеток в многоклеточных образованиях связывают с переходом клеток в непролиферативное состояние. Однако наличие чувствительной

сублинии, утратившей конfluenceнтную резистентность, позволяет объяснять механизм резистентности опухолевых клеток в конfluenceнтной культуре не только отсутствием в них пролиферативной активности, поскольку A431-S также как и A431 дикого типа не пролиферируют в конfluenceнтной культуре, но при этом являются чувствительными к действию TRAIL. Наличие сублинии A431-R, резистентной в растущей культуре, говорит о множественности возможных механизмов резистентности опухолевых клеток. Кроме того, в нашей работе был исследован уровень экспрессии белков участвующих в реализации механизма апоптоза. В результате исследования было показано, что одним из возможных механизмов возникновения конfluenceнтной резистентности является нарушение проведения сигнала с лиганд-рецепторного комплекса. Полученные результаты указывают на высокий и разнообразный потенциал защиты опухолевых клеток от гибели, что необходимо учитывать при разработке новых противоопухолевых стратегий.

MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data

S.V. Kazakov¹, V.I. Ulyantsev¹, V.B. Dubinkina^{2,3}, A.V. Tyakht², D.G. Alexeev^{2,3}

¹ *ITMO University*

49 Kronverksky Pr., Saint-Petersburg, Russia

² *Research institute of physico-chemical medicine*

1a Malaya Pirogovskaya, Moscow, Russia

² *Moscow institute of physics and technology (State University)*

9 Institutskiy per., Dolgoprudny, Moscow Region, Russia

e-mail: svkazakov@rain.ifmo.ru

Since the emergence of high-throughput genomic sequencing technologies, a huge volume of data has been accumulated describing complex microbial communities (microbiota). Efficient statistical analysis of such data – including identification of taxonomic and functional composition, community richness and similarity – requires dimension reduction.

Common methods for feature extraction from shotgun metagenomic data include:

- reference-based methods implicating alignment of sequence reads against a catalog of reference sequences;
- methods based on de novo assembly with subsequent analysis of the long contigs;
- abstract composition-based methods including k-mer spectrum analysis, neural networks, Markov models, etc.

Despite the popularity of the available approaches, most of them have inherent disadvantages that limit their scope of applicability. For instance, reference-based methods require a representative database of known genomes, however, many microbial branches of the tree of life still lack such sequences. Composition-based methods do not require a reference base; but the interpretation of the resulting features is often unclear.

We have developed a novel approach *MetaFast* based on semi-assembly. It does not require a priori knowledge about the taxa possibly included in the microbiota. Other advantages over the

above-mentioned methods are rather small system requirements and interpretability of the extracted features.

To test the applicability of our method, several datasets consisting of simulated and real metagenomes were used. Comparison analysis showed that in general *MetaFast* produces a dissimilarity matrix similar to those provided by reference-based approaches.

APACHE TAVERNA: SCIENTIFIC WORKFLOW MANAGEMENT SYSTEM

D. Karyakin¹, A. Williams², S. Soiland-Reyes², R. Haines²

¹ *National Research Saratov State University
83 Astrakhanskaya Street, Saratov, 410012*

² *Apache Taverna*

e-mail: samhane.me@gmail.com, tel. +7 917 215 75 96

Taverna is an open source domain independent Workflow Management System - a suite of tools used to design and execute scientific workflows.

Workflow Management System (WMS) is a piece of software that provides an infrastructure to setup, execute, and monitor scientific workflows. In other words, the WMS provide an environment where in silico experiments can be defined and executed.

Workflow and databundle viewers are useful for workflow management system. They help to understand what steps are in a workflow, what data is processed on each step.

The Taverna suite is written in Java and includes the Taverna Engine (used for enacting workflows) that powers both Taverna Workbench (the desktop client application) and Taverna Server (which executes remote workflows). Taverna is also available as a Command Line Tool for faster execution of workflows from a terminal without the overhead of a GUI [1]. Although Taverna was conceived for bioinformatics, its user base also encompasses domains such as astronomy, digital preservation, biodiversity and virtual physiology. Taverna has been an open source project since 2003, developed by the myGrid consortium and originally led by the University of Manchester and EMBL-EBI. In October 2014, Taverna became an incubating project at the Apache Software Foundation [2].

Taverna Databundle is a file format returned by the Taverna Server containing the results of a Taverna workflow run and its intermediate values and provenance metadata. GSOC project proposes to create a web-based presentation of a workflow run.

1. Apache Taverna - Introduction. <http://taverna.incubator.apache.org/introduction/>

2. Apache Taverna: Sustaining research software at the Apache Software Foundation. https://github.com/stain/bosc2015-apachetaverna/blob/master/soilandreyes_bosc.md

Изучение генетических основ социального поведения на основе анализа генов селекционированных линий лис

В.И. Киргизова^{1,2}, Т.В. Андреева^{1,2}, М.А. Генаев², А.О. Брагин², А.П. Григоренко^{1,2,5}, Ф.Е. Гусев^{1,2}, К.В. Гунбин², П.С. Деменков², Д.А. Афонников², А.Ю. Гольцов¹, Н.И. Ершов², Н.А. Колчанов², Е.И. Рогаев^{1,2,5}

¹Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия (*iogen@vigg.ru*); ²Центр нейробиологии и нейрогенетики мозга, Институт цитологии и генетики СО РАН, Новосибирск, Россия;

³Department of Psychiatry, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, Worcester, USA
vitalina.kirgizova@gmail.com, тел. 8 (962) 913-50-00

Линии серебристо-черных лисиц (*Vulpes vulpes*), селекционированные на дружелюбное или агрессивно-враждебное поведение по отношению к человеку, являются уникальной моделью для изучения процессов доместикации и механизмов формирования различных типов социального поведения. Отбор проводится на протяжении 50 поколений на базе Института цитологии и генетики СО РАН в Новосибирске. С использованием платформы Illumina HiSeq2000 получены представительные библиотеки геномной ДНК клеток крови и мозга, и проведено глубокое секвенирование полных геномов с высоким покрытием ручной, агрессивной, неселекционированной лис и дикого типа лисы. Проведено аннотирование белок-кодирующих генов и отобраны гены-кандидаты с генетическими вариациями, по которым различаются ручные и агрессивные линии лисиц, которые затем были подтверждены стандартным методом секвенирования по Сэнгеру на расширенной выборке. Выявленные нами аминокислотные замены в ионных каналах, статистически специфичные только для агрессивных или ручных лисиц позволяют предположить сигнальные пути, вовлечённые в «эволюцию» социального поведения.

Работа поддерживается Правительством Российской Федерации (Грант № 14.В25.31.0033).

Исследование микробиоты полости рта у детей при терапии кариеса жевательной группы зубов

Климова Е.А.

Санкт-Петербургский государственный университет
Университетская наб., 7-9, Санкт-Петербург, 199034
e-mail: biberdent@gmail.com

Кариес в нашей стране является одним из самых распространённых заболеваний и проблема лечения данной патологии в жевательной группе зубов остается одной из актуальных проблем в стоматологии [Максимовский Ю.М., Митронин А.В., 2011; Максимовский Ю.М., 2003; Brown L.] et al, 2002, Кузьмина Э.М., 1998].

Установлено, что два тесно связанных вида стрептококков группы Mutans, названные *S. mutans* и *S. sobrinus*, взаимосвязаны с кариесом зубов человека. Их кислотопродуцирующие свойства и потенциал способности существования в кислотной среде напрямую связаны с кариесогенным потенциалом с кариесогенным потенциалом этих бактерий [Conrads G. et al., 2014].

Задачи исследования: рассмотреть существующие методы подхода к выбору стоматологических материалов; на основании молекулярно-генетических методов провести исследование микробиоты участков поражения и реставраций в зависимости от применяемых материалов; исследовать молекулярно-эпидемиологический профиль штаммов *S.mutans* у пациентов крупного промышленного города

Методы исследования включают клинические методы у пациентов 2х возрастных групп (определение стоматологического статуса, включающее экстраоральный и интраоральный осмотры, клинические тесты, включающие зондирование, перкуссию, температурную пробу, высушивание, окрашивание) и лабораторные - метод штриховых посевов, метод двухслойного агара, выделение геномной ДНК, полимеразная цепная реакция, электрофорез ДНК, геномное секвенирование.

Впервые получен блок новых данных, свидетельствующих об информативности определения активности показателей микробиоты и ротовой жидкости для оценки уровня биодеградации стоматологических материалов. Осуществлен молекулярно-генетический анализ штаммов *S.mutans* с целью выявления

актуальных штаммов, циркулирующих в Санкт-Петербурге, что является важным для последующих профилактических мероприятий. Сделаны предварительные выводы о том, что ряд реставрационных материалов изменяют показатели активности показателей биологической среды полости рта.

Совершенствование метода микроядерного тестирования для измерения генетической токсичности и исследования геномной стабильности

Левицкая Ольга Сергеевна, Крылов Николай Валерьевич,
Бодунков Николай Евгеньевич, Капырин Николай Игоревич

*Российский государственный аграрный университет – МСХА имени
К.А. Тимирязев ул. Тимирязевская, 49, Московский авиационный
институт Волоколамское шоссе, д. 4, Россия, Москва, oliavelevits@mail.ru*

В настоящее существует потребность оценки влияния факторов различной природы (как экзогенных: радиация, химические токсиканты, биологические агенты; так и эндогенных: старение, аутоимунные заболевания, наследственные болезни) на генетический аппарат организма. Наиболее надёжным и информативным для оценки генетической токсичности является метод микроядерного тестирования, однако сейчас он не может быть широко распространён и универсализирован, ввиду большого влияния человеческого фактора при проведении подсчета и относительно высокой трудоёмкости., а также отсутствия доступных и обширных баз данных, позволяющих сравнивать результаты и выводить закономерности.

Микроядра – это микроскопические структуры, находящиеся в цитоплазме клетки, их частота встречаемости возрастает пропорционально количеству хромосомных аберраций. Подсчет микроядер ведётся методом световой микроскопии с элементами механизации шаговыми двигателями и автоматизации распознавания при помощи технического зрения.

Цель: Создание не инвазивного, автоматического экспресс микроядерного теста для широкого внедрения в научную и медицинскую практику. Мониторинг среды на генетическую токсичность в высоком разрешении и исследование геномной стабильности.

В настоящее время исследования ведутся по следующим направлениям: поиск оптимальных методов окраски препаратов, обучение интеллектуальной системы распознава-

ния микроядер (техническое зрение), разработка механизма миниатюрной установки, исследование возможных методов идентификации микроядер в препаратах иными путями – биоэлектрическими, акустическими, фазово-контрастной микроскопии.

На данный момент идёт испытание рабочего прототипа и программного обеспечения для распознавания, основанное на стандартных для микроядерного теста принципах: при помощи световой микроскопии ведётся подсчёт микроядер в мазках крови, окрашенных по Романовскому-Гимза при увеличении 800х. Скорость подсчета при нашей методике уже превышает стандартную в ~10 раз. Так же исключение пропусков полей, сокращение фактора усталости и рассеянного внимания увеличивает достоверность метода.

Таким образом, рабочий прототип установки для микроядерного теста уже демонстрирует повышенную скорость и большую точность при обработке данных. Дальнейшее совершенствование может позволить вывести метод в широкую практику

Исследование ассоциации полиморфных вариантов генов цитокинов и факторов роста с бесплодием в супружеских парах

Т.А. Мараховская, Е.В. Машкина

Южный федеральный университет, Академия биологии и биотехнологии им. Дмитрия Иосифовича Ивановского, Просп. Стачки 194/1, г. Ростов-на-Дону, Россия, 344090.
tmarakhovskaya@mail.ru

Нарушение баланса провоспалительных и противовоспалительных цитокинов рассматривают как важное патогенетическое звено в развитии иммунологического бесплодия.

Целью работы было исследование частоты регистрации полиморфизма *-31С-Т* гена *IL-1b*, *-174G-С* гена *IL-6*, *-592С-А* и *-819С-Т* гена *IL-10*, *-308G-А* гена *TNF*, *-634G-С* гена *VEGFA*, *G915С* гена *TGFb1* среди мужчин и женщин из бесплодных супружеских пар для выявления значимых межгенных взаимодействий при помощи алгоритма снижения размерности (Multifactor Dimensionality Reduction, MDR).

Материалом для исследования послужили образцы ДНК мужчин и женщин из бесплодных супружеских пар (всего 69 образцов, из них 28 парных образцов). Контрольную группу составили 108 образцов ДНК фертильных мужчин и женщин, из них 32 парных образца.

Для супружеских пар показана значимость взаимодействия генов *IL-1b* и *IL-10* ($p=0.014$). Согласно данной модели повышенный риск развития бесплодия характерен для пар, в которых хотя бы у одного из супругов в генотипе присутствует полиморфизм либо *IL-1b*, либо *IL-10*.

Для мужчин из бесплодных супружеских пар была выявлена значимая модель взаимодействия четырех генов - *IL-1b*, *IL-10*, *VEGFA*, *TGFb1* ($p<0.0001$). Для мужчин гомозиготных по аллели *915G* гена *TGFb1*, аллели *-634G* гена *VEGFA*, а так же аллели *-592С* гена *IL-10* и при наличии аллели *-31Т* гена *IL-1b* характерен повышенный риск развития бесплодия.

При сравнении групп женщин из бесплодных пар и контроля определена оптимальная модель взаимодействия генов

IL-1b и *VEGFA* (($p=0.016$). Согласно данной модели пониженный риск развития бесплодия характерен для женщин со следующими генотипами: -31TT *IL-1b* / -634GG *VEGFA*; -31CT *IL-1b* / -634GC *VEGFA*; -31TT *IL-1b* / -634GC *VEGFA*.

Изучение генетического разнообразия земляники садовой (*Fragaria ananassa*) с использованием молекулярных маркеров SSR-типа

О.А. Межнина, О.Ю. Урбанович

Институт генетики и цитологии НАН Беларуси
Республика Беларусь, 220072, г. Минск, ул. Академическая, 27
e-mail: olgamezhnina@gmail.com

Данное исследование направлено на изучение генетического потенциала сортов земляники садовой, культивируемых в Беларуси, и разработку методов их ДНК-идентификации. Материалом для исследования служили 70 сортов земляники садовой, возделываемой в Республике Беларусь. Для изучения генетического разнообразия были подобраны SSR-маркеры, специфичные к геному рода *Fragaria*.

Для каждого маркера определялась длина аллелей и количество полиморфных фрагментов у каждого сорта. Максимальное количество аллелей было выявлено с помощью маркера UFFa3-D11. Количество обнаруженных в нем аллелей составило 9. Маркеры FG7ab и FG7cd выявили равное число аллелей – 7. Частота распространения аллелей в изученных локусах была разная. Значение PD, рассчитанное для каждого локуса, колебалось от 0,74 для локуса FG7ab до 0,87 для локуса UFFa3-D11. Среднее значение PIC для 3 SSR-локусов составило 0,75.

Данные о составе аллелей были использованы для построения дендрограммы филогенетического сходства сортов земляники садовой. В общей сложности при анализе 3 локусов микросателлитных последовательностей среди 70 сортов земляники садовой было выявлено 50 различных генотипов. Необходимо отметить, что сорта белорусской селекции не сформировали отдельного кластера и оказались близки по составу аллелей к иностранным сортам. Вероятно, это объясняется использованием в комбинациях скрещивания селекционных форм и образцов, генетически сходных с сортами иностранного происхождения. Для более детальной характеристики генотипов земляники садовой необходимо использовать дополнительные молекулярные маркеры и оценить их эффективность для целей ДНК-идентификации.

Изучение полиморфизма генов устойчивости класса NBS-LRR пшеницы с использованием NGS

А.А. Ненарокова

Московский государственный университет имени М.В.Ломоносова,
биологический факультет

119234, Россия, Москва, Ленинские горы, д. 1, стр. 12

e-mail: a.nenarokova@gmail.com, тел.: +7(963)765-58-67

Пшеница мягкая (*Triticum aestivum*) является важнейшим культурным растением. Геном *T. aestivum* обладает сложной структурой (это аллогексапloid, несущий 3 субгенома) и большими размерами (около 17 млрд п.н.), что сильно усложняет работу по секвенированию и аннотации генома этого растения. Несмотря на усилия Международного консорциума по секвенированию генома пшеницы (IWGSC), включающего более 1000 исследователей, проект по сборке генома *T. aestivum* до сих пор находится в draft-стадии. Это делает актуальным изучение с помощью секвенирования нового поколения (NGS) отдельных групп генов *T. aestivum*. Одной из таких групп являются гены семейства NBS-LRR – наиболее обширный и изученный класс генов устойчивости растений к патогенам (R-генов). Для их исследования используют метод NBS-профайлинга – амплификацию с использованием вырожденных праймеров, комплементарных консервативным областям NBS-домена. В данной работе впервые была исследована эффективность использования NBS-профайлинга в комбинации с последующим NGS для видов родов *Triticum* и *Aegilops* (ближайший род к *Triticum*). Показано, что при использовании этого метода для данных видов было получено значительное количество неспецифических последовательностей. Для обработки данных низкого качества такого рода была разработана и оптимизирована программная система. Было изучено разнообразие генов NBS-LRR у 176 образцов, относящихся к 10 видам рода *Aegilops* и различным сортам 23 видов рода *Triticum*. Было выявлено 446 полиморфных сайтов в 120 NBS-LRR-генах, несущих 525 вариантов замен. Было показано, что NBS-LRR-гены даже близких видов рода *Triticum* обладают значительными различиями.

Анализ лизосомных болезней накопления, связанных с мутациями

Д.А. Зубцов¹, С.П. Коваленко¹, О.М. Плотникова¹

¹ Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)»
141700, Россия, Московская обл., г. Долгопрудный, Институтский пер., д.9
e-mail: plotnikova@phystech.edu, +7(916)-245-25-19.

В данной работе рассмотрена связь наследственного аутосомно-рецессивного заболевания – болезни Краббе с нуклеотидной мутацией ДНК в гене GALC 14q хромосомы. В ходе работы были проанализированы геномы 1092 людей – участников проекта 1000 геномов. На основании анализа базы данных «human gene mutations database BIOBASE» и литературного анализа было выявлено 227 случаев однонуклеотидных мутаций. В результате работы была построена модель встречаемости болезни Краббе на примере популяции участников проекта 1000 геномов. Для оценки теоретической встречаемости пациентов, больных болезнью Краббе были использованы основы статистического анализа, а анализ проводился при помощи создания специальной программы.

Анализ, проведенный в данной работе, показывает, что частота встречаемости болезни составляет 1 больной на 12 000 человек, что в 8 раз превышает медицинские данные (1 на 100 000). Данное отличие может быть вызвано тем, что болезнь Краббе часто неправильно диагностируют, в связи с тем, что она имеет общие симптомы, при этом являясь редкой болезнью. В результате пациенты получают неправильное лечение, а также, в результате частых случаев ранней смертности (большинство детей, болеющих Краббе, умирают в возрасте до 2 лет) осложняется изучение данного заболевания.

Финансирование исследования: НИОКР «Разработка технологии и организация производства системы анализа генетических заболеваний с использованием массового параллельного секвенирования» Б Шифр «3.2-Генетика».

The method of analysis inherited interspersed and satellite repeats using genome sequencing data and application on genome sequencing data of DNA samples patients with schizophrenia family's trio

M.S. Protasova^{1,2}, F.E. Gusev^{1,2}, T.V. Andreeva^{1,2}, A.P. Grigorenko^{1,2}, E.I. Rogaev^{1,2,3}

¹ *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia, Gubkina, 3*

² *Center for Brain Neurobiology and Neurogenetics, Institute of Cytology and Genetics, RAS,*

Novosibirsk, Russia, Prospekt Lavrentyeva, 10

³ *Department of Psychiatry, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School,*

Worcester, USA

e-mail: maria2100@list.ru, tel. +7 (499)-135-43-14

Insertions of interspersed repeat into human genome are highly polymorphic and may alter gene expression, as they contain signaling sequences and change distance between exons; therefore they can participate in disease development. We developed a bio-informatical approach to identify repeat insertions into personal genomes using whole genome sequencing data. Our method starts with alignment of raw sequences to human reference genome (e.g. with BWA) and then reads (and read pairs) with abnormal mapping pattern are selected: 1) only fraction of a single read is mapped or 2) paired read is not mapped or mapped into a distant repeat location. Further, we identify the repeat class and breakpoint position using RepeatMasker, Bedtools and a handful of in-house scripts, which are organized into a pipeline. This approach is capable to locate insertions for broad spectrum of different repeat classes (SINE, LINE, LTR, DNA-transposons, satellite) and we used it for two family trios, each with schizophrenic offspring. Analysis revealed multiple inherited insertions of retrotransposons (further confirmed by PCR) in both parents and children; however, no de novo mutations were found. The method is convenient to use for WGS data and can identify both inherited and somatic variants.

This work was supported by the Russian Foundation for Basic Research (RFBR № № 14-04-01907 A).

Компьютерный анализ и обработка экспрессионных данных

А.М. Спицина¹, Ю.Л. Орлов²

¹Новосибирский государственный университет
Россия, 630090, г. Новосибирск, ул. Пирогова, 2

²Институт Цитологии и Генетики Сибирского отделения Российской академии наук

Россия, 630090, г. Новосибирск, пр-т Академика Лаврентьева, 10
e-mail: anastasia.spitsina@gmail.com

За последние годы в различных базах данных (BioGPS, GEO NCBI) был накоплен большой массив экспериментальных данных, полученных с помощью ДНК-микрочипов и имеющих большое значение для медицины и статистики. Эти данные, как правило, имеют значительный объем и требуют компьютерной обработки. Для работы с такими данными был разработан программный комплекс на языке C++. Он предназначен для статистического анализа и предобработки данных экспрессии, и имеет такие опции, как расчет и построение профилей тканеспецифичности, фильтрация генов по имеющейся информации о расположении на хромосомах, визуализация связи генов с помощью коэффициентов корреляции (как линейной, так и ранговой). Визуализация строится в форме генной сети с помощью разработанного скрипта на языке Java.

Было проведено исследование корреляций экспрессии генов в составе генных сетей циркадного ритма и регуляции холестерина, а также генов, отвечающих за агрессивное поведение мышей. Полученные с помощью программы результаты были проанализированы и сравнены с имеющейся в базе данных STRING информацией о совместной экспрессии генов. Проведена реконструкция генных сетей и визуализация связей генов из рассмотренных выборок.

В работе были использованы данные микрочипов Affymetrix для человека, мыши *Mus musculus* и крысы *Rattus norvegicus*, а также данные RNA-Seq. По результатам, полученным с помощью программы, был проведен сравнительный анализ экспрессионных данных и выявлены структурные особенности генов с высокой экспрессией.

Данная работа была выполнена при поддержке гранта РФФИ №14-04-01906 и бюджетного проекта ИЦиГ СО РАН VI.61.1.2.

Использование технологии RNA-seq для анализа некодирующих РНК гена *SGMS1* человека

И.Б. Филиппенков, Е.О. Калиниченко, О.Ю. Сударкина, Л.В. Дергунова

*Федеральное государственное бюджетное учреждение науки
Институт молекулярной генетики Российской академии наук
пл. Академика Курчатова, 2, Москва, 123182
e-mail: Filippenkov@img.ras.ru, тел. 7 (499)196-1858, факс 7
(499)196-0221*

Современные методы исследования транскриптома показали, что многие РНК транскрибируются с межгенных и интронных регионов генома, не кодируют белок, но участвуют в функционировании генов. Для направленного изучения транскрипции отдельных участков ДНК используется технология высокопроизводительного секвенирования «RNA CaptureSeq», основанная на обогащении РНК искомыми транскриптами. Наше исследование направлено на изучение структурно-функциональной организации жизненно важного гена сфингомиелинсинтазы 1 (*SGMS1*) человека, длина которого составляет 320 тпн. Более 98% нуклеотидной последовательности гена приходится на интроны. Ранее нами было показано, что в разнообразии транскриптов гена *SGMS1* участвуют альтернативный сплайсинг, внутренние промоторы, локализованные в интронах, и альтернативное интронное полиаденилирование. В настоящем исследовании для глубокого анализа транскрипции гена *SGMS1* человека и поиска некодирующих РНК, участвующих в его функционировании, применен метод «RNA CaptureSeq». Для обогащения образцов РНК лобной коры и крови человека использовали зонды Agilent SureSelect, специфичные последовательности гена *SGMS1*. В результате работы выявлено большое количество интронных транскриптов гена *SGMS1*. В настоящее время ведется анализ результатов секвенирования и поиск новых транскриптов гена *SGMS1*.

В целях полнотранскриптомного анализа экспрессии генома в изучаемых тканях в норме и условиях патологии пла-

нируется использование технологии RNA-seq с последующим применением методов биоинформатического анализа.

Работа получила финансовую поддержку Программы Российской академии наук «Молекулярная и клеточная биология» и Программы поддержки ведущих научных школ.