

ОТЗЫВ

официального оппонента на диссертацию Богомолова Антона Геннадьевича на тему: «Разработка метода визуализации хромосомоспецифичных последовательностей ДНК при проведении FISH», представленной на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Актуальность избранной темы. В настоящее время, методы анализа хромосом на основе флуоресцентной гибридизации *in situ* (FISH) находят широкое применение в молекулярно-цитогенетической диагностике хромосомных аномалий и в сравнительном анализе хромосом различных видов. Трудности вызывает наличие в геномах эукариот, в т.ч. человека, большого количества диспергированных повторяющихся последовательностей ДНК, продуцирующих при гибридизации *in situ* интенсивный флуоресцентный сигнал, который должен быть подавлен для успешной идентификации материала хромосом и хромосомных районов. Однако, использование супрессионной гибридизации *in situ* может оказаться неэффективным по причине большого размера генома (наличия большого количества повторяющихся последовательностей ДНК в нем) или отсутствия необходимого количества Cot-1 ДНК. Таким образом, наиболее актуальным методом становится компьютерный анализ экспериментальных изображений.

Данная работа служит ярким примером грамотного подхода к разработке и использованию нового метода компьютерного анализа изображений многоцветной FISH. Метод открывает новые возможности в исследовании хромосом тех организмов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть выполнена.

Научная новизна исследования, полученных результатов, выводов и рекомендаций, сформулированных в диссертации. К наиболее важным научным

результатам исследования, характеризующим его новизну, могут быть отнесены:

- Разработан и апробирован оригинальный компьютерный метод VISSIS (visualization of specific signal in silico), который позволяет выделить сигналы хромосомспецифичных последовательностей при проведении многоцветной флуоресцентной гибридизации *in situ* без супрессии гибридизации повторяющихся последовательностей.
- Впервые была решена задача идентификации материала целых хромосом и крупных хромосомных районов у 19 видов саранчовых с помощью гибридизации *in situ* ДНК-проб, полученных из целых хромосом или хромосомных районов.
- На примере анализа FISH-изображений с ДНК-пробами из хромосом человека 18 и 19, контрастных по содержанию длинных и коротких диспергированных повторов, продемонстрирована возможность анализировать закономерности распределения повторяющихся последовательностей, а также получить данные о возможной локализации районов, богатых генами.

Значимость для науки и практики полученных автором результатов. Теоретическая значимость исследования состоит в том, что:

- впервые установлена гомеология крупных С-негативных районов хромосом в геномах девятнадцати видов саранчовых семейства Pamphagidae с большим количеством повторяющихся последовательностей.
- установлено, что эффективность применения метода VISSIS для анализа результатов FISH зависит от различий в содержании разных типов повторяющихся последовательностей ДНК в хромосомах, из которых были получены ДНК-пробы. FISH с парами ДНК-проб из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК, позволяет достичь сопоставимых с CISS-гибридизацией результатов.

- обнаружена отрицательная корреляция между отношением сигнал/шум на обработанных изображениях и отношением содержания диспергированных повторов SINE/LINE в хромосомах, из которых были получены ДНК-пробы.

Практическая значимость исследования заключается в том, что разработанный метод VISSIS может заменить стадию супрессии гибридизации повторяющихся последовательностей при проведении двухцветной FISH или улучшить ее результаты в случаях неполной супрессии повторяющихся последовательностей. Перспективным также является использование метода VISSIS в ходе молекулярно-цитогенетических исследований хромосом тех организмов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть выполнена.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации. Работа Богомолова А.Г. построена по общепринятой схеме и состоит из введения, обзора литературы, двух глав, описывающих материалы, методы, результаты собственных исследований и их обсуждение, а также заключения, выводов и списка литературы. Список литературы включает в себя 189 источников, из них только 10 на русском языке, остальные 179 на английском языке, что говорит о хорошем владении автором зарубежной литературой. Из общего числа 16 источников или примерно 8,5% относятся к последним 5 годам (2015-2019 гг.). Работа изложена на 139 страницах и содержит 12 таблиц и 34 рисунков.

Структура и логика изложения соответствуют поставленным в диссертации задачам исследования, нацеленным на создание компьютерного метода визуализации сигналов, продуцируемых хромосомоспецифичными последовательностями ДНК, при проведении двухцветной FISH районо- и хромосомоспецифичных ДНК-проб, а также оценку влияния особенностей состава повторяющихся последовательностей в хромосомах на результаты компьютерной обработки с использованием данного метода:

1. Разработать метод и программное обеспечение для визуализации и анализа сигналов хромосомоспецифичных последовательностей ДНК, полученных при проведении двухцветной FISH ДНК-проб из целых хромосом.
2. Апробировать метод на изображениях хромосом человека, полученных после проведения двухцветной FISH различных хромосомоспецифичных ДНК-проб с супрессией и без супрессии гибридизации повторяющихся последовательностей.
3. Провести сравнительный анализ результатов компьютерной обработки созданным методом с результатами применения CISS-гибридизации.
4. Оценить зависимость результатов компьютерной обработки от количественного содержания разных типов диспергированных повторов в хромосомах, из которых были получены ДНК-пробы.
5. Апробировать разработанный метод для идентификации материала целы хромосом и крупных хромосомных районов у видов (описторхид и саранчовых), геном которых в несколько раз отличается от генома человека по размеру.

Для решения поставленных задач автор опирается на обширную теоретическую и методологическую базу. Во введении обоснована актуальность темы исследования, сформулированы цель и задачи исследования, охарактеризованы научная новизна, теоретическая и практическая ценность исследования, сформулированы два положения, выносимые на защиту, содержащие наиболее существенные результаты работы, обладающие научной новизной.

Обзор литературы (стр. 13) очень подробен и обстоятелен, охватывает как описание типов ДНК-проб и их применения (стр. 13-14), схему проведения использующих FISH исследований (стр. 14-21), проблему повторяющихся последовательностей ДНК (стр. 21-24), методы экспериментального улучшения сигнала (стр. 24), методы подавления гибридизации повторяющихся последовательностей (стр. 24-28), методы для уникальных последовательностей (стр. 28-30), метод элиминации повторов из ДНК-клонотек (стр. 30-34), особенности анализа видов с известной последовательностью генома (стр. 34-37), так и описание

особенностей получения FISH-изображений (стр. 37-39), методов предобработки (стр. 39-40), сегментации (стр. 41-51), классификации объектов (стр. 52-54), а также описание предложенного в литературе компьютерного метода и его недостатков (стр. 55-56). Приводится собственное заключение (стр. 56-57).

В Главе 2 (стр. 58) описана разработка компьютерного метода визуализации хромосомоспецифичных последовательностей ДНК. Рассмотрена общая схема метода (стр. 62), а также его этапы, такие как сегментация изображений (стр. 62), калибровка интенсивностей сигналов (стр. 66), классификация объектов (стр. 69), выделение специфичного сигнала (стр. 79). На вход алгоритма подаются цветные изображения, в которых два канала содержат результаты FISH ДНК-проб, полученных из разных индивидуальных хромосом или хромосомных районов, а третий канал – результаты DAPI-окрашивания. В основе метода лежит допущение, что сигналы, обусловленные повторяющимися последовательностями, присутствующими в обеих ДНК-пробах, должны иметь близкие значения интенсивности. Сигнал хромосомоспецифичных последовательностей оценивается как разница между суммарным сигналом и суммой неспецифичного и фонового сигналов, которые оцениваются по результатам анализа FISH-изображения с другой ДНК-пробой. В разделе, посвященном постобработке изображений (стр. 80), рассмотрены ошибки классификации первого и второго рода. Для применения метода разработанная программа визуализации хромосомоспецифичных последовательностей VisualCS (стр. 82).

Главе 3 (стр. 85) посвящена применению разработанного метода. Описано получение предоставленных автору коллегами оригинальных изображений результатов FISH хромосомоспецифичных ДНК-проб с хромосомами человека (стр. 85), описторхид (стр. 86), саранчовых (стр. 87), которые существенно различаются по размеру генома и по доле повторяющихся последовательностей в нем. Разработанный метод позволил успешно идентифицировать материал целых хромосом и отдельных хромосомных районов, в том числе у саранчовых, для которых ранее эту задачу решить не удавалось. В случае неполной супрес-

сии гибридизации повторяющихся последовательностей в изображениях целых хромосом у человека метод позволил значительно повысить относительный уровень интенсивности сигнала с хромосомоспецифичных последовательностей ДНК (стр. 89). Приведена таблица с рекомендациями по выбору пар хромосомоспецифичных ДНК-проб для достижения максимальной элиминации сигнала повторяющихся диспергированных последовательностей ДНК посредством компьютерной обработки FISH-изображений методом VISSIS (стр. 103). Сравнение результатов CISS-гибридизации и компьютерной обработки FISH-изображений (стр. 105) показало, что компьютерная обработка дает сопоставимые результаты в случаях, когда ДНК-пробы созданы из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК. Важное отличие метода VISSIS от метода RENS, предложенного ранее, заключается в том, что он не зависит от субъективного мнения исследователя (стр. 108). Применение CISS-гибридизации для хромосомного анализа описторхид затруднено ввиду отсутствия достаточного количества исходного материала для получения фракции высокоповторяющейся ДНК. Компьютерная обработка дает возможность идентифицировать материалы соответствующих хромосом (стр. 109). Применение метода VISSIS позволяет идентифицировать материал хромосом и крупных хромосомных районов у саранчовых, а эффективность метода может быть сопоставима с результатами CISS-гибридизации на хромосомах человека (стр. 112). В заключении диссертационной работы (стр. 117) кратко обобщены основные результаты исследования. По результатам работы сформулированы обоснованные выводы (стр. 120).

Таким образом, на основе достаточного анализа предметной области, адекватной постановки научной проблемы и задач исследования, корректного применения современных подходов и методов получены вполне достоверные и обоснованные результаты.

Замечания по содержанию и оформлению диссертации.

В обзоре литературы имеется неточность в употреблении терминов. Например, на стр 19 автор пишет «плотного гибридного сигнала», но о какой плотности идет речь непонятно.

На стр. 38 автор утверждает, что «обычно ... выбирают 8-битные числа, допускающие диапазон значений от 0 до 255», однако, такое представление является лишь самым простым, разрешения в 12 и 16 бит не редкость. Далее утверждается, что «интенсивность ... коррелирует с количеством последовательностей ДНК». Если это так, то следует сначала изучить природу этой корреляции, ее силу и статистическую значимость. Обычным допущением, однако является, что интенсивность пропорциональна количеству проб.

В Таблице 1.1 на стр. 44 автор приводит точность метода в %. Однако, остается непонятным, какая это точность – средняя, максимальная или какая-то еще. То же относится к таб. 1.2, 1.3, 1.4. Для характеристики метода следует использовать среднюю точность по набору данных, с обязательным указанием среднеквадратичного отклонения. При этом сравнение требуется проводить по соответствующему статистическому критерию для средних.

На стр. 58 автору следовало пояснить является ли фон в формуле 2.1 случайным.

На стр. 63, рис 2.4 не описано для какого из указанных на стр. 62 методов изображено T.

На стр 63 требуется пояснить медианный фильтр с каким структурным элементом используется.

На стр. 68, формула 2.7 автор предлагает использовать среднюю интенсивность, однако, медиана представляется лучшим кандидатом. Было бы полезно привести соответствующую гистограмму для пояснения.

На стр. 70, рис 2.7 должен быть указан угол бета.

На стр 71 автор пишет: «Значения Rip сортируются по модулю в порядке убывания: $|R3| \leq |R2| \leq |R1|$ », однако это порядок возрастания, когда следующее больше предыдущего.

На стр 76 автор пишет «... $w3$ равным 0.36 (рисунок 2.10)», что неверно, надо 3.6.

Использованный автором метод подбора весов требует более четкого обоснования. При подборе $w1$ принимается $w3=1$, которое потом становится 3.6, однако найденное $w1=0.07$ не меняется. Естественно было бы продолжить последовательный подбор до достижения сходимости, т. е. пока изменения весов не станут незначительными.

На стр. 76 автор предлагает воспользоваться критерием Ирвина, что следовало бы разъяснить подробнее (напр., см. обсуждение <http://arhiuch.ru/st7.html>):

1. Как проверялась предполагаемая нормальность выборки.
2. Как влияет на точность результатов то, что критерий предполагает использование среднеквадратичного отклонения генеральной выборки, а применение на практике выборочного СКО не подтверждено. Кроме того, выборочное СКО для 4 объектов неточно.
3. Для разных распределений – разные таблицы уровней. Какую таблицу уровней автор предлагает применять и почему?
4. Рассматривались ли другие критерии для выбросов?

На стр. 84 следовало пояснить по какой причине в списке операционных систем нет Windows 8 и 8.1, а также Windows сервер разных версий. Также автору следовало указать под какими лицензиями распространяются разработанные программы и где и как их можно скачать.

На стр. 98, таблица 3.2 некоторые ошибки равны в точности 0, следовало прокомментировать это в тексте.

Также замечены опечатки:

Стр 4 Cot-1 ДНК – фракци~~юя~~ ДНК

Стр 21 может использовать — может быть использован

Стр 31 ДНК-коне~~о~~теки

Стр 34 Данные секвенирования открыва~~ю~~т

Стр 35 последовательностей в интересующем~~го~~ исследователя

Стр 38 Цветное изображение является двумерным~~й~~ массивом векторов: (каждому

Стр 68 выравнивание уровней интенсивностей сигналов может быть достигну~~тоа~~

Отмеченные недостатки, впрочем, не являются критическими, не умаляют ценности работы и не влияют на общую положительную оценку диссертационной работы Богомолова А.Г.

Заключение о соответствии диссертации критериям, установленным Положением о порядке присуждения ученых степеней. Диссертационное исследование Богомолова А.Г. выполнено на актуальную тему, представляет собой законченную научную работу, имеет теоретическую и практическую ценность.

Основные результаты диссертации изложены в 16 научных работах, из них шесть в рецензируемых научных журналах, в том числе пять – во входящих в перечень ВАК по специальности 03.01.09 и десять работ в сборниках тезисов конференций, апробированы научной общественностью и получили положительную оценку. Получено два авторских свидетельства.

Автореферат и публикации соискателя отражают основное содержание диссертации.

Таким образом, диссертация Богомолова Антона Геннадьевича является научно-квалификационной работой, характеризуется новизной, содержит решение актуальной задачи, имеющей теоретическое и практическое значение, что соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к диссертациям на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика, а ее автор заслуживает присуждения искомой ученой степени.

Ведущий научный сотрудник
НИЛ «Математическая биология и биоинформатика», ИПММ, федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого», кандидат биологических наук



К.Н. Козлов

Козлов Константин Николаевич
195251, Санкт-Петербург, Политехническая ул., д. 29, кор. НИК (АФ), пом. А.3.28,

Телефон: +7-812-290-9642, +7-904-518-8581

Адрес электронной почты: kozlov_kn@spbstu.ru

Место и адрес работы: НИЛ «Математическая биология и биоинформатика»
ИПММ ФГАОУ ВО «СПбПУ»

Должность: в.н.с. НИЛ «Математическая биология и биоинформатика» ИПММ

Вх 2171/40
29.04.2019

