

На правах рукописи

Богомоллов Антон Геннадьевич

**РАЗРАБОТКА МЕТОДА ВИЗУАЛИЗАЦИИ
ХРОМОСОМОСПЕЦИФИЧНЫХ
ПОСЛЕДОВАТЕЛЬНОСТЕЙ
ДНК ПРИ ПРОВЕДЕНИИ FISH**

03.01.09 Математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание учёной степени

кандидата биологических наук

Новосибирск – 2019

Работа выполнена в ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» в лаборатории морфологии и функции клеточных структур, г. Новосибирск, Россия

Научный руководитель **Рубцов Николай Борисович**
д.б.н., профессор., г.н.с. лаб. морфологии и функции клеточных структур ФГБНУ ФИЦ «Институт цитологии и генетики СО РАН»

Официальные оппоненты **Бажан Сергей Иванович**
д.б.н., доцент, заведующий теоретическим отделом ФБУН ГНЦ ВБ «Вектор», р.п. Кольцово, Новосибирская область.

Козлов Константин Николаевич
к.б.н., ведущий научный сотрудник научно-исследовательской лаборатории Математической биологии и биоинформатики Института прикладной математики и механики ФГАОУ ВО «СПбПУ», г. Санкт-Петербург.

Ведущее учреждение: ФГБУН Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук (ИМКБ СО РАН), г. Новосибирск

Защита диссертации состоится «__» _____ 2019 г. на утреннем заседании диссертационного совета Д 003.011.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН» в конференц-зале Института по адресу:

пр. ак. Лаврентьева 10, г. Новосибирск, 630090,

тел +7 (383) 3634906, факс +7(383) 3331278.

e-mail: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института: www.bionet.nsc.ru.

Автореферат разослан «__» _____ 2019 г.

Ученый секретарь
диссертационного совета,
доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Новые методы хромосомного анализа, основанные на флуоресцентной гибридизации *in situ* (FISH), заметно изменили представление о строении и структурно-функциональной организации генома. Выявление и идентификация гомологии протяженных участков хромосом в настоящее время являются актуальной задачей как при проведении молекулярно-цитогенетической диагностики хромосомных аномалий, так и в исследованиях, посвященных сравнительному анализу хромосом различных видов (Jensen, 2014; Riegel, 2014; Shakoori, 2017). Но решение этих задач с применением FISH осложнено присутствием в геномах большинства видов эукариот значительного количества диспергированных повторяющихся последовательностей ДНК, которые создают препятствие для выявления сигналов, обусловленных хромосомо- и районспецифичными последовательностями ДНК (Dorman et al., 2013; Trifonov et al., 2017). В этой связи становится актуальной разработка методов, позволяющих четко дифференцировать специфичные сигналы последовательностей ДНК при проведении FISH.

Для молекулярно-цитогенетического анализа хромосом млекопитающих эта проблема решается, в основном, с помощью супрессионной гибридизации *in situ* (Chromosome In Situ Suppression hybridization - CISS-гибридизация). Метод заключается в предварительном отжиге ДНК проб в присутствии избытка ДНК из фракции повторяющихся последовательностей (Cot-1 ДНК). К сожалению, этот метод оказался недостаточно эффективным при исследовании хромосом видов, размер генома которых в несколько раз превышает размер генома человека. Геном таких видов содержит большое количество повторяющихся последовательностей, что не позволяет провести полную супрессию их гибридизации *in situ* (Houben et al., 2001; Schubert et al., 2001; Han et al., 2015). Применение CISS-гибридизации также оказалось затруднительным в случае отсутствия источника для получения необходимого количества Cot-1 ДНК (Trifonov et al., 2017). Кроме того, CISS-гибридизация не позволяет выявлять хромосомоспецифичные кластеры повторов из-за супрессии гибридизации повторяющихся последовательностей ДНК.

Перечисленная совокупность проблем, возникающих при идентификации и анализе индивидуальных хромосом и хромосомных районов методом FISH с использованием ДНК-проб, полученных из целых хромосом или протяженных хромосомных районов, делает актуальными разработку и

реализацию новых подходов и методов для анализа результатов FISH.

Цели и задачи диссертационной работы. Целью работы является создание компьютерного метода визуализации сигналов, продуцируемых хромосомоспецифичными последовательностями ДНК, при проведении двухцветной FISH районо- и хромосомоспецифичных ДНК-проб, а также оценка влияния особенностей состава повторяющихся последовательностей в хромосомах на результаты компьютерной обработки с использованием данного метода.

Для достижения цели был поставлен и решён ряд задач:

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 целых хромосом и крупных хромосомных районов у видов (описторхид и саранчовых), геном которых в несколько раз отличается от генома человека по размеру.

Научная новизна работы. Разработан оригинальный компьютерный метод VISSIS (visualization of specific signal *in silico*), который позволяет выделить сигналы хромосомоспецифичных последовательностей при проведении многоцветной флуоресцентной гибридизации *in situ* без супрессии гибридизации повторяющихся последовательностей. Метод открывает новые возможности в исследовании хромосом тех организмов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть выполнена. С помощью разработанного метода впервые проведена идентификация гомологичных хромосом и крупных хромосомных районов у 19 близкородственных видов саранчовых с

помощью гибридизации *in situ* ДНК-проб, полученных из целых хромосом или хромосомных районов.

Теоретическая и практическая значимость работы. Разработанный метод VISSIS может заменить стадию супрессии гибридизации повторяющихся последовательностей при проведении двухцветной FISH или улучшить ее результаты в случаях неполной супрессии повторяющихся последовательностей. Это позволяет сократить объем используемой Cot-1 ДНК в ходе проведения FISH-диагностики хромосомных патологий человека и тем самым снизить стоимость проведения анализа. Кроме того, метод расширяет возможности применения FISH в исследованиях по сравнительной цитогенетике и геномике. Особенно перспективным является использование метода в ходе молекулярно-цитогенетических исследований хромосом тех организмов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть выполнена.

Положения, выносимые на защиту

1. Разработан компьютерный метод выявления хромосоמו- и районспецифичных последовательностей ДНК в эухроматиновых районах хромосом при флуоресцентной гибридизации *in situ* ДНК-проб из целых хромосом или хромосомных районов без супрессии повторяющихся последовательностей.
2. В результате использования данного метода впервые установлена гомеология крупных С-негативных районов хромосом в геномах девятнадцати видов саранчовых семейства *Pamphagidae* с большим количеством повторяющихся последовательностей.

Апробация работы. Материалы работы вошли в отчет по гранту Российского фонда фундаментальных исследований (№ 16-34-00498 мол_а; руководитель проекта Богомолов А.Г.). Результаты данного исследования были представлены в виде устных и стендовых докладов на 8 научных конференциях: международная конференция, посвященная 100-летию со дня рождения академика АН СССР Д.К. Беляева (2017, г Новосибирск), международные конференции по биоинформатике, структуре и регуляции генома (BGRS 2010, BGRS 2012, BGRS 2014, BGRS 2016, BGRS 2018, г. Новосибирск), Международная научная конференция Научного парка СПбГУ "Трансляционная биомедицина: современные методы междисциплинарных исследований в аспекте внедрения в практическую

медицину" (2016, г. Санкт Петербург), 19-я международная хромосомная конференция (2013, г. Болонья, Италия).

Объём и структура диссертации. Диссертация изложена на 139 страницах машинописного текста, содержит 34 рисунка и 12 таблиц. Список литературы включает 189 ссылок. Диссертация состоит из введения, литературного обзора, описания разработанного метода, описания результатов его апробации, заключения, выводов и списка литературных источников.

Публикации. По теме диссертации опубликовано 16 работ, из них 6 статей в рецензируемых научных журналах (в том числе 5 статей, опубликованных в журналах, входящих в перечень ВАК по специальности 03.01.09) и 10 тезисов конференций. Получено два авторских свидетельства.

Личный вклад автора. Основные результаты, изложенные в диссертации, получены автором самостоятельно: разработан компьютерный метод, проведена его апробация на изображениях FISH ДНК-проб с хромосомами человека, двух видов описторхид и 19-ти видов саранчовых, проанализированы результаты вычислительных экспериментов. Изображения с результатами флуоресцентной гибридизации *in situ* были предоставлены Задесенец К.С., Карамышевой Т.В, Джетыбаевым И.Е. и обработаны автором диссертации.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

Обзор литературы состоит из четырех разделов и заключения. Первый раздел содержит описание метода FISH и связанных с ним понятий (ДНК-проба, флуорохром, псевдоцвет и т.д.). Также в нем описывается проблема выявления специфичного сигнала последовательностей ДНК, характерных для отдельных хромосомных районов или целых хромосом. В разделе 1.2 дается характеристика уникальных и повторяющихся последовательностей ДНК, а также их классификация у эукариот. Раздел 1.3 дает обзор существующих методов улучшения выявления специфичного FISH-сигнала ДНК-проб, полученных из целых хромосом или хромосомных районов. Последний раздел посвящен описанию основных терминов цифровой обработки изображений, стандартных задач обработки FISH-изображений хромосом и методов их решения, а также варианта определения сигналов хромосомспецифичных последовательностей ДНК с помощью компьютерной обработки.

Глава 2. Компьютерный метод визуализации хромосомоспецифичных последовательностей ДНК (метод *Visualization of specific signal in silico* – VISSIS)

Глава посвящена описанию разработанного метода VISSIS (Богомолов и др., 2012). В основе метода лежит допущение, что сигналы, обусловленные повторяющимися последовательностями, присутствующими в парах ДНК-проб (далее **неспецифичный сигнал**), и фоновые сигналы должны иметь близкие значения интенсивности. Тогда интенсивности этих сигналов можно оценить по результатам гибридизации с другой ДНК-пробой. А интенсивность сигнала хромосомоспецифичных последовательностей (далее **специфичный сигнал**) оценивается как разница между сигналом ДНК-пробы в пикселе изображения и суммой неспецифичного и фонового сигналов. Общая схема метода проиллюстрирована на рисунке 1.

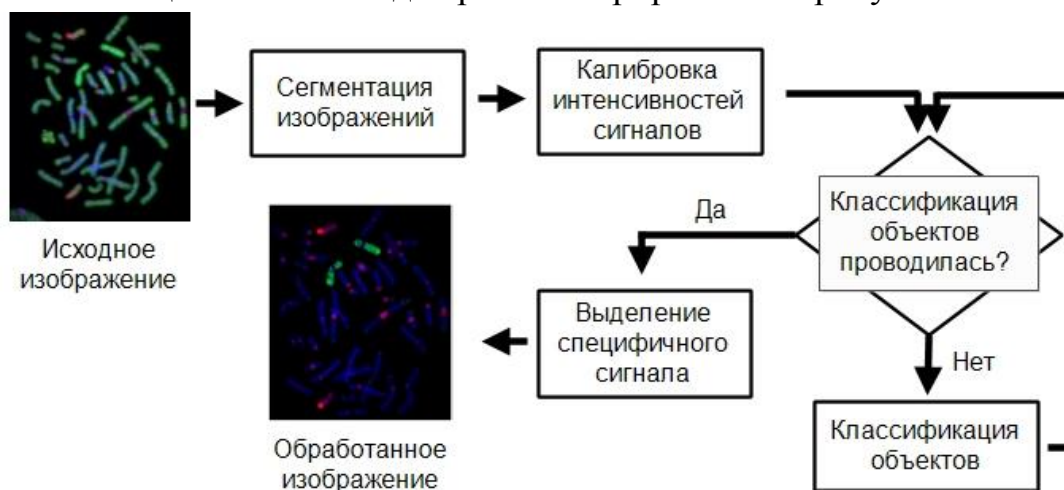


Рисунок 1. Основные этапы разработанного метода.

Выделение на изображении участков, занятых хромосомами, выполняется в несколько этапов. Вначале проводится пороговая сегментация изображения (Otsu, 1979; Calvard, Ridler, 1978; Zack et al., 1977; Каруга et al., 1985). После этого ложные объекты (ядра, различные артефакты) исключаются из обработки. Заключительным этапом сегментации является разделение кластеров хромосом на отдельные хромосомы с помощью фильтра средней разности (Poon et al., 1999) или метода водораздела (Wenzhong, Xiaohui, 2010).

Интенсивности неспецифичных сигналов в пикселе изображения могут не совпадать (Rens et al., 2006). Можно выделить две основные причины возникновения различий: особенности состава повторяющихся последовательностей ДНК в хромосомах, из которых получены ДНК-пробы;

особенности регистрации изображений (разное время экспозиции, характеристики флуорохромов и т.д.).

Разная представленность типов повторяющихся последовательностей в исходных хромосомах приводит к окрашиванию хромосомных районов с различной интенсивностью разными ДНК пробами и, как следствие, сохранению неспецифичного сигнала после компьютерной обработки. Особенную трудность для компьютерного анализа вызывают области с большой разницей интенсивности сигналов в пикселях (далее **яркие области**). В случае наличия на изображении таких областей следует проводить их отдельную компьютерную обработку. Сегментация ярких областей проводится путем задания интервалов, ассоциируемых с ними, на гистограмме отношений интенсивностей сигналов ДНК-проб.

Выравнивание уровней интенсивностей сигналов достигается путем умножения значения интенсивности сигналов одной из ДНК-проб на поправочный коэффициент k_p . Коэффициент вычисляется по формуле:

$$k_p = \frac{\sum_{p \in \Omega} i_a(p) i_b(p)}{\sum_{p \in \Omega} i_b(p)^2} \quad (1)$$

где k_p – поправочный коэффициент; Ω - все множество пикселей, принадлежащих объектам класса S_{ns} (объекты, которые не содержат ДНК, специфичную хромосомам a и b), за исключением пикселей из ярких областей; $i_a(p)$ - интенсивность сигнала в пикселе p с ДНК-пробы, полученной из хромосомы a ; $i_b(p)$ - интенсивность сигнала в пикселе p с ДНК-пробы, полученной из хромосомы b .

Метод распознает на изображении три класса объектов: классы S_a и S_b включают объекты, содержащие ДНК, специфичную для хромосом a и b , из которых получены ДНК-пробы, класс S_{ns} включает другие объекты, в пикселях которых зарегистрированы только неспецифичный и фоновый сигналы. Информация о результатах классификации используется для корректной оценки поправочного коэффициента k_p и на этапе постобработки изображений. В качестве характеристик объектов для классификации используются значения логарифмов отношения средних интенсивностей сигналов от разных флуорохромов в различных областях объекта (каждый объект разделяется на три области). Объекты одного класса в пространстве признаков хорошо отделяются от других классов (рисунок 2). Поэтому классификация объектов осуществляется с помощью метода ближайшего соседа (Sampat et al., 2004).

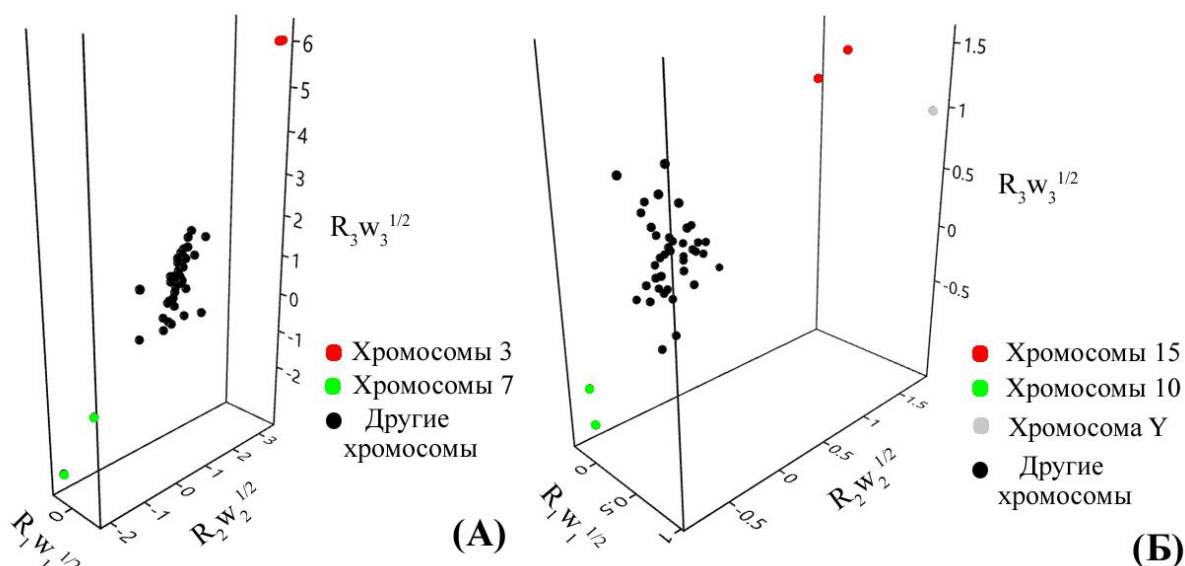


Рисунок 2. Пример расположения объектов в пространстве признаков. Каждая точка соответствует объекту, а по осям откладываются значения признаков объекта с учетом их значимости ($w_1=0.07$; $w_2=1$; $w_3=3.6$). **А** – Расположение объектов, обнаруженных на изображении CISS-гибридизация ДНК-проб из хромосом 3 и 7 человека. **Б** – Расположение объектов, обнаруженных на изображении FISH ДНК-проб из хромосом 10 и 15 человека.

Классификатор вычисляет евклидовы расстояния между вектором признаков неизвестного (еще не классифицированного) объекта и векторами прототипов классов. Объект относится к тому классу, прототип которого оказывается ближайшим. Класс C_{ns} представлен вектором, у которого значение каждого признака равно нулю. Для других классов первоначально в качестве прототипов выбираются объекты с минимальной и максимальной суммой значений признаков. Затем проводится повторная классификация объектов, но в качестве прототипов классов C_a и C_b устанавливаются вектора со средними значениями характеристик ранее отнесенных к ним объектов.

Заключительным этапом компьютерной обработки является визуализация специфичного сигнала. Оценка интенсивности специфичного сигнала проводится с учетом значений интенсивности в соседних пикселях. Полученный специфичный сигнал является слабо интенсивным и для лучшей его визуализации применяется линейное контрастирование.

Разработанный метод VISSIS реализован в программе VisualCS (“Visualization Chromosome-specific Signals”, авторское свидетельство №2018662647, выдано 12.10.18).

Глава 3. Применение компьютерного метода визуализации хромосомоспецифичных последовательностей

В ходе апробации разработанный метод был применен для обработки набора изображений мейотических и митотических хромосом видов, существенно различающихся по размеру генома и доле повторяющихся последовательностей в геноме (описторхиды, млекопитающие, саранчовые).

Первоначально метод был успешно протестирован на изображениях FISH ДНК-проб из целых хромосом человека с супрессией гибридизации повторяющихся последовательностей (Богомолов и др., 2012; Гайнер и др., 2017). Указанные изображения очень удобны для тестирования, так как позволяют однозначно определить класс объектов из-за низкого уровня интенсивности сигналов диспергированных повторяющихся последовательностей ДНК. Обработанные изображения визуально не сильно отличались от исходных в случаях полной супрессии гибридизации повторяющихся последовательностей. Но при проведении CISS-гибридизации с недостаточной супрессией повторяющихся последовательностей, дополнительное применение метода VISSIS значительно улучшало выявление специфичного сигнала. Такая комбинация методов может использоваться для снижения затрат на проведения диагностики хромосомных патологий человека.

Дальнейшие исследования, основанные на цифровом анализе нескольких сотен изображений, подтвердили возможность идентифицировать хромосомы человека без супрессии гибридизации повторяющихся последовательностей (рисунок 3).

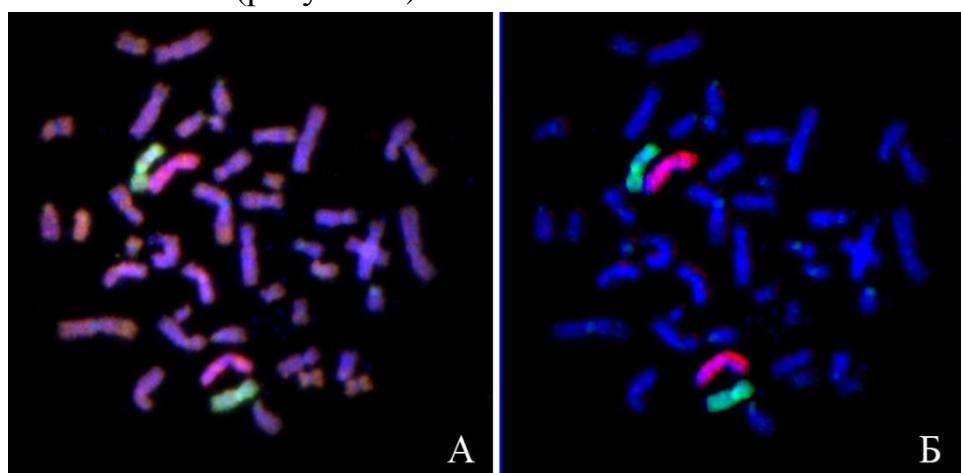


Рисунок 3. FISH на хромосомах человека без супрессии гибридизации повторяющихся последовательностей с ДНК-пробами из хромосом человека 7 (зеленый) и 3 (красный). А – Необработанное исходное изображение. Б – Результат применения метода VISSIS.

Однако, эффективность элиминирования сигнала диспергированных повторяющихся последовательностей различалась для разных пар ДНК-проб. Было выдвинуто предположение, что оставшийся после компьютерной обработки неспецифичный сигнал связан с особенностью состава повторяющихся последовательностей в хромосомах, из которых были получены ДНК-пробы. Для его проверки была проведена оценка доли SINE (Short interspersed nuclear element) и LINE (Long interspersed nuclear element) последовательностей в хромосомах (данные взяты из базы данных Genome Browser USCS (<https://genome-euro.ucsc.edu/index.html>)). Эти последовательности имеют разную представленность в хромосомных районах. Поэтому с увеличением различий по соотношению количества SINE/LINE последовательностей в хромосомах, из которых получены ДНК-пробы, ожидается ухудшение элиминирования неспецифичного сигнала.

Оценка различия хромосом по соотношению SINE/LINE последовательностей производилась с помощью параметра $DChr$, который вычислялся по формуле:

$$DChr(a, b) = \sqrt{(s_a - s_b)^2 + (l_a - l_b)^2}, \quad (2)$$

где a и b – номера хромосом; s_a и s_b – доли SINE последовательностей в хромосомах a и b ; l_a и l_b – доли LINE последовательностей в хромосомах a и b .

В качестве оценки эффективности метода VISSIS использовали отношение сигнал/шум (SNR):

$$SNR = \frac{\mu_s}{\mu_{ns} + 1} \quad (3)$$

где μ_s – среднее значение интенсивности специфичного сигнала; μ_{ns} – среднее значение интенсивности неспецифичного сигнала.

Анализ изображений хромосом человека выявил отрицательную корреляцию значения SNR на обработанных изображениях с оценкой различий хромосом, из которых были получены ДНК-пробы (Пирсона, Спирмена и Кендалла, $p < 0.01$). На рисунке 4 показаны значения SNR на обработанных изображениях FISH ДНК-проб из хромосомы 19 при различных вариантах второй ДНК-пробы.

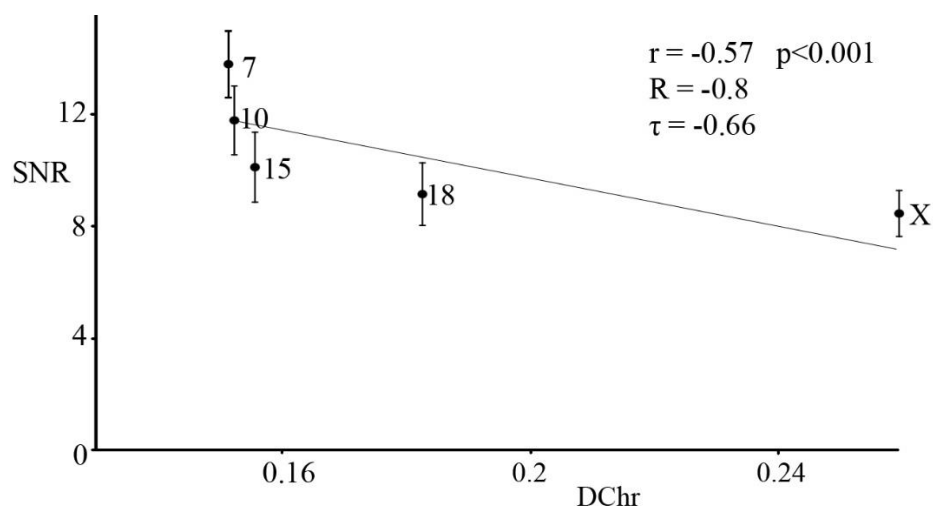


Рисунок 4. Корреляция отношения сигнал/шум (SNR) на обработанных изображениях FISH ДНК-проб из хромосомы 19 человека при различных вариантах второй ДНК-пробы и оценки различия хромосом (*DChr*). Номера хромосом, из которых изготовлена вторая ДНК-проба, указаны справа от значений SNR на графике. r , R и τ – значения коэффициентов корреляции Пирсона, Спирмена и Кендалла, соответственно.

Также был проведен сравнительный анализ средних профилей интенсивности, построенных для хромосом после компьютерной обработки изображений FISH ДНК-проб из хромосом 18 и 19 человека, с распределением SINE последовательностей. Указанные хромосомы значительно отличаются по относительному содержанию в них диспергированных повторов. Исходя из выдвинутого в работе предположения, профиль интенсивности неспецифичного сигнала на изображениях FISH ДНК-проб из хромосомы 19 должен соответствовать распределению SINE последовательностей.

Профили интенсивности сигнала на хромосомах приводились к одному размеру (профиль состоял из 50 точек), затем вычислялся средний профиль для гомологичных хромосом в наборе изображений. Длина профиля выбрана на основе анализа размеров самой маленькой хромосомы в наборе FISH-изображений. Для построения графика распределения последовательностей ДНК (профиль распределения) хромосомы были разбиты на 50 фрагментов, и в каждом фрагменте оценивалось количество соответствующих последовательностей. Нормализация графиков проводилась по методике (Legrand et al., 2008). Профили были построены для хромосом 1, 3, 7, 12, 16 и 17 по 14 изображениям хромосом (результаты для хромосомы 1 и 3 проиллюстрированы на рисунке 5).

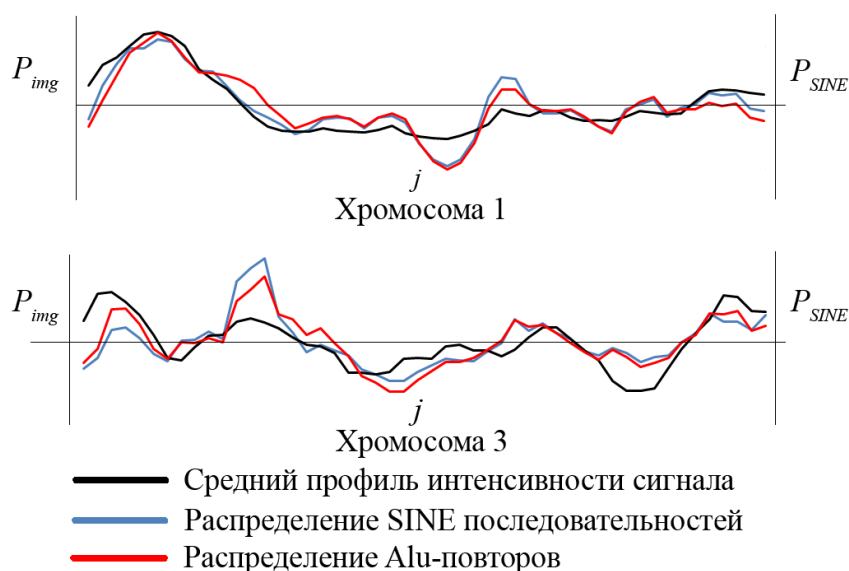


Рисунок 5. Сравнительный анализ средних профилей интенсивности, построенных для хромосом 1 и 3 на обработанных изображениях FISH ДНК-проб из хромосом 18 и 19, с данными о распределении SINE последовательностей в соответствующих хромосомах. P_{img} – значение интенсивности сигнала в точке профиля для среднего профиля интенсивности. P_{SINE} – количество повторов в хромосомном фрагменте для распределения SINE последовательностей в хромосоме. j – номер точки профиля интенсивности/распределения.

Количественную оценку различий между профилями (RMSD – root-mean-square deviation) проводили по формуле:

$$RMSD = \sqrt{\frac{\sum_{j=1}^N (P_{img}(j) - P_{SINE}(j))^2}{N}} \quad (4)$$

где $P_{img}(j)$ – значение среднего профиля интенсивности в точке j ; $P_{SINE}(j)$ – значение профиля распределения SINE последовательностей в точке j ; N – длина профилей (в данном случае $N=50$).

Среднее значение оценки RMSD и ее стандартное отклонение при сравнении средних профилей интенсивности и распределения Alu-повторов (RMSD_Alu) составили 0.73 ± 0.19 , а при сравнении средних профилей интенсивности и распределения SINE последовательностей (RMSD_SINE) – 0.73 ± 0.12 . Чтобы определить, являются ли полученные профили одинаковыми, было проведено сравнение средних профилей интенсивности, построенных по описанному выше способу на изображениях FISH хромосом человека с мечеными Alu-последовательностями, с данными по распределению Alu-повторов из базы данных Genome Browser USCS (RMSD_AluAlu). Так как оба профиля отражают распределение Alu-последовательностей в хромосомах, то их

можно считать одинаковыми. Профили были построены для хромосом 1, 3, 6, 7, 11, 12, 16 и 17 по десяти изображениям хромосом (результаты для хромосомы 1 и 3 проиллюстрированы на рисунке 6). Среднее значение и стандартное отклонение $RMSD_AluAlu$ составили 0.76 ± 0.18 .

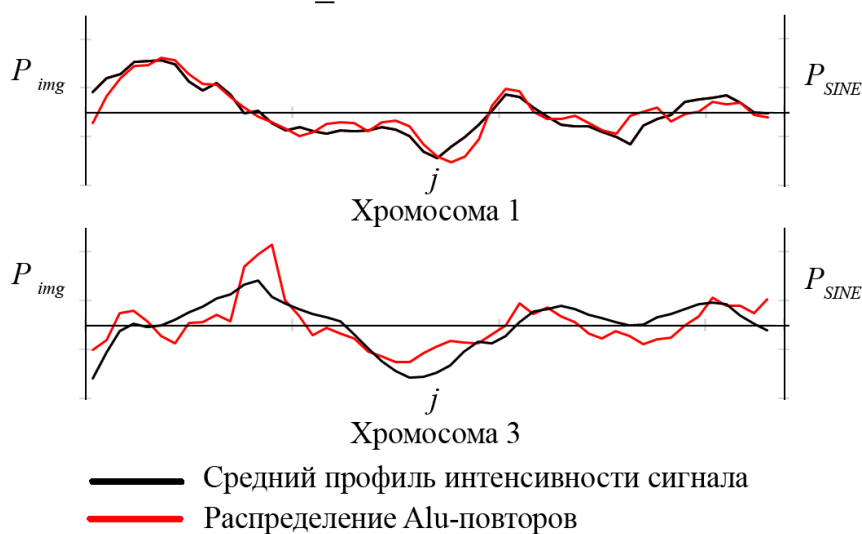


Рисунок 6. Сравнительный анализ средних профилей интенсивности, построенных для хромосом 1 и 3 на изображениях FISH хромосом человека с мечеными Alu-последовательностями, с данными о распределении Alu-повторов в соответствующих хромосомах. P_{img} – нормированное значение интенсивности сигнала в точке профиля для среднего профиля интенсивности. P_{SINE} – количество повторов в хромосомном фрагменте (значение нормировано) для распределения Alu-последовательностей в хромосоме. j – номер точки профиля.

Сравнение набора значений оценок $RMSD_AluAlu$, $RMSD_SINE$, $RMSD_Alu$ не выявило значимых различий медиан (критерий Краскела–Уоллиса, $p > 0.8$). Другими словами, полученные данные говорят в пользу того, что оставшийся после компьютерной обработки неспецифичный сигнал совпадает с распределением SINE последовательностей в хромосомах. Эта особенность метода может быть полезна для исследования генома малоизученных видов. Подбирая ДНК-пробы из хромосом, контрастных по содержанию повторяющихся последовательностей, можно проанализировать закономерности распределения повторяющихся последовательностей, а также получить данные о возможной локализации районов, богатых генами, поскольку существует взаимозависимость распределения Alu-повторов и генов на хромосоме (Pidpala et al., 2008).

На основе анализа относительного содержания различных типов повторов в разных хромосомах человека и оценки эффективности применения метода VISSIS установлены пары хромосом, для которых компьютерная обработка результатов FISH ДНК-проб из этих пар хромосом

позволяет надежно определять сигналы, формируемые хромосомоспецифичными последовательностями (Таблица 1).

Таблица 1 – Рекомендованные пары хромосом человека для проведения двухцветной FISH без супрессии гибридизации повторяющихся последовательностей для достижения максимальной эффективности компьютерной обработки методом VISSIS

Номер хромосомы <i>a</i> , из которой была получена первая ДНК-проба	Хромосомы, рекомендованные для получения второй ДНК-пробы
1	7, 9, 10*, 11, 12, 15, 20
2	3, 4, 5, 6, 7, 8*, 9, 10, 11, 18
3	2, 4, 5, 6*, 7, 8, 11
4	2, 3, 5*, 6, 8
5	2, 3, 4, 6*, 8, 11
6	2, 3*, 4, 5, 7, 8, 11, 18
7	1, 2, 3, 6, 8, 9, 10*, 11, 12
8	2, 3*, 4, 5, 6, 7, 10, 11, 18
9	1, 2, 7, 10, 14*, 15, 18
10	1, 2, 7*, 8, 9, 11, 12, 20
11	1, 2, 3, 5, 6, 7*, 8, 10, 12
12	1, 7*, 10, 11, 20
13	14, 18*
14	9*, 13, 15, 18
15	1, 9, 14*
16	17*
17	16*
18	2, 6, 8, 9*, 13, 14
19	17*
20	1* 10, 12
21	13*
22	16*
X	4*
Y	21*

Звездочкой (*) отмечены хромосомы, наиболее близкие к хромосоме *a* по соотношению количества SINE/LINE последовательностей

В работе было проведено сравнение FISH-изображений, полученных в результате CISS-гибридизации и в результате компьютерной обработки методом VISSIS и методом, предложенным Rens с соавторами (Rens et al., 2006) (далее метод RENS). Компьютерная обработка дает сопоставимые с CISS-гибридизацией результаты в случаях, когда ДНК-пробы созданы из хромосом, имеющих сходный состав повторяющихся последовательностей

ДНК (рисунок 7). Это делает компьютерные методы хорошей альтернативой CISS-гибридизации.

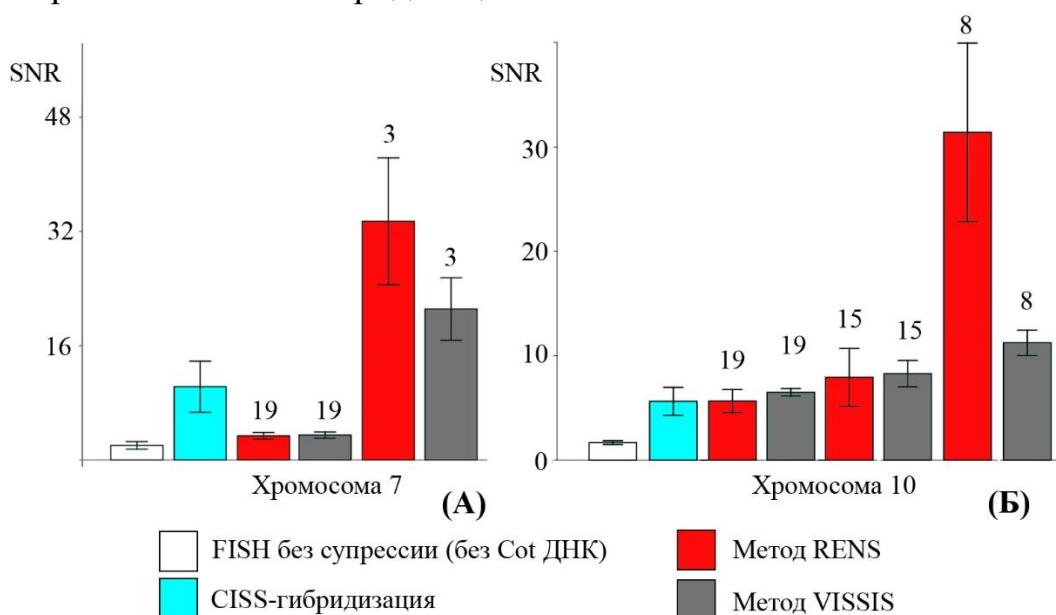


Рисунок 7. Сравнение эффективности подходов к усилению интенсивности сигнала с хромосомспецифичных последовательностей при проведении FISH ДНК-проб из хромосом 7 (А) и 10 (Б). Для методов RENS и VISSIS над каждым столбцом диаграммы указан номер хромосомы, из которой получена вторая ДНК-проба.

Метод VISSIS демонстрирует менее высокие значения SNR, чем метод RENS. Однако в отличие от метода RENS, метод VISSIS не зависит от субъективного мнения исследователя. Еще одним важным отличием метода VISSIS является снижение или полное исключение интенсивности неспецифичного сигнала на всех хромосомах, в том числе на объектах класса *Ca* и *Cb* (Богомолов и др., 2014).

Перспективным направлением применения метода VISSIS является молекулярно-цитогенетические исследования хромосом у видов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть проведена. Так, апробация метода VISSIS для анализа результатов FISH ДНК-проб, полученных из индивидуальных хромосом описторхид (*Metorchis xanthosomus*, *Opisthorchis felineus*), показала, что этот метод позволяет получать качественное специфическое окрашивание хромосом у видов, размер генома которых значительно уступает размеру генома млекопитающих (рисунок 8) (Богомолов и др., 2012). Вероятно, хромосомы этих видов обеднены диспергированными повторяющимися последовательностями.

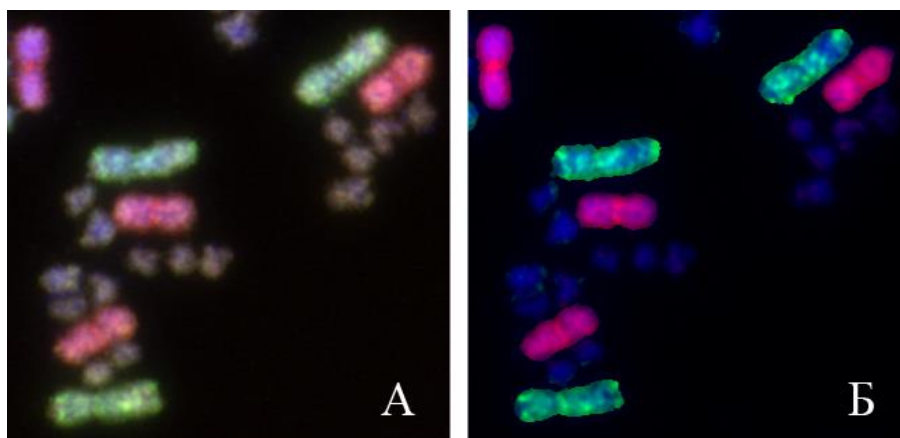


Рисунок 8. Пахитенные хромосомы. FISH ДНК-проб из хромосомы 1 (зеленый) и хромосомы 2 (красный) вида *M. xanthosomus*. **А** – Необработанное исходное изображение. **Б** – Результат применения метода VISSIS.

Более впечатляющими являются результаты использования метода VISSIS для идентификации гомологичных районов хромосом у саранчовых, геном которых по размеру в несколько раз превышает геном человека (рисунок 9) (Jetybayev et al., 2017).

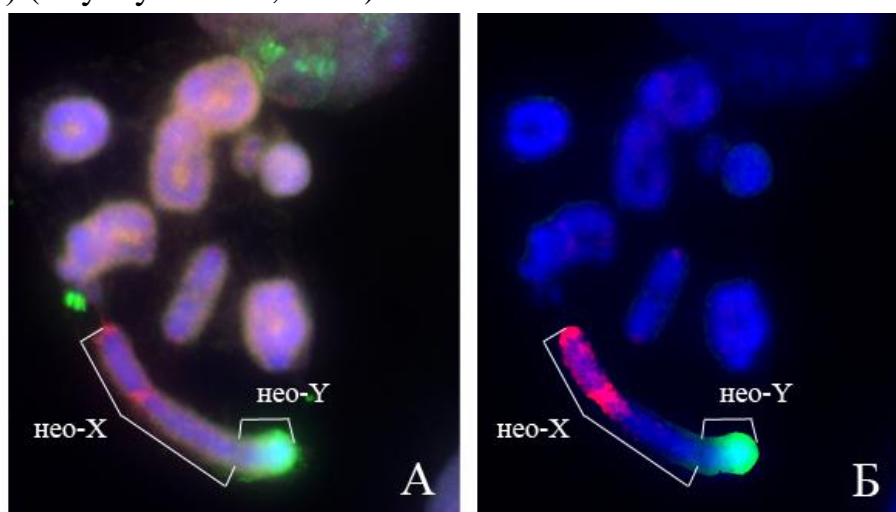


Рисунок 9. FISH микродиссекционных ДНК-проб из короткого плеча нео-Х хромосомы (красный сигнал) и нео-У хромосомы (зеленый сигнал) вида *Nocaracris tardus* с мейотическими хромосомами этого вида. **А** – Необработанное исходное изображение. **Б** – Результат применения метода VISSIS.

Ранее у исследованных видов саранчовых не удавалось выявлять специфичные сигналы в С-негативных хромосомных районах после FISH с ДНК-пробами, полученными из индивидуальных хромосом, даже при проведении CISS-гибридизации. Это обусловлено высоким содержанием в хромосомных районах диспергированных повторяющихся последовательностей.

Применение метода VISSIS для анализа хромосом 19 видов семейства *Pamphagidae* позволило проследить эволюцию нео-половых хромосом. На основе данных компьютерного анализа результатов FISH микродиссекционных ДНК-проб с митотическими и мейотическими хромосомами этих видов, было сделано заключение, что саранчовые семейства *Pamphagidae* представляют собой уникальную группу, в которой нео-Y хромосомы находятся на разных стадиях эволюции, включающих возникновение новых интеркалярных С-позитивных блоков и потерю С-негативных районов, происходящих из предковой аутосомы (Jetybayev et al., 2017).

ВЫВОДЫ

1. Разработаны компьютерный метод VISSIS (visualization of specific signal *in silico*) и компьютерная программа VisualCS (Visualization Chromosome-specific Signals), позволяющие визуализировать сигнал хромосомоспецифичных последовательностей ДНК при проведении FISH с ДНК-пробами, полученными из целых хромосом или протяженных хромосомных районов.
2. Показано, что при анализе изображений хромосом человека, полученных после проведения FISH с использованием хромосомоспецифичных ДНК-проб, метод VISSIS позволяет корректно выделять сигналы хромосомоспецифичных последовательностей ДНК, элиминируя полностью или значительно снижая интенсивность сигналов диспергированных повторов.
3. 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. Эти различия могут приводить к сохранению сигнала повторяющихся последовательностей после компьютерной обработки изображений. FISH с парами ДНК-проб из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК, позволяет достичь сопоставимых с CISS-гибридизацией результатов.
4. 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. При анализе изображений FISH ДНК-проб из 18 и 19 хромосом человека, контрастных по содержанию LINE- и SINE-

повторов, показана возможность использования метода VISSIS для анализа распределения в хромосомах различных типов диспергированных повторяющихся последовательностей.

5. Показано, что предложенный метод VISSIS позволяет визуализировать сигналы хромосомоспецифичных последовательностей ДНК при исследовании хромосом организмов, для которых CISS-гибридизация не дает удовлетворительных результатов (хромосомы саранчовых) или не может быть выполнена по объективным причинам (хромосомы описторхид).
6. Применение метода VISSIS впервые позволило установить гомеологию крупных С-негативных районов в половых хромосомах нео-Х и нео-У при сравнительном анализе 19 видов саранчовых семейства *Pamphagidae*.

СПИСОК ОСНОВНЫХ РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ:

1. Jetybayev I.Y., Bugrov A. G., Buleu O. G., **Bogomolov, A. G.**, Rubtsov, N. B. Origin and Evolution of the Neo-Sex Chromosomes in Pamphagidae Grasshoppers through Chromosome Fusion and Following Heteromorphization // *Genes*. – 2017. – V. 8. – №. 11. – P. 323-353 (Импакт фактор Web of Science 3.191).
2. Barkovskaya M.Sh., **Bogomolov A.G.**, Knauer N.Yu., Rubtsov N.B., Kozlov V.A. Development of software and modification of Q-FISH protocol for estimation of individual telomeres length in immunopathology // *Journal of Bioinformatics and Computational Biology*. – 2017 – V 15 – № 2 – P. 1650041 (Импакт фактор Web of Science 0.991).
3. Karamysheva T.V., Torgasheva A.A., Yefremov Y.R., **Bogomolov A.G.**, Liehr T., Borodin P.M., Rubtsov N.B. Spatial organization of fibroblast and spermatocyte nuclei with different B-chromosome content in Korean field mouse, *Apodemus peninsulae* (Rodentia, Muridae) // *Genome*. – 2017 – V. 60 – N 10 – P. 815-824 (Импакт фактор Web of Science 1.892).
4. **Богомолов А.Г.**, Карамышева Т.В., Рубцов Н.Б. Флуоресцентная гибридизация *in situ* ДНК-проб, полученных из индивидуальных хромосом и хромосомных районов // *Молекулярная биология*. – 2014 – V. 48 – №6 – С. 881–890 (Импакт фактор Web of Science 0.977).
5. **Богомолов А.Г.**, Задесенец К.С., Карамышева Т.В., Подколодный Н.Л., Рубцов Н.Б. Визуализация хромосомоспецифичных последовательностей ДНК при проведении FISH микродиссекционных ДНК-проб с метафазными хромосомами // *Вавиловский журнал генетики и селекции*. – 2012 – Т. 16. – №2. – С. 202-211 (индексируется в Scopus).