

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»

На правах рукописи

Богомолов Антон Геннадьевич

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

03.01.09. - математическая биология, биоинформатика

Диссертация на соискание ученой степени

кандидата биологических наук

Научный руководитель:

д.б.н., профессор

Рубцов Николай Борисович

Новосибирск 2019

ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ	2
СЛОВАРЬ СОКРАЩЕНИЙ.....	4
ВВЕДЕНИЕ	6
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ.....	13
1.1 ГИБРИДИЗАЦИИ НУКЛЕИНОВЫХ КИСЛОТ <i>IN SITU</i> . ТИПЫ ДНК-ПРОБ И ИХ ПРИМЕНЕНИЕ.....	13
1.2 ПОВТОРЯЮЩИЕСЯ И УНИКАЛЬНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДНК. ПРОБЛЕМА ИДЕНТИФИКАЦИИ МАТЕРИАЛА ХРОМОСОМ И ХРОМОСОМНЫХ РАЙОНОВ	21
1.3 ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ УЛУЧШЕНИЯ ВЫЯВЛЕНИЯ СПЕЦИФИЧНОГО FISH-СИГНАЛА ДНК-ПРОБ, ПОЛУЧЕННЫХ ИЗ ЦЕЛЫХ ХРОМОСОМ ИЛИ ХРОМОСОМНЫХ РАЙОНОВ	24
1.3.1 Методы подавления гибридизации повторяющихся последовательностей	24
1.3.2 Метод «дифференциального удерживания уникальных последовательностей».....	28
1.3.3 Относительное обогащение ДНК-пробы уникальными последовательностями	29
1.3.4 Методы элиминации из ДНК-клонотек повторяющихся последовательностей ДНК.....	30
1.3.5 Особенности анализа видов с секвенированным геномом	34
1.4 КОМПЬЮТЕРНЫЙ АНАЛИЗ FISH-ИЗОБРАЖЕНИЙ	37
1.4.1 Предобработка изображений.....	39
1.4.2 Сегментация изображений	41
1.4.3 Классификация объектов	52
1.4.4 Определение сигналов хромосомоспецифичных последовательностей ДНК с помощью компьютерной обработки FISH-изображений.....	54
1.5 ЗАКЛЮЧЕНИЕ	56
ГЛАВА 2. КОМПЬЮТЕРНЫЙ МЕТОД ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК (МЕТОД VISUALIZATION OF SPECIFIC SIGNAL <i>IN SILICO</i> – VISSIS).....	58
2.1 ОСНОВНЫЕ ПРИНЦИПЫ ОБРАБОТКИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ, ЛЕЖАЩИЕ В ОСНОВЕ МЕТОДА	58
2.2 СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЯ	62

2.3 КАЛИБРОВКА ИНТЕНСИВНОСТЕЙ СИГНАЛОВ FISH. СЕГМЕНТАЦИЯ ОБЛАСТЕЙ С БОЛЬШОЙ РАЗНИЦЕЙ ИНТЕНСИВНОСТИ СИГНАЛОВ В ПИКСЕЛЯХ	66
2.4 КЛАССИФИКАЦИЯ ОБЪЕКТОВ	69
2.5 ВЫДЕЛЕНИЕ СПЕЦИФИЧНОГО СИГНАЛА	79
2.6 ПОСТОБРАБОТКА ИЗОБРАЖЕНИЙ.....	80
2.7 ПРОГРАММА ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ: VISUALCS	82
2.8 ЗАКЛЮЧЕНИЕ	83
ГЛАВА 3. ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО МЕТОДА ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ.....	85
3.1 ПОЛУЧЕНИЕ ИЗОБРАЖЕНИЙ РЕЗУЛЬТАТОВ FISH РАЙОНО- И ХРОМОСОМОСПЕЦИФИЧНЫХ ДНК-ПРОБ	85
3.1.1 <i>Получение изображений результатов FISH хромосомоспецифичных ДНК-проб с хромосомами человека.....</i>	85
3.1.2 <i>Получение изображений результатов FISH хромосомоспецифичных ДНК-проб с хромосомами описторхид.....</i>	86
3.1.3 <i>Получение изображений результатов FISH районо- и хромосомоспецифичных ДНК-проб с хромосомами саранчовых</i>	87
3.2 ПРИМЕНЕНИЕ МЕТОДА VISSIS ДЛЯ ИДЕНТИФИКАЦИИ МАТЕРИАЛА ЦЕЛЫХ ХРОМОСОМ У ЧЕЛОВЕКА.....	89
3.3 СРАВНЕНИЕ РЕЗУЛЬТАТОВ CISS-ГИБРИДИЗАЦИИ И КОМПЬЮТЕРНОЙ ОБРАБОТКИ FISH-ИЗОБРАЖЕНИЙ.....	105
3.4 ПРИМЕНЕНИЕ МЕТОДА VISSIS ДЛЯ ИДЕНТИФИКАЦИИ ХРОМОСОМ У ОПИСТОРХИД	109
3.5 ПРИМЕНЕНИЕ МЕТОДА VISSIS ДЛЯ ИДЕНТИФИКАЦИИ ХРОМОСОМ И КРУПНЫХ ХРОМОСОМНЫХ РАЙОНОВ У САРАНЧОВЫХ	112
3.6 ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДА И ЕГО ОГРАНИЧЕНИЯ	116
ЗАКЛЮЧЕНИЕ	117
ВЫВОДЫ	120
СПИСОК ЛИТЕРАТУРЫ	122

СЛОВАРЬ СОКРАЩЕНИЙ

BAC – (*Bacterial Artificial Chromosome*) бактериальные искусственные хромосомы

Cot-1 ДНК – фракцию ДНК, которая ренатурируется час при 60°C с начальной концентрацией 89 мг/мл.

CGH – (*Comparative Genome Hybridization*) сравнительная геномная гибридизация

CISS-гибридизация – (*Chromosomal In Situ Suppression Hybridization*) супрессионная гибридизация *in situ*

CSTS – (*Correct Separation of Touching Chromosomes*) параметр, определяющий эффективность метода разделения находящихся в контакте хромосом, равный отношению правильно разделенных хромосом к общему числу хромосом в кластерах.

DAPI – (4,6-diamidino-2-phenyl-indole) 4', 6'-диамидино-2-фенилиндол

DChr – оценка различия хромосом по соотношению относительного количества коротких и длинных диспергированных повторяющихся последовательностей в них

DOP-PCR – (*Degenerate Oligonucleotide Primed Polymerase Chain Reaction*) полимеразная цепная реакция с частично вырожденным праймером

FISH – (*Fluorescence In Situ Hybridization*) флуоресцентная гибридизация *in situ*

FISH-изображение – изображение с результатами флуоресцентной гибридизации *in situ*

GISH – (*Genomic In Situ Hybridization*) геномная гибридизация *in situ*

LINE – (*Long Interspersed Repeated Element*) длинные диспергированные повторы

NS – (*Non-specific Signals*) неспецифичные сигналы

Pre-ISH – (*Preparative In Situ Hybridization*) препаративная гибридизация *in situ*

RENS – компьютерный метод, разработанный Ренсом и его коллегами в 2006 году для определения сигналов хромосомоспецифичных последовательностей ДНК

RMSD – (*root-mean-square deviation*) количественная оценка различий между профилями интенсивности сигналов

scFISH пробы – (*single-copy FISH probes*) разработанные ДНК-пробы с высоким содержанием уникальных последовательностей

SINE – (*Short Interspersed Repeated Element*) короткие диспергированные повторы

SS – (*Specific Signals*) специфичные сигналы

VISSIS – (*VI*sualization of *S*pecific *S*ignal *I*n *S*ilico) разработанный в настоящей работе компьютерный метод

кДНК (комплементарная ДНК) – последовательность ДНК, синтезированная на основе молекулы РНК с использованием обратной транскриптазы

Класс объектов C_a – объекты, содержащие ДНК, специфичную для хромосомы a

Класс объектов C_b – объекты, содержащие ДНК, специфичную для хромосомы b

Класс объектов C_{ns} – объекты, которые не содержат ДНК, специфичную для хромосом a и b . Регистрируемый сигнал связан только с повторяющимися последовательностями ДНК.

ПЦР – полимеразная цепная реакция

ВВЕДЕНИЕ

Актуальность работы

Развитие методов детекции и визуализации в хромосомах эукариот индивидуальных последовательностей ДНК значительно расширило возможности цитогенетического анализа. 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. Однако 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. В этой связи становится актуальной разработка методов, позволяющих четко дифференцировать специфичные сигналы последовательностей ДНК при проведении FISH.

В настоящее время как в научных исследованиях, так и в клинической практике для идентификации материала целых хромосом и отдельных хромосомных районов важным инструментом является FISH (Jensen, 2014; Riegel, 2014; Shakoori, 2017). В большинстве случаев для успешного применения FISH требуется подавление или элиминация флуоресцентного сигнала диспергированных повторяющихся последовательностей ДНК. Поскольку такие последовательности составляют значительную часть генома человека и геномов других видов эукариот (Евгеньев, 2007; Pidpala et al., 2008), они в большом количестве присутствуют не только в хромосомоспецифичных, но и в районспецифичных ДНК-пробах. Проведение FISH с метафазными хромосомами без супрессии гибридизации повторяющихся последовательностей ДНК приводит к появлению интенсивного сигнала не только в тех районах или хромосомах, которые были использованы для создания ДНК-проб, но также и во многих других районах и хромосомах.

Для молекулярно-цитогенетического анализа хромосом млекопитающих эта проблема решается, в основном, с помощью супрессионной гибридизации *in situ*

(Chromosome In Situ Suppression hybridization - CISS-гибридизация). Подход 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* (Fuchs et al., 1996; 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 специфичных ДНК-проб, делает актуальным разработку и реализацию новых подходов и методов для анализа результатов FISH. Одним из таких подходов является компьютерный анализ микроскопических изображений многоцветной FISH, позволяющий элиминировать флуоресцентные сигналы, продуцируемые мечеными диспергированными повторяющимися последовательностями. Использование компьютерного анализа могло бы заменить стадию супрессии гибридизации повторяющихся последовательностей при проведении FISH и тем самым расширить применение FISH в исследованиях по сравнительной цитогенетике и геномике, а также удешевить метод молекулярно-цитогенетической диагностики при анализе хромосомных патологий человека.

Цели и задачи исследования

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 заявленной цели были поставлены следующие задачи:

- 1) Разработать метод и программное обеспечение для визуализации и анализа сигналов хромосомспецифичных 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 хромосомспецифичных 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 и практическая ценность

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. Разработанный метод может также быть применен для улучшения результатов CISS-гибридизации, особенно 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. Перспективным 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.

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. Ранее у этих видов не удавалось идентифицировать районы, содержащие хромосомоспецифичные последовательности ДНК из-за большого количества повторяющихся последовательностей ДНК в их геномах.

Показано, что на результаты компьютерной обработки влияют особенности состава повторяющихся последовательностей хромосом, из которых были получены ДНК-пробы. Чем сильнее различия в содержании различных типов повторяющихся последовательностей, в хромосомах или хромосомных районах, из которых были получены ДНК-пробы, тем менее эффективным является элиминация сигнала повторяющихся последовательностей из итогового изображения. В случае хромосом млекопитающих различия в эффективности элиминации сигнала повторяющихся последовательностей могут проявляться на уровне R- и G-бэндов хромосом. Эта уникальная особенность созданного метода будет полезна в исследовании геномов малоизученных видов. Подбирая пары 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. Принципиально, такая возможность была продемонстрирована на примере анализа FISH-изображений с ДНК-пробами из хромосом человека 18 и 19, контрастных по содержанию длинных и коротких диспергированных повторов.

Положения, выносимые на защиту

1. Разработан компьютерный метод выявления хромосомо- и районоспецифичных 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 C-негативных районов хромосом в геномах девятнадцати

видов саранчовых семейства *Pamphagidae* с большим количеством повторяющихся последовательностей.

Апробация работы

Материалы работы вошли в отчёт по гранту Российского фонда фундаментальных исследований (№ 16-34-00498 мол_а; руководитель проекта А.Г. Богомолов). 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).

В рамках диссертации опубликовано 16 работ, из них шесть статей в рецензируемых научных журналах (пять статей в рецензируемых научных журналах, входящих в список ВАК по специальности 03.01.09. - "математическая биология, биоинформатика"), и получено два авторских свидетельства.

Публикации

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 – № 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 – Т. 48 – №6 – С. 881–890 (Импакт фактор Web of Science 0.977).

5) **Богомолов А.Г.**, Задесенец К.С., Карамышева Т.В., Подколотный Н.Л., Рубцов Н.Б. Визуализация хромосомспецифичных 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).

Авторские свидетельства:

1. **Богомолов А.Г.**, Подколотный Н.Л., Рубцов Н.Б. “Визуализация хромосомспецифичных последовательностей ДНК при проведении FISH (ВизуалКС) / Visualization chromosome-specific signals in FISH-images (VisualCS)”, 2018

2. **Богомолов А.Г.**, Барковская М.Ш., Рубцов Н.Б. “Оценка длины теломерных районов хромосом (MeTeLen) / Measurement of telomere length on chromosomes (MeTeLen)”, 2016

Оба результата интеллектуальной деятельности внедрены в научно-практическую деятельность ИЦиГ СО РАН и используются в Лаборатории морфологии и клеточных структур (Акты внедрения результатов интеллектуальной деятельности №2018/3 и №2018/4)

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.

Личный вклад соискателя. Основные результаты, изложенные в диссертации, получены автором самостоятельно: разработан компьютерный метод, проведена его апробация на изображениях FISH ДНК-проб с хромосомами человека, 2-ух 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.

Благодарности. Автор выражает искреннюю благодарность руководителю диссертации д.б.н. Рубцову Н.Б. Автор благодарит к.б.н. Карамышеву Т.В., к.б.н. Джетыбаева И.Е., к.б.н. Задесенец К.С. за предоставленные FISH-изображения. Автор благодарит к.т.н. Деменкова П.С., д.б.н. Орлова Ю.Л., к.б.н. Афонникова Д.А., к.б.н. Лавреху В.В, Подкольного Н.Л за консультации и плодотворные научные дискуссии.

Глава 1. ОБЗОР ЛИТЕРАТУРЫ

1.1 Гибридизации нуклеиновых кислот *in situ*. Типы ДНК-проб и их применение

Последние два десятилетия продемонстрировали значительные успехи в развитии микроскопической техники, регистрации и обработке микроскопических изображений. В значительной степени это обусловлено разработкой и внедрением в практику молекулярно-цитогенетических исследований и медицинской диагностики хромосомных патологий человека новых методов визуализации и компьютерной обработки изображений, новых методов молекулярно-цитогенетического анализа и создания новых флуорохромов (Peng et al., 2012). Одним из наиболее широко используемых методов современной цитогенетики стал метод флуоресцентной гибридизации *in situ* (fluorescence *in situ* hybridization – FISH) (Liehr et al., 2004; Shakoori, 2017). Метод FISH позволил перейти от изучения морфологии хромосом к исследованию локализации и распределения в них конкретных последовательностей ДНК. В отличие от методов дифференциального окрашивания хромосом, он позволяет точно определить присутствие и локализацию ДНК (или РНК), характеризующуюся индивидуальной последовательностью нуклеотидов, непосредственно в клетке или клеточном ядре, на метафазных хромосомах, в интерфазных ядрах, на растянутых хроматиновых фибриллах или нитях ДНК.

Принцип гибридизации *in situ* заключается во взаимодействии между денатурированными нитями ДНК и мечеными денатурированными фрагментами ДНК (их называют пробами) непосредственно на цитологическом препарате. Далее в работе под пробами будут подразумеваться меченые последовательности ДНК, хотя частично описанные ниже рассуждения справедливы и для РНК. При наличии комплементарности на гомологичном участке цитологического препарата образуются стабильные молекулярные гибриды (дуплексы) между ДНК исследуемого образца и ДНК пробы. В дальнейшем, после отмывки оставшейся несвязанной ДНК использованной пробы, места локализации дуплексов определяются с помощью выявления специальных элементов, связанных с ДНК пробы.

В первых экспериментах в качестве специальных элементов для выявления дуплексов на хромосомах применялись радиоактивные изотопы (Pardue and Gall, 1969; John et al., 1969). Главными недостатками их использования являлось высокая трудоемкость микроскопического анализа и низкая точность локализации ДНК-проб (Levsky, Singer, 2003).

В 1980-х годах появились методы нерадиоактивного мечения, основанные на использовании флуорохромов (Bauman et al., 1980; Langer et al., 1981). **Флуорохромы (или флуорофоры)** – это специальные вещества, способные расходовать часть энергии поглощенного света на излучение света определенной длины волны при возвращении из возбужденного состояния в стабильное. Замена радиоактивного изотопа флуорохромами не только сделала методы безопаснее, но также обеспечила более высокую точность локализации и открыла путь к одновременному использованию большого числа ДНК-проб (Swiger, Tucker, 1996; Moter, Göbel, 2000; Halder et al., 2004; Shakoori, 2017). В настоящее время радиоактивные изотопы полностью вытеснены флуорохромами (Moter, Göbel, 2000).

Нуклеотиды ДНК пробы связываются либо напрямую с флуорохромом (прямое мечение ДНК), либо через промежуточные репортерные молекулы (непрямое мечение ДНК) (Riegel, 2014, Liehr, 2017). При прямом мечении после завершения гибридизации проводится только отмывка несвязанной меченой ДНК перед проведением микроскопии. В случае непрямого мечения для определения локализации дуплексов необходима дополнительная процедура окрашивания. Достоинством непрямого мечения ДНК-проб является более высокий уровень сигнала. Это объясняется присутствием на репортерных молекулах нескольких молекул флуорохрома. Кроме того, в некоторых системах детекции существует возможность каскадного усиления сигнала (Liehr, 2017).

Можно выделить три основных этапа при проведении исследований, использующих FISH (схема перечисленных ниже этапов проиллюстрирована на рисунке 1.1):

- 1) **Создание ДНК-пробы** – выделение или синтез ДНК-фрагментов, которые будут использоваться в качестве ДНК-пробы, и введение в них специальных элементов;

2) **Непосредственная гибридизация ДНК-пробы с ДНК цитологического препарата**, включающая денатурацию ДНК, последующую совместную ренатурацию и гибридизацию, отмывку оставшейся свободной меченой ДНК. Обычно этот этап включает обработку цитологического препарата, делающую его ДНК доступной для гибридизации с меченой ДНК соответствующей ДНК пробы;

3) **Детекция меченой ДНК** - визуализация районов связывания ДНК препарата и ДНК пробы при помощи флуоресцентного микроскопа. Для варианта непрямого мечения ДНК-проб этап включает в себя связывание репортерных элементов с флуорохромами либо другие варианты их маркирования флуорохромами.

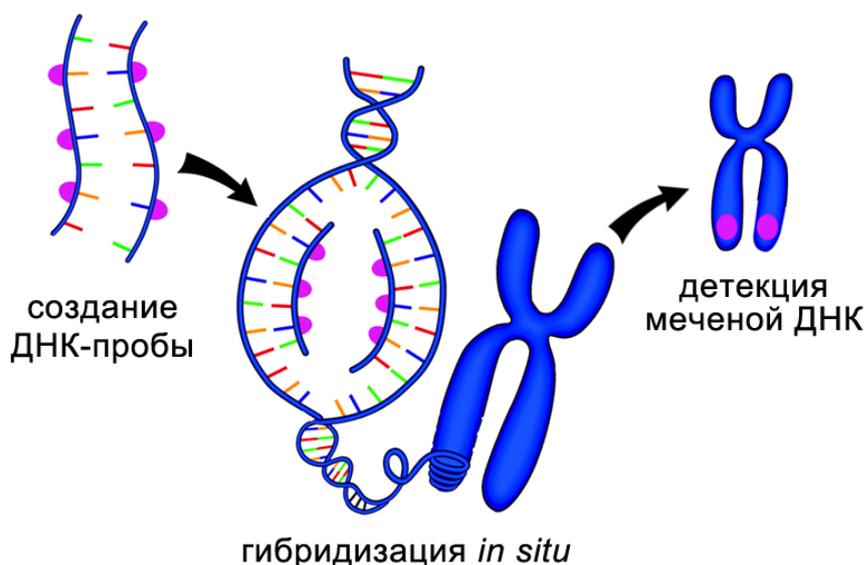


Рисунок 1.1. Основные этапы флуоресцентной гибридизации *in situ*. Рисунок сделан на основании статьи Уипполда и Перри (Wippold, Perry, 2007).

Обычно под “ДНК-пробой” понимают фрагмент или набор фрагментов ДНК, помеченных таким образом, чтобы была возможность их визуализации при микроскопии препарата после проведения гибридизации *in situ*. Фрагменты ДНК в пробе могут быть как клонированными фрагментами ДНК, так и совокупностью фрагментов, отвечающим некоторому общему требованию. Разнообразие таких требований очень велико. Например, это может быть набор фрагментов ДНК, полученный из более крупного клонированного фрагмента ДНК, из продукта полимеразной цепной реакции, из конкретного хромосомного района, выделенного микродиссекцией метафазных хромосом, из целых индивидуальных хромосом,

фрагменты ДНК из конкретной фракции геномной ДНК и, наконец, представляющие всю геномную ДНК (Halder et al., 2004; Liehr, 2017). Методы применения ДНК-проб и решаемые с их помощью задачи могут значительно отличаться в зависимости от набора фрагментов ДНК, содержащихся в ДНК-пробе, и вариантов их мечения.

ДНК-пробы, основой которых являются клонированные фрагменты ДНК, широко используются в диагностике хромосомных патологий для определения числа хромосом в пренатальной и доимплантационной диагностике, детекции микроделеций и транслокаций при диагностике врожденных аномалий и диагностике онкологических заболеваний. В зависимости от используемого вектора (плазмиды, космиды, фазмиды, фаги, искусственные хромосомы бактерий и дрожжей) размер клонированного фрагмента может варьировать от сотни до более чем миллиона пар оснований (Kjeldsen, Kolvraa, 2002; Halder et al., 2004; Shakoori, 2017). Обычно такие ДНК-пробы используются для идентификации генов, определенных участков хромосом или кластеров повторяющихся последовательностей (локуспецифичные ДНК-пробы) (Kjeldsen, Kolvraa, 2002; Shakoori, 2017). Клонированный фрагмент может представлять собой уникальную или повторяющуюся последовательность ДНК, а также одновременно содержать как уникальные последовательности, так и повторы.

Помимо клонированных фрагментов ДНК к ДНК-пробам с известным составом следует отнести ДНК-пробы, полученные с помощью полимеразной цепной реакции (ПЦР). В настоящее время с помощью ПЦР можно амплифицировать фрагменты, размером превышающие 20 тысяч пар оснований. Такую возможность предоставляет метод ПЦР длинных фрагментов (Cheng et al., 1994). При наличии последовательностей скаффолдов секвенированного генома не составляет больших проблем подобрать пары праймеров для амплификации конкретных участков генома (Rogan et al., 2001). Перспективность такого подхода связано с конструированием ДНК-проб, содержащих только уникальные последовательности (см. раздел 1.3.5).

С помощью ПЦР также могут быть получены ДНК-пробы для локализации кластеризованных повторов: теломерной ДНК (Ijdo et al., 1991), рибосомальной ДНК (Abbo et al., 1994) и ДНК прицентромерного гетерохроматина (Nietzel et al.,

2001). Такие пробы используются не только в клинической диагностике, но и в исследованиях, посвященных изучению эволюции кариотипа в различных таксонах.

Обычно с увеличением размера клонированного фрагмента увеличивается количество присутствующих в нем повторяющихся последовательностей, которые при проведении FISH будут давать дополнительные сигналы в многочисленных районах хромосом. Уже при использовании таких ДНК-проб возникает проблема подавления гибридизации фрагментов ДНК, содержащих повторяющиеся последовательности (Sealey et al., 1985; Schwarzacher 2003). Эта проблема становится намного более значимой при использовании ДНК-проб, созданных из протяженных хромосомных районов, целых хромосом или даже всего генома. В большинстве случаев их использование основано на супрессии гибридизации повторяющихся последовательностей либо других способов элиминации их сигнала из полученных микроскопических изображений. Исключением являются случаи гибридизации с ДНК-пробами, состоящими из ДНК одного из родительских геномов организма, возникшего на основе межвидовой гибридизации (амфиполиплоида), и случаи кросс-гибридизации для некоторых пар видов с использованием ДНК-проб, содержащими ДНК одного из видов (Muller et al., 1998; Raina, Rani, 2001; Devi et al., 2005; Markova, Vyskot, 2009; Rampin et al., 2012; Silva, Souza, 2013).

В качестве примера использования кросс-гибридизации можно привести гибридизацию *in situ* ДНК-проб, полученных из целых хромосом двух видов гиббонов, с хромосомами человека. Указанная гибридизация стала отдельным методом анализа внутрихромосомных перестроек у человека и получила собственное название RxFISH (Muller et al., 1998). Слабый сигнала повторяющихся последовательностей ДНК можно объяснить различиями в составе диспергированных повторяющихся последовательностей у гиббонов и человека.

Пробы, состоящие из ДНК одного из родителей амфиполиплоидов, используются для проведения геномной гибридизации *in situ* (Genomic *In Situ* Hybridization - GISH) (Schwarzacher et al., 1989). Геномы таких амфиполиплоидов (как твердые и мягкие пшеницы) включают хромосомы, пришедшие из разных родительских видов, в которых существенно отличаются диспергированные повторы. При проведении GISH основной гибридизационный сигнал дают именно эти последовательности ДНК, что позволяет надежно определять принадлежность

конкретных хромосом к группе хромосом, происходящих от одного из родительских геномов. GISH позволят успешно выявлять транслокации между хромосомами, относящимися к разным геномам, инсерции в хромосомы других групп (Raina, Rani, 2001; Devi et al., 2005; Markova, Vyskot, 2009; Rampin et al., 2012; Silva, Souza, 2013).

В остальных случаях FISH с ДНК-пробами, в состав которых входит ДНК из протяженных районов хромосом или даже всего генома, основан на выявлении сигнала уникальных последовательностей ДНК. Полногеномные ДНК-пробы представляют собой меченую фрагментированную суммарную геномную ДНК. Фрагментирование требуется для обеспечения эффективной гибридизации меченой ДНК *in situ*. Кроме непосредственных проблем, связанных с гибридизацией слишком длинных фрагментов ДНК, существуют также проблемы доступности ДНК цитологического препарата для крупных молекул (Schwarzacher, 2003). Следует напомнить, что в недалеком прошлом полногеномные ДНК-пробы часто использовались для изучения распределения в хромосомах повторяющихся последовательностей, однако, в последние годы число таких исследований резко сократилось во многом из-за сложности интерпретации полученных результатов. На результаты FISH полногеномных ДНК-проб большое влияние оказывает как количество разных типов повторов, так и их локализация в хромосомах (рассеянные, тандемные или смешанные варианты локализации). Учитывая наличие в геноме множества различных повторов и их различного распределения по геному суммарная картина гибридизации *in situ* полногеномных ДНК-проб оказывается очень сложной для анализа.

Существенно большую информацию можно получить, используя для FISH клонированные фрагменты ДНК, содержащие конкретные повторяющиеся последовательности. В связи с этим, полногеномные ДНК-пробы в настоящее время используются преимущественно для выявления нарушения баланса хромосомных районов (сравнительная геномная гибридизация, проводимая на препаратах метафазных хромосом) или анализа вариации числа копий (copy number variation – CNV) относительно небольших участков генома (сравнительная геномная гибридизация на микрочипах) (Pita et al., 2014; Durmaz et al., 2015). Сравнительная геномная гибридизация базируется на гибридизации уникальных фрагментов ДНК,

и в классическом варианте тоже требует супрессии гибридизации повторяющихся последовательностей (Kallioniemi et al., 1992; Weiss et al., 1999).

Для настоящего исследования наибольший интерес представляют хромосомоспецифичные и районспецифичные ДНК-пробы. В недалеком прошлом их получали либо объединением десятков и сотен протяженных клонированных фрагментов ДНК (обычно в бактериальных искусственных хромосомах), либо видоспецифической *inter-SINE* ПЦР (Nelson et al., 1991) ДНК межвидовых гибридов соматических клеток, содержащих лишь одну интересующую исследователя хромосому нужного вида. Однако, вследствие развития методов сиквенс-независимой ПЦР (ПЦР с использованием линкер-адаптеров, ПЦР с частично вырожденным праймером, системы полногеномной амплификации) и методов выделения индивидуальных хромосом были разработаны новые технологии создания хромосомоспецифичных и районспецифичных ДНК-проб. ДНК изолированных хромосом (с помощью сортировки проточной цитометрией или микродиссекции метафазных хромосом) амплифицировали с помощью сиквенс-независимой ПЦР (Ludecke et al., 1989; Telenius et al., 1992; Deleye et al., 2017). Несмотря на то, что полученные ДНК-пробы содержали лишь часть уникальной ДНК интересующей хромосомы или хромосомного района, этого оказалось достаточно для получения плотного гибридизационного сигнала на соответствующих хромосомах даже после практически полной супрессии гибридизации повторяющихся последовательностей. Хромосомы оказывались плотно окрашенными, что привело к появлению в молекулярной цитогенетике нового термина: “хромосомный пэинтинг” (chromosome painting) (Pinkel et al., 1988). В отличие от различных методов хромосомного окрашивания, основанных на связывании красителя с белками или ДНК хромосом, хромосомный пэинтинг окрашивает хромосомы в соответствии с составом уникальных последовательностей их ДНК.

Использование районо- и хромосомоспецифичных проб нашло широкое применение как в диагностике хромосомных патологий человека, онкологических заболеваниях, так и в исследованиях эволюции хромосом млекопитающих (Pita et al., 2014; Shakoori, 2017; Yang et al., 2017). К сожалению, если вопрос супрессии гибридизации повторяющихся последовательностей ДНК при диагностике

хромосомных патологий человека является, в основном, вопросом стоимости проводимой диагностики, то в исследованиях, в которые оказывались вовлечены другие виды, традиционная CISS-гибридизация часто не могла быть проведена из-за невозможности получения фракции повторяющихся последовательностей (малый размер и малое число доступных особей исследуемого вида) (Rens et al., 2006) либо не позволяла получить удовлетворительных результатов из-за очень высокого содержания повторяющихся последовательностей в анализируемых хромосомах (Fuchs et al., 1996; Houben et al., 2001; Schubert et al., 2001; Han et al., 2015). Таким образом, актуальной стала задача разработки альтернативных подходов к элиминации гибридационного сигнала повторяющихся последовательностей.

В настоящее время FISH является важным инструментом в научных исследованиях и клинической практике (Devi et al., 2005; Jensen, 2014; Shakoori, 2017). Метод FISH в значительной степени способствовал также накоплению новых знаний в различных областях исследований, таких, как филогенетические и таксономические, молекулярная токсикология, цитогенетика, локализация генов, анализ экспрессии генов, радиационная биология и т.д. В клинической практике метод широко используется для обнаружения анеуплоидии и частых хромосомных патологий в дородовой и послеродовой диагностике и в цитогенетике онкологических заболеваний. Существенным преимуществом метода FISH является то, что он позволяет прямое исследование последовательностей ДНК на цитологических препаратах не только митотических хромосом, но и препаратах неделящихся клеток и срезов тканей (Kjeldsen, Kolvraa, 2002; Riegel 2014; Bakker et al., 2015). При отсутствии делящихся клеток проводится определения числа копий исследуемых фрагментов ДНК или копий хромосом в интерфазных ядрах. Более подробную информацию об особенностях и областях применения метода FISH можно найти в обзорных статьях (Ried et al., 1998; Kjeldsen, Kolvraa, 2002; Langer et al., 2004; Geurts, de Jong, 2013; Pita et al., 2014; Durmaz et al., 2015; Shakoori, 2017) и специализированной книге (Fluorescence In Situ Hybridization (FISH). Application Guide, 2017).

Регистрация и отдельная запись сигналов различных флуорохромов с одного поля зрения открыли новые возможности для анализа изображений FISH. Для визуализации результатов микроскопии и их последующей обработки широко

используются «псевдоцвета». Так как, регистрация интенсивности сигнала каждого отдельного флуорохрома обычно осуществляется в отдельном канале, то для ее визуализации на микроизображении может использоваться любой удобный исследователю цвет. Он обычно не совпадает с настоящим цветом данного флуорохрома и называется «псевдоцветом».

В результате широкого применения компьютерной обработки микроизображений понятие «псевдоцвета» было значительно расширено. При проведении 24-х цветной FISH, позволяющей на одной метафазной пластинке идентифицировать материал всех хромосом человека, псевдоцвет присваивался не только сигналам отдельных флуорохромосом, но также их комбинациям (Speicher et al., 1996; Garini et al., 1996; Tanke et al., 1999). Этот же подход был реализован при разработке многоцветного бэндинга хромосом, основанного на межвидовой гибридизации *in situ* (RxFISH) (Muller et al., 1998). При проведении M-бэндинга хромосом (Chudoba et al., 1999) различные псевдоцвета присваиваются даже не комбинациям флуорохромов, а соотношениям их интенсивностей. Сравнительный анализ интенсивностей гибридизационного сигнала ДНК-проб, полученных из геномной ДНК клеток пациента и здорового донора лег в основу сравнительной геномной гибридизации (CGH) (Kallioniemi et al., 1992; Weiss et al., 1999). В этом случае изменения соотношения интенсивностей указывает на нарушение генетического баланса в анализируемых районах хромосом.

Несмотря на достигнутые успехи в применении FISH с районо- и хромосомоспецифичными ДНК-пробами и широкое распространение данного метода, ряд проблем выявления специфичного сигнала последовательностей ДНК, характерных для отдельных хромосомных районов или целых хромосом, до настоящего времени остается не до конца решенным. Одной из таких проблем является выделение сигнала уникальных последовательностей ДНК у видов с высоким содержанием диспергированных повторяющихся последовательностей в геноме.

1.2 Повторяющиеся и уникальные последовательности ДНК. Проблема идентификации материала хромосом и хромосомных районов

Одно из первых упоминаний терминов “повторяющиеся” и “уникальные” последовательности ДНК встречается в работе Бриттена и Кона (Britten, Kohne,

1968). Авторы этой статьи измерили кинетику реассоциации фрагментов ДНК у разных организмов. Различия в скоростях реассоциации разных последовательностей ДНК связаны с их разной представленностью в геномной ДНК. Последовательности, представленные в геноме всего один раз, назвали **уникальными**. А нуклеотидные последовательности, представленные в геноме некоторым количеством одинаковых или незначительно различающихся копий, и, как следствие, быстрее находившие комплементарную цепь, назвали **повторяющимися последовательностями**. Такое определение охватывает огромное количество разнообразных элементов ДНК, отличающиеся по структуре и происхождению.

Повторяющиеся последовательности могут быть разделены на две большие группы: тандемные и диспергированные повторяющиеся последовательности (Padeken et al., 2015). **Тандемные повторы** - это короткие повторяющиеся нуклеотидные последовательности, которые расположены вплотную друг за другом (тандемно). Вторая группа состоит из повторяющихся нуклеотидных последовательностей, которые рассеяны (диспергированы) по всему геному. Эти последовательности могут изменять свое положение в геноме или имели ранее такую возможность. Последовательности ДНК способные перемещаться внутри генома называют **мобильными генетическими элементами**. На основании молекулярных механизмов, используемых мобильными генетическими элементами для перемещения в геноме, их разделяют на два больших класса: **ретротранспозоны и ДНК-транспозоны** (Евгеньев, 2007; Wicker et al., 2007; Padeken et al., 2015;). Для перемещения ДНК-транспозонов, как правило, характерен механизм вырезания и вставки (cut and paste) с участием закодированного в них фермента транспозазы. Ретротранспозоны используют для своего перемещения механизмы, в которых важную роль играет процесс обратной транскрипции, - синтез ДНК на матрице РНК при помощи обратной транскриптазы. Такой механизм иногда называют копировать и вставить (copy and paste). Значительная часть мобильных элементов представлена в виде неавтономных вариантов, которые не способны к самостоятельному переносу. Их перемещение может происходить за счет транскрипции автономными элементами соответствующего семейства генома.

Повторяющиеся последовательности были найдены во всех организмах эукариот (Kazazian, 2004; Евгеньев, 2007; Munoz-Lopez, Garcia-Perez, 2010). Эти последовательности могут составлять значительную часть генома (Pidpala et al., 2008; Munoz-Lopez, Garcia-Perez, 2010). К примеру, у млекопитающих они занимают больше 40% генома, а их доля в геноме растений составляет 50-90% (Kazazian, 2004; Евгеньев, 2007; de Koning et al., 2011; Сергеева, Салина, 2011). В среднем, доля повторяющихся последовательностей в геноме пропорциональна его размеру, и большие геномы, как правило, содержат большую долю (Евгеньев, 2007).

Присутствие большого количества повторяющихся последовательностей ДНК в хромосомах является серьезной проблемой при проведении FISH (Dorman et al., 2013). Фрагменты ДНК, в совокупности входящие в состав ДНК-пробы, должны быть гомологичны достаточно протяженному району хромосомы для их успешного выявления на цитологическом препарате. С увеличением размера “окрашиваемого” хромосомного района растет и количество представленных в ДНК-пробе повторяющихся последовательностей, и, как результат, - интенсивность сигнала в других районах хромосом (Landegent et al., 1987; Dorman et al., 2013). Этот альтернативный источник сигнала может иметь почти такую же интенсивность, что и у интересующего исследователя сигнала хромосомспецифичных последовательностей (Craig et al., 1997; Trifonov et al., 2017). На рисунке 1.2 видно, что ДНК-пробы хромосом 10 и 15 связываются не только с ДНК соответствующих хромосом, но и с гомологичными повторяющимися последовательностями на других хромосомах.

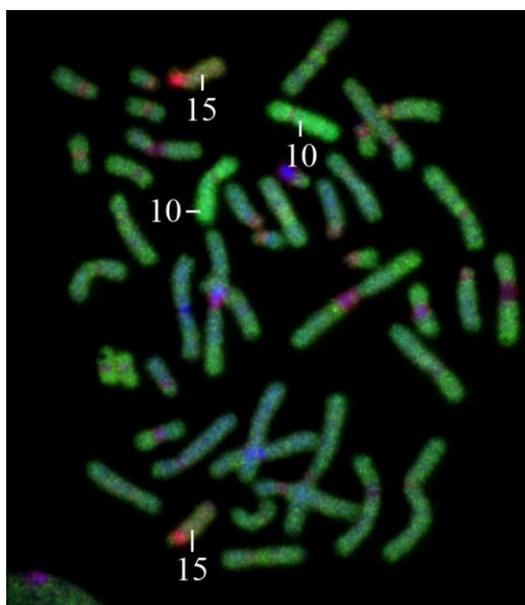


Рисунок 1.2. Двухцветная FISH ДНК-проб хромосом человека 15 (красный) и 10 (зеленый) с метафазными хромосомами. Цифровая микрофотография получена с помощью микроскопа AxioPlan 2 Imaging (Zeiss, Германия) с набором фильтров 49 (Zeiss, Германия), SP 101 FITC (CHROMA, США), SP 101 FITC (CHROMA, США), CCD-камерой (CV M300, JAI Corporation, Япония). Рисунок предоставлен сотрудником ИЦиГ СО РАН Карамышевой Т.В.

Отметим, что эта проблема цитогенетического анализа характерна в целом для всех эукариот. Причем, чем больше доля повторяющихся последовательностей в геноме исследуемых организмов, тем сложнее распознавать хромосомы и хромосомные районы, из которых была получена соответствующая ДНК-проба. Наличие в геноме исследуемого вида большого количества диспергированных повторов может сделать практически невозможным проведение FISH с ДНК-пробами, полученными из индивидуальных хромосом и их районов.

1.3 Экспериментальные методы улучшения выявления специфического FISH-сигнала ДНК-проб, полученных из целых хромосом или хромосомных районов

При анализе результатов FISH более существенным является не абсолютная интенсивность специфического сигнала, а его превышение над фоновым уровнем, который обусловлен как гибридизацией повторяющихся последовательностей ДНК-пробы с соответствующими повторами, так и техническими проблемами регистрации сигнала (Рубцов и др., 2003). Увеличение относительной интенсивности специфического сигнала может быть достигнуто либо уменьшением интенсивности сигнала диспергированных повторяющихся последовательностей, либо усилением интенсивности сигнала хромосомоспецифичных последовательностей ДНК (последовательностей, представленных только в конкретной хромосоме) (Rogan et al., 2001; Knoll et al., 2003, Dorman et al., 2013). Первая стратегия, как правило, основывается на подавлении гибридизации повторяющихся последовательностей. Вторая – на увеличении доли уникальных последовательностей в ДНК-пробе (количества уникальных последовательностей в ней).

1.3.1 Методы подавления гибридизации повторяющихся последовательностей

Первые попытки по снижению интенсивности сигнала повторяющихся последовательностей ДНК были связаны с выведением части присутствующих в ДНК-пробе повторов из гибридизации с ДНК хромосом на цитологическом препарате. В 1985 Силей с соавторами (Sealey et al., 1985) предложили

предварительно реассоциировать меченую ДНК с фрагментированной геномной ДНК изучаемого вида. В этом случае высоко- и умеренно повторяющиеся последовательности геномной ДНК с достаточно большой вероятностью связываются с меченой ДНК пробы, что позволяет блокировать их участие в последующей гибридизации. Такой подход был использован авторами для выявления уникальных и мало повторяющихся последовательностей при проведении саузерн-блот гибридизации (Sealey et al., 1985).

При использовании фрагментированной геномной ДНК существует вероятность, которой обычно пренебрегают, – связывания на этапе предгибридизации фрагментированной ДНК и уникальных последовательностей, представленных в ДНК-пробе (данный недостаток был отмечен в работах (Lichter et al., 1988) и (Zwick et al., 1997)).

Более эффективным оказалось проведение предварительной гибридизации меченой ДНК с Cot ДНК соответствующего вида (Nisson et al., 1991; Zwick et al., 1997; Schwarzacher, 2003). Термин “Cot” был введен Бриттенем и Коном (Britten, Kohne, 1968) как параметр, характеризующий фракцию нуклеиновых кислот по их кинетике реассоциации.

Значение Cot вычисляется по следующей формуле:

$$Cot = \frac{C_o * t}{4.98} \quad (1.1)$$

где C_o молярная концентрация продукта; t – время инкубации.

Позже Бриттен с коллегами ввел стандартную характеристику “Cot-1 ДНК” означающий фракцию ДНК, которая денатурируется час при 60°C с начальной концентрацией 89 мг/мл (Britten et al., 1976).

Чтобы получить Cot-1 ДНК, геномную ДНК фрагментируют ультразвуком, денатурируют и проводят ренатурацию до тех пор, пока значение Cot (см. выше формулу 1.1) не будет равно единице. Затем одноцепочечные фрагменты и участки ДНК гидролизуют с помощью экзонуклеазы (Landegent et al., 1987; Craig et al., 1997; Trifonov et al., 2017). Полученные последовательности ДНК содержат высоко и умеренно повторяющиеся последовательности ДНК и не содержат уникальных.

Первыми, кто применил немеченые Cot-1 ДНК для подавления гибридизации повторяющихся последовательностей ДНК-проб с ДНК цитологического препарата, были Ландегент и его коллеги (Landegent et al., 1987). Такой подход был реализован

при проведении FISH ДНК-проб, приготовленных на базе фрагментов ДНК человека, клонированных в бактериальных искусственных хромосомах (ВАС). Годом позже были опубликованы данные об использовании Cot-1 ДНК для предварительного отжига с денатурированной ДНК ДНК-проб, приготовленных из ДНК библиотек индивидуальных хромосом (Lichter et al., 1988). Этот прием получил очень широкое распространение при проведении разнообразных исследований и приобрел собственное название - CISS-гибридизация (*Chromosomal In Situ Suppression Hybridization*) (Lichter et al., 1988; Cremer et al., 1988). До настоящего времени CISS-гибридизация является традиционным подходом к решению проблемы подавления сигнала диспергированных повторяющихся последовательностей при проведении FISH с ДНК-пробами, созданными из индивидуальных хромосом или их протяженных районов (Boyle et al., 2011; Dorman et al., 2013).

К сожалению, разработка метода CISS-гибридизации не решила все проблемы проведения FISH районо- или хромосомоспецифичных ДНК-проб. Использование Cot-1 ДНК в CISS-гибридизации сталкивается, как минимум, с двумя проблемами:

1) В ряде случаев CISS-гибридизация не может быть проведена из-за трудностей получения Cot-1 ДНК (Rens et al., 2006). Для супрессии гибридации повторяющихся последовательностей требуется большой объем Cot-1 ДНК. К примеру, при проведении FISH с хромосомами млекопитающих требуется 50-100-кратный избыток Cot-1 ДНК относительно меченой ДНК (Lichter et al., 1988). Поскольку выход Cot-1 ДНК в зависимости от доли гетерохроматина в геноме составляет 10-15% от исходной геномной ДНК (Trifonov et al., 2017), для проведения FISH с одной хромосомоспецифичной ДНК-пробой (~0.4 мкг меченой ДНК) требуется 0.2-0.3 мг геномной ДНК. Для некоторых видов практически невозможно получить требуемого количества геномной ДНК. Наиболее часто такая проблема возникает при работе с мелкими организмами, получение большого количества которых затруднено. Отметим, что проблема доступности биологических образцов также актуальна для получения фрагментированной геномной ДНК.

2) Супрессия гибридации повторяющихся последовательностей приводит к исчезновению сигнала со всех повторяющихся последовательностей, в том числе и с хромосомоспецифичных (Rens et al., 2006). Доля хромосомоспецифичных

последовательностей невелика в геноме человека. По этой причине связывание этих последовательностей с Cot-1 ДНК не приводит к сильным искажениям результатов FISH. Но для других видов отсутствие сигнала с хромосомоспецифичных повторяющихся последовательностей может быть существенным недостатком.

На результат CISS-гибридизации влияет множество факторов, одним из которых является уровень супрессии гибридизации повторяющихся последовательностей. Сигнал в С-позитивных блоках гетерохроматина хромосом (участках ДНК, насыщенных повторами) может быть и достаточно интенсивным, и полностью отсутствовать (Рубцов и др., 2003; Rens et al., 2006; Trifonov et al., 2017). Также следует иметь в виду, что при неполной супрессии может появляться интенсивный сигнал в С-позитивных районах различных хромосом (Bolzer et al., 1999; Trifonov et al., 2017).

В некоторых случаях интенсивность сигнала FISH в районах, содержащих кластеры повторяющихся последовательностей, может отражать особенности источников ДНК, вовлеченных в исследование образца, ДНК которого использовалась для получения Cot-1 ДНК, и образца, хромосомы которого анализируются в эксперименте. Как правило, это обусловлено значительными вариациями в размере С-позитивных районов хромосом. Так у человека размер прицентромерных С-позитивных районов хромосом 1, 9, 16 и дистального района q-плеча Y-хромосомы может отличаться более, чем на порядок. Более того, в этих районах может значительно отличаться и состав ДНК даже у гомологичных хромосом. Разница в С-позитивных блоках в исследуемом образце и образце, из которого была получена Cot-1 ДНК, может быть достаточной для неполного подавления одного из типов повторов (Trifonov et al., 2017). Последующая FISH выявит интенсивный сигнал последовательностей, гомологичных этим повторам. К примеру, если Cot-1 ДНК человека была получена из тканей женщины, то после проведения супрессивной FISH районспецифичных ДНК-проб интенсивно окрашивается не только соответствующий район хромосомы, но и теломерный С-гетерохроматиновый блок длинного плеча Y-хромосомы (Рубцов и др., 2003). Схожая ситуация может возникнуть при использовании Cot-1 ДНК из образцов мужчины с делецией С-позитивного района Y хромосомы.

Как видно из выше изложенного, даже с хромосомами человека, кариотип и геном которого хорошо изучены, супрессия гибридизации повторяющихся последовательностей может привести к проблемам в интерпретации полученных результатов. Еще более сложной может оказаться интерпретация результатов CISS-гибридизации с хромосомами малоизученных видов животных и растений.

CISS-гибридизация часто не дает удовлетворительных результатов при исследовании на организмах с большим размером генома. В геномах таких видов количество повторяющихся последовательностей слишком велико, что не позволяет провести полную супрессию их гибридизации *in situ* (Fuchs et al., 1996; Houben et al., 2001; Schubert et al., 2001; Han et al., 2015). Кроме того, содержание уникальных последовательностей в участках хромосом может оказаться слишком низким для того, чтобы дать достаточно плотный интенсивный специфичный сигнал (Fuchs et al., 1996; Schubert et al., 2001). Проблемы проведения CISS-гибридизации с хромосомами растений и насекомых с большим размером генома до настоящего времени остаются не решенными.

1.3.2 Метод «дифференциального удерживания уникальных последовательностей»

В 1998 году был разработан метод «дифференциального удерживания уникальных последовательностей» (Hozier et al., 1998). Чтобы уменьшить долю повторяющихся последовательностей, связанных с ДНК цитологического препарата, было предложено проводить отмывку в условиях, обеспечивающих сохранение основной части дуплексов уникальных последовательностей, но приводящих к диссоциации повторяющихся последовательностей. Причинами нестабильности гибридов повторяющихся последовательностей, по-видимому, является дивергенция повторяющихся последовательностей ДНК. Вариабельность последовательностей ДНК внутри семейств повторов приводит к неполной гибридизации ДНК-пробы с ДНК препарата и, как следствие, к снижению стабильности дуплексов.

В отличие от CISS-гибридизации метод «дифференциального удерживания» позволяет обойтись без такой предварительной обработки ДНК-проб, как предгибридизация с конкурирующей ДНК. В связи с этим, он применим ко многим объектам исследования вне зависимости от доступности Cot-1 ДНК

соответствующего вида. К сожалению, использование метода «дифференциального удерживания» не позволяет достичь того уровня относительной интенсивности специфичного сигнала, который характерен для стандартной CISS-гибридизации (см. результаты применения метода в статье (Hozier et al., 1998)).

1.3.3 Относительное обогащение ДНК-пробы уникальными последовательностями

Проблема гибридизации районо- и хромосомоспецифичных ДНК-проб может быть решена созданием ДНК-проб, содержащих только уникальные последовательности, или, по крайней мере, в значительной степени обогащенных ими. При разработке методов сиквенс-независимой полимеразной цепной реакции для создания хромосомоспецифичных библиотек ДНК всегда использовались праймеры, которые обеспечивали предпочтительную амплификацию уникальных последовательностей. Это учитывали при выборе праймера для ПЦР с частично вырожденным праймером (Degenerate Oligonucleotide-Primed PCR – DOP-PCR) (Telenius et al., 1992), а также при выборе рестриктаз, используемых в ходе линкер-адаптер ПЦР (Linker-adapter PCR – LA-PCR) (Ludecke et al., 1989; Morozkin et al., 2011). И в том, и в другом случае протокол полимеразной цепной реакции предполагал предпочтительное вовлечение в амплификацию ГЦ-обогащенных участков генома, т.е. участков, обогащенных генами и уникальными последовательностями ДНК.

Несмотря на достигаемое перечисленными выше способами увеличение доли уникальных последовательностей в ДНК-пробах, содержание в них диспергированных повторов остается относительно высоким. Проведение FISH с полученными ДНК-пробами приводит к достаточно интенсивному сигналу на всех хромосомах. Для получения специфичного окрашивания целевой хромосомы или ее района обычно требуется проведение FISH в варианте CISS-гибридизации.

ДНК-пробы, практически не содержащие повторы, могут быть получены из кДНК библиотек с помощью метода препаративной гибридизации *in situ* (Preparative *in situ* hybridization, Pre-ISH) (Hozier et al., 1994). Идея метода заключается в следующем: 1) суммарная кДНК фрагментируется, обшивается праймерами и амплифицируется; 2) проводится гибридизация *in situ* амплифицированного материала с метафазными хромосомами; 3) с помощью микроманипуляционных

процедур собирается материал нужной хромосомы или интересующего хромосомного района; 4) из собранного материала выделяется ДНК, которая затем амплифицируется в присутствии соответствующих праймеров и метится в дополнительных циклах ПЦР.

В итоге получают ДНК-пробы с высокой долей уникальных последовательностей, которые гомологичны транскрибируемой ДНК соответствующей хромосомы или ее исследуемого района. К сожалению, создание таких высокоспецифичных ДНК-проб - процесс трудоемкий и дорогостоящий.

1.3.4 Методы элиминации из ДНК-клонотек повторяющихся последовательностей ДНК

Отдельное направление исследований представляет собой разработка методов элиминации повторяющихся последовательностей из ранее приготовленных ДНК-клонотек. Повторяющиеся последовательности могут быть исключены из ДНК-проб с помощью метода магнитного разделения макромолекул или с помощью супрессии амплификации повторяющихся ДНК в ПЦР.

Магнитная сепарация использует внешнее магнитное поле для эффективного разделения частиц с магнитным моментом из суспензии (Berensmeier, 2006; Safarik, Safarikova, 2009). Для того, чтобы молекулы ДНК стали чувствительны к магнитному полю, формируются комплексы этих молекул с магнитными нано- или микрочастицами. Очень часто для этих целей в повторяющиеся последовательности вводится биотин, проводится их гибридизация с ДНК ДНК-библиотеки, а затем дуплексы повторов удаляются с помощью магнитных частиц, покрытых стрептавидином. При добавлении таких магнитных частиц в раствор ДНК дуплексы повторов связываются с ними за счет взаимодействия биотин-стрептавидин.

Возможно “негативное” и “позитивное” разделение с помощью магнитных частиц (Bolzer et al., 1999). При позитивном отборе продуктом на выходе считаются молекулы ДНК, связанные с магнитными частицами (Chen-Liu et al., 1995; Rouquier et al., 1995; Overmyer et al., 1996; Masabanda, Griffin, 2003), при негативном – не связанные молекулы (Craig et al., 1997).

Рассмотрим принцип магнитной сепарации на примере работы Крейга с соавторами (Craig et al., 1997). Авторы предложили исключать повторяющиеся последовательности из набора ДНК-проб с помощью аффинной хроматографии. На

первом шаге ДНК исходной ДНК-клонотеки гибридизуется с Cot-1 ДНК, меченой биотином. Затем с помощью аффинной хроматографии дуплексы, связавшиеся с покрытыми стрептавидином магнитными шариками, отделяются от ДНК исходной ДНК-клонотеки (негативный отбор). Очищенный таким способом от повторяющихся последовательностей материал амплифицируется в ПЦР с соответствующими праймерами, что позволяет после введение в него метки получать ДНК-пробы с увеличенной долей уникальных последовательностей. Схема эксперимента представлена на рисунке 1.3.

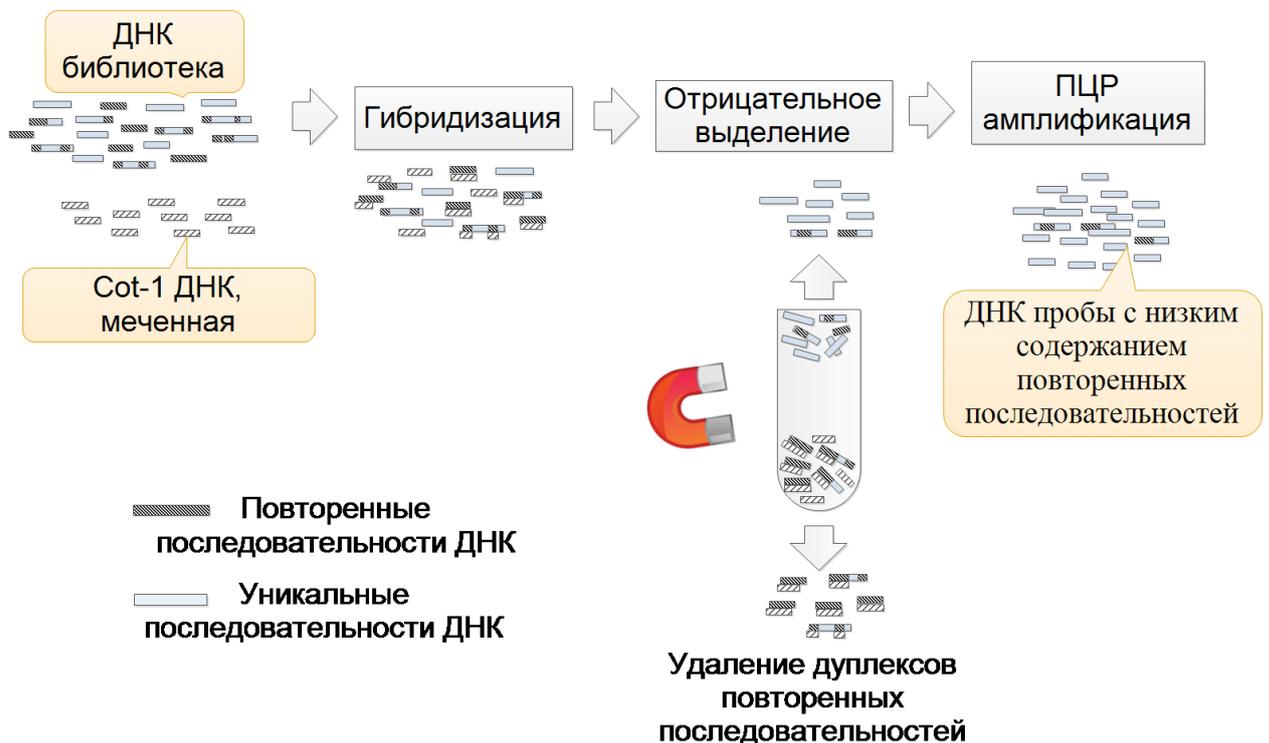


Рисунок 1.3. Метод построения ДНК-проб с малым содержанием повторяющихся последовательностей с помощью негативного магнитного разделения. Рисунок сделан на основе статьи (Craig et al., 1997).

Различными исследователями было показано, что использование таких ДНК-проб увеличивает относительную интенсивность сигнала уникальных последовательностей, почти так же эффективно, как и при CISS-гибридизации (Bolzer et al., 1999; Rauch et al., 2000). Следует учитывать, что в процессе магнитного разделения часть уникальных последовательностей исчезает вместе с повторяющимися последовательностями на стадии удаления дуплексов. Сильные потери сказываются на представленности уникальных последовательностей интересующей хромосомы или ее района в полученной ДНК-пробе. Последующая FISH с такими ДНК-пробами может привести к относительному усилению сигнала

в С-позитивных районах за счет сохранившихся в ДНК-пробе повторов, гомологичных многократно повторяющимся последовательностям в этих районах хромосом, и существенному снижению интенсивности специфичного сигнала в С-негативных районах.

В 2005 году Дуган и его коллеги предложили способ элиминации из ДНК-клонотек повторяющихся последовательностей с помощью подавления амплификации повторяющихся ДНК (Dugan et al., 2005). Для затруднения амплификации повторяющихся последовательностей, представленных в ДНК-пробах, при проведении ПЦР в реакционную смесь были добавлены Cot-1 ДНК. Cot-1 ДНК связываются с гомологичными последовательностями в ДНК-пробах, тем самым предотвращают связывание праймеров с ними и значительно снижают эффективность ПЦР в заблокированных районах. Продукт амплификации повторно метится и может быть использован в качестве ДНК-пробы.

При таком подходе Cot ДНК может быть вовлечена в амплификацию в случае неполной гибридизации как с ДНК-пробами, так и с другими последовательностями Cot ДНК (гибриды повторяющихся последовательностей). Не исключено, что последующее использование библиотек ДНК приведёт к незначительному усилению сигнала с ДНК соответствующих хромосом и хромосомных районов. Кроме того, могут возникнуть трудности при интерпретации результатов FISH.

Другой способ создания ДНК-проб с малым содержанием повторяющихся последовательностей был предложен в работе (Swennenhuis et al., 2012). Основные этапы подхода показаны на рисунке 1.4.

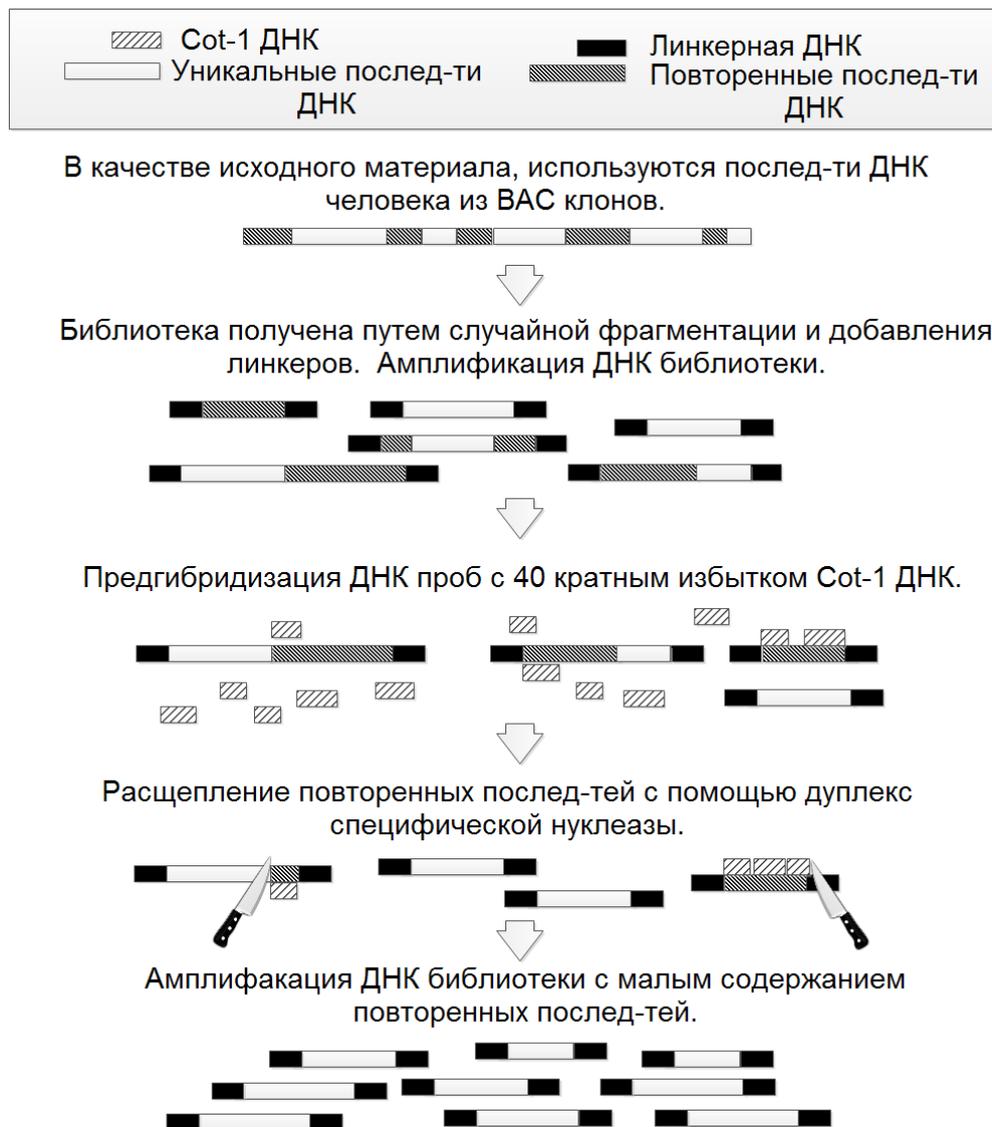


Рисунок 1.4. Этапы создания ДНК библиотеки с малым содержанием повторяющихся последовательностей, указанные в статье. Рисунок сделан на основе статьи (Swennenhuis et al., 2012).

Подход был разработан для работы с ДНК-клонотеккой, полученной из комплекта ВАС, но может с успехом применяться для ДНК-клонотек, полученных другими способами. В проведенном исследовании ДНК ВАС фрагментировали и фланкировали последовательностями, которые в дальнейшем могли быть использованы в качестве праймеров в ПЦР. До проведения ПЦР полученную ДНК денатурировали, а затем проводили отжиг в присутствии избытка Cot-1 ДНК. Последующая обработка нуклеазой, специфично расщепляющей двухцепочечную ДНК, предотвращала амплификацию фрагментов ДНК, содержащих повторы. Таким образом, амплификация ДНК и введение метки позволили получить ДНК-пробу, обедненную повторяющимися последовательностями. К сожалению, и этот подход

также не обеспечивает полного удаления из ДНК-пробы повторов и приводит к некоторому уменьшению представленности в ней уникальных последовательностей. На эти недостатки указывали сами авторы работы.

Следует подчеркнуть, что создание ДНК-проб с малым содержанием повторов оправдано при их дальнейшем многократном использовании в молекулярно-цитогенетических исследованиях. Иначе получение таких ДНК-проб и их последующие применение не дает значимых преимуществ по сравнению с CISS-гибридизацией. Также отметим, что использование Cot-1 ДНК для создания ДНК-проб с малым содержанием повторов приводит к трудностям, характерным для CISS-гибридизации (см. раздел 1.3.1). В частности, элиминация повторяющихся последовательностей из ДНК-проб не всегда можно провести из-за сложности получения Cot-1 ДНК.

1.3.5 Особенности анализа видов с секвенированным геномом

Информация о структуре генома исследуемого организма позволяет, с одной стороны, улучшить результаты рассмотренных выше методов, с другой, – разработать принципиально новые подходы. Например, затраты на проведение CISS-гибридизации могут быть снижены за счет замены Cot-1 ДНК на фрагментированную ДНК из районов хромосом, обогащенных повторяющимися последовательностями (Chung et al., 1997). Использование такого подхода предполагает корректную подборку районов хромосом в качестве источников конкурирующей ДНК. Указанное решение позволяет исключить сигнал последовательностей, гомологичных ДНК участков хромосом, из которых она получена.

Данные секвенирования открывает дополнительные возможности амплификации геномной ДНК. Примером может служить IRS-PCR (interspersed repetitive sequence system PCR) (Nelson et al., 1989). Указанный метод ПЦР широко используется для видоспецифической амплификации ДНК хромосом в гибридах соматических клеток. Его результаты полностью зависят от особенностей распределения диспергированных повторов в геноме. Метод предполагает вовлечение в амплификацию последовательностей, расположенных в геноме между диспергированными повторами. Использование праймеров, комплементарных концам Alu-последовательностей при амплификации ДНК человека или концам B1-

последовательностей в случае ДНК мыши, и направленных наружу от них, позволяет исключить значительную часть диспергированных повторов. Однако, IRS-PCR позволяет вовлечь в амплификацию только те фрагменты ДНК, которые локализованы между короткими диспергированными повторами (short interspersed elements – SINE), расположенными недалеко друг от друга. Это приводит к частичной и выборочной амплификации геномной ДНК: у млекопитающих из амплификации исключается ДНК конститутивного гетерохроматина, практически не содержащая последовательностей SINE; в меньшей степени в амплификацию вовлекается ДНК G-бэндов хромосом, как обедненная последовательностями SINE. Полученные таким способом ДНК-пробы при гибридизации с метафазными хромосомами не дают сигнала в C-позитивных районах хромосом и менее интенсивно окрашивают G-позитивные районы. Так же как и предыдущие методы относительного обогащения ДНК-проб уникальными последовательностями, IRS-PCR не исключает амплификацию повторяющихся последовательностей, и полученные с ее помощью ДНК-пробы при гибридизации *in situ* требуют проведения супрессии гибридизации повторяющихся последовательностей.

Одно из перспективных направлений улучшения результатов FISH связано с разработкой ДНК-проб, содержащих только специфичные уникальные последовательности, которые выбираются после биоинформатического анализа секвенированного и собранного генома. Впервые эта идея была реализована в 2001 году (Rogan et al., 2001). Роган и его коллеги предложили конструировать ДНК-пробы (single-copy FISH probes – scFISH пробы) для указанного участка генома с помощью специально разработанного программного обеспечения. Оно позволяло определять локализацию повторяющихся последовательностей в интересующего исследователя районе и исключать эти повторы из дальнейшего анализа. Для амплификации части оставшейся ДНК подбирались специальные праймеры. Полностью процесс разработки и получения scFISH проб представлен на рисунке 1.5.

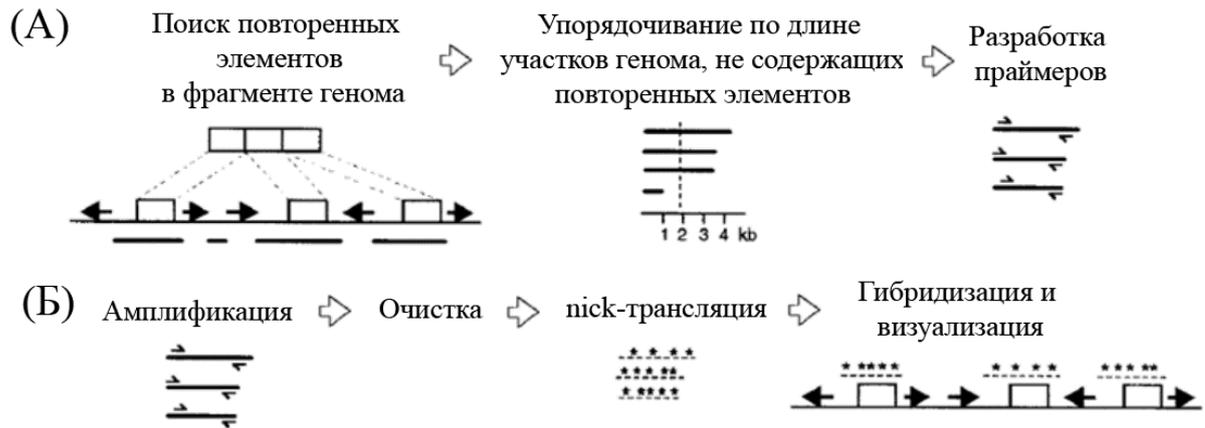


Рисунок 1.5 Схема процесса получения scFISH проб. **А** – Анализ последовательностей в отмеченном фрагменте генома. Последовательности ДНК в указанном участке генома сравниваются с базой данных повторяющихся последовательностей для определения позиции повторяющихся элементов (стрелки слева или справа), тем самым очерчивая границы уникальных интервалов (сплошные линии). Найденные последовательности затем сравниваются с последовательностями генома для определения потенциальных семейств генов, псевдогенов и других последовательностей с низким уровнем содержания повторов. Подбираются праймеры для амплификации фрагментов ДНК, длина которых превышает заданный порог (в данном случае 2 кб). **Б** – Уникальные участки амплифицируются с помощью ПЦР, очищаются и помечаются с помощью nick-трансляции. Рисунок взят из статьи (Rogan et al., 2001) с переводом обозначений на русский язык.

Предложенный принцип был использован в ряде исследований (Navin et al., 2006; Boyle et al., 2011; Beliveau et al., 2012; Han et al., 2015). Позже было показано, что при создании scFISH-проб не обязательно исключать все повторяющиеся последовательности. Анализ уровня их гомологии с известными семействами повторов позволял определить, какие из них могут быть включены в состав ДНК-пробы (Dorman et al., 2013).

Разработанные ДНК-пробы имеют два преимущества по сравнению с обычными ДНК-пробами:

- 1) Они оптимизированы по размеру, что обеспечивает их лучшее проникновение в фиксированные клетки и ткани;
- 2) scFISH-пробы высокоспецифичны и не содержат повторяющихся последовательностей, что делает возможным более полную их гибридизацию с ДНК-мишенью (Rogan et al., 2001; Beliveau et al., 2012).

Первые работы по конструированию ДНК-проб с помощью программного обеспечения позволяли окрашивать небольшие районы хромосом (около 100 тысяч пар нуклеотидов) (Rogan et al., 2001; Navin et al., 2006). Дальнейшее развитие этой

технологии привело к созданию ДНК-проб, способных “окрашивать” не только более крупные районы, но и С-негативные районы целых хромосом (Boyle et al., 2011; Beliveau et al., 2012; Han et al., 2015). При этом не возникало полного покрытия сигналом всей хромосомы. В части районов сигналы, сливаясь, давали сплошную зону окрашивания, в части – представляли собой набор отдельных сигналов, достаточно плотно покрывающих участок хромосомы. Отсутствие сигнала в С- позитивных районах объясняли как проблемами их секвенирования, так и составом их ДНК, состоящей преимущественно из повторяющихся последовательностей (в отличие от ДНК-проб, содержащих преимущественно ДНК из эухроматиновых районов (Knoll, Rogan, 2003)). Неполное покрытие сигналом всей хромосомы делает использование scFISH-проб в качестве хромосомоспецифичных ДНК-проб менее эффективным, чем CISS-гибридизация стандартных хромосомоспецифичных ДНК-проб. Однако scFISH-пробы могут быть очень полезны для визуализации конкретных хромосомных районов.

1.4 Компьютерный анализ FISH-изображений

Технология гибридизации *in situ* подразумевает анализ микроскопических изображений. Долгое время эта процедура проводилась цитогенетиками вручную (визуальная интерпретация изображений). Развитие компьютерных технологий полностью изменило способы анализа хромосом (Peng et al., 2012). Разработка методов компьютерной обработки результатов микроскопии после проведения CISS-гибридизации определила появление целого ряда новых методов хромосомной диагностики аномалий хромосом человека, включая 24-цветную FISH (Speicher et al., 1996; Schroeck et al., 1996), М-бэндинг (Chudoba et al., 1999); RxFISH (Muller et al., 1998); сравнительную геномную гибридизацию (CGH) (Kallioniemi et al., 1992; Weiss et al., 1996). В 2006 году был предложен подход к решению проблемы исключения сигнала повторяющихся последовательностей с помощью цифровой обработки FISH-изображений (Rens et al., 2006). Этот подход, а также важные понятия из цифровой обработки, основные задачи анализа FISH-изображений и их решения будут рассмотрены в этом разделе.

Для генерации цифровых изображений используются специальные регистрирующие устройства, которые преобразуют испускаемый флуорохромами свет в электрический заряд (аналоговый сигнал). В настоящее время в качестве

такого устройства обычно используется матрица цифровой камеры (Schier et al., 2011). Матрица состоит из сенсоров, каждый из которых имеет свои координаты и может с определенной периодичностью сообщать контроллеру о количестве принятых фотонов света. В дальнейшем аналоговый сигнал преобразуется с помощью процедуры дискретизации и квантования в цифровую форму. Результатом таких преобразований является отображение непрерывных значений интенсивности сигнала в целые числа. Обычно для представления интенсивности выбирают 8-битные числа, допускающие диапазон значений от 0 до 255.

Изображение, полученное в результате регистрации сигнала от одного флуорохрома называется **монохромным**. Такое изображение представляет собой двумерный массив целых чисел, каждый элемент которого отражает количество фотонов, зарегистрированных в соответствующих сенсорах (Ljosa, Carpenter, 2009; Stuurman, Swedlow, 2012). Элементы этого массива называют **пикселями изображения** (pixel производное от “picture element”). Так как флуорохромы связаны с ДНК-пробой, то интенсивность сигнала в каждом пикселе FISH-изображения коррелирует с количеством последовательностей ДНК в цитологическом препарате, гомологичных последовательностям в ДНК-пробах, находящемся в соответствующем месте.

При регистрации сигналов с нескольких флуорохромов формируются цветные изображения. **Цветное изображение** является двумерный массив векторов. (каждому пикселу соответствует вектор значений). В наиболее часто встречающемся случае цветных изображений вектор содержит три элемента: красная, зеленая и синяя компонента пиксела. А цифровое изображение описывается серией из трех или более монохромных изображений. Цветное изображение может быть сформировано за счет регистрации в отдельном канале интенсивности сигнала каждого отдельного флуорохрома. Более сложным вариантом формирования изображения является задание цветов в зависимости от комбинации сигналов флуорохромов (Speicher et al., 1996; Gariniy et al., 1996; Tanke et al., 1999).

В большинстве случаев при приготовлении препаратов хромосом они сохраняются в виде относительно компактных, распластанных на плоскости структур, в то время как остальные элементы цитоплазмы превращаются в тонкую и слабоокрашиваемую пленку. Но высота хромосом обычно невелика и не превышает

половины микрона (считается, что чем ниже их высота, тем лучше прошло распластывание хромосом на стекле) (Hisco et al., 1997).

Поэтому при обработке изображений FISH метафазных хромосом используются методы анализа двумерных цветных изображений, хотя в микроскопии существуют методы, использующие трехмерное и четырехмерное представление данных (позволяющей следить за изменением трехмерной организации хромосомы в живой клетке во времени) (Cohen, 2014). Выбор методов цифровой обработки и очередность их использования для извлечения количественной информации диктуется специфическими особенностями решаемой задачи (Jang et al., 2014). Обычно процесс анализа изображений хромосом включает в себя следующие этапы: предобработка, сегментация изображений, подсчет характеристик, классификация объектов (Munot et al., 2012; Nair et al., 2015). Рассмотрим каждый из этапов в отдельности.

1.4.1 Предобработка изображений

Качество изображения с результатами флуоресцентной гибридизации *in situ* зависит от многих факторов: локальных условий гибридизации, условий съемки, характеристик оптической системы, характеристик регистрирующих устройств и т.д. (Waters, 2009). На любом из этапов формирования изображения могут возникнуть явления, которые приводят к различным искажениям регистрируемой интенсивности сигнала. Это затрудняет как их визуальный анализ человеком, так и автоматическую компьютерную обработку. Наиболее частыми вариантами искажений является слабый контраст изображения, изменение интенсивности сигнала из-за шумов или неравномерного освещения, пространственный сдвиг изображения в одном или нескольких каналах при цветном изображении. Необходимо отметить, что большинство из них можно избежать, соблюдая протокол проведения FISH и правильную настройку микроскопа.

Полное восстановление исходного изображения по искаженному является сложной задачей, так как требует точного знания математической модели формирования изображения и искажений (Визильтер и др., 2010, с. 108). По этой причине на практике применяются методы, направленные на устранения конкретных искажений. Целью этапа предобработки является построение некоторого приближения исходного изображения и повышение качества

изображения для упрощения последующей цифровой обработки. Распространенными задачами данного этапа при анализе изображений хромосом являются повышение контраста изображений и уменьшение шумов, присутствующих на них (Yan, 2011).

Шум на изображении проявляется в виде отклонения значения интенсивности от истинного (действительного) значения. Существуют большое количество методов, разработанных для борьбы с различными типами шумов (такие методы называют методами фильтрации). Обычно делаются предположения в пользу определенной модели шумов и применяются методы фильтрации, позволяющие исключить такого рода шум. Некоторые подходы к оценке типа шума на изображениях можно найти в работе (Гонсалес и Вудс, 2005, с. 343-345).

Наиболее популярным методом подавления шумов на FISH-изображениях является медианный фильтр (Wang et al., 2008a; Nair et al., 2015; Shirazi et al., 2016). Метод, несмотря на свою простоту, эффективен в борьбе с импульсными шумами (результатом влияния импульсных шумов является замена значения интенсивности сигнала в некоторых пикселей изображения на случайную величину) и при этом сохраняет границы хромосом.

Другой распространенной проблемой анализа хромосом является слабый контраст изображений и, соответственно, нечеткие границы хромосом (Yan, 2011; Nair et al., 2015). Увеличение контрастности изображения облегчает определение границ хромосом и бэндов. **Бэндом** называют участок хромосомы, отличающийся по интенсивности окрашивания от соседних (An International System for Human Cytogenetic Nomenclature, 2013). Таким образом, предобработка повышает точность сегментации и классификации объектов.

Контрастность всего изображения может быть увеличена методом эквализации гистограммы (Гонсалес и Вудс, 2005, с. 150-158). Но более оправданным является усиления непосредственно границ хромосом и бэндов за счет применения алгоритмов выделения контуров изображений или вейвлет анализа. Краткое описание указанных методов можно найти в обзорных статьях (Yan, 2011; Nair et al., 2015).

1.4.2 Сегментация изображений

Сегментация изображения обычно определяется как разбиение исходного изображения на области, каждая из которых удовлетворяет некоторому критерию схожести. В хромосомном анализе под сегментацией понимается разделение изображения на области, занятые отдельными хромосомами и фоном (участки изображения вне хромосом) (Nair et al., 2015).

Изображения хромосом по интенсивности сигнала отличаются от фонового уровня (Wang et al., 2008a; Arachchige, 2014). Поэтому на гистограмме интенсивности такого изображения, как правило, видно два пика. Один из них соответствует фону, а другой - объекту (Netten et al., 1997; Roshtkhari, Setarehdan, 2008; Arachchige et al., 2010; Jahani, Setarehdan, 2011). Очевидный способ выделения объектов (хромосом) из окружающего фона состоит в выборе значения порога, разграничивающего моды гистограммы интенсивности. Пиксели со значением интенсивности в пределах определенного диапазона отмечаются как принадлежащие объекту (хромосоме). Пиксели, выходящие за этот диапазон, принадлежат фону (Гонсалес и Вудс, 2005, с. 851). Методы, позволяющие определить такой порог, называются методами пороговой сегментации.

Полетти и соавторы провели в 2012 году сравнение одиннадцати методов пороговой сегментации (Poletti et al., 2012a). Авторы провели предварительную сегментацию изображений метафазных хромосом вручную и сравнили результат с результатами применения методов пороговой сегментации. Лучшие результаты продемонстрировали методы локальной пороговой сегментации, особенно метод адаптивной сегментации (AdT - "adaptive thresholding") (Grisan et al., 2009). Метод разбивает исходное изображение на подобласти, в каждой из которых для сегментации используется свое значение порога. Полученные области соединяют с помощью билинейной интерполяции. Преимущество методов локальной пороговой сегментации можно объяснить тем, что они менее чувствительны к некоторым шумам и неравномерному освещению (Гонсалес и Вудс, 2005, с. 858-859).

Методы сегментации с глобальным порогом также продемонстрировали неплохие результаты при сегментации хромосом. В этой группе лучшие результаты показал метод Отсу (Otsu, 1979). Отметим, что этот метод часто используют в качестве первого этапа сегментации хромосом (Minaee et al., 2014).

Но даже в случаях оптимально выбранного порога, сегментация не всегда дает хорошие результаты. В первую очередь это связано с проблемами качества изображения: неоднородность интенсивности фоновой флуоресценции, высокая изменчивость интенсивности сигнала на хромосомах, наложение хромосом и их близкое расположение в метафазной пластинке (Ji, 1994; Restif, 2006; Jahani et al., 2011; Poletti et al., 2012a; Minaee et al., 2014). Поэтому пороговую сегментацию часто рассматривают в качестве лишь первого шага сегментации хромосом (Poletti et al., 2012a). Стоит отметить, что на первом этапе можно применять и другие методы сегментации, однако, эффективность и простота реализации именно методов пороговой сегментации, делает их более предпочтительными.

Последующие шаги обработки применяются не к целому изображению, а к отдельным связным областям на нем (объектам). Под **связной компонентой изображения или маской объекта** понимают такое множество пикселей на изображении, в котором:

- все пиксели однородны по некоторому признаку. Очень часто это ненулевое значение интенсивности сигнала.
- между любыми двумя пикселями, принадлежащими данной области, существует непрерывный путь, состоящий из пикселей, также принадлежащих данной области и являющихся при этом «соседями» в смысле заданного отношения соседства.

Процедура поиска связных компонент заключается в формировании маркированного изображения. Каждый пиксель такого изображения содержит метку связной компоненты, к которой он принадлежит. Метка представляет собой некоторый идентификатор, используемый в качестве уникального имени объекта (Шапиро, Стокман, 2006, с. 85).

После применения процедуры сегментации на изображении метафазной пластинки обнаруживаются три класса объектов: отдельные хромосомы, кластеры хромосом и ложные объекты (Ji, 1994; Arora, Dhir, 2016).

Ложные объекты — это объекты, которые необходимо удалить из дальнейшей обработки. К этому классу объектов можно отнести ядра клеток, не полностью попавшие в кадр хромосомы, посторонние флуоресцирующие объекты и

т.п. Часто указанные объекты распознаются на основе анализа их геометрических характеристик: размера и округлости (Ji, 1994; Sreejini et al., 2012).

Кластеры хромосом– объекты, содержащие несколько хромосом. Обычно они образуются в результате наложения или соприкосновения разных хромосом. Кроме того, кластеры могут появиться при неоптимальном пороге сегментации.

Вторым шагом сегментации является разделение кластеров на отдельные хромосомы. Этот этап включает в себя решение следующих задач: обнаружение кластеров хромосом на изображении, разделение кластеров на однородные области, соединение однородных областей в отдельные хромосомы. Выделяют три группы методов разделения кластеров хромосом: геометрические, спектральные и спектрально-геометрические (Karvelis et al., 2010; Zhao, Kong, 2013; Zhao et al., 2013).

Первая группа методов (геометрические) основана на анализе формы объектов. Часто задача разделения кластеров сводится к поиску точек разреза и соединения их разделительными линиями. **Точками разреза** называют пиксели, расположенные на границе хромосомного кластера рядом с областью соприкосновения или наложения хромосом. В таблице 1.1 перечислены основные методы этой группы и их ограничения. Ниже приведено краткое описание каждого из методов. Для сравнения эффективности методов в таблице приводятся значения CSTC (correct separation of touching chromosome):

$$CSTC = 100 \frac{ChrsCorrectlySeparated}{ChrsInClusters} \quad (1.2)$$

где *ChrsCorrectlySeparated* – количество правильно разделенных хромосом; *ChrsInClusters* – количество хромосом в кластерах.

Таблица 1.1 – Геометрические методы разделения кластеров хромосом

Описание метода	Размер тестовой выборки	Точность метода (CSTC, %)	Ограничения
<p>Разделение соприкасающихся хромосом с помощью “светлого” пути. Наложенные хромосомы разделяются с помощью точек разреза, расположенных рядом с точками ветвления скелета объекта. Граница кластера хромосом максимально вогнута в этих точках (Ji, 1994)</p>	<p>695 изображений метафазных хромосом. Общие количество хромосом составило 30998. Из них 10811 образуют кластеры.</p>	<p>84%</p>	<p>Работает с кластерами, состоящими только из двух хромосом</p>
<p>В качестве точек разреза рассматриваются пиксели, в которых вогнутость границы объекта максимальна. Окончательный выбор точек и линий разреза осуществляется на основе упрощенного представления форм хромосом и проверки различных вариантов наложения хромосом (Agam, Dinstein, 1997).</p>	<p>124 изображения кластеров хромосом.</p>	<p>Общая точность составляет 82%. Для кластеров, состоящих из двух хромосом 87%</p>	<p>Плохо работает на кластерах, содержащих небольшие хромосомы, либо состоящих из хромосом с большой зоной соприкосновения.</p>
<p>В работе рассматриваются несколько вариантов сегментации/разделения хромосом с помощью методов водораздела. В одном из вариантов оставшиеся после сегментации большие объекты разделяются с помощью “светлого” пути (Karvelis et al., 2005).</p>	<p>Общие количество хромосом составило 940. В том числе 396 соприкасающихся и 29 пересекающихся хромосом.</p>	<p>Типы водораздела: не рекурсивный – 47%, рекурсивный – 82%. Комбинация рекурсивного метода водораздела и метода “светлого” пути – 92%</p>	<p>Возможно ошибочное разделение единичного объекта (пересегментация).</p>

Описание метода	Размер тестовой выборки	Точность метода (CSTC, %)	Ограничения
<p>В качестве точек разреза используются экстремумы функции кривизны границы кластера, расположенные рядом с центром области пресечения хромосом. Центр определяется с помощью диаграмм Вороного, а окончательный вариант точек разреза - с помощью триангуляции Делоне. (Srisang et al., 2011).</p>	35 изображений наложенных хромосом	80%	Работает только с X-образными кластерами.
<p>Разделение соприкасающихся хромосом с помощью “светлого” пути. В более сложных случаях рассматривались различные варианты разделения хромосом в зависимости от формы кластера (Moallem et al., 2013).</p>	26 изображений наложенных хромосом и 60 соприкасающихся хромосом	Разделение соприкасающихся хромосом – 70%, разделение наложенных хромосом – 67%	Работает с кластерами, состоящими только из двух хромосом.
<p>Метод применяет триангуляцию Делоне ко всем пикселям на границе кластера хромосом. Вершины больших по площади треугольников являются точками разреза (Munot et al., 2013)</p>	60 изображений наложенных хромосом	75-100% Для кластеров, состоящих из двух хромосом – 100%	Плохо работает на кластерах, содержащие сильно изогнутые или небольшие хромосомы. Не работает с I-образными кластерами.
<p>Точки разреза определяются на основе дескрипторов границ. Значение функция кривизны в этих точках выше среднего, а сумма расстояний до других граничных пикселей наименьшая (Minae et al., 2014)</p>	25 изображений метафазных хромосом. Общее количество хромосом составило 1150. Из них 62 образуют кластеры	91.9%	Трудозатратность

Автор первого из перечисленных в таблице методов предложил подход для анализа результатов G-дифференциального окрашивания хромосом. Джи рассматривал пиксели, в которых граница кластера хромосом максимально вогнута, как потенциальные точки разреза и пытался их соединить разделительной линией, проходящей через области с высокой интенсивностью (Ji, 1994). Построенную разделительную линию назвали “светлым” путем (от английского “pale path”). Такой подход позволяет разделить соприкасающиеся хромосомы. В случае кластера, состоящего из наложенных хромосом, для разделения использовался скелет объекта (множество точек, равноудаленных от границы объекта). Автор предположил, что точки ветвления скелета (точки с количеством соседей больше трех) указывают на область наложения хромосом. Для разделения кластера хромосом проводится поиск точек ветвления и четырех точек разреза рядом с ней. Описанный подход к разделению объектов может быть легко адаптирован для анализа изображения хромосом, окрашенных красителем DAPI (краситель 4', 6'-диамидино-2-фенилиндол. Аббревиатура от английского названия “4,6-diamidino-2-phenyl-indole”). Для этого изображения необходимо предварительно инвертировать, либо проводить “светлый” путь через области с низкой интенсивностью.

Алгоритм, предложенный Джи, был модифицирован в работе Муаллем с коллегами (Moallem et al., 2013). Авторы учитывали тип кластера (X-образная и T-образная форма кластеров) для разделения наложенных хромосом. Форма кластера определялась с помощью концевых точек скелета кластера. Каждому типу кластера соответствуют разное количество точек разреза и различные варианты проведения разделительных линий. Оптимальный вариант разделения кластера и последующие объединения регионов одной хромосомы выбирался на основе анализа формы полученных объектов с данными об одиночных хромосомах (критерий схожести). К сожалению, с помощью ранее предложенных методов нельзя разделить кластеры, содержащие больше двух хромосом (de Faria et al., 2005; Zhao, Kong, 2013).

Хорошие результаты были получены с помощью комбинации метода “светлого” пути и метода водораздела (Karvelis et al., 2005). Метод водораздела построен на аналогии затопления геологического ландшафта. В топографическом представлении изображения, значения интенсивности сигнала в каждом пикселе

представляют собой значения высоты поверхности. Метод проводит “затопление” с локальных минимумов и постепенно добавляет к ним пиксели на основе значения интенсивности (наращивание бассейнов). Для препятствия слиянию двух бассейнов проводится линия водораздела (границы областей). Следует отметить, что метод водораздела довольно часто используется для разделения кластеров хромосом (Wenzhong, Xiaohui, 2010).

Агам и Динштейн также использовал пиксели, в которых вогнутость границы объекта максимальна, для построения всевозможных разделительных линий (Agam, Dinstein, 1997). Особенностью их работы является упрощенные представления формы хромосомы в виде прямоугольников для выбора точек и линий разреза. Метод работает с любыми кластерами хромосом, однако, его точность снижается с увеличением количества хромосом в кластере. Кроме того, возникают трудности разделения хромосом с большой зоной соприкосновения и кластеров, которые содержат хромосомы небольшого размера (de Faria et al., 2005).

Точность разделения кластеров можно повысить за счет информации о бэндах хромосом. К примеру, поиск совпадений фрагментов хромосом в заранее известных шаблонах различных хромосом (Charters, Graham, 1999). Но это сужает область применения метода до анализа хромосом конкретного вида.

Точки разреза могут быть найдены с помощью триангуляция Делоне. Впервые такой подход применили Срисанг и его коллеги (Srisang et al., 2011). В работе на первом этапе проводился анализ функции кривизны границы объектов. Экстремумы этой функции считались потенциальными точками разреза. Окончательно в качестве точек разреза выбирались точки, расположенные наиболее близко к центру области пересечения хромосом. Центр указанной области определялся с помощью диаграмм Вороного, а возможные варианты точек разреза - с помощью триангуляции Делоне. Как отмечают сами авторы, метод работает только с X-образными кластерами, и не работает T- и I-образными вариантами. Идея использования триангуляции Делоне получила свое развитие в работе Мунота и его соавторов (Munot et al., 2013). Авторы применили триангуляцию ко всем пикселям, расположенных на границе кластера хромосом, и заметили, что каждой области пересечения хромосом соответствуют два треугольника наибольшего размера. Вершины таких треугольников являются точками разреза.

Оригинальный подход по разделению кластеров хромосом, работающий на любых типах изображений, предложил Мина с коллегами (Minae et al., 2014). Кластеры определяются при помощи дескрипторов областей, в которых указаны отношение главных осей, анализ выпуклости, подсчет концевых точек скелета. Точки разреза определяются на основе дескрипторов границы: оценка функции кривизны в пикселе и сумма расстояний до других пикселей границы объекта. Авторы метода считают, что в точках разреза значения функции кривизны выше среднего значения, а сумма расстояний до других пикселей границы - наименьшая.

С появлением многоцветной FISH новым источником информации для разделения хромосом стал цвет. Гомологичные пары хромосом в FISH-изображениях имеют идентичные спектральные характеристики (“окрашивается” в определенный цвет) за счет комбинации флуорохромов. Спектральные методы для разделения кластеров хромосом используют информацию об интенсивности сигналов различных флуорохромов. Задача разделения сводится к классификации объектов (Schwartzkopf et al., 2005; Karvelis et al., 2008; Sreejini et al., 2012) или пикселей (Schwartzkopf et al., 2001; Sampat et al., 2002; Schwartzkopf et al., 2002; Choi et al., 2004; Choi et al., 2006a) на основе цветовой информации. Методы отличаются алгоритмами классификации и выбранными признаками объектов (характеристики, на основании которых проводится классификация).

В таблице 1.2 приведены результаты применения методов и краткое их описание. Точность классификации оценивалась как:

$$Accuracy = 100 \frac{ObjsCorrectlyClassificated}{CountOfObjs} \quad (1.3)$$

где *ObjsCorrectlyClassificated* – количество правильно классифицированных объектов или пикселей; *CountOfObjs* – общее количество объектов или пикселей.

Исходя из данных, приведенных в таблице 1.2, можно было бы сделать вывод о превосходстве методов классификации пикселей над методами классификации регионов. Однако, в действительности это не так (Sreejini et al., 2012; Zhao, Kong, 2013). Завышенная точность методов классификации пикселей объясняется малым размером тестовых выборок.

Таблица 1.2 – Спектральные методы разделения кластеров хромосом

Описание метода	Размер тестовой выборки (количество изображений метафазных хромосом)	Точность классификации (Accuracy, %)
Классификация пикселей хромосом на основе цветовых характеристик. Обнаружение и разделение кластеров осуществляется с помощью энтропии Шеннона (Schwartzkopf et al., 2001; Schwartzkopf et al., 2002).	200	70-80%
Байесовская классификация пикселей изображения (Sampat et al., 2002).	5	91.4%
Классификация пикселей изображения по методу максимального подобия (Choi et al., 2004).	5	89.8%
Классификация пикселей хромосом разными способами: байесовская классификация, метод К-ближайших соседей (при k=1, k=3, k=7, k=9) (Sampat et al., 2005).	5	Байесовская классификация - 83% Метод К-ближайших соседей 90.5% (при k=7)
Классификация объектов с использованием функции максимального подобия. При обнаружении кластеров проводится классификация его пикселей и их объединения в целые хромосомы на основе значений функции максимального подобия (Schwartzkopf et al., 2005).	200	Общая точность составила 91.9%. Точность классификации соприкасающихся хромосом – 77%, а наложенных – 34%.
Классификация пикселей масок хромосом на основе нечеткой логики (Choi et al., 2006a).	9	88%
Для более точного определения границ хромосом объединяются результаты двух методов: метода пороговой сегментации Отсу и варианта метода водораздела, работающий с цветными изображениями. Байесовская классификация объектов. Объединение близкорасположенных объектов одного класса в один объект (Karvelis et al., 2008).	183	82.5%
Сегментация изображения с помощью метода водораздела. Байесовская классификация объектов. Объединение близкорасположенных объектов одного класса (Sreejini et al., 2012).	40	84.2%

Главный недостаток спектральных методов заключается в том, что они не могут разделить кластеры, образованные из гомологичных хромосом (Zhao, Kong, 2013). Устранить указанный недостаток и повысить точность разделения можно с использованием геометрической информации.

Методы, использующие геометрическую и цветовую информацию, называются спектрально-геометрическими. Обычно эти методы улучшают точность разделения хромосом за счет комбинации геометрической и цветовой информации.

К примеру, Карвелис и его коллеги изменили способ построения разделительных линий в кластерах хромосом (Karvelis et al., 2010). Вместо традиционного поиска “светлого” пути линии разреза проходят через области с высоким значением градиента. Такой способ построения назвали градиентным путем.

Choi и его коллеги разработали новый подход к разделению кластеров хромосом (Choi et al., 2006b). Авторы определяют тип кластера хромосом на основе анализа границ и формы его скелета: X-образный, T-образный, I-образный. В зависимости от типа формы кластера выдвигаются различные гипотезы наложения хромосом и их последующего разделения. Для каждой гипотезы проводится оценка ее вероятности на основе размера и цветовой информации районов кластера хромосом, которые в случае верности рассматриваемой гипотезы являются хромосомами. Выбирается наиболее вероятная гипотеза.

Чжао и его соавторы предложили использовать цветовую информацию на стадии объединения отдельных районов кластера в целые хромосомы (Zhao et al., 2013). Области относили к одному классу на основе цветовых характеристик и размеров объединенных областей. Позже этот метод был модифицирован (Zhao, Kong, 2013). В новом варианте цветовая информация использовалась не только для объединения схожих областей, но и для разделения хромосом на однородные по цвету области с помощью алгоритмов кластеризации и классификации.

Как видно из приведенных примеров, использование геометрической и цветовой информации значительно упрощает процедуру разделения кластеров хромосом. При этом точность разделения остается на достаточно высоком уровне (см. Таблицу 1.3).

Таблица 1.3 – Спектрально-геометрические методы разделения кластеров хромосом

Описание метода	Размер тестовой выборки	Точность метода (CSTC, %)
<p>Определяется тип формы кластера хромосом на основе его скелета и границ. В зависимости от типа формы кластера выдвигаются различные гипотезы наложения хромосом и их последующего разделения. Выбирается наиболее вероятной гипотезы наложения хромосом (Choi et al., 2006b)</p>	<p>Общие количество хромосом составило 562. Из них 134 образуют кластеры.</p>	<p>Общая точность составляет 96.7%. Для кластеров, состоящих из двух или трех хромосом – 89%. При большем количестве хромосом в кластере – 67%</p>
<p>Разделение кластеров с помощью метода водораздела и градиентного пути. Объединение объектов в целые хромосомы проводится на основе результатов байесовской классификации объектов (Karvelis et al., 2010)</p>	<p>1178 изображений соприкасающихся хромосом и 189 наложенных хромосом</p>	<p>Разделение соприкасающихся хромосом – 90.6%, разделение наложенных хромосом – 80.4%</p>
<p>Геометрическое разделение кластеров: выполняется поиск точек разреза и соединение их линией. Объединение объектов осуществляется в зависимости от значений функции схожести объектов (учитывается площадь объекта и спектральная информация). Точность метода повышается за счет алгоритма поиска гомологичных хромосом (Zhao et al., 2013)</p>	<p>200 изображений кластеров хромосом</p>	<p>Разделение соприкасающихся хромосом – 79%, разделение наложенных хромосом – 73%. Для кластеров, состоящих из двух хромосом точность составила 80%. Для кластеров, состоящих из трех хромосом – 74%. При большем количестве хромосом – 71%</p>
<p>Разбиение кластеров на однородные объекты на основе цветовой информации (алгоритм классификации и многоканального объединения районов). Объединение объектов и поиск гомологичных хромосом проводится также как в работе (Zhao et al., 2013).</p>	<p>200 изображений кластеров хромосом</p>	<p>Разделение соприкасающихся хромосом – 85%, разделение наложенных хромосом – 79%. Для кластеров, состоящих из двух хромосом 93% Для кластеров, состоящих из трех хромосом 85%. При большем количестве хромосом – 83%</p>

1.4.3 Классификация объектов

Под **классификацией** понимают процедуру принятия решения о принадлежности объекта к одному из классов согласно данным ряда измерений отличительных **характеристик (признаков)** объектов. Соответственно, **классом** называют совокупность объектов, обладающих общими свойствами (схожими характеристиками) (Гонсалес и Вудс, 2005, с. 983-984).

Очень часто в литературе под задачей классификации хромосом подразумевается задача кариотипирования. Кариотипирование является стандартной задачей цифровой обработки. Для ее решения были испробованы многие современные подходы:

- классификация на основе нейронных сетей (Lerner, 1998; Badawi et al., 2003; Moradi, Setarehdan, 2006; Wang et al., 2008a; Wang et al., 2009; Rungruangbaiyok, Phukpattaranont, 2010; Ming, Tian, 2010; Poletti et al., 2012b);
- байесовская классификация (Karvelis et al., 2008; Srejini et al., 2012);
- классификация с помощью функции максимального подобия (Schwartzkopf et al., 2005);
- классификация на основе нечеткой логики (Choi et al., 2006a);
- метод опорных векторов (Kou et al., 2002; Markou et al., 2012; Arachchige et al., 2013);
- метод k- ближайших соседей (Sampat et al., 2005; Serbanescu, 2010).

В таблице 1.4 указана точность различных подходов классификации хромосом. Точность оценивалась по формуле 1.3. Более детальное сравнение подходов представлено в статьях (Wang et al., 2009; Nair et al., 2015; Abid, Hamami, 2016).

В основном для классификации хромосом используют три категории характеристик: геометрические признаки, особенности распределения бэндов на хромосоме и цветовую информацию (Jahani, Setarehdan, 2012; Zhao et al., 2013; Mohammadi, 2014; Pravina, 2015). Геометрические признаки базируются на морфологических особенностях хромосом: размеры хромосомы и положение центромеры. Длина хромосомы (количество точек в срединной линии) и центромерный индекс (отношение короткого плеча хромосомы ко всей ее длине) являются наиболее важными и широко используемыми геометрическими

характеристиками (Yan, Li, 2012; Jahani, Setarehdan, 2012; Mohammadi, 2014; Nair et al., 2015; Abid, Namami, 2016). Эти характеристики вычисляются с использованием параметра **серединная (центральная) линия**, которая представляет собой геометрическое место точек плоскости, равноудаленных от границы объекта. Более подробную информацию о методах построения серединной линии можно найти в работах (Groen et al., 1989; Piper, Granum, 1989; Lerner et al., 1995; Kao et al., 2008; Arachchige et al., 2013; Abid, Namami, 2016).

Таблица 1.4 –Подходы к классификации хромосом и их точность

Группа методов (подход)	Точность методов классификации из этой группы (Accuracy, %)
Классификация на основе нейронных сетей	61.3-98%
Байесовская классификация	72.3-84.2%
Классификация с помощью функции максимального подобия	69%
Классификация на основе нечеткой логики	89%
Метод опорных векторов	90-93%
Метод k- ближайших соседей	79.1-90.5%

Применение только геометрических характеристик не дает надежной идентификации хромосом внутри кариотипа (Schwartzkopf et al., 2005; Markou et al., 2012). Для повышения точности идентификации используется также анализ распределения бэндов хромосомы. При дифференциальном окрашивании хромосом бэнды обычно представляют собой однородные полосы, идущие поперек хромосом. Каждой хромосоме соответствует полосчатая структура (комплекс поперечных меток, или бэндов), отражающая линейную неоднородность хромосомы и специфичная для гомологичных пар хромосом и их участков (Nair et al., 2015). Важными характеристиками становятся расположение бэндов на хромосоме, их интенсивность и ширина (Moradi, Setarehdan, 2006). Традиционно обнаружение бэндов (построение полосчатой структуры) осуществляется с помощью **профиля**

плотности. Для его построения проводятся перпендикуляры к серединной линии и для каждого перпендикуляра вычисляется среднее значение интенсивности сигналов в принадлежащих ему пикселях (Piper, Granum, 1989; Abid, Hamami, 2016). Профиль плотности состоит из полученных средних значений. Различные варианты характеристик, основанных на профиле плотности, описаны в работах (Groen et al., 1989; Piper, Granum, 1989; Stanley et al., 1998; Wang et al., 2008b; Sethakulvichai et al., 2012).

Оценка геометрических характеристик и построения профиля плотности достаточно трудоемкий процесс (Schwartzkopf et al., 2005; Pravina, 2015). Кроме того, для сильно изогнутых и наложенных хромосом возникают дополнительные проблемы их вычисления (Schwartzkopf et al., 2005). Несмотря на это, классификация с использованием профиля плотности, центрального индекса и длины хромосомы остается одним из лучших решений для обработки монохромных изображений.

Цветовая информация используется при классификации хромосом в многоцветных FISH-изображениях, в которых хромосомы одного класса имеют идентичные цветовые характеристики (одинаковые интенсивности от разных флуорохромов). Ранее в разделе 1.4.2 упоминались методы классификации, базирующиеся на цветовых характеристиках. Часто в качестве характеристик используется либо интенсивность сигнала от каждого флуорохрома (методы классификации каждой точки хромосомы), либо среднее значение интенсивности сигналов и ее стандартное отклонение для пикселей хромосом (классификация районов хромосом) (Choi et al., 2004; Schwartzkopf et al., 2005; Sreejini et al., 2012). Цветовой информации достаточно для разделения хромосом на классы, но точность классификации можно повысить за счет геометрических характеристик (Zhao et al., 2013).

1.4.4 Определение сигналов хромосомоспецифичных последовательностей ДНК с помощью компьютерной обработки FISH-изображений

В 2006 году был предложен принципиально новый подход, позволяющий устранить часть недостатков подхода CISS-гибридизации и исключить сигнал с диспергированных повторяющихся последовательностей *in silico* (Rens et al., 2006) (далее метод RENS). Идея метода основана на том, что каждая хромосома- или

район- специфичная ДНК-проба содержит сходные повторяющиеся последовательности. При гибридизации *in situ* с двумя различными ДНК-пробами, сигналы, обусловленные повторяющимися последовательностями, присутствующими в обеих ДНК-пробах, должны иметь близкие интенсивности. Соответственно, отношение интенсивности этих сигналов должно быть приблизительно равно единице. На практике же отношение интенсивностей сигналов может отличаться от единицы, так как регистрируемая интенсивность сигнала зависит от многих факторов.

Предлагаемый метод включает построение гистограммы отношений интенсивностей сигналов ДНК-проб в каждом пикселе изображения. Чтобы учесть возможное несоответствие интенсивностей сигналов, авторы подхода предложили задавать интервал значений соотношений интенсивностей, ассоциируемый с повторяющимися последовательностями. Границы интервала предполагается устанавливать интерактивно. На рисунке 1.6 представлена гистограмма отношений интенсивностей сигналов, на которой указан интервал с границами T_1 – T_2 , соответствующий повторяющимся последовательностям.

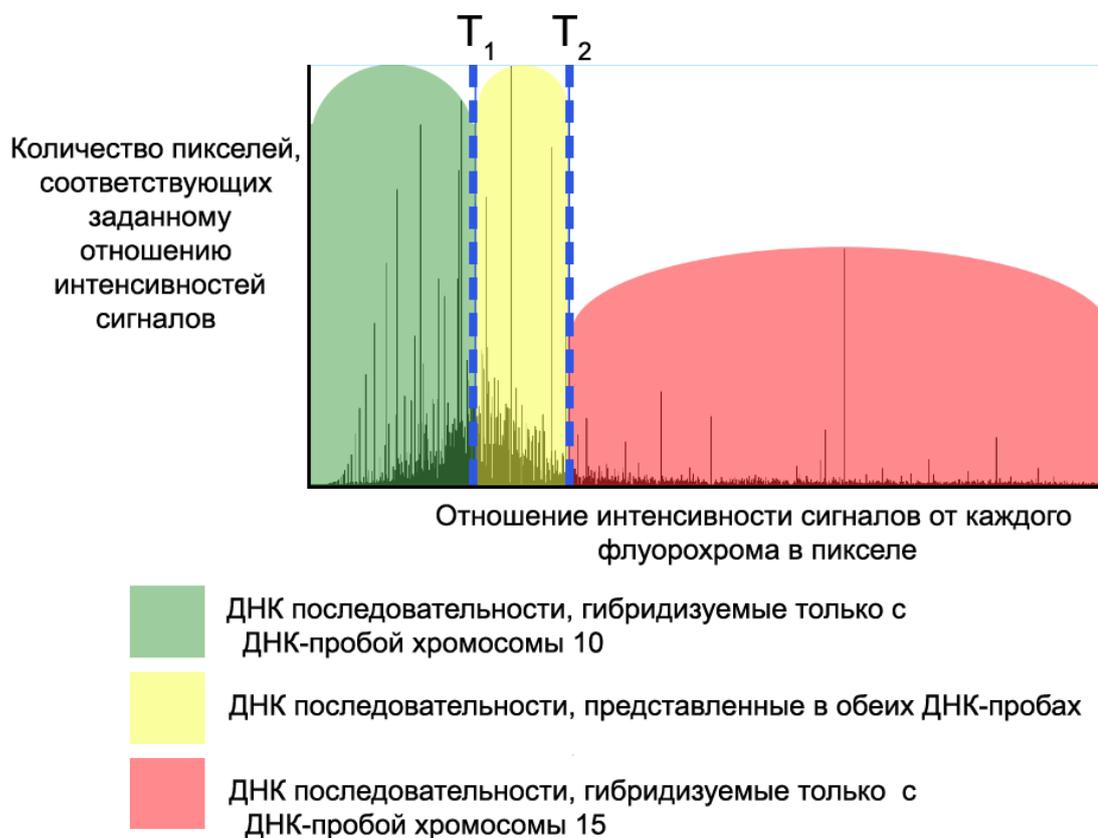


Рисунок 1.6. Гистограмма отношений интенсивностей сигналов от каждого флуорохрома и заданные на ней пороги. Гистограмма построена для изображения, приведенного на рисунке 1.2. Рисунок подготовлен на основе материала в статье (Rens et al., 2006).

В дальнейшем значения этих границ использовали для классификации сигналов в каждом пикселе микроскопического изображения. Сигналам в пикселях, где соотношение их интенсивностей попадало в этот интервал, присваивали нулевое значение интенсивности. Сигнал в пикселях, где соотношение было меньше T_1 , относили к специфичному сигналу первой ДНК-пробы, а если соотношение было больше T_2 – к специфичному сигналу второй ДНК-пробы.

Существенным недостатком указанного подхода является зависимость результатов обработки изображений от субъективного мнения исследователя. Пользователь сам задает границы, определяя, что на образце, по его мнению, является повторяющимися последовательностями, а что нет.

1.5 Заключение

Несмотря на значительные успехи в развитии молекулярных технологий, некоторые задачи молекулярной цитогенетики до сих пор остаются нерешенными. Например, задача идентификации материала целых хромосом и крупных хромосомных районов у видов, размер генома которых в несколько раз превышает размер генома человека. Проблема связана с наличием в хромосомах повторяющихся последовательностей, что приводит к регистрации сигнала на всех хромосомах при проведении FISH районо- или хромосомоспецифичных ДНК-проб. Чем больше доля повторяющихся последовательностей в геноме исследуемых организмов, тем сложнее распознавать хромосомы и хромосомные районы, из которых была получена соответствующая ДНК-проба.

Лучшим решением проблемы идентификации целых хромосом и их протяженных районов на сегодняшний день считается создание ДНК-проб, состоящих только из уникальных последовательностей. К сожалению, получение таких ДНК-проб – процесс дорогостоящий и трудоемкий.

Результатом поиска альтернативных решений проблемы является множество методов, позволяющих увеличить относительную интенсивность сигнала FISH уникальных последовательностей по сравнению с интенсивностью сигналов с повторяющихся последовательностей. Самым популярным и эффективным из них является CISS-гибридизация. К сожалению, данный подход не может быть проведен для хромосом некоторых видов эукариот из-за трудностей получения Cot-1 ДНК.

Кроме того, он, как и многие другие методы визуализации специфичных последовательностей ДНК, недостаточно эффективен при исследовании хромосом видов, размер генома которых в несколько раз превышает размер генома человека. Перечисленные ограничения и недостатки CISS-гибридизации, делают актуальным разработку и реализацию новых подходов для анализа результатов проведения FISH ДНК-проб из индивидуальных хромосом и хромосомных районов.

Глава 2. КОМПЬЮТЕРНЫЙ МЕТОД ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК (МЕТОД VISUALIZATION OF SPECIFIC SIGNAL *IN SILICO* – VISSIS)

2.1 Основные принципы обработки микроскопических изображений, лежащие в основе метода

Проблема идентификации целых хромосом и их протяженных районов может быть решена с помощью цифровой обработки FISH-изображений. В работе создан оригинальный компьютерный метод визуализации сигналов, продуцируемых хромосомоспецифичными последовательностями ДНК при проведении двухцветной FISH районо- и хромосомоспецифичных ДНК-проб (Богомолов и др., 2012). Метод получил название *visualization of specific signal in silico* (VISSIS).

Метод VISSIS работает с цветными 24-битными изображениями метафазных хромосом (цвет представляется с использованием 256 уровней интенсивности для каждой из трёх компонент пикселя). Изображения интегрируют сигналы трех разных каналов: два канала с результатами FISH от пары ДНК-проб, полученных из разных индивидуальных хромосом или хромосомных районов, и один канал с результатами DAPI-окрашивания (пример такого изображения приведен на рисунке 2.1).

Для объяснения идеи метода рассмотрим изображения результатов FISH с одной из ДНК-проб (монохромное изображение). Сигнал ДНК-пробы в каждом пикселе можно представить, как сумму интенсивностей сигналов:

$$i = i_{ss} + i_{ns} + bg \quad (2.1)$$

где i – интенсивность сигнала с микродиссекционной ДНК-пробы в пикселе изображения; i_{ss} – интенсивность сигнала с последовательностей, представленных только в хромосоме, из которой были получены ДНК-пробы (далее специфичный сигнал); i_{ns} – интенсивность сигнала с диспергированных повторяющихся последовательностей (далее неспецифичный сигнал); bg – интенсивность фонового сигнала.

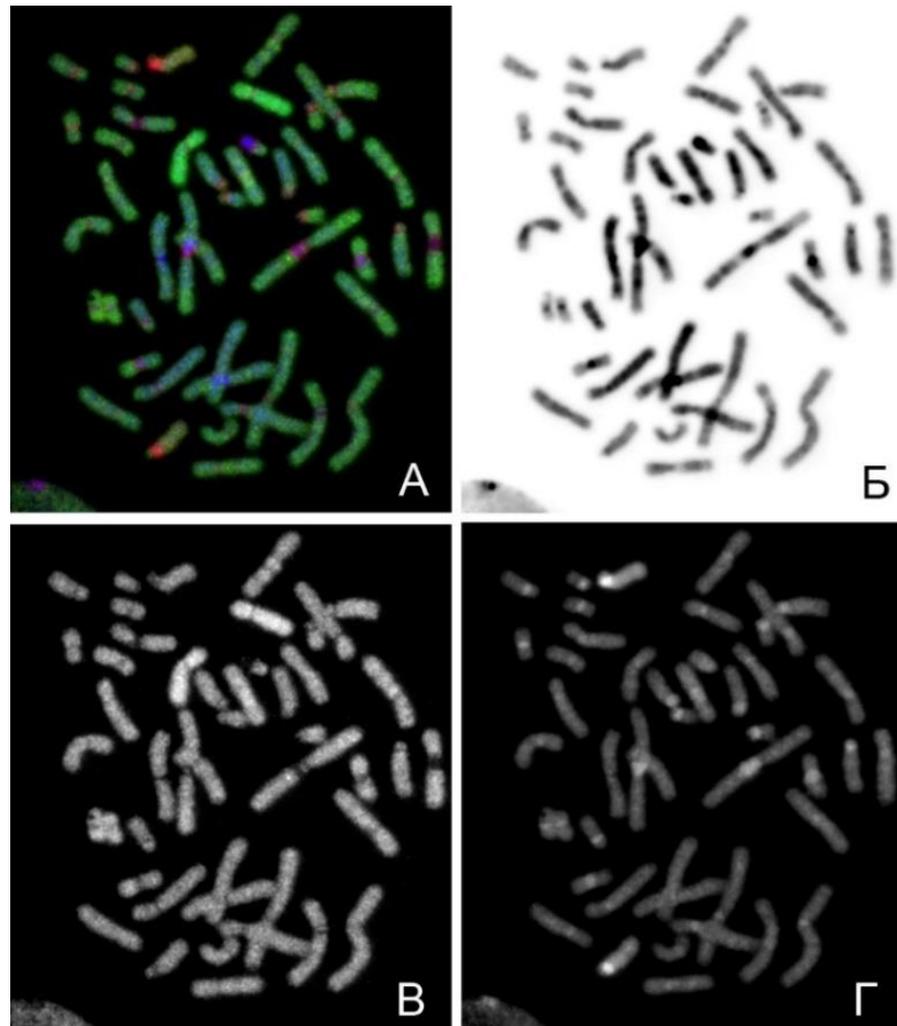


Рисунок 2.1. Пример анализируемого изображения. **А** – двухцветная FISH хромосом человека с ДНК-пробами из 10 (зеленый) и 15 (красный) хромосом. **Б** – инвертированный DAPI-бэндинг (инвертированный синий канал на результирующем изображении А). **В** – результаты FISH ДНК-проб из хромосомы 10 (зеленый канал). **Г** – результаты FISH ДНК-проб из хромосомы 15 (красный канал). Изображения предоставлены сотрудником ИЦиГ СО РАН Карамышевой Т.В.

Источниками фонового сигнала являются неспецифическая сорбция ДНК-пробы на поверхность цитологического препарата и тело хромосомы, клеточный дебрис (остатки клеточной мембраны, цитоплазмы и т.д.) и другие флуоресцирующие объекты, а также различные аппаратные шумы. Как правило, интенсивность фонового сигнала намного меньше интенсивности специфического и неспецифического сигналов.

Таким образом, районам хромосом, из которых была получена ДНК-проба, соответствует сумма специфического, неспецифического и фонового сигнала, тогда как районам остальных хромосом – только неспецифический и фоновый сигнал (рисунок 2.2).

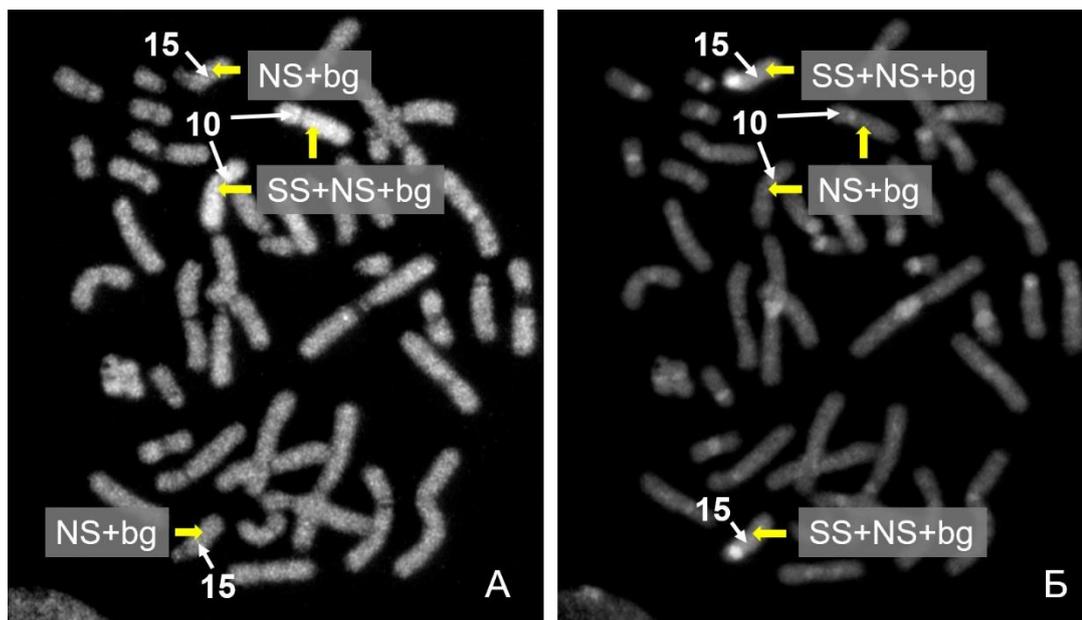


Рисунок 2.2. Типы сигналов, формирующие результирующий сигнал на хромосомах человека. **А** – FISH хромосом человека с ДНК-пробами из хромосомы 10. На хромосоме 10 зарегистрирован не только неспецифичный сигнал (NS) и фоновый сигнал (bg), но и сигнал с хромосомоспецифичных последовательностей (SS). **Б** – FISH хромосом человека с ДНК-пробы из хромосомы 15. На данном изображении на хромосоме 10 зарегистрирован только неспецифичный и фоновый сигналы, в отличие от хромосомы 15, где присутствуют все три типа сигналов.

Сигнал с хромосомоспецифичных последовательностей можно оценить, как разницу между сигналом ДНК-пробы и суммой неспецифичного и фонового сигнала (см. формулу 2.1). В основе метода VISSIS лежит допущение, что относительный уровень фонового и неспецифичного сигнала приблизительно одинаковый для двух разных ДНК-проб. Поэтому в обработке микроскопических изображений оценка интенсивности этих сигналов может быть получена в результате анализа FISH-изображения с другой ДНК-пробой.

Соответственно, интенсивность специфичного сигнала в каждом пикселе изображения может быть вычислена как:

$$\begin{aligned} i_{ssa} &= i_a - i_b \\ i_{ssb} &= i_b - i_a \end{aligned} \quad (2.2)$$

где i_a – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы a ; i_b – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы b ; i_{ssa} – рассчитанная интенсивность сигнала с хромосомоспецифичных последовательностей, представленных в микродиссекционной ДНК-пробе, полученной из хромосомы a ; i_{ssb} – рассчитанная интенсивность сигнала с хромосомоспецифичных последовательностей, представленных в микродиссекционной ДНК-пробе, полученной из хромосомы b .

Сравнение интенсивности сигналов с разных ДНК-проб является корректным, так как ДНК-пробы в рассматриваемом пикселе изображения находились в идентичных локальных условиях гибридизации (плотность упаковки материала хромосом, возможность проникновения ДНК-проб) и регистрации сигнала (уровень освещения препарата). В этом случае интенсивности сигналов зависят в большей степени от состава ДНК использованных ДНК-проб. Верность этого предположения показана в методах сравнительной геномной гибридизации и многоцветного бэндинга хромосом (Liehr, Glaser, Kosyakova, 2017; Liehr, Kosyakova, Weise, 2017).

Необходимо учесть, что регистрируемая интенсивность сигнала зависит от многих факторов, в частности, от времени экспозиции. Поэтому относительный уровень интенсивности сигналов, обусловленных повторяющимися последовательностями, присутствующими в ДНК-пробах, могут отличаться (со схожей проблемой столкнулись авторы работы (Rens et al., 2006)). Для выравнивания уровня интенсивностей сигналов необходимо проведение процедуры калибровки интенсивностей сигналов. При количественной обработке сигналов также необходимо принять меры к повышению устойчивости результатов обработки к различным искажениям интенсивности сигнала и определить области изображения, к которым должна быть применена компьютерная обработка.

На основании перечисленных требований к цифровой обработке изображений, в предлагаемый метод были включены следующие этапы: сегментация изображения, классификация объектов, калибровка интенсивностей сигналов, выделение специфического сигнала. Общая схема метода проиллюстрирована на рисунке 2.3. Подробно перечисленные этапы будут описаны в следующих разделах.

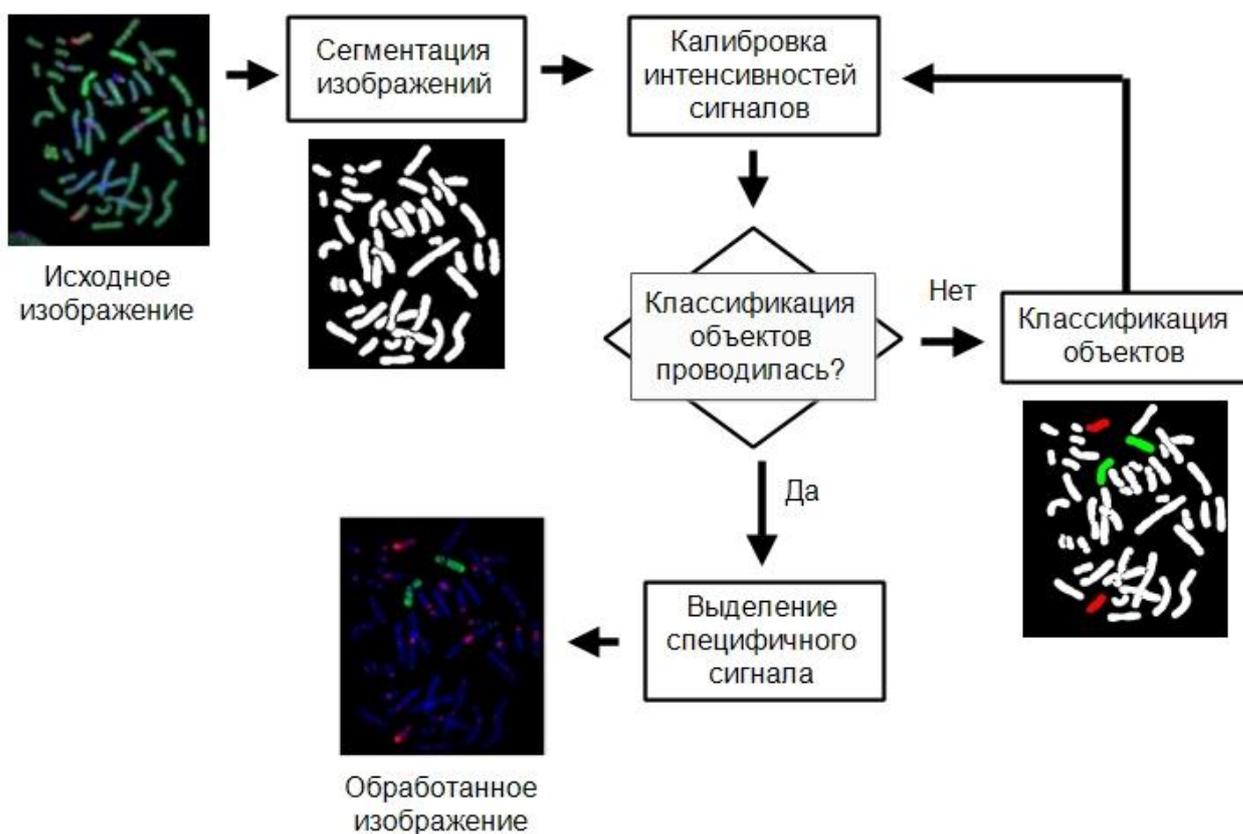


Рисунок 2.3. Основные этапы метода VISSIS.

2.2 Сегментация изображения

Сегментация выделяет на изображении участки, занятые хромосомами (далее такие области называются объектами). Последующие этапы метода VISSIS используют и обрабатывают информацию только в пикселях обнаруженных объектов. Таким образом, сегментация позволяет сократить количество вычислений и уменьшает влияния различных шумов на оценку интенсивности специфического сигнала.

Для сегментации используются результаты окрашивания красителем DAPI. На таких изображениях хромосомы имеют более высокие значения интенсивности сигналов, чем другие районы изображения (рисунок 2.4). В качестве первого шага сегментации выбирается порог на гистограмме интенсивности. Пиксели со значением интенсивности выше этого порога относятся к хромосомам. Пиксели, в которых значение интенсивности сигнала ниже порогового, относятся к фону. Автоматический выбор порога осуществляется одним из методов: метод Отсу (Otsu, 1979), метод интерактивного выделения (Calvard, Ridler, 1978), алгоритм Triangle (Zack et al., 1977), метод Капура (Kapura et al., 1985).

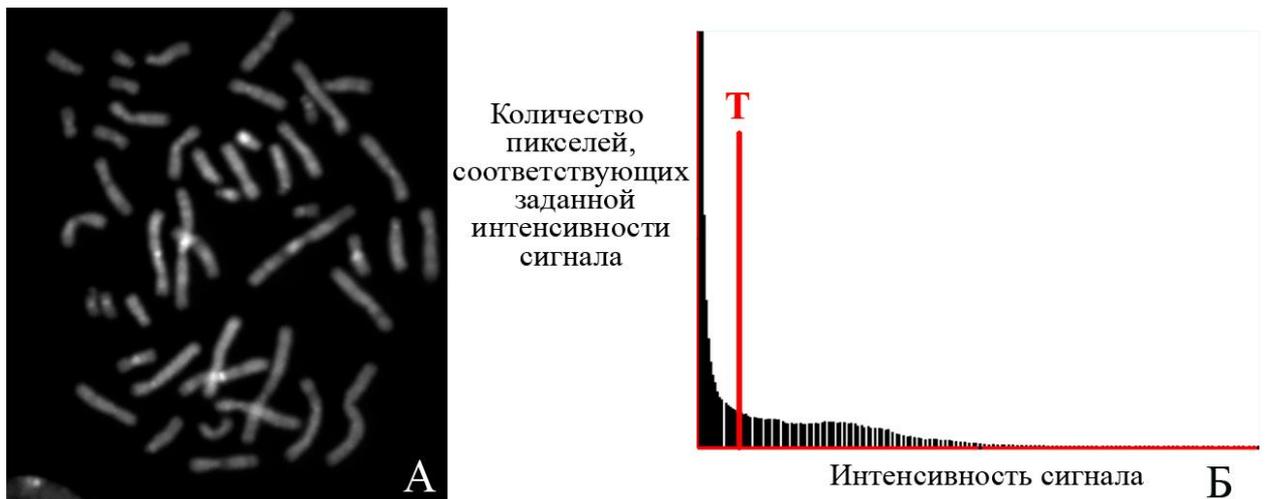


Рисунок 2.4. Изображение DAPI-окрашивания метафазных хромосом человека. А – Результаты окрашивания красителем DAPI. Б – гистограмма интенсивности. Т- порог сегментации.

Пороговая сегментация очень чувствительна к шуму на изображении (Zimmer, 2012; Arachchige, 2014). В частности, неравномерное освещение может превратить хорошо разделяемую гистограмму интенсивности в такую, которую невозможно эффективно разделить с помощью единого глобального порога (Гонсалес и Вудс, 2005, с. 858-859). В подобных ситуациях для выделения областей лучше использовать методы с локальным порогом. В этом случае алгоритм пороговой сегментации принимает следующий вид: исходное изображение разбивается на области размером 10×10 пикселей и для каждой области устанавливается свой порог на гистограмме интенсивности, построенной по пикселям рассматриваемой области. Порог вычисляется одним из перечисленных выше методов автоматического выбора порога.

Чтобы повысить устойчивость результатов пороговой сегментации к различным шумам, проводится также предварительная фильтрация DAPI-изображения с помощью медианного фильтра (Гонсалес и Вудс, 2005, с. 194-196). Похожие рекомендации для повышения точности сегментации были даны в работах (Karvelis et al., 2008; Zimmer, 2012; Shirazi et al., 2016).

Поиск объектов (связных компонент) на обработанном изображении осуществляется методом рекурсивной маркировки (Шапиро и Стокман, 2006, с. 86). Найденные объекты можно разделить на три группы: отдельные хромосомы, кластеры хромосом и ложные объекты.

Ложные объекты не представляют интереса для исследователя и должны быть исключены из дальнейшего анализа. К ложным объектам относятся: не полностью вошедшие в кадр хромосомы, ядра клеток и различные артефакты. Часть таких объектов может быть обнаружена автоматически с помощью дескрипторов областей: площади и эксцентриситета (Ji, 1994). Площадь объекта обычно оценивается как количество пикселей, принадлежащих ему. Эксцентриситет является мерой близости формы объекта к кругу. Он равен нулю для круглого объекта и единице для объекта в форме линии. Значение эксцентриситета вычисляется по формуле:

$$E = \sqrt{\frac{2C}{U_x + U_y + C}},$$

$$U_x = \frac{1}{12} + \frac{1}{N} \sum_{P(x,y) \in \Omega} (x - x_c)^2, \quad U_y = \frac{1}{12} + \frac{1}{N} \sum_{P(x,y) \in \Omega} (y - y_c)^2,$$

$$U_{xy} = \frac{1}{12} + \frac{1}{N} \sum_{P(x,y) \in \Omega} (x - x_c)(y - y_c), \quad C = \sqrt{(U_x - U_y)^2 + 4U_{xy}^2} \quad (2.3)$$

где E – эксцентриситет; U_x , U_y , U_{xy} и C – вспомогательные переменные; x_c и y_c – координаты геометрического центра объекта (пиксель, в котором минимизируется сумма расстояний до пикселей объекта); N – количество пикселей, относящихся к объекту; Ω – все множество пикселей, принадлежащих объекту; $P(x,y)$ – пиксель с координатами (x,y) .

Координаты геометрического центра объекта определяются по формуле:

$$x_c = \frac{1}{N} \sum_{P(x,y) \in \Omega} x, \quad y_c = \frac{1}{N} \sum_{P(x,y) \in \Omega} y \quad (2.4)$$

где N – количество пикселей, относящихся к объекту, Ω – все множество пикселей, принадлежащих объекту, $P(x,y)$ – пиксель с координатами (x,y) .

Объекты автоматически отмечаются как ложные, если их площадь в три раза меньше, либо в пять раз больше средней площади объектов на изображении. Ядрами считаются объекты, в четыре раза превышающие среднюю площадь объектов на изображении, и со значением эксцентриситета меньше 0.45 (параметры подобраны эмпирически на 20 изображениях интерфазных ядер).

Заключительным шагом сегментации является разделение кластеров на отдельные хромосомы. Разделение хромосом проводится фильтрами средней разности (Poou et al., 1999) или методом водораздела (Wenzhong, Xiaohui, 2010). В некоторых случаях на этой стадии наблюдалось ошибочное разделение целых хромосом на отдельные объекты (эффект пересегментации). Если количество таких

объектов мало, то объединение частей одной хромосомы лучше проводить интерактивно (с помощью специального интерфейса программы, в которой реализован метод VISSIS, вручную указываются на изображении фрагменты одной хромосомы). В более сложных случаях проводится автоматическое объединение с помощью классификации объектов и последующем объединением близко расположенных объектов одного класса. Схожая идея лежит в основе методов спектрального разделения кластеров хромосом (Karvelis et al., 2008; Srejini et al., 2012). Расстояние между объектами оценивается, как манхэттенское расстояние между крайними пикселями и геометрическими центрами объектов:

$$d(A, B) = |x_A - x_B| + |y_A - y_B| \quad (2.5)$$

где $d(A, B)$ – расстояние между объектами A и B ; x_A и y_A – координаты пикселей объекта A на изображении; x_B и y_B – координаты пикселей объекта B на изображении.

Крайними пикселями считаются пиксели с наибольшими или наименьшими значениями координат по оси x и y . Пример расположения крайних пикселей в объекте представлен на рисунке 2.5.

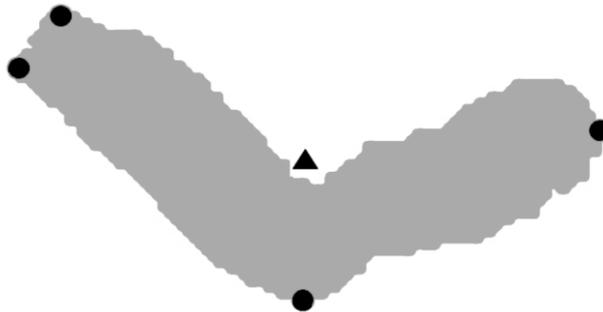


Рисунок 2.5. Расположение крайних пикселей и геометрического центра на объекте. Точки, принадлежащие объекту, обозначены серым цветом. ● – крайние пиксели объекта. ▲ – геометрический центр.

Для каждого объекта рассчитываются расстояния между его пикселями (крайние пиксели, геометрический центр объекта) и пикселями других объектов. Если кратчайшее расстояние между объектами одного класса меньше определенного порога, то объекты объединяются. Стоит обратить внимание на то, что автоматическая процедура приводит к повторному объединению ранее разделенных хромосом одного класса в кластер. Необходимо заметить, что разделение всех кластеров является желательным, но не обязательным условием для корректной работы метода. Основное требование, предъявляемое к объектам - отсутствие

кластеров, в состав которых входят исследуемые хромосомы (т.е. хромосомы, из которых были получены ДНК-пробы).

На рисунке 2.6 суммированы основные шаги сегментации и последовательность их выполнения.

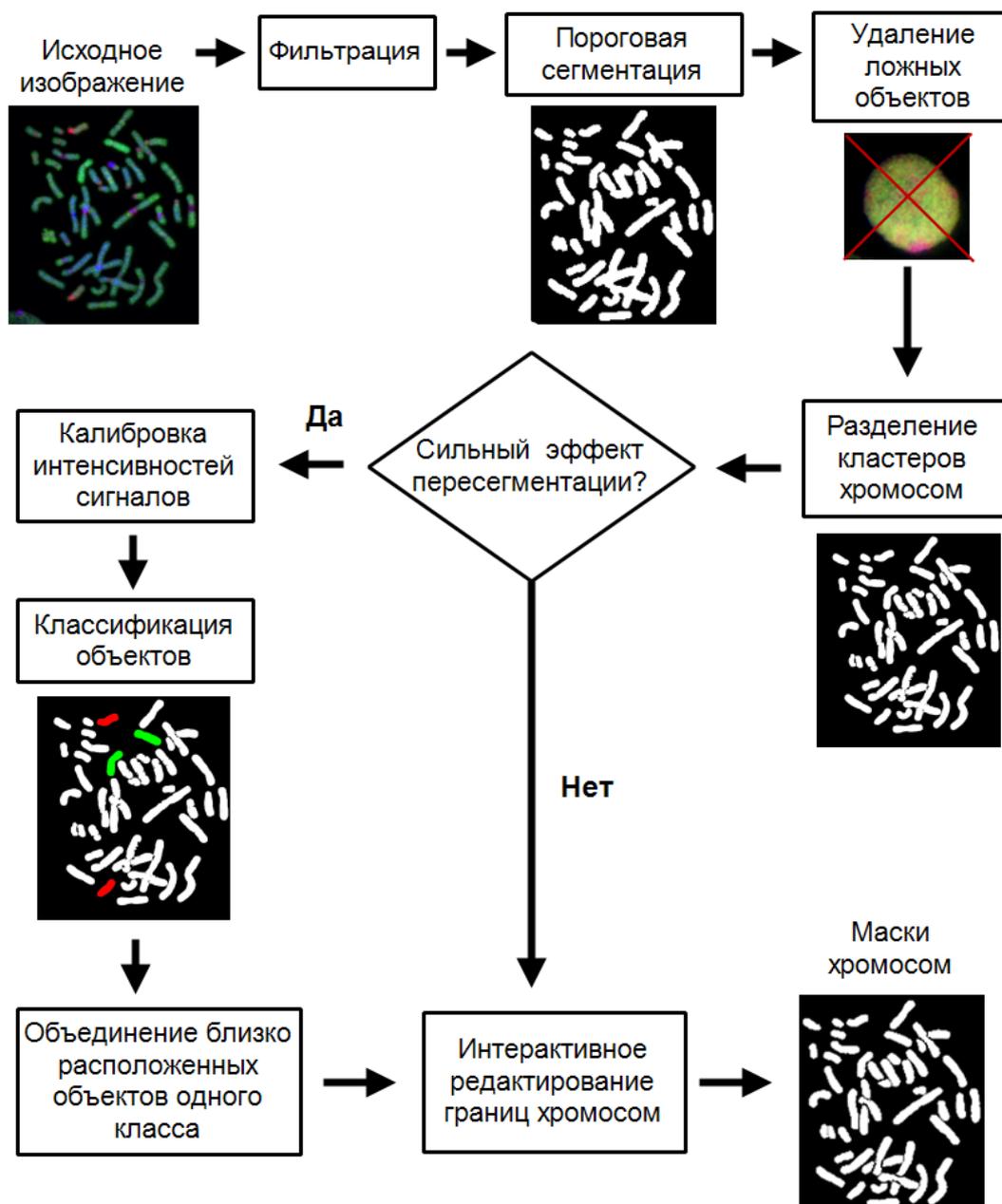


Рисунок 2.6. Основные этапы сегментации изображения.

2.3 Калибровка интенсивностей сигналов FISH. Сегментация областей с большой разницей интенсивности сигналов в пикселях

Интенсивности неспецифичных сигналов могут не совпадать. Можно выделить две основные причины возникновения различий. Первая причина связана

с особенностями регистрации изображений: разное время экспозиции, характеристики флуорохромов и т.д. Это делает необходимым проведение процедуры калибровки интенсивностей сигналов с разных ДНК-проб перед их сравнением.

Вторая причина обусловлена особенностями состава повторяющихся последовательностей ДНК в хромосомах, из которых получены ДНК-пробы (см. раздел 3.2). Разная представленность типов повторяющихся последовательностей в исходных хромосомах приводит к окрашиванию хромосомных районов с различной интенсивностью разными ДНК пробами. Особенную трудность для компьютерного анализа вызывают кластеры повторяющихся последовательностей ДНК. Большое количество повторяющихся последовательностей одного типа в кластерах может привести к слиянию сигналов с этих последовательностей ДНК и сплошному окрашиванию участка хромосомы после компьютерной обработки даже при небольших различиях в количестве последовательностей данного типа в хромосомах, из которых были получены ДНК-пробы. Области изображений хромосом, в пикселях которых наблюдаются сильные различия интенсивности сигналов от разных флуорохромов (далее яркие области), могут являться причиной некорректной калибровки интенсивностей сигналов, возникновения ошибок классификации объектов, и низкого контраста обработанных изображений. Поэтому нужно обращать особое внимание на выбор комплектов ДНК-проб. Появление ярких областей можно избежать подбором ДНК-проб из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК. В разделе 3.2 приведены рекомендации по выбору пар ДНК-проб, созданных из целых хромосом человека.

В случае наличия на изображении ярких областей следует проводить их отдельную компьютерную обработку. Сегментация ярких областей проводится с помощью гистограммы отношений интенсивностей сигналов ДНК-проб. Пикселям ярких областей, как правило, являются пиксели с наименьшими и наибольшими значениями отношения интенсивности сигналов (крайние точки на гистограмме). Интервалы значений интенсивностей сигналов, ассоциируемые с яркими областями, задаются интерактивно на гистограмме.

Для проведения калибровки интенсивности сигналов FISH используется упрощенная модель регистрации, предложенная Ретифом (Restif, 2006). В этом случае искажения интенсивности сигналов представлено в виде:

$$\begin{aligned} i_a &= \alpha_a (i_{ssa} + i_{ns} + bg), \\ i_b &= \alpha_b (i_{ssb} + i_{ns} + bg) \end{aligned} \quad (2.6)$$

где i_a – регистрируемая интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы a ; i_b – регистрируемая интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы b ; α_a и α_b – коэффициенты усиления сигнала, связанные с параметрами регистрации изображений; i_{ssa} – интенсивность сигнала с хромосомоспецифичных последовательностей, представленных в микродиссекционной ДНК-пробе, полученной из хромосомы a ; i_{ssb} – интенсивность сигнала с хромосомоспецифичных последовательностей, представленных в микродиссекционной ДНК-пробе, полученной из хромосомы b ; i_{ns} – интенсивность неспецифичного сигнала; bg – интенсивность фонового сигнала.

В соответствии с этой моделью, выравнивание уровней интенсивностей сигналов может быть достигнута путем умножения значения интенсивности сигналов с ДНК-пробы из хромосомы b на α_a/α_b . В дальнейшем это отношение будет называться поправочным коэффициентом (kr). Значения коэффициента может быть получено непосредственно из анализа интенсивностей сигналов ДНК-проб в пикселях объектов, которые не содержат ДНК, специфичную хромосомам a и b (объекты класса Cns). Регистрируемый сигнал в этих объектах обусловлен диспергированными повторяющимися последовательностями ДНК, поэтому отношение α_a/α_b можно вычислить как отношение средних интенсивностей сигналов ДНК-проб:

$$kr = \frac{\mu_a}{\mu_b} \quad (2.7)$$

где kr – поправочный коэффициент; μ_a – средняя интенсивность сигналов с ДНК-пробы, полученной из хромосомы a , в пикселях объектов класса Cns ; μ_b – средняя интенсивность сигналов с ДНК-пробы, полученной из хромосомы b , в пикселях объектов класса Cns .

Усреднение интенсивности сигналов в формуле 2.7 снижает влияние небольших отклонений интенсивности сигналов, связанных с шумами, на значение поправочного коэффициента. Но среднее значение чувствительно к экстремально малым или большим значениям интенсивности сигнала одной из ДНК-проб. По этой причине для корректной оценки поправочного коэффициента kr требуется исключить из обработки пиксели ярких областей.

Более точный подход к оценке коэффициента kp основан на анализе уравнения линейной регрессии без свободного члена, построенного для значений интенсивностей FISH-сигналов. Коэффициент наклона прямой является поправочным коэффициентом и вычисляется по формуле:

$$kp = \frac{\sum_{p \in \Omega} i_a(p) i_b(p)}{\sum_{p \in \Omega} i_b(p)^2} \quad (2.8)$$

где kp - поправочный коэффициент; Ω - все множество пикселей, принадлежащих объектам класса Cns , 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 .

2.4 Классификация объектов

Классификация является важным этапом метода VISSIS. Информация о классе объекта используется для корректной калибровки интенсивности сигналов, а также на этапе постобработки изображений. Каждый объект на изображении относится к одному из трех классов:

- Ca – объекты, содержащие ДНК, специфичную для хромосомы a ;
- Cb – объекты, содержащие ДНК, специфичную для хромосомы b ;
- Cns – объекты, которые не содержат ДНК, специфичную для хромосом a и b . Регистрируемый сигнал связан только с повторяющимися последовательностями ДНК.

Под хромосомами a и b подразумеваются хромосомы или их районы, из которых получены первая и вторая ДНК-пробы, соответственно.

Так как разработанный метод VISSIS не направлен на изучение хромосом конкретного вида, то большинство характеристик и методов классификации хромосом не могут быть использованы на этом этапе обработки (ранее в разделе 1.4.3 отмечалось, что задача классификации хромосом сводится к решению задачи кариотипирования). На данном этапе объекты различных классов могут быть разделены с помощью отношения интенсивности сигналов от разных флуорохромов. Каждый объект на изображении разделяется на три части, и для каждой из частей вычисляется логарифм отношения средних интенсивностей сигналов:

$$R_{ip} = \log_2 \left(\frac{\mu a(ip)}{\mu b(ip)} \right) \quad (2.9)$$

где R_{ip} – отношение средних интенсивностей сигналов от разных флуорохромов, вычисленное для области объекта ip ; ip – номер одной из областей, на которые был разделен объект; μa – средняя интенсивность сигналов с микродиссекционной ДНК-пробы, полученной из хромосомы a , в пикселях области ip ; μb – средняя интенсивность сигналов с микродиссекционной ДНК-пробы, полученной из хромосомы b , в пикселях области ip .

Разделение объектов на области (части) проводится с помощью оси инерции объекта. Ось инерции – это ось, вокруг которой объект может вращаться с минимальной инерцией. Она определяется как прямая, проходящая через геометрический центр объекта под углом наклона β к оси X (рисунок 2.7).

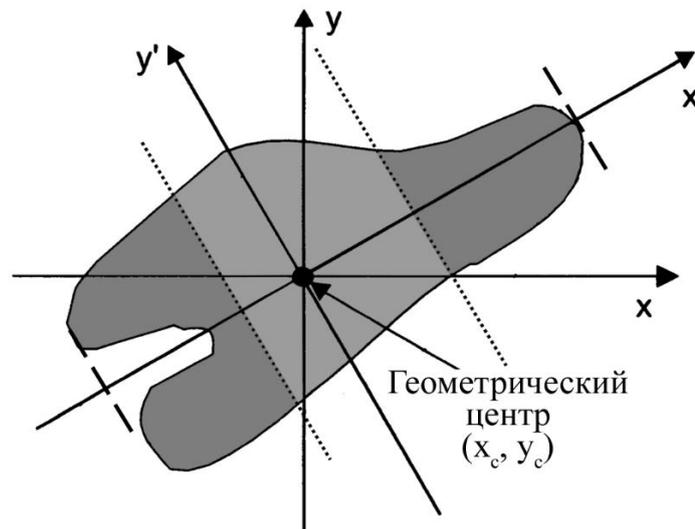


Рисунок 2.7. Разделение объектов с помощью оси инерции. Пунктиром обозначены границы проекции объекта на ось инерции x' . Точечные линии разделяют области объекта.

Угол β вычисляется по формуле:

$$\beta = \begin{cases} \frac{180}{\pi} \arctg \left(\frac{U_y - U_x + C}{2U_{xy}} \right), & \text{если } U_y > U_x \\ \frac{180}{\pi} \arctg \left(\frac{2U_{xy}}{U_y - U_x + C} \right), & \text{если } U_y \leq U_x \end{cases} \quad (2.10)$$

где β – угол наклона оси инерции; U_x , U_y , U_{xy} и C – вспомогательные переменные, которые вычисляются так же, как в формуле 2.3.

Проекция объекта на ось разбивается на три равных отрезка. Разделительные линии, которые разделяют объект на области, проводятся через общие точки полученных отрезков и перпендикулярны оси инерции (рисунок 2.7).

Значения R_{ip} сортируются по модулю в порядке убывания: $|R_3| \leq |R_2| \leq |R_1|$. Нумерация областей в зависимости от величины $|R_{ip}|$ позволяет проводить сравнение значений R_{ip} у гомологичных хромосом в зависимости от их расположения на изображении. Другими словами, оценка отношения интенсивности сигналов проводится для одинаковых областей у гомологичных хромосом.

Разделение объектов на области не случайно. Выбранная характеристика (значение логарифма отношения средних интенсивностей сигналов ДНК-проб) для классификации весьма чувствительна к экстремально малым или большим значениям интенсивности сигналов. Аномальные значения интенсивности сигнала одной из ДНК-проб приводят к отклонению значений характеристик объекта от ожидаемых для такого класса объектов отношений интенсивностей сигналов. Как ранее отмечалось в разделе 2.3, различия интенсивностей сигналов ДНК-проб могут быть связаны с различиями в содержании разных типов повторяющихся последовательностей ДНК в хромосомах, из которых были получены ДНК-пробы. Наиболее сильные различия в интенсивности FISH-сигналов наблюдаются в С-позитивных районах хромосом, что объясняется наличием в этих районах кластеров повторяющихся последовательностей ДНК. Обычно указанные хромосомные районы меньше областей, на которые разбиваются объекты, и, как правило, содержатся в области $ip=1$ (значение R_{ip} по модулю в этой области самое большое). Хотя в некоторых случаях часть пикселей крупных С-позитивных районов могут находиться в области $ip=2$. Таким образом, полученные характеристики R_{ip} имеют разную значимость для процедуры классификации. Разбиение объектов на области позволяет снизить влияние С-позитивных районов хромосом и других районов с большой разницей интенсивности FISH-сигналов в пикселях на результаты классификации объектов.

Существуют трудности разбиения небольших по размеру объектов на области (длина проекции таких объектов на ось инерции составляет меньше 12 пикселей). В качестве характеристики для таких объектов используется отношение средних интенсивностей сигналов целого объекта:

$$R_1 = R_2 = R_3 = \log_2 \left(\frac{\mu a}{\mu b} \right) \quad (2.11)$$

где μa – средняя интенсивность сигналов с ДНК-пробы, полученной из хромосомы a ; μb – средняя интенсивность сигналов с ДНК-пробы, полученной из хромосомы b .

Таким образом, любой объект представлен набором чисел $\{R_1, R_2, R_3\}$, в котором каждое число соответствует значению отношения интенсивности сигналов разных районов объекта. Математически объект эквивалентен вектору и задается точкой в пространстве признаков.

В методе классы также представлены векторами признаков. Такие вектора называются прототипами соответствующих классов. Класс *Cns* представлен вектором, у которого значение каждого признака равно нулю. Это объясняется тем, что сигналы повторяющихся последовательностей, присутствующих в обеих ДНК-пробах, имеют близкие значения интенсивности (Rens et al., 2006), а объекты класса *Cns* содержат ДНК, гомологичную только этим последовательностям ДНК. Объект, у которого сумма значений признаков максимальна, является прототипом класса *Ca*, а объект, у которого сумма значений признаков минимальна, прототипом класса *Cb*.

Объекты одного класса в пространстве признаков соответствуют обособленному множеству точек, обладающему свойствами хорошей отделимости. На рисунке 2.8 приведены реальные данные по расположению объектов в пространстве признаков. Для классификации объектов применяется метод ближайшего соседа (Samrat et al., 2004; Гонсалес и Вудс, 2005, с. 989-990). Классификатор вычисляет евклидовы расстояния между вектором признаков неизвестного (еще не классифицированного) объекта и векторами прототипов классов (уравнение 2.12). Объект относится к тому классу, прототип которого оказывается ближайшим. Расстояние между векторами вычисляется по формуле:

$$d(A, B) = \sqrt{w_1(R_{A,1} - R_{B,1})^2 + w_2(R_{A,2} - R_{B,2})^2 + w_3(R_{A,3} - R_{B,3})^2} \quad (2.12)$$

где $d(A, B)$ – евклидово расстояние в пространстве признаков между объектами A и B ; $R_{A,ip}$ – логарифм отношения средних интенсивностей сигналов в области ip объекта A ; $R_{B,ip}$ – логарифм отношения средних интенсивностей сигналов в области ip объекта B ; w_1, w_2 и w_3 – коэффициенты, отражающие значимость характеристик.

Естественная, с геометрической точки зрения, евклидова мера расстояния может оказаться бессмысленной, если признаки измерены в разных единицах. По этой причине для корректного расчёта расстояния между точками в качестве характеристик используется логарифм отношения интенсивности сигналов (см. формулы 2.9 и 2.11).

Подбор коэффициентов w_1, w_2 и w_3 осуществлялся на основе результатов классификации объектов на изображениях FISH ДНК-проб из пар хромосом

человека 7 и 19, 10 и 15, 10 и 19, 15 и 19, 18 и 19, 19 и X (тестовая выборка изображений). Изображения из первого набора (FISH ДНК-проб из 10 и 15 хромосомы) содержали умеренное количество хромосомных районов, в которых наблюдаются сильные различия в интенсивности FISH-сигналов. Остальные наборы изображений содержали большое количество хромосомных районов с сильными различиями интенсивности.

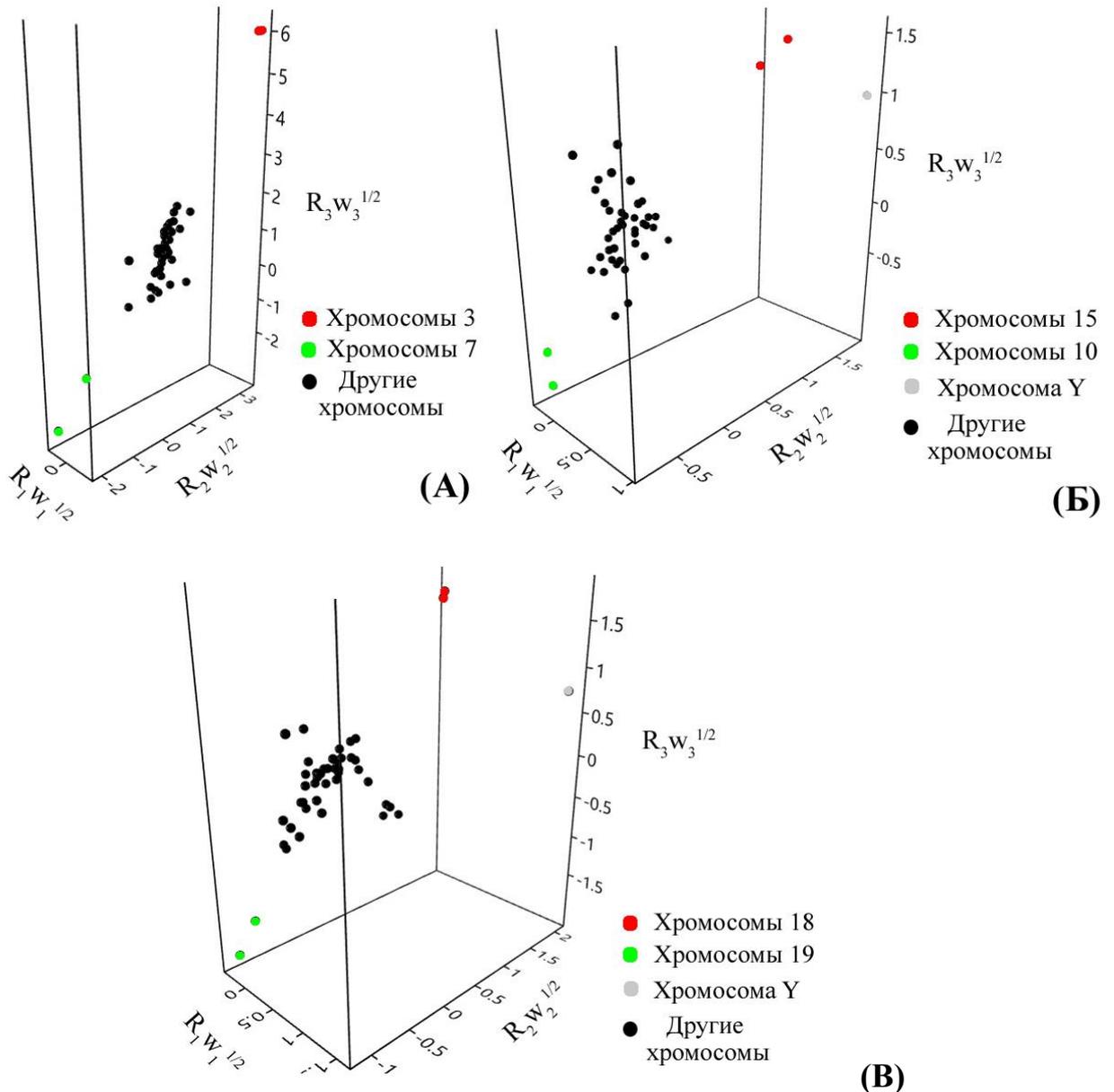


Рисунок 2.8. Пример расположения объектов в пространстве признаков. Каждая точка соответствует объекту, а по осям откладываются значения признаков объекта с учетом их значимости ($w_1=0.07$; $w_2=1$; $w_3=3.6$). **А** – Расположение объектов, обнаруженных на изображении FISH ДНК-проб из 3 и 7 хромосомы с супрессией гибридизации повторяющихся последовательностей ДНК. **Б** – Расположение объектов, обнаруженных на изображении FISH ДНК-проб из 10 и 15 хромосомы. **В** – Расположение объектов, обнаруженных на изображении FISH ДНК-проб из 18 и 19 хромосомы.

Метафазные хромосомы в рассматриваемых наборах изображений включали Y-хромосому человека. Эта хромосома содержит крупный C-позитивный район и часто является причиной ошибок классификации. При вычислении характеристик пиксели C-позитивного района присутствуют во всех трех областях, на которые разделили Y-хромосому. Впоследствии классификатор ошибочно отмечает Y-хромосому как объект класса Ca или Cb . В некоторых случаях различия интенсивностей сигналов в пикселях Y-хромосомы настолько велики, что она выбиралась в качестве прототипа одного из классов.

Результаты классификации оценивали с помощью ошибок первого (E_1) и второго рода (E_2). Ошибкой классификации первого рода (ошибка недопредсказания) считается ошибка, при которой классификатор отмечает объект класса Ca или Cb как объект класса Cns . Ошибкой классификации второго рода (ошибка перепредсказания) считается ошибка, при которой классификатор относит объект к классам Ca или Cb , хотя верным классом должен быть Cns .

Оценку ошибок классификация проводили по формулам 2.13 и 2.14.

$$E_1 = \frac{FNns}{Na + Nb} \quad (2.13)$$

где $FNns$ - число хромосом, ошибочно отнесенных к классу Cns ; Na – количество объектов класса Ca ; Nb - количество объектов класса Cb .

$$E_2 = \frac{FNab}{Nns} \quad (2.14)$$

где $FNab$ – число хромосом, которые содержат только неспецифичный сигнал, но были отнесены к одному из классов со специфичным сигналом; Nns - количество объектов класса Cns .

Характеристика объекта в области $ip=1$ (R_1) является наименее значимой для процедуры классификации. Это связано с тем, что значение R_1 зависит не только от класса объекта, но и от наличия хромосомных районов, в которых наблюдаются сильные различия в интенсивности FISH-сигналов. На рисунке 2.9 показан график изменения ошибок E_1 и E_2 в зависимости от величины w_1 .

Минимальные значения ошибок первого и второго рода наблюдается при w_1 равном 0.07 ($E_1=0.0506$ и $E_2=0.0544$). Последующее уменьшение значения w_1 приводит к существенному увеличению ошибки E_2 .

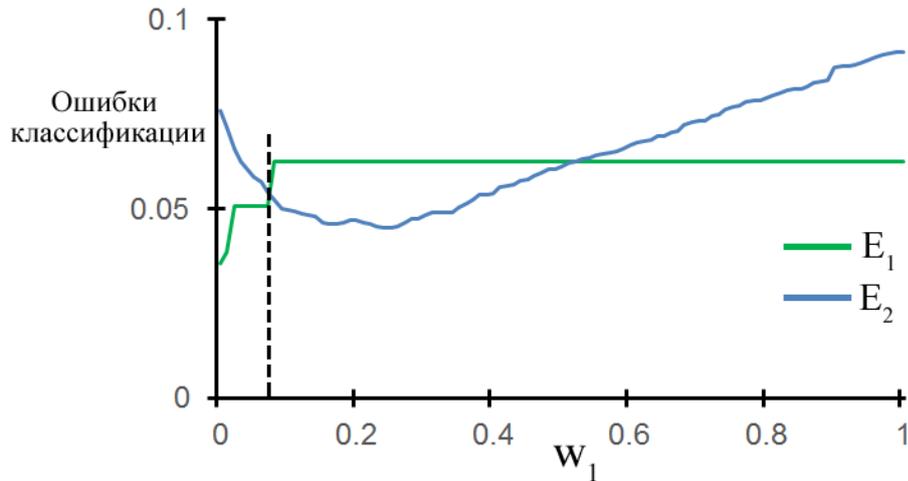


Рисунок 2.9. Изменение ошибок классификации объектов в зависимости от величины w_1 . Значения w_2 и w_3 фиксированы и равны единице. Всего проанализировано 84 изображения FISH ДНК-проб из хромосом человека 7 и 19, 10 и 15, 10 и 19, 15 и 19, 18 и 19, 19 и X. Пунктиром обозначено значение $w_1=0.07$.

Дальнейшее уменьшение ошибок классификации проводилось за счет изменения коэффициента w_3 . Отношение средних интенсивностей сигналов от разных флуорохромов в пикселях области $ip=3$ (R_3) зависит только от класса объекта. Поэтому характеристика объекта R_3 имеет важное значение для классификации. Минимальные значения ошибок классификации достигаются при w_3 равным 0.36 (рисунок 2.10). Ошибки первого и второго рода при таком значении w_3 составили соответственно 0.0149 и 0.039.

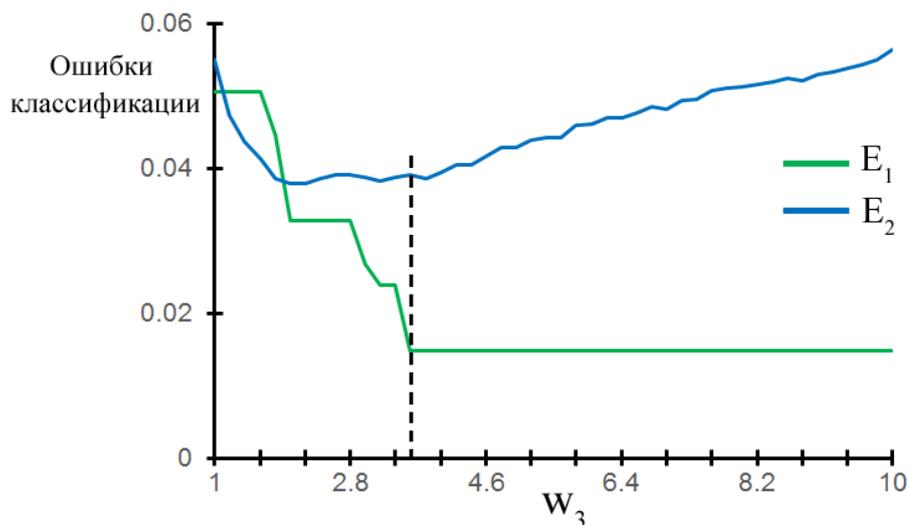


Рисунок 2.10. Изменение ошибок классификации объектов в зависимости от величины w_3 . Значения w_1 и w_2 фиксированы ($w_1=0.07$, $w_2=1$). Всего проанализировано 84 изображения FISH ДНК-проб из хромосом человека 7 и 19, 10 и 15, 10 и 19, 15 и 19, 18 и 19, 19 и X. Пунктиром обозначено значение $w_3=3.6$.

Последующие увеличение коэффициента w_3 не повлияло на ошибку первого рода. Поскольку этот тип ошибок оказывает существенное влияние на этапы калибровки интенсивностей сигналов и выделение специфичного сигнала, необходимо понимать причины их возникновения. Результаты анализа изображений, в которых произошла ошибочная классификация, рассмотрены более подробно в разделе 3.2. Таким образом, оптимальные значения коэффициентов w_1 , w_2 и w_3 в формуле 2.12 равняются 0.07, 1 и 3.6, соответственно.

Для уменьшения влияния первоначального выбора прототипов классов S_a и S_b на результаты классификации применяется процедура повторного определения классов объектов. В качестве прототипов классов S_a и S_b выбираются вектора со средними значениями характеристик ранее отнесенных к ним объектов. Такие вектора называются векторами математического ожидания классов или центром тяжести класса (Шапиро, Стокман, 2006).

В некоторых случаях количество классов может отличаться от трех. Например, когда вместо одной из ДНК-проб используется меченая Cot-1 ДНК или фрагментированная геномная ДНК. Проверка количества классов на изображении проводится с помощью критерия Ирвина (Irwin, 1925). Рассматриваемый критерий используется для поиска в элементах выборки грубых ошибок (аномальных значений, выбросов). Согласно этому критерию, аномальной считается точка ряда Y_t , отстоящая от предыдущей точки Y_{t-1} , на величину, большую среднеквадратического отклонения ($\lambda_t > \lambda_q$):

$$\lambda_t = |Y_t - Y_{t-1}| / \sigma, \quad \lambda_t > \lambda_q \quad (2.15)$$

где σ - среднеквадратическое отклонение, рассчитываемое по исследуемой выборке с учётом сомнительного значения; λ_q - критическое (табличное) значение.

Объекты одного класса образуют компактно локализованные подмножества в пространстве признаков. Отсюда можно сделать вывод, что множество объектов можно разбить на два класса, если между ними существует достаточно большой промежуток в пространстве признаков. Поэтому в общих чертах процедуру оценки количества классов можно свести к поиску достаточно больших отрезков между точками в пространстве признаков. Эти расстояния значительно отличаются от внутриклассовых расстояний и могут рассматриваться как выбросы (рисунок 2.8).

Процедура оценки количества классов на изображении выглядит следующим образом:

1) Вычисляем расстояния d_i между ближайшими объектами в пространстве признаков по формуле 2.8. Схема алгоритма подсчета расстояний приведена на рисунке 2.11.

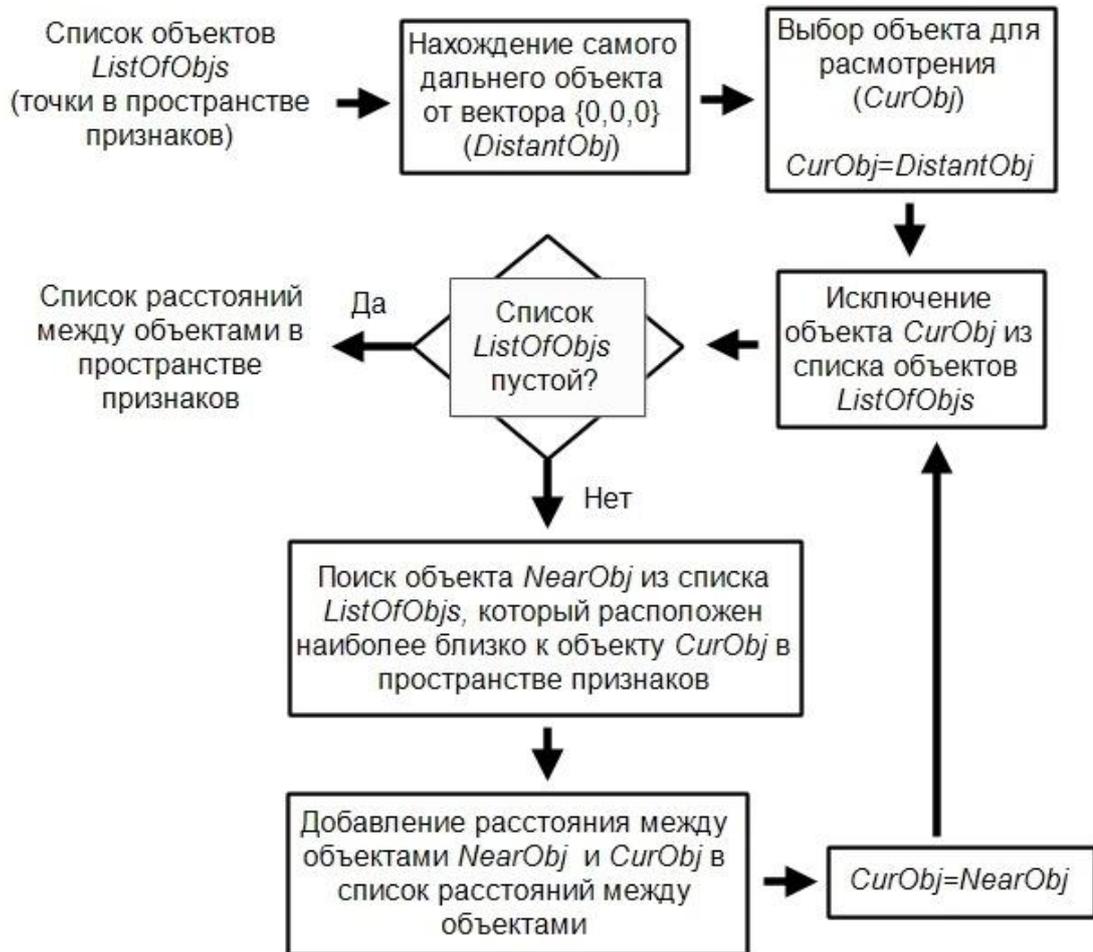


Рисунок 2.11. Блок-схема алгоритма поиска расстояний между объектами в пространстве признаков.

2) Упорядочиваем полученные расстояния $d_1 \leq d_2 \leq \dots d_{n-1} \leq d_n$. На изображение можно обнаружить объекты трех классов, поэтому расстояния d_{n-1} и d_n являются потенциальными выбросами.

3) Проверяем значение d_{n-1} для выборки $d_1 \leq d_2 \leq \dots d_{n-2} \leq d_{n-1}$ с помощью критерия Ирвина. Если элемент d_{n-1} является выбросом, то на изображении присутствуют три класса объектов (так как наблюдается два экстремально больших значения d_n и d_{n-1}), если нет, то переходим к следующему шагу.

4) Проверяем значение d_n для выборки $d_1 \leq d_2 \leq \dots d_{n-2} \leq d_n$ с помощью критерия Ирвина. Если элемент d_n является выбросом, то на изображении присутствуют два

класса объектов, в противном случае изображение содержит только объекты класса *Cns*.

Уровень значимости критерия Ирвина установлен равным 0.05. Процедура оценки количества классов проводится только когда количество обнаруженных объектов на изображение больше четырех.

Результаты процедуры можно рассматривать как показатель качества изображения для дальнейшего анализа распределения специфичного сигнала. Распознавание меньшего количества классов может указывать на плохое качество FISH или на использование пар ДНК-проб из хромосом, контрастных по содержанию типов повторяющихся последовательностей. В первом случае расстояние между объектами разных классов в пространстве признаков невелико из-за недостаточно интенсивного специфичного сигнала. Во втором случае более интенсивное окрашивание хромосомных районов одной из ДНК-проб приводит к расположению части объектов класса *Cns* между объектами классов *Ca* или *Cb* и основной группы объектов класса *Cns*, и, как следствие, неверной оценки межклассового расстояния. Общая схема процесса классификации объектов приведена на рисунке 2.12.

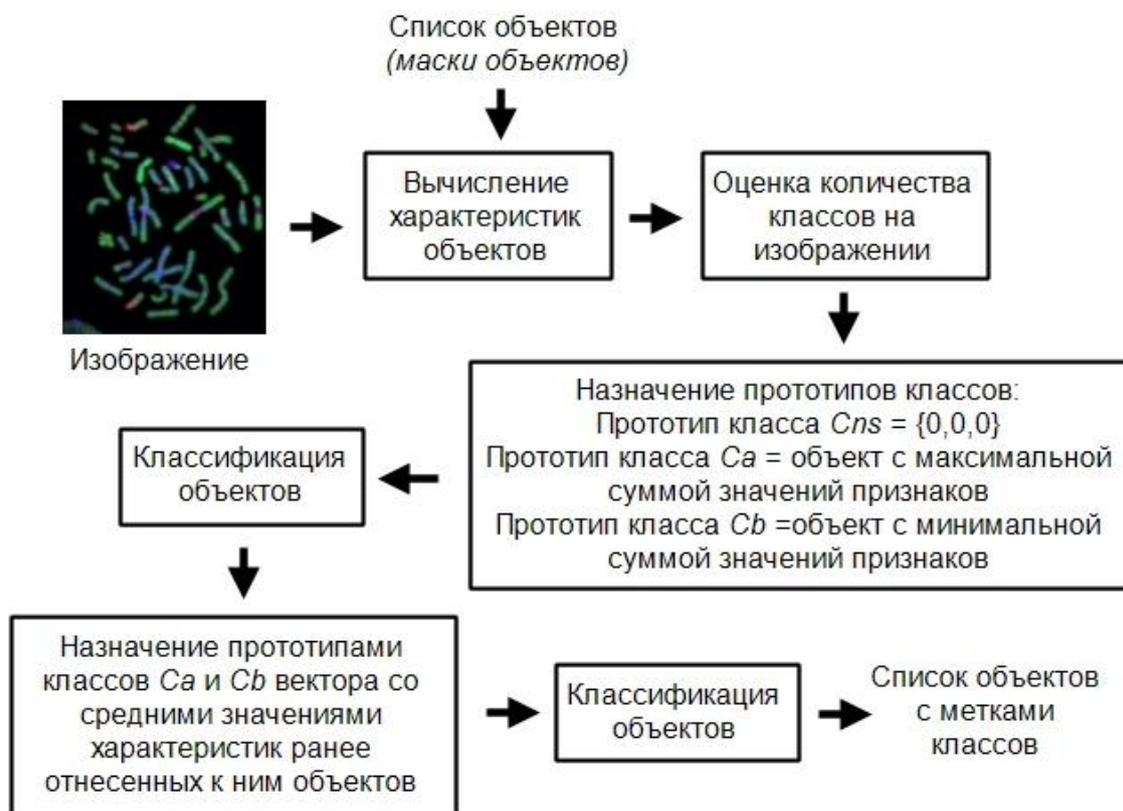


Рисунок 2.12. Основные этапы классификации объектов.

2.5 Выделение специфичного сигнала

Выделение специфичного сигнала в пикселях объектов проводится с учетом значений интенсивности в соседних пикселях:

$$R(x, y) = \frac{\sum_{u=-1}^1 \sum_{v=-1}^1 z(u, v) (i_a(x + u, y + v) - i_b(x + u, y + v))}{\sum_{u=-1}^1 \sum_{v=-1}^1 z(u, v)}$$

$$z(u, v) = \begin{cases} 2, & \text{если } u = v = 0 \\ 1, & \text{если } u \neq 0 \vee v \neq 0 \end{cases}$$

$$i'_a(x, y) = \begin{cases} R(x, y), & \text{если } R(x, y) \geq 0 \\ 0, & \text{если } R(x, y) < 0 \end{cases}$$

$$i'_b(x, y) = \begin{cases} 0, & \text{если } R(x, y) \geq 0 \\ -R(x, y), & \text{если } R(x, y) < 0 \end{cases} \quad (2.16)$$

где $i_a(x, y)$ – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы a , в пикселе (x, y) ; $i_b(x, y)$ – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы b , в пикселе (x, y) ; $R(x, y)$ – оценка различий интенсивности сигналов ДНК-проб в пикселе (x, y) ; $z(u, v)$ – коэффициенты, отражающие вклад пикселя в будущие значения интенсивности. $i'_a(x, y)$ и $i'_b(x, y)$ – новые значения интенсивности сигналов в пикселе (x, y) .

Анализ соседних пикселей позволяет уменьшить влияние импульсного шума на результаты компьютерной обработки. Полученный специфичный сигнал является слабо интенсивным и для лучшей его визуализации используется линейное контрастирование. Задача контрастирования заключается в "растягивании" реального динамического диапазона значений интенсивности на всю возможную шкалу значений. Новые значения интенсивности сигнала вычисляются по формуле:

$$i(x, y) = \frac{L_{max}(i(x, y) - i_{min})}{(i_{max} - i_{min})} \quad (2.17)$$

где $i(x, y)$ – интенсивность сигнала в пикселе (x, y) ; i_{min} – минимальное значение интенсивности сигналов в пикселях объектов; i_{max} – максимальное значение интенсивности сигналов в пикселях объектов. При вычислении i_{max} пиксели ярких областей не учитываются; L_{max} – максимальное допустимое значение интенсивности сигнала для изображения.

Новое значение интенсивности в пикселях ярких областей может выйти за допустимый диапазон значений ($i > L_{max}$). Чтобы исключить такие ситуации, значениям интенсивности, превышающим верхнюю границу диапазона, присваивается максимально допустимое значение интенсивности. Напомним, что метод работает с цветными изображениями, состоящими из трех монохромных 8-

битовых изображений. Поэтому максимально допустимое значение равно 255 ($L_{max}=255$).

2.6 Постобработка изображений

Идеальным результатом компьютерной обработки является наличие на изображении только специфичного сигнала. На практике часть неспецифичного сигнала сохраняется, а его интенсивность зависит от выбора комплекта ДНК-проб при проведении FISH (см. раздел 3.2). Постобработка упрощает экспертизу интерпретацию результатов. Задача постобработки заключается в выборе порогов отдельно для каждого канала, которые разделяют множество пикселей объектов на два класса по уровню интенсивности FISH-сигналов. Сигналы со значением интенсивности выше установленного порога считаются специфичными, а ниже – неспецифичными.

Порог задается на основе анализа распределений интенсивности сигналов в пикселях объектов разных классов. В качестве распределения интенсивности специфичного сигнала используется гистограмма интенсивности, построенная по пикселям объектов класса S_a или S_b . Выбор класса зависит от обрабатываемого монохромного FISH-изображения. К примеру, для FISH-изображений с ДНК-проб из хромосомы a распределение интенсивности специфичного сигнала вычисляется с помощью объектов класса S_a . Для удобства изложения последующего материала обозначим выбранный класс объектов как класс S_s . Гистограмма интенсивности, построенная по пикселям объектов других классов, рассматривается в качестве распределения интенсивности неспецифичного сигнала. Указанные гистограммы интенсивности позволяют оценить ошибки классификации пикселей для заданного порога (рисунок 2.13).

Ошибка классификации первого рода E_1 показывает долю пикселей со специфичным сигналом, ошибочно отнесенным к неспецифичному сигналу:

$$E_1 = N_{abT} / N_{ab} \quad (2.18)$$

где N_{abT} – количество пикселей объектов класса S_s с меньшей интенсивностью, чем заданный порог; N_{ab} – общее количество пикселей объектов класса S_s .

Ошибка второго рода E_2 показывает долю пикселей с неспецифичным сигналом, ошибочно отнесенным к специфичному сигналу:

$$E_2 = N_{nsT} / N_{ns} \quad (2.19)$$

где N_{nsT} – количество пикселей объектов, не относящихся к классу C_s , с интенсивностью больше заданного порога; N_{ns} – общее количество пикселей объектов, не относящихся к классу C_s .

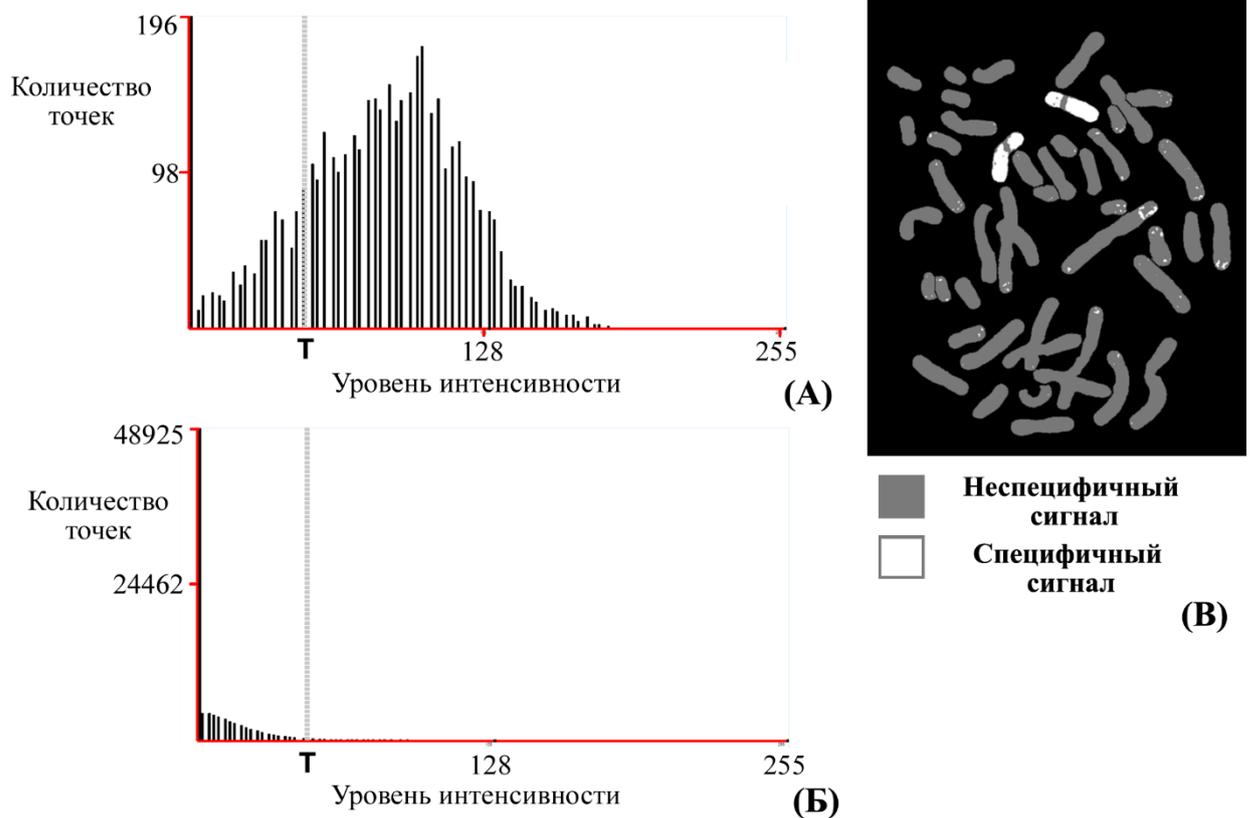


Рисунок 2.13. Постобработка изображения FISH с ДНК-пробой из хромосом человека 10. **А** – гистограмма интенсивности, построенная по пикселям хромосомы 10 (объект класса C_s). **Б** – гистограмма интенсивности, построенная по пикселям хромосом за исключением хромосом 10 (объекты, не относящиеся к классу C_s). **В** – псевдоизображение, построенное для заданного порога $T=47$. При установленном пороге ошибки классификации ошибки первого и второго рода: $E_1 = 0.18$ и $E_2=0.009$.

Выбор порога производится таким образом, чтобы минимизировать ошибки классификации пикселей. По умолчанию устанавливается порог, при котором разность ошибок первого и второго рода по модулю минимальна.

После задания порога пикселям изображения могут быть присвоены псевдоцвета, в зависимости от класса, к которому был отнесен пиксель (создание псевдоизображения). Другим вариантом результата постобработки является

присвоение нулевой интенсивности сигналам, которые были классифицированы как неспецифичные.

2.7 Программа визуализации хромосомоспецифичных последовательностей: VisualCS

Программа VisualCS (**VISUAL**ization **C**hromosome-specific **S**ignals) реализует разработанный метод визуализации специфичных сигналов и вариант компьютерной обработки FISH-изображений, предложенные Ренсом и его коллегами (Rens et al., 2006), а также инструменты анализа распределений интенсивности сигналов на изображениях (гистограммы и профили интенсивности). Программа зарегистрирована в Роспатенте (авторское свидетельство №2018662647, выдано 12.10.2018). VisualCS написана на языке C# (платформа .NET) и предназначена для работы на персональном компьютере с операционной системой WINDOWS (Windows XP, Windows Vista, Windows 7 и Windows 10).

На вход программа получает FISH-изображения в одном из графических форматов (TIFF, BMP, JPEG). Для ее корректной работы изображения должны включать следующую информацию: зеленый и красный каналы содержат результаты FISH двух ДНК-проб, полученных из разных индивидуальных хромосом или хромосомных районов, а синий канал – результаты DAPI-окрашивания.

Выходными данными VisualCS являются обработанные изображения. Также программа позволяет сохранять / загружать следующую информацию об объектах: границы хромосом и областей с различной интенсивностью FISH-сигналов, результаты классификации объектов. Перечисленные данные подаются/выгружаются в текстовых файлах (.txt). Каждый объект в файле хранится в виде отдельной строки. Строка имеет следующий формат:

$$\begin{aligned}
 & Type, Class, x_1, y_1, x_2, y_2, \dots, x_n, y_n, \\
 Type & = \begin{cases} 0, & \text{если объект является хромосомой} \\ 1, & \text{если объект является яркой областью,} \end{cases} \\
 Class & = \begin{cases} \text{пусто,} & \text{если } Type = 1 \\ [1,3], & \text{иначе} \end{cases} \quad (2.20)
 \end{aligned}$$

где $Type$ – тип маски объекта; $Class$ – класс объекта (для хромосом); $x_1, y_1, x_2, y_2, \dots, x_n, y_n$ – координаты пикселей объекта на изображении.

Автоматическая обработка изображений не всегда дает удовлетворительные результаты. Улучшить результаты можно за счет изменения параметров по

умолчанию в настройках программы. Примерами таких параметров является количество областей, на которые необходимо разделить изображение при локальной сегментации, коэффициенты w_1 , w_2 и w_3 , отражающие значимость характеристик объектов, параметры ложных объектов и другие.

Кроме того, в программе предусмотрена интерактивная корректировка результатов отдельных этапов компьютерной обработки. В качестве примера на рисунке 2.14 проиллюстрирована форма ручного разделения масок объектов. Пользователю предлагается провести разделительную линию.



Рисунок 2.14. Интерактивное разделение масок объектов в программе VisualCS.

2.8 Заключение

В настоящей главе представлено описание разработанного метода визуализации хромосомспецифичных последовательностей ДНК на FISH-изображениях (метод VISSIS). Метод работает на цветных изображениях, включающих в себя два канала с результатами FISH ДНК-проб, полученных из разных индивидуальных хромосом или хромосомных районов, и один канал с результатами DAPI-окрашивания. В основе метода лежит допущение, что сигналы, обусловленные повторяющимися последовательностями, присутствующими в обеих ДНК-пробах, должны иметь близкие значения интенсивности. Сигнал хромосомспецифичных последовательностей оценивается как разница между

суммарным сигналом и суммой неспецифичного и фонового сигналов. Интенсивности неспецифичного и фонового сигналов оцениваются на основе анализа результатов FISH-изображения с другой ДНК-пробой.

Метод VISSIS включает в себя следующие основные этапы: выделение объектов, которые являются хромосомами; калибровка интенсивностей сигналов FISH; классификация объектов в зависимости от значений отношения интенсивности сигналов FISH в пикселях объектов; выделение сигнала хромосомоспецифичных последовательностей. Метод был реализован в виде программы VisualCS. Результаты апробации метода на изображениях хромосом человека и видов, размер генома которых в несколько раз отличается от генома человека (описторхиды, саранчовые), описаны в следующей главе.

Глава 3. ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО МЕТОДА ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

3.1 Получение изображений результатов FISH районо- и хромосомоспецифичных ДНК-проб

В ходе апробации разработанный метод был применен для обработки набора изображений хромосом видов, существенно различающихся по размеру генома и доле повторяющихся последовательностей в геноме (описторхиды, млекопитающие, саранчовые). Все экспериментальные процедуры проводили сотрудники лаборатории морфологии и функции клеточных структур ИЦиГ СО РАН: Карамышева Татьяна Витальевна (изображения хромосом человека), Джетыбаев Ильяс Еркинович (изображения хромосом саранчовых) и Задесенец Кира Сергеевна (изображения хромосом описторхид). Оригинальные изображения предоставлены перечисленными коллегами.

3.1.1 Получение изображений результатов FISH хромосомоспецифичных ДНК-проб с хромосомами человека

Цитологические препараты хромосом человека готовили по материалам статьи (Henegariu et al., 2001). Сбор интересующего хромосомного материала и последующую его амплификацию в полимеразной цепной реакции с частично вырожденным праймером проводили согласно протоколу (Rubtsov et al., 2000b). Мечение ДНК проводили в дополнительных циклах ПЦР в присутствии дезоксинуклеотидтрифосфата (dTTP), конъюгированного с флуорохромами Alexa Fluor 488-5-dUTP, TAMRA5-dUTP, Spectrum Orange. Общую окраску ДНК осуществляли красителем DAPI. Наборы ДНК-проб были созданы для анализа хромосомных патологий человека, но в рамках текущей работы FISH ДНК-проб проводили на метафазные хромосомы здорового человека.

FISH осуществляли как по стандартному протоколу CISS-гибридизации (Lichter et al., 1988), так и без супрессии гибридизации повторяющихся последовательностей. Проведена FISH с метафазными хромосомами человека с использованием следующих пар микродиссекционных ДНК-проб:

- FISH пар ДНК-проб, полученных из следующих хромосом человека 3 и 7, 7 и 19, 8 и 10, 10 и 15, 10 и 19, 12 и 15, 13 и 18, 15 и 19, 18 и 19, 19 и X, без супрессии гибридизации повторяющихся последовательностей;
- CISS-гибридизация пар ДНК-проб из хромосом человека 1 и 2, 3 и 7, 10 и 15;

Пары ДНК-проб подбирали с учетом соотношения количества коротких и длинных диспергированных повторяющихся последовательностей (short interspersed elements – SINE, long interspersed elements – LINE) в исходных хромосомах. FISH-изображения использовали для апробации разработанного метода, а также для оценки зависимости результатов компьютерной обработки от содержания разных типов диспергированных повторов в хромосомах, из которых были получены ДНК-пробы.

Также были получены изображения результатов FISH на метафазных хромосомах человека с мечеными Alu-последовательностями. Указанные изображения использовали для анализа распределения Alu-последовательностей на хромосомах. Приготовление ДНК-пробы из фрагмента ДНК, гомологичного части Alu-повтора, и проведение FISH осуществляли по протоколу (Lichter et al., 1990). Мечение Alu-последовательностей проводили в дополнительных циклах ПЦР в присутствии dTTP, конъюгированного с флуорохромом Spectrum Orange.

Микроскопический анализ был выполнен в ЦКП микроскопического анализа биологических объектов СО РАН на микроскопе AxioPlan 2 Imaging microscope (Zeiss, Germany) с набором фильтров 49 (Zeiss, Germany), SP101 FITC (CHROMA, USA) и SP103v1 Cy3tmv1 (CHROMA, USA), CCD-камерой (CV M300, JAI Corporation, Japan). Для регистрации использовали программное обеспечение ISIS4 (METASystems GmbH).

3.1.2 Получение изображений результатов FISH хромосомоспецифичных ДНК-проб с хромосомами описторхид

Цитологические препараты хромосом описторхид готовили по методике, описанной Задесенец с соавторами (Zadesenets et al., 2012). Микродиссекция хромосом и последующая амплификация ДНК диссектированного материала были выполнены по стандартному протоколу (Rubtsov et al., 2000a). Мечение полученных продуктов ПЦР проводили AlexaFluor488, биотином или диоксигенином в

дополнительных циклах ПЦР. Общую окраску ДНК осуществляли красителем DAPI. FISH проводили по протоколу (Pinkel et al., 1986).

Была выполнена FISH ДНК-проб, полученных из хромосом 1 и 2 вида *Metorchis xanthosomus*, и FISH ДНК-проб, полученных из хромосом 1 и 2 вида *Opisthorchis felineus*, с мейотическими хромосомами исходных видов без супрессии гибридизации повторяющихся последовательностей.

Микроскопический анализ препаратов хромосом выполняли на микроскопе AxioPlan 2 Imaging microscope (Zeiss, Germany) с набором фильтров 49 (Zeiss, Germany), SP101 FITC (CHROMA, USA) и SP103v1 Cy3.5v1 (CHROMA, USA), CCD-камерой (CV M300, JAI Corporation, Japan), программное обеспечение ISIS4 (METASystems GmbH).

Изображения были получены в рамках изучения особенностей структурно-функциональной организации хромосом описторхид для исследования, проводимого Задесенец К.С. Виды описторхид рассматривались как пример видов, для которых ранее не удавалось провести CISS-гибридизацию из-за трудностей получения Cot-1 ДНК.

3.1.3 Получение изображений результатов FISH районо- и хромосомспецифичных ДНК-проб с хромосомами саранчовых

Препараты хромосом были приготовлены по материалам статьи (Bugrov et al., 2004). Приготовление микродиссекционных ДНК-проб проводили по измененному стандартному протоколу (Rubtsov et al., 2000a) (с модификациями в протоколе можно ознакомиться в статье (Jetybaev et al., 2017)). Мечение ДНК проводили в дополнительных циклах ПЦР в присутствии dUTP, конъюгированного с флуорохромами Tamra-5-dUTP и Fluorescein-12-dUTP. Общую окраску ДНК осуществляли красителем DAPI. FISH выполняли по протоколу (Pinkel et al., 1986). Проведены следующие FISH микродиссекционных ДНК-проб с митотическими и мейотическими хромосомами саранчовых:

- FISH ДНК-проб из нео-Y хромосомы и короткого плеча нео-X хромосомы *Nocaracris cyanipes* без супрессии гибридизации повторяющихся последовательностей;

- FISH ДНК-проб из нео-Y хромосомы и короткого плеча нео-X хромосомы *Nocaracris tardus* без супрессии гибридизации повторяющихся последовательностей;
- FISH ДНК-проб из нео-Y хромосомы и короткого плеча нео-X хромосомы *Nocaracris rubripes* без супрессии гибридизации повторяющихся последовательностей;
- FISH ДНК-проб из нео-Y хромосомы и короткого плеча нео-X хромосомы *Paranocarodes tolunayi tolunayi* без супрессии гибридизации повторяющихся последовательностей;
- FISH ДНК-проб из проксимальной части нео-Y хромосомы и короткого плеча нео-X хромосомы *Asiotmethis heptapotamicus* без супрессии гибридизации повторяющихся последовательностей.

Также была выполнена кросс-гибридизация полученных ДНК-проб на другие виды саранчовых семейства *Pamphagidae*:

- FISH ДНК-проб из нео-половых хромосом видов *Nocaracris cyanipes*, *Nocaracris tardus*, *Nocaracris rubripes* и *Paranocarodes tolunayi tolunayi* на хромосомы видов *Nocaracris citripes*, *Nocaracris cyanipes*, *Nocaracris idrisi*, *Nocaracris rubripes*, *Nocaracris furvus furvus*, *Nocaracris sureyana*, *Nocaracris tardus*, *Paranocarodes anatoliensis anatoliensis*, *Paranocarodes karabagi*, *Paranocarodes tolunayi tolunayi*, *Paranocarodes turkmen*.
- FISH ДНК-проб из нео-половых хромосом вида *Asiotmethis heptapotamicus* на хромосомы видов *Asiotmethis turritus*, *Asiotmethis limbatus*, *Asiotmethis muricatus*, *Glyphotmethis adaliae*, *Glyphotmethis holtzi pulchripes*, *Glyphotmethis efe*, *Glyphotmethis dimorphus dimorphus*.
- FISH ДНК-проб из нео-половых хромосом видов *Nocaracris cyanipes* на хромосомы вида *Asiotmethis muricatus*

Микроскопический анализ был выполнен в ЦКП микроскопического анализа биологических объектов СО РАН на микроскопе AxioImager.M1 (Zeiss, Germany) с набором фильтров 49, 46HE, 43HE (Zeiss, Germany), CCD-камерой ProgRes MF (JenaOptik, Jena, Germany). Для регистрации использовали программное обеспечение ISIS5 (METASystems GmbH).

Изображения были получены в рамках изучения молекулярной эволюции Y-хромосомы у семейства *Pamphagidae* для исследования механизмов эволюции нео-половых хромосом саранчовых, проводимого Джетыбаевым И.Е. Ранее не удавалось провести идентификацию материала целых хромосом и крупных хромосомных районов у видов семейства *Pamphagidae* с помощью FISH из-за большого количества повторяющихся последовательностей ДНК в геномах саранчовых. Изображения хромосом саранчовых рассматривались как пример удачного использования метода для анализа хромосом видов, для которых CISS-гибридизация не дает удовлетворительных результатов.

3.2 Применение метода VISSIS для идентификации материала целых хромосом у человека

Геном человека является одним из наиболее изученных геномов сложных организмов. Кроме того, в литературе накоплен большой объем данных по применению FISH для идентификации районов хромосом этого вида. Указанные причины делают хромосомы человека удобным объектами для апробации метода VISSIS.

Первоначально метод был успешно протестирован на изображениях FISH с супрессией повторяющихся последовательностей (обработка микроскопических изображений, полученных в результате CISS-гибридизации) (Богомоллов и др., 2012; Гайнер и др., 2017). Такие изображения очень удобны для тестирования этапа классификации объектов. Класс объектов на изображении легко определить из-за низкого уровня интенсивности сигналов диспергированных повторяющихся последовательностей ДНК. Как и ожидалось, метод правильно определил количество классов и безошибочно классифицировал объекты. Его применение привело к улучшению полученных изображений, особенно в случае неполной супрессии гибридации повторяющихся последовательностей. На рисунке 3.1 приведено сравнение результатов FISH, проведенных по стандартному протоколу CISS-гибридизации, до и после компьютерной обработки.

Как видно из рисунков 3.1 А и Б, обработанные изображения FISH с полной супрессией повторяющихся последовательностей визуально практически не отличаются от исходных изображений (компьютерная обработка изображений не дала значимых преимуществ в идентификации материала хромосом). Но в случае

неполной супрессии гибридизации повторяющихся последовательностей метод позволил значительно повысить относительный уровень интенсивности сигнала с хромосомоспецифичных последовательностей ДНК (рисунки 3.1 В и Г) (отношения сигнал/шум на изображениях увеличилось в 4-6 раз). Таким образом, разработанный метод может использоваться для уменьшения требуемого объема Cot-1 ДНК в ходе проведения FISH-диагностики и тем самым снизить стоимость проведения диагностики хромосомных патологий человека.

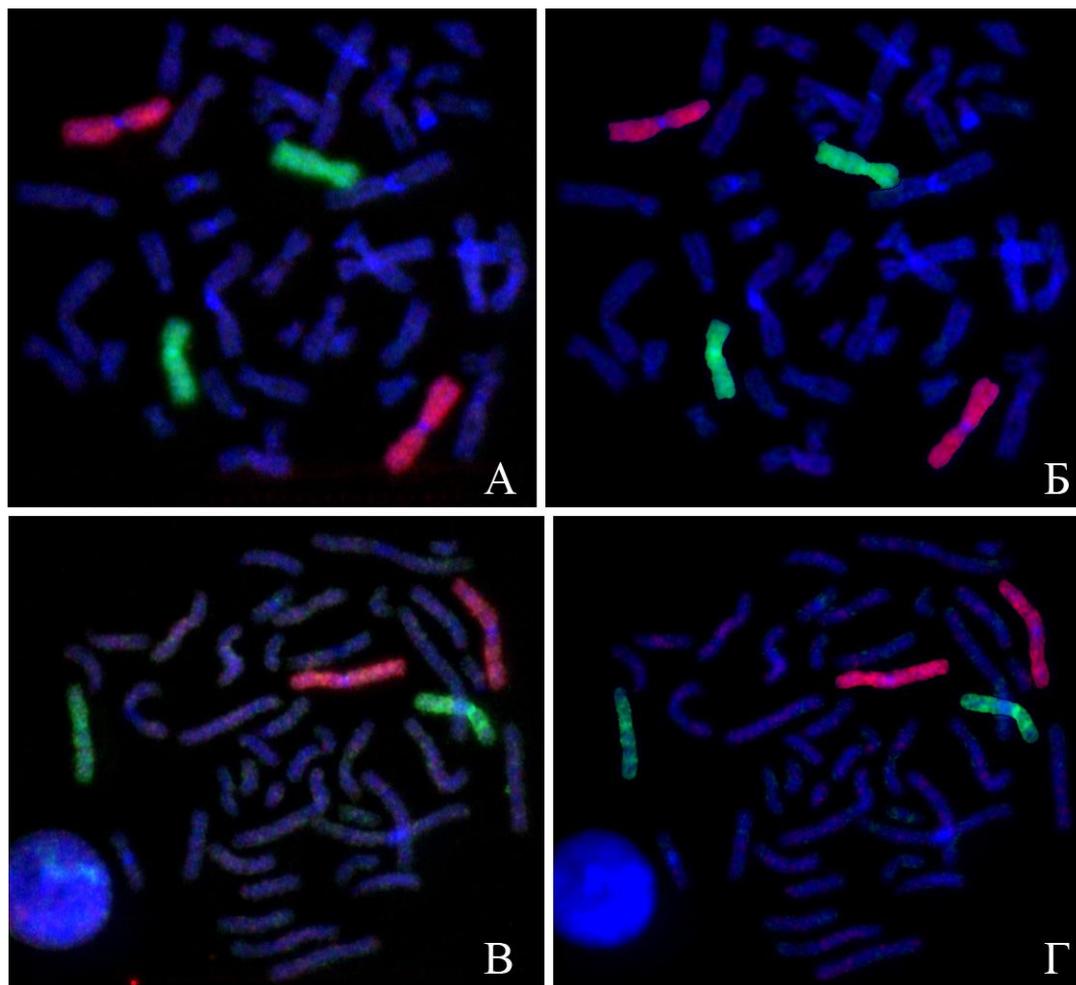


Рисунок 3.1. CISS-гибридизация с ДНК-пробами, полученными из хромосом человека 7 (зеленый) и 3 (красный). **А** – Полная супрессия повторяющихся последовательностей. **Б** – Результат компьютерной обработки изображения **А** с помощью метода VISSIS. **В** – Неполная супрессия повторяющихся последовательностей. **Г** – Результат компьютерной обработки изображения **В** с помощью метода VISSIS.

Дальнейшие исследования, основанные на цифровом анализе нескольких сотен изображений, подтвердили возможность идентифицировать хромосомы человека без супрессии гибридизации повторяющихся последовательностей. Однако, эффективность элиминирования сигнала диспергированных

повторяющихся последовательностей различалась для разных пар ДНК-проб. В качестве наглядного примера на рисунках 3.2 и 3.3 приведены результаты компьютерной обработки FISH ДНК-проб двух пар хромосом человека: 3 и 7, 7 и 19.

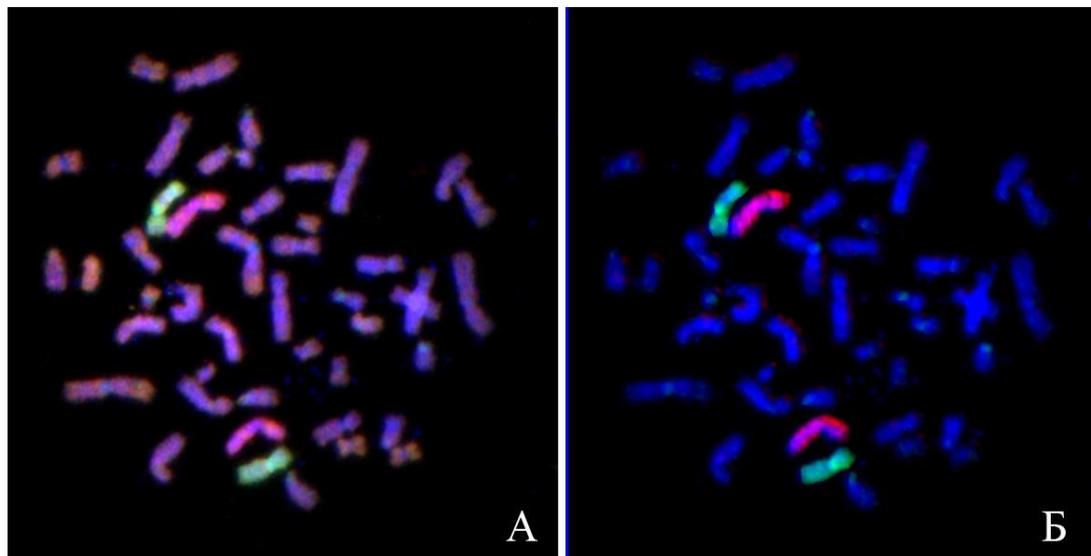


Рисунок 3.2. Двухцветная FISH на хромосомах человека без супрессии гибридизации повторяющихся последовательностей с ДНК-пробами хромосом человека 7 (зеленый) и 3 (красный). **А** – Необработанное исходное изображение. **Б** – Результат применения метода VISSIS. Оценка различия хромосом по соотношению количества коротких и длинных диспергированных повторяющихся последовательностей равна $D_{Chr(7, 3)} = 0.025$.

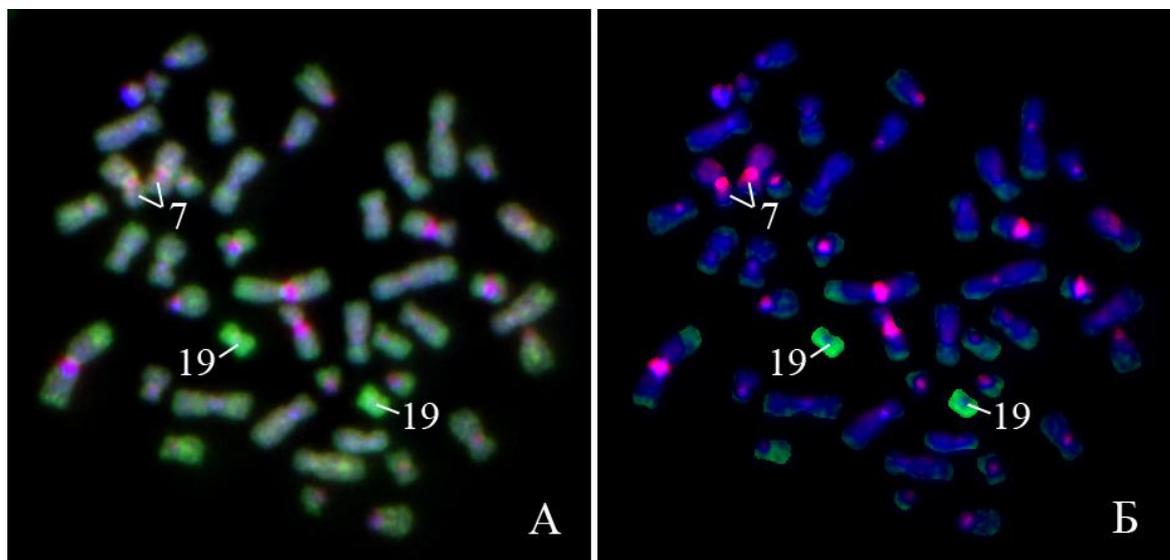


Рисунок 3.3. Двухцветная FISH на хромосомах человека без супрессии гибридизации повторяющихся последовательностей с ДНК-пробами хромосом человека 7 (красный) и 19 (зеленый). **А** – Необработанное исходное изображение. **Б** – Результат применения метода. Оценка различия хромосом по соотношению количества коротких и длинных диспергированных повторяющихся последовательностей равна $D_{Chr(7, 19)} = 0.156$.

На основе визуального анализа обработанных изображений было сделано предположение, что оставшийся после компьютерной обработки неспецифичный сигнал связан с особенностью состава повторяющихся последовательностей в

хромосомах, из которых были получены ДНК-пробы. Действительно, разработанный метод предполагает, что сигналы, обусловленные повторяющимися последовательностями, присутствующими в обеих ДНК-пробах, должны иметь близкие значения интенсивности. На практике хромосомные районы окрашиваются с различной интенсивностью разными ДНК-пробами. Это связано с тем, что у человека и других видов млекопитающих и птиц в разных хромосомах и разных районах их хромосом содержится разное количество повторяющихся последовательностей различных типов. Более того, отличается и их соотношение. Чем сильнее различия в содержании типов повторяющихся последовательностей, представленных в ДНК-пробах, тем более существенным может быть разница в интенсивности окрашивания хромосомных районов и, как следствие, более интенсивный неспецифичный сигнал, выявляемый на обработанном изображении.

Для проверки предположения была проведена оценка доли SINE (Short interspersed nuclear element) и LINE (Long interspersed nuclear element) последовательностей в парах хромосом, из которых были получены ДНК-пробы. Известно, что эти последовательности имеют разную представленность в хромосомных районах: G-позитивные районы обогащены повторами LINE, а G-негативные – SINE. В соответствии с выдвинутым в работе предположением, чем сильнее различаются хромосомы, из которых получены ДНК-пробы, по соотношению количества SINE/LINE последовательностей, тем хуже будет элиминирование сигнала диспергированных повторяющихся последовательностей. Данные по представленности SINE/LINE последовательностей в хромосоме взяты из базы данных Genome Browser USCS (<https://genome-euro.ucsc.edu/index.html>). Общая картина распределения хромосом человека в зависимости от доли SINE/LINE в них отображена на рисунке 3.4.

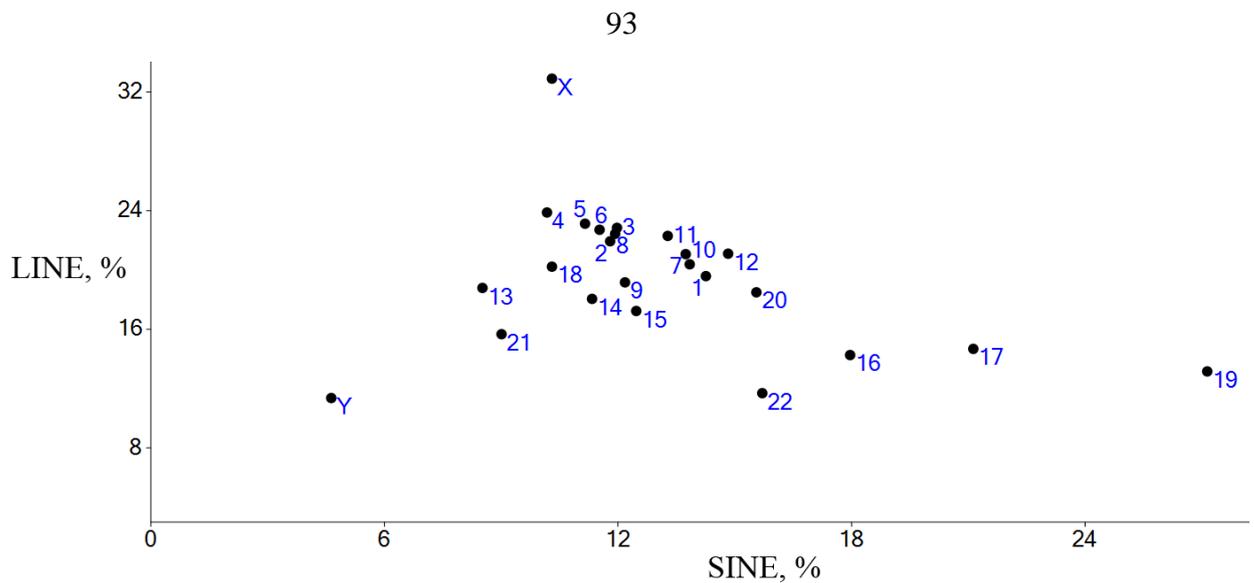


Рисунок 3.4. Распределение хромосом человека в зависимости от типов повторяющихся последовательностей. По оси ординат – доля LINE (Long interspersed nuclear element) последовательностей в хромосоме, по оси абсцисс – доля SINE (Short interspersed nuclear element) последовательностей в хромосоме. Изображение создано на основе информации из базы данных Genome Browser USCS (<https://genome-euro.ucsc.edu/index.html>).

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

$$DChr(a, b) = \sqrt[2]{(s_a - s_b)^2 + (l_a - l_b)^2} \quad (3.1)$$

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 (signal to noise ratio – SNR):

$$SNR = \frac{\mu_s}{\mu_{ns} + 1} \quad (3.2)$$

где μ_s – среднее значение интенсивности специфического сигнала (сигнал на хромосомах, содержащих ДНК, гомологичную хромосоме, из которой была получена ДНК-проба); μ_{ns} – среднее значение интенсивности неспецифического сигнала (сигнал на хромосомах, содержащих ДНК, гомологичную только повторяющимися последовательностям). Ранее подобный подход к оценке эффективности визуализации специфического сигнала был предложен в работе (Bolzer et al., 1999).

Результаты вычисления отношения сигнал/шум для различных FISH-изображений приведены в таблице 3.1.

Таблица 3.1 – Среднее значение отношения сигнал/шум (SNR) и его стандартное отклонение после компьютерной обработки FISH ДНК-проб из хромосом человека

ДНК-пробы из хромосом человека <i>a</i> и <i>b</i>		SNR		<i>DChr(a,b)</i>	Количество FISH-изображений
Номер хромосомы <i>a</i>	Номер хромосомы <i>b</i>	Для результатов FISH ДНК-пробы из хромосомы <i>a</i>	Для результатов FISH ДНК-пробы из хромосомы <i>b</i>		
3	7	8.96±1.61	20.55±3.99	0.025	11
7	19	3.57±0.41	10.15±1.25	0.1556	12
8	10	17.56±2.42	11.25±1.22	0.0281	5
10	15	8.28±1.25	8.2±0.62	0.0343	25
10	19	6.5±0.36	13.85±1.19	0.1513	7
12	15	9.36±2.84	7.51±0.86	0.0452	35
13	18	17.74±1.66	15.73±3.19	0.0229	13
15	19	6.21±0.95	11.79±1.23	0.1522	11
18	19	6.46±1.3	9.15±1.12	0.1826	19
19	X	8.4±0.82	6.28±0.92	0.2594	7

Полученные данные указывают на зависимость эффективности метода от выбора пары ДНК-проб при проведении FISH. Наблюдается отрицательная корреляция между отношением сигнал/шум (SNR) и оценкой различий хромосом *DChr* ($p < 0.01$). На рисунке 3.5 приведены значения SNR на обработанных изображениях FISH ДНК-проб из 10, 15 и 19 хромосом человека при различных вариантах вторых ДНК-проб.

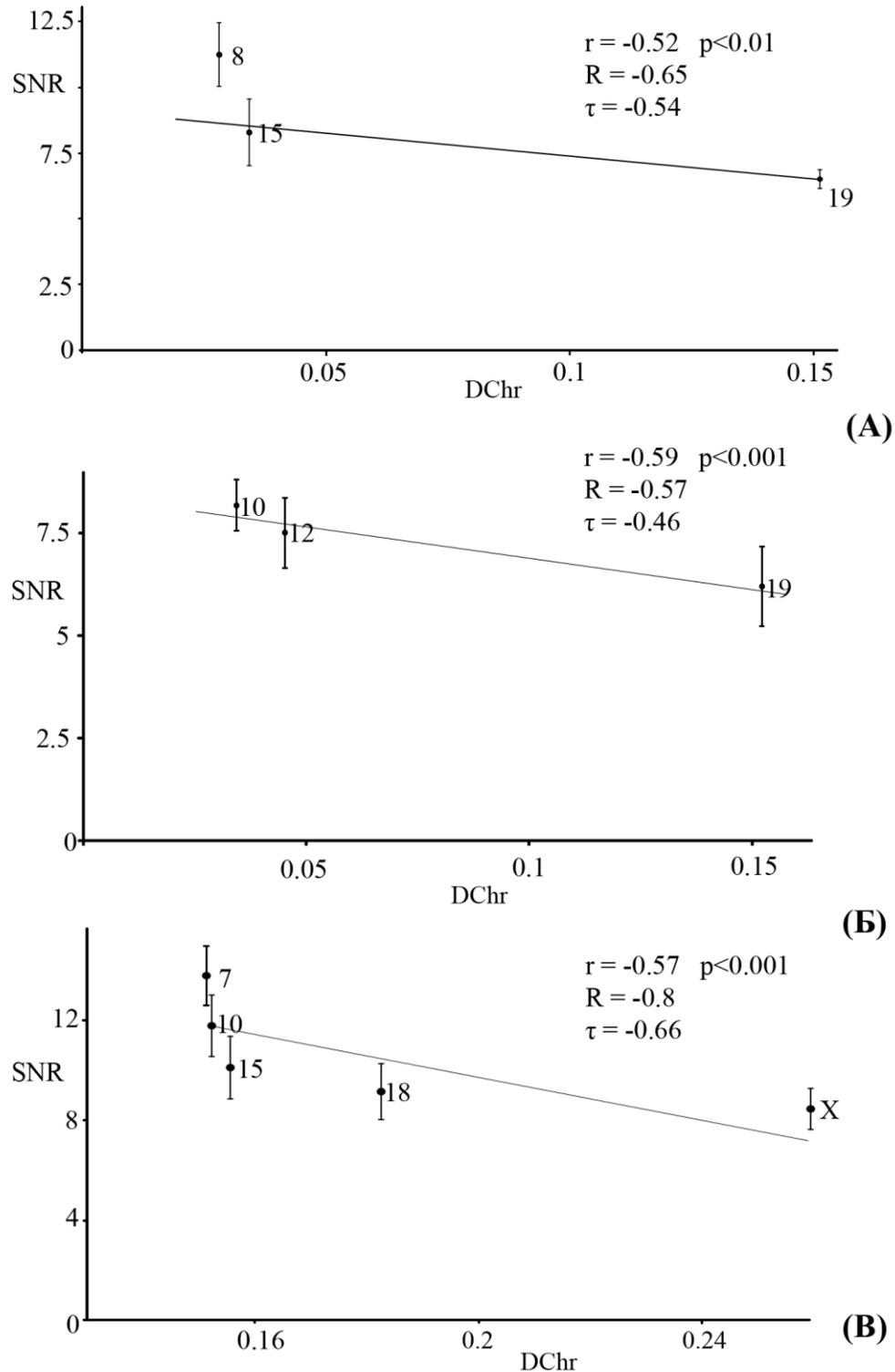


Рисунок 3.5. Корреляция отношения сигнал/шум (SNR) и оценки различия хромосом (*DChr*). **А** – SNR на обработанных изображениях FISH ДНК-проб из 10 хромосомы человека при различных вариантах второй ДНК-пробы. **Б** – SNR на обработанных изображениях FISH ДНК-проб из 15 хромосомы человека при различных вариантах второй ДНК-пробы. **В** – SNR на обработанных изображениях FISH ДНК-проб из 19 хромосомы человека при различных вариантах второй ДНК-пробы. На графике указано среднее значение и стандартное отклонение SNR. 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.

Стоит отметить, что значения SNR чувствительны к изменениям интенсивности сигналов на изображении. Отклонения интенсивности от реальных значений могут быть связаны с ошибками классификации и неправильной оценкой поправочного коэффициента. Поэтому была проведена проверка результатов классификации и, при необходимости, их исправление для корректного вычисления SNR.

Таблица 3.2 – Результаты классификации объектов на изображении FISH ДНК-проб из хромосом человека

Для результатов FISH ДНК-пробы из хромосомы <i>a</i> и <i>b</i>		Ошибка первого рода (E_1)	Ошибка второго рода (E_2)	Количество объектов класса S_a и S_b на изображениях	Общее количество хромосом на изображениях
Номер хромосомы <i>a</i>	Номер хромосомы <i>b</i>				
3	7	0	0.0102	28	322
7	19	0.0625	0.0318	48	552
8	10	0	0.0106	18	206
10	15	0	0.0324	100	1150
10	19	0	0.0306	28	322
12	15	0	0.0463	139	1609
13	18	0	0	53	620
15	19	0	0.0537	44	491
18	19	0.0132	0.0514	76	874
19	X	0.0476	0.0408	21	322

Примечание – Класс S_a включает объекты, содержащие ДНК, специфичную для хромосомы *a*. Класс S_b включает объекты, содержащие ДНК, специфичную для хромосомы *b*.

С другой стороны, ошибки классификации можно рассматривать как оценку точности метода. В таблице 3.2 приведены результаты классификации объектов на изображениях FISH ДНК-проб из целых хромосом человека.

Ожидаемым результатом является рост количества ошибок с увеличением величины различий хромосом, из которых получены ДНК-пробы (рисунок 3.6).

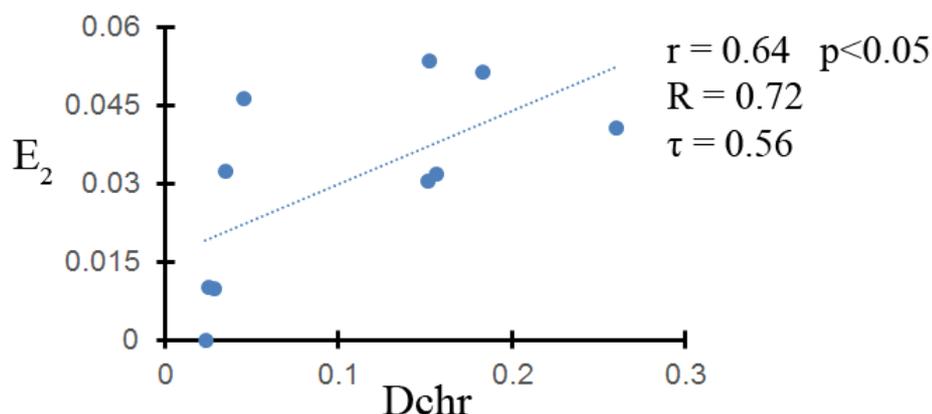


Рисунок 3.6. График зависимости ошибок классификации от величины различий хромосом, из которых получены ДНК-пробы, по соотношению количества SINE/LINE последовательностей. Точечной линией указана линия тренда. r , R и τ – значения коэффициентов корреляции Пирсона, Спирмена и Кендалла, соответственно.

Как упоминалось ранее, компьютерная обработка результатов FISH ДНК-проб из хромосом, контрастных по содержанию LINE- и SINE-повторов, приводит к сохранению на обработанном изображении интенсивного неспецифичного сигнала. Интенсивные сигналы повторяющихся последовательностей являются причиной ошибок классификатора, который относит объекты класса *Cns* к другим классам (ошибки классификации второго рода).

В некоторых случаях наблюдается неправильная классификация объектов класса *Ca* и *Cb* (ошибки классификации первого рода). Ошибки объясняются наличием на хромосомах больших по размеру областей, в которых значения отношения интенсивностей сигналов от разных флуорохромов не характерны для такого класса объектов. Как следствие, объекты располагаются в пространстве признаков ближе к объектам класса *Cns*, чем к объектам своего класса. Появление таких областей может быть следствием наложения хромосом, неравномерного освещения метафазных хромосом источником света, слабой представленности ДНК-пробы в хромосоме (плохое качество FISH) и т.д.

В работе также проводили 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 (рисунок 3.7).

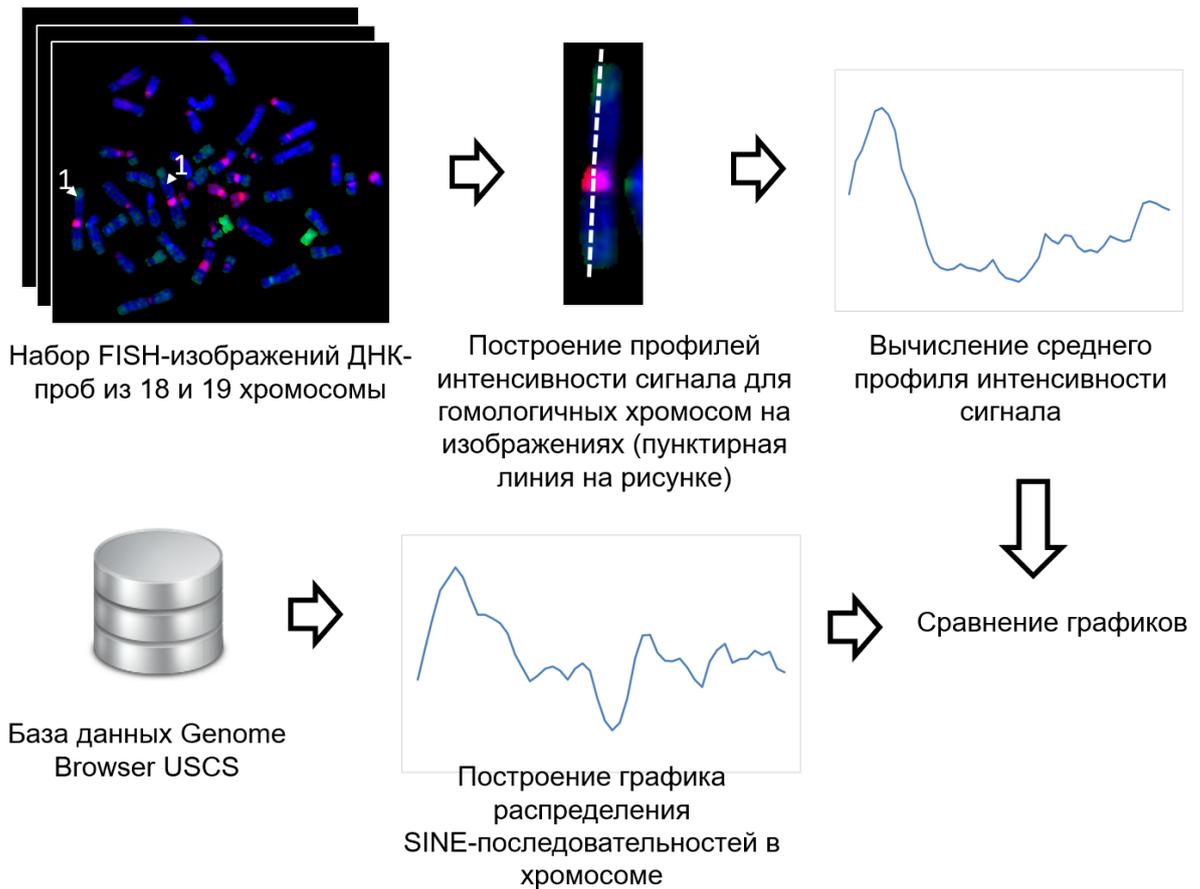


Рисунок 3.7. Схема эксперимента по сравнению средних профилей интенсивности, построенных для хромосом на обработанных изображениях, с распределением SINE последовательностей.

Хромосомы 18 и 19 человека значительно отличаются по относительному содержанию в них разных типов диспергированных повторов: хромосома 18 обогащена LINEs, тогда как хромосома 19 – SINEs (рисунок 3.4). Поэтому было выдвинуто предположение, что профиль интенсивности неспецифичного сигнала на изображениях хромосом после применения VISSIS должен соответствовать распределению SINE или LINE последовательностей.

Профили интенсивности сигналов проводили вдоль длины хромосом. Полученные профили приводились к одному размеру, затем вычислялся средний профиль для гомологичных хромосом в наборе изображений. Длину профиля установили равным 50 точкам на основе анализа размеров самой маленькой хромосомы в наборе FISH-изображений.

Данные по распределению Alu-повторов (один из наиболее обширных членов класса SINE) и SINE последовательностей в конкретной хромосоме были взяты из базы данных Genome Browser UCSC (<https://genome-euro.ucsc.edu/index.html>). Для

удобства вычислений графики распределений (профили распределений) также состояли из 50 точек. Для их построения хромосомы были разбиты на 50 фрагментов. Затем оценивалось количество SINE последовательностей и Alu-повторов в каждом фрагменте.

Нормализацию профилей проводили по методике, описанной в работе (Legrand et al., 2008). Новое значение профилей вычисляли как:

$$P(j) = \frac{P(j) - \mu_p}{\sigma_p} \quad (j = 1, \dots, N) \quad (3.5)$$

где j – номер точки в профиле; N – количество точек в профиле (в данном случае $N=50$); $P(j)$ – значения профиля интенсивности или профиля распределения повторов в точке j ; μ_p – среднее значение профиля; σ_p – стандартное отклонение значений профиля.

Профили были построены для хромосом 1, 3, 7, 12, 16 и 17 по 14 изображениям хромосом. Результаты проиллюстрированы на рисунке 3.8. Из рисунка видно, что профили интенсивности и графики распределения последовательностей SINE, построенные для одной хромосомы, имеют схожую форму.

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

$$RMSD = \sqrt{\frac{\sum_{j=1}^N (P_{img}(j) - P_{SINE}(j))^2}{N}} \quad (3.6)$$

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 $N=50$). Аналогичный подход к сравнению профилей интенсивностей был предложен в работе (Roshtkhari, Setarehdan, 2008).

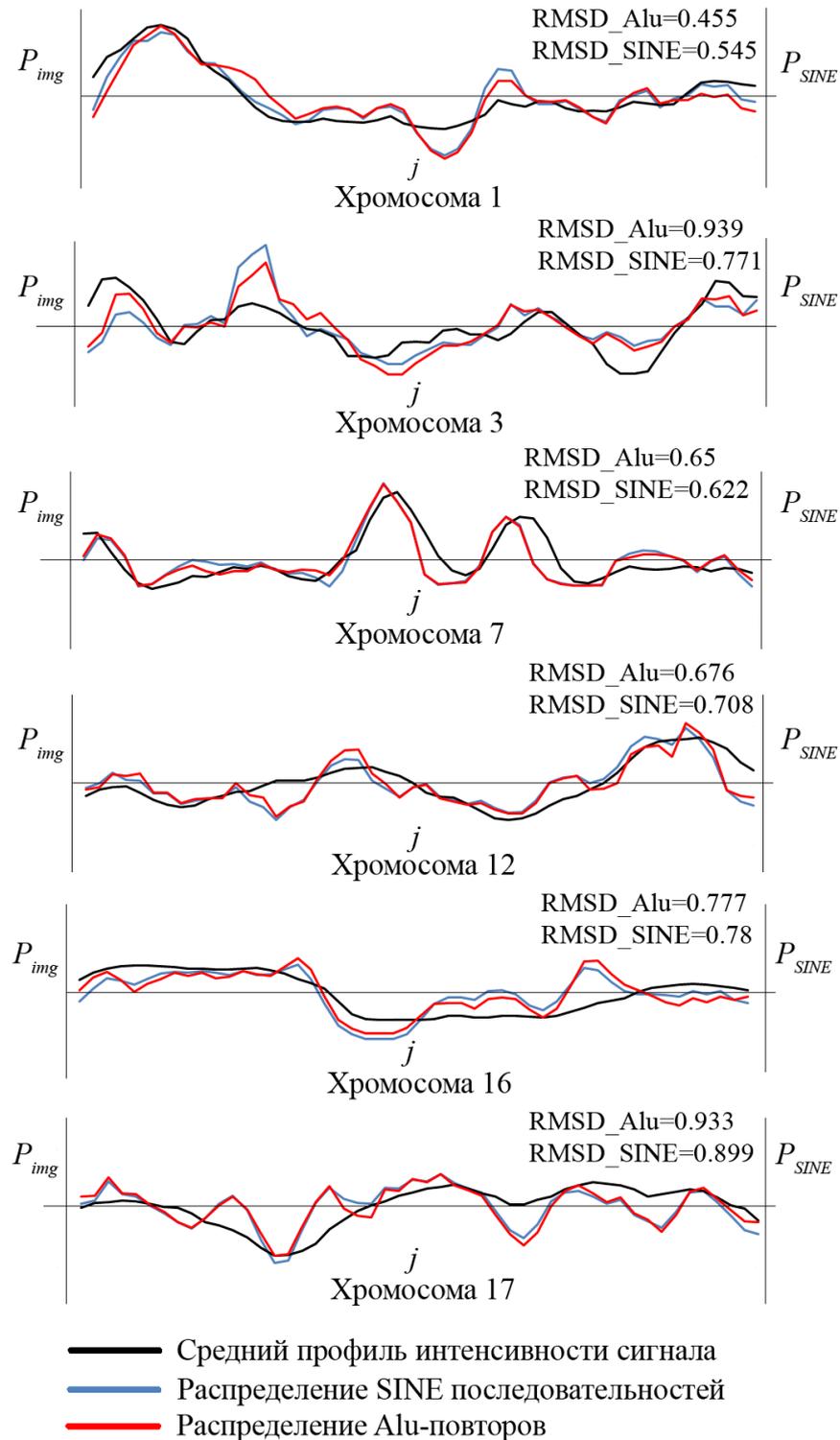


Рисунок 3.8. Сравнительный анализ средних профилей интенсивности, построенных для хромосом 1, 3, 7, 12, 16 и 17 на обработанных изображениях FISH ДНК-проб из хромосом 18 и 19, с данными о распределении SINE последовательностей в соответствующих хромосомах. P_{img} – нормированное значение интенсивности сигнала в точке профиля для среднего профиля интенсивности. P_{SINE} – количество повторов в хромосомном фрагменте (значение нормировано) для распределения SINE последовательностей в хромосоме. j – номер точки профиля. RMSD_Alu и RMSD_SINE – количественные оценки различий среднего профиля интенсивности, с распределением Alu-повторов и SINE последовательностей, соответственно.

Среднее значение оценки RMSD и ее стандартное отклонение при сравнении средних профилей интенсивности и распределения Alu-последовательностей (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, проводили 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 (<https://genome-euro.ucsc.edu/index.html>) (RMSD_AluAlu). Так как оба профиля отражают распределение Alu-последовательностей в хромосомах, то их можно считать одинаковыми. Оценку различий между профилями (RMSD_AluAlu) в дальнейшем использовали в качестве диапазона значений RMSD, при котором профили можно считать одинаковыми. Профили были построены для хромосом 1, 3, 6, 7, 11, 12, 16 и 17 по десяти изображениям хромосом (рисунок 3.9). Среднее значение и стандартное отклонение RMSD_AluAlu составили 0.76 ± 0.18 .

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 (средние профили интенсивности совпадают с профилями распределения SINE последовательностей), а главной причиной его сохранения являются различия в окрашивании хромосомных районов разными ДНК-пробами.

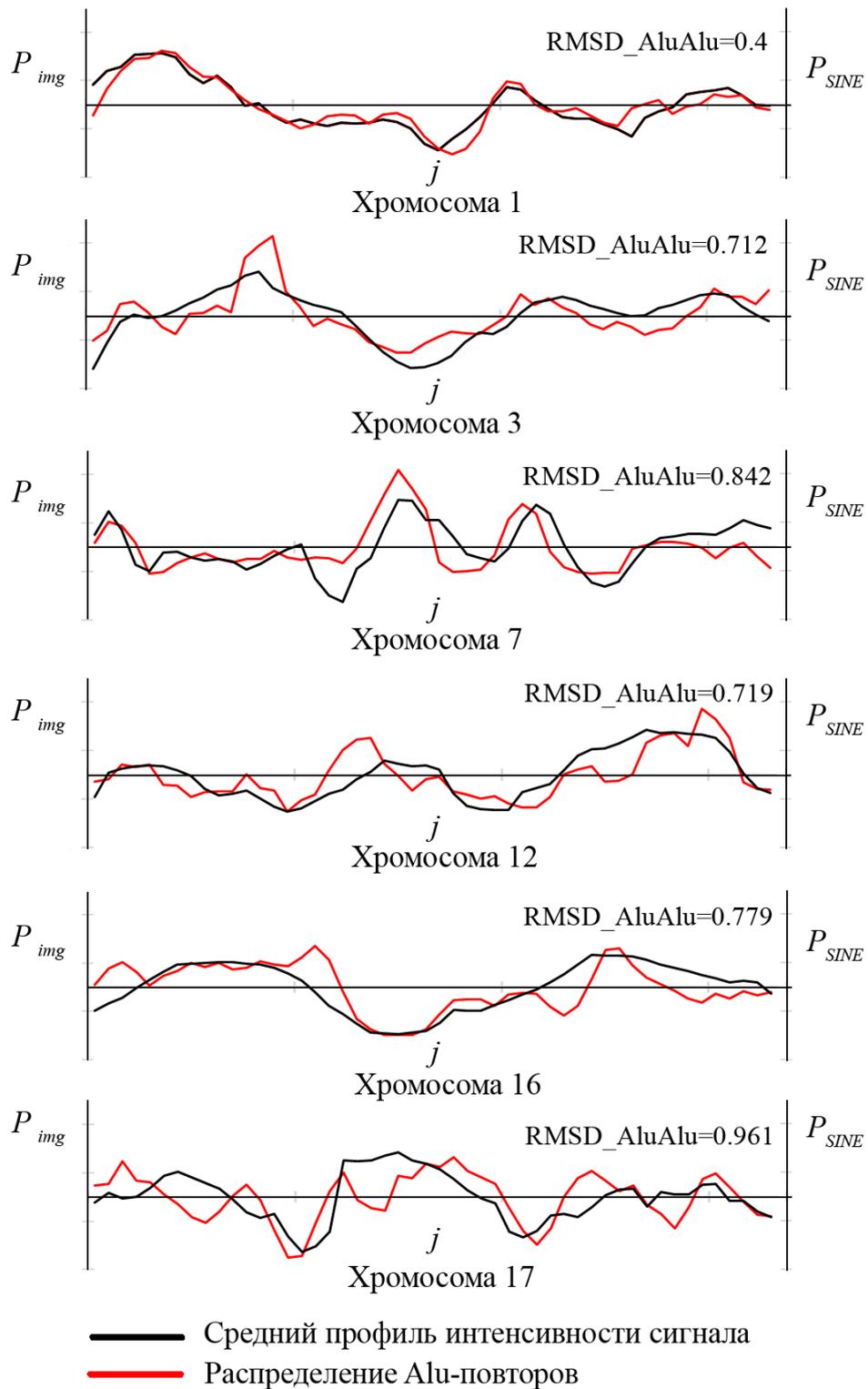


Рисунок 3.9. Сравнительный анализ средних профилей интенсивности, построенных для хромосом 1, 3, 7, 12, 16 и 17 на изображениях FISH хромосом человека с мечеными Alu-последовательностями, с данными о распределении Alu-повторов в соответствующих хромосомах. P_{img} – нормированное значение интенсивности 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. $RMSD_AluAlu$ – количественные оценки различий среднего профиля интенсивности с распределением Alu-повторов.

Исходя из вышеизложенного материала, можно сделать вывод, что при использовании метода VISSIS необходимо учитывать особенности состава повторяющихся последовательностей хромосом, из которых получены ДНК-пробы. Подбор комплектов ДНК-проб из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК, позволяет значительно уменьшить интенсивность сигнала диспергированных повторяющихся последовательностей и проанализировать распределение хромосомоспецифичных последовательностей на самих хромосомах после компьютерной обработки.

В таблице 3.3 указаны рекомендации по выбору пар хромосомоспецифичных ДНК-проб для достижения максимальной элиминации сигнала повторяющихся диспергированных последовательностей ДНК посредством компьютерной обработки FISH-изображений методом VISSIS.

Таблица 3.3 – 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*

Номер хромосомы <i>a</i> , из которой была получена первая ДНК-проба	Хромосомы, рекомендованные для получения второй ДНК-пробы
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 <i>a</i> по соотношению количества SINE/LINE последовательностей	

Для каждой хромосомы человека *a* приведены данные не только о хромосоме, имеющей наиболее сходный с ней состав SINE/LINE последовательностей (оценка *Dchr* минимальна), но также о хромосомах, для которых оценка различий по соотношению количества SINE/LINE последовательностей с хромосомой *a* не превышала 0.03 ($Dchr \leq 0.03$). Верхняя граница *Dchr* (значение 0.03) выбрана на основе визуального анализа обработанных FISH-изображений, данных по классификации объектов и значений SNR. Обработанные FISH-изображения, полученные с ДНК-пробами из хромосом со значениями *Dchr* меньше 0.03, содержат слабоинтенсивный неспецифичный сигнал и позволяют легко идентифицировать материал исходных хромосом. Кроме того, при таких значениях *Dchr* наблюдается малое количество ошибок классификации и большие значения SNR, (см. рисунок 3.6 и таблицу 3.1).

Размер гетерохроматиновых районов может сильно варьировать внутри вида, что приводит к различному соотношению количества SINE/LINE последовательностей в хромосомах у разных особей. У человека размер прицентромерных С-позитивных районов хромосом 1, 9, 16 и дистального района q-плеча Y-хромосомы может отличаться более, чем на порядок. Большие размеры блоков гетерохроматина могут являться причиной неудовлетворительных результатов компьютерной обработки при анализе перечисленных хромосом.

Приведенные выше рекомендации относятся к анализу хромосом с небольшими по размеру C-позитивными районами.

ДНК-пробы из хромосом, контрастных по содержанию типов повторяющихся последовательностей, могут быть использованы для анализа закономерностей распределения повторяющихся последовательностей в исследовании геномов малоизученных видов. У млекопитающих пары хромосом, сильно различающиеся по содержанию SINE/LINE последовательностей, можно выбрать на основе дифференциального окрашивания хромосом. Анализ размеров и количества бэндов на хромосомах позволяет оценить долю SINE и LINE последовательностей в них. Отметим, что существует взаимозависимость распределения Alu-повторов и генов на хромосоме (Grover et al., 2003; Pidpala et al., 2008). Поэтому при небольшой модификации метода VISSIS его можно также использовать для получения данных о возможной локализации районов, богатых генами.

Необходимо подчеркнуть, что информация о составе повторяющихся последовательностей в хромосомах исследуемого вида не всегда известна. Это может привести к ошибочной интерпретации результатов применения VISSIS при использовании ДНК-проб из хромосом, контрастных по содержанию типов повторяющихся последовательностей ДНК. В программе VisualCS предусмотрена постобработка изображений, которая позволяет оценить вероятность обнаружения сигнала заданной интенсивности на объектах разных классов. Более подробную информацию о процедуре обработки можно найти в разделе 2.6. Предполагается, что такая обработка упростит анализ изображений.

3.3 Сравнение результатов CISS-гибридизации и компьютерной обработки FISH-изображений

В настоящее время основным подходом к решению проблемы идентификации хромосомного материала с помощью FISH является CISS-гибридизация (Lichter et al., 1988). По этой причине новые подходы к улучшению выявления специфичного сигнала, как правило, сравниваются с результатами CISS-гибридизации. В работе было проведено сравнение FISH-изображений, полученных в результате CISS-гибридизации и в результате компьютерной обработки методом VISSIS и методом RENS (Rens et al., 2006).

Среднее значение отношения сигнал/шум на изображениях с результатами CISS-гибридизации колеблется от 5 до 18 (смотрите данные таблицы 3.4). Эти данные согласуются с опубликованной ранее информацией (Bolzer et al., 1999).

Таблица 3.4 – Среднее значение отношения сигнал/шум (SNR) и ее стандартное отклонение для CISS-гибридизации с различными ДНК-пробами

Номер хромосомы, из которой была получена ДНК-проба	SNR	Количество проанализированных изображений
1	6.45±0.84	9
2	5±0.54	9
3	10.91±2.8	18
7	11.19±2.94	18
10	5.64±1.34	24
15	18±4.4	24

Среднее значение отношения сигнал/шум, определяющее эффективность разработанного компьютерного метода VISSIS, колеблется от 3.6 до 20.6, в зависимости от комплекта ДНК-проб (смотрите данные таблицы 3.1). Эта зависимость также наблюдается у компьютерного метода RENS (рисунок 3.9). Оба метода предполагают, что сигналы, обусловленные повторяющимися последовательностями, присутствующими в ДНК-пробах, должны иметь близкие значения интенсивности. На практике разное количество типов повторяющихся последовательностей в хромосомах, из которых получены ДНК-пробы, приводит к различной интенсивности сигналов FISH в хромосомных районах. После цифровой обработки изображений сигналы повторяющихся последовательностей сохраняются в этих районах. Поэтому эффективность элиминации сигнала повторяющихся диспергированных последовательностей ДНК в значительной степени зависит от выбранных пар ДНК-проб.

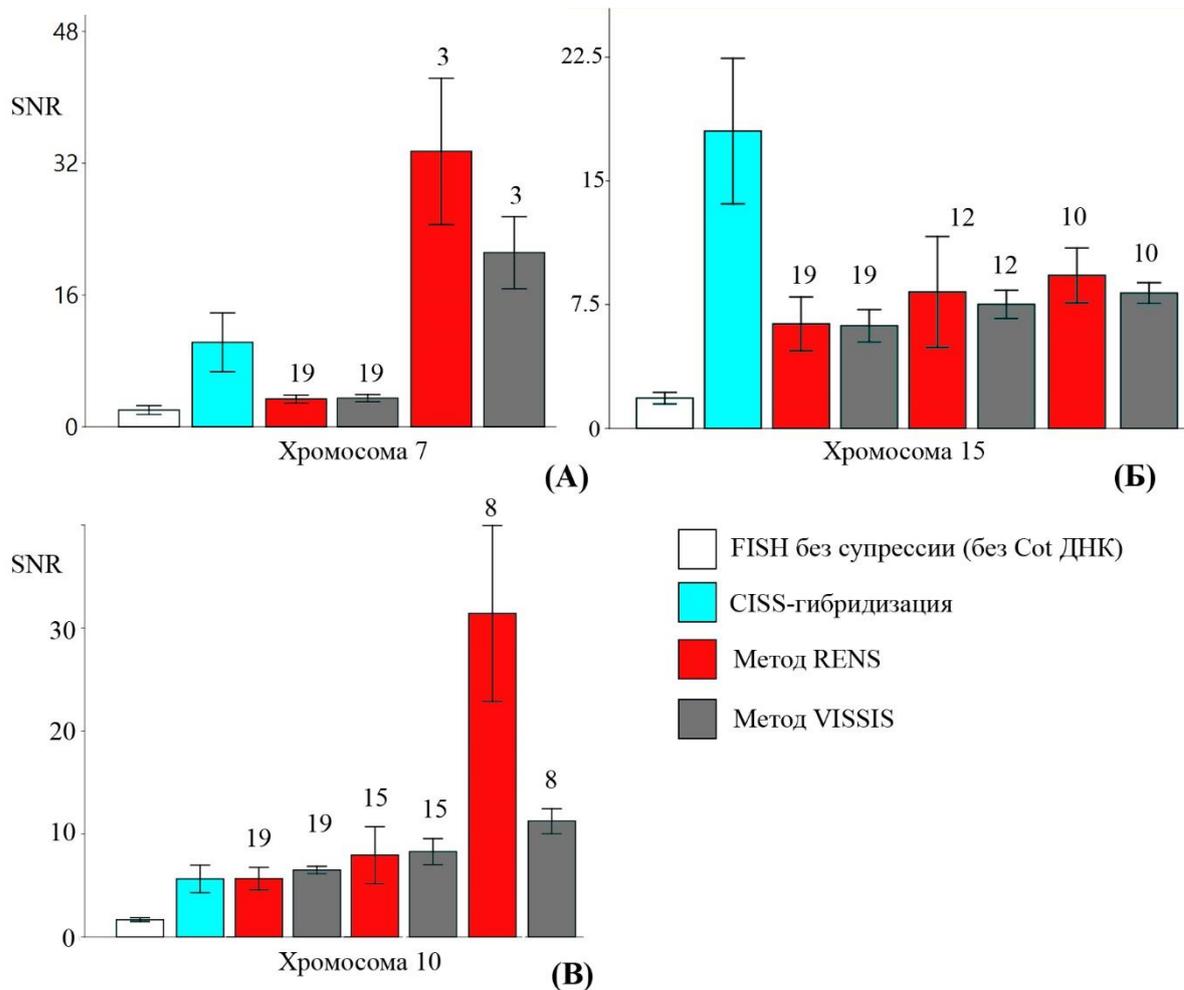


Рисунок 3.10. Сравнение эффективности подходов к усилению интенсивности сигнала с хромосомспецифичных последовательностей при проведении FISH ДНК-проб из хромосом 7 (А), 15 (Б) и 10 (В). Для методов RENS и VISSIS над каждым столбцом диаграммы указан номер хромосомы, из которой получена вторая ДНК-проба. Оценки различия хромосом по соотношению количества SINE/LINE последовательностей равны: $DChr(7, 19) = 0.156$; $DChr(7, 3) = 0.025$; $DChr(15, 19) = 0.1522$; $DChr(12, 15) = 0.0452$; $DChr(10, 15) = 0.0343$; $DChr(8, 10) = 0.0281$; $DChr(10, 15) = 0.1513$.

На рисунке 3.10 показаны значения SNR для CISS-гибридизации ДНК-проб из 7, 10 и 15 хромосомы человека, а также значения SNR до и после компьютерной обработки методами VISSIS и RENS изображений FISH ДНК-проб из 7, 10 и 15 хромосом при различных вариантах второй ДНК-пробы. Из рисунка видно, что компьютерная обработка дает сопоставимые с CISS-гибридизацией результаты в случаях, когда ДНК-пробы созданы из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК. Это делает компьютерные методы хорошей альтернативой CISS-гибридизации. Заметим, что оба компьютерных метода не требуют использования Cot-1 ДНК, что позволяет им избежать недостатков, свойственных CISS-гибридизации.

Метод VISSIS демонстрирует более низкие значения SNR, чем метод RENS. Такое различие объясняется, с одной стороны, наличием большого количества пикселей с нулевой интенсивностью сигнала на обработанных методом RENS изображениях. С другой стороны, метод VISSIS снижает или полностью исключает интенсивность неспецифичного сигнала на всех хромосомах, в том числе на объектах класса *Ca* и *Cb* (эта разница между методами хорошо показана на рисунке 3.11).

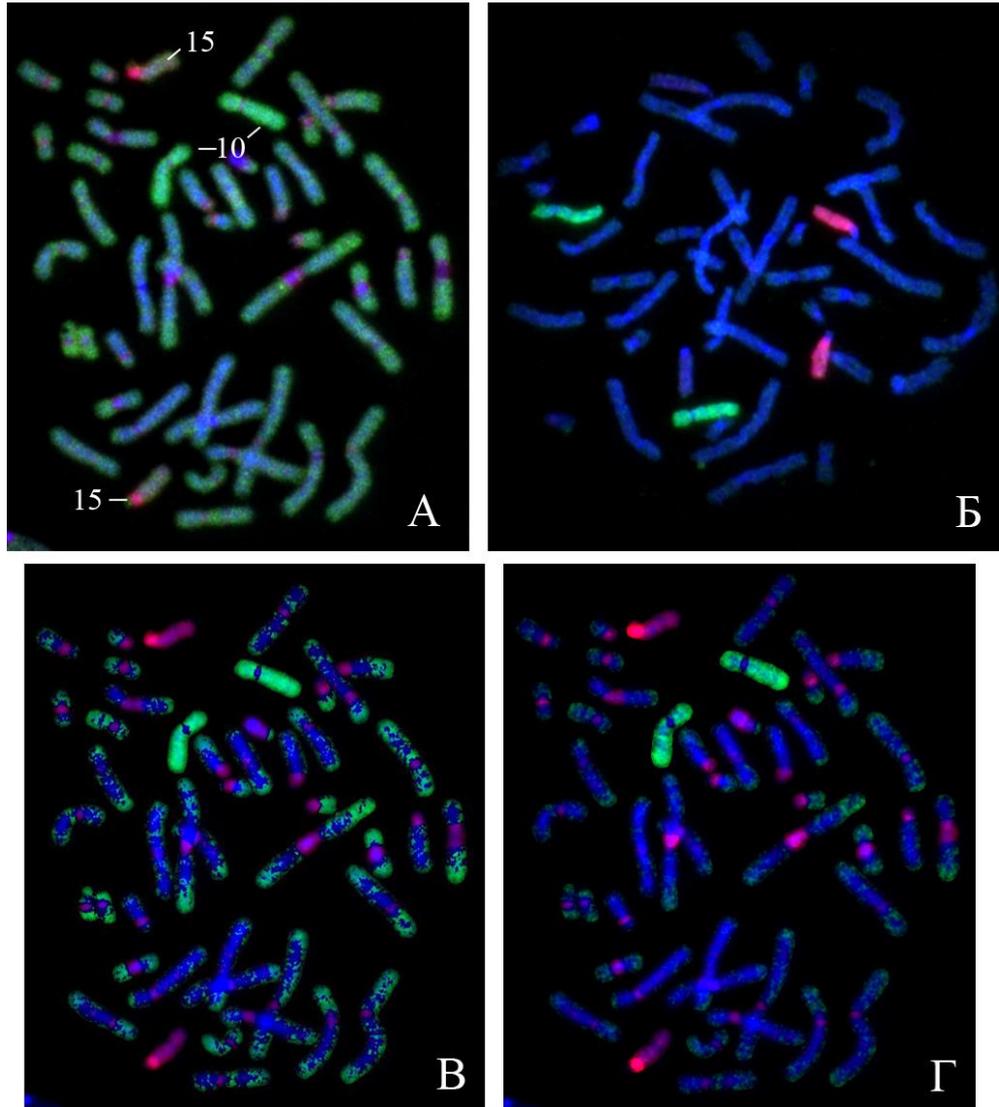


Рисунок 3.11. Сравнение методов выделения специфичного сигнала на результатах FISH ДНК-проб из 10 (зеленый) и 15 (красный) хромосом человека. **А** – FISH без супрессии повторяющихся последовательностей. **Б** – CISS гибридизация. **В** – Результат компьютерной обработки изображения А с помощью метода, предложенного Ренсом и его коллегами (Rens et al., 2006). **Г** – Результат компьютерной обработки изображения А с помощью разработанного метода VISSIS.

Важным отличием метода VISSIS от метода RENS заключается в том, что он не зависит от субъективного мнения исследователя (Богомолов и др., 2014). Также

при использовании метода VISSIS не требуется вручную устанавливать настройки для определения сигналов, обусловленных гибридизацией диспергированных повторяющихся последовательностей. Напротив, компьютерная обработка методом RENS основана на задании экспертом для каждого изображения интервала значений соотношений интенсивностей сигналов ДНК-проб, ассоциируемого с повторяющимися последовательностями (раздел 1.4.4).

Методы RENS и VISSIS могут быть использованы для улучшения результатов CISS-гибридизации. Цифровая обработка снижает требуемый объем Cot-1 ДНК для успешного проведения хромосомного пэйнтинга, и тем самым снижает стоимость проведения CISS-гибридизации. Но наиболее перспективным представляется использование компьютерных методов в ходе молекулярно-цитогенетических исследований хромосом у видов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть проведена.

3.4 Применение метода VISSIS для идентификации хромосом у описторхид

Размер геномов описторхид составляет 300-400 миллионов пар оснований, что почти в 10 раз меньше генома человека. Несмотря на небольшие размеры генома описторхид, диспергированные повторяющиеся последовательности представлены в их хромосомах в достаточном количестве для того, чтобы при проведении FISH ДНК-проб, созданных из целых хромосом или протяженных хромосомных районов, появлялся сигнал на всех хромосомах (Zadesenets et al., 2012). Применение CISS-гибридизации для хромосомного анализа описторхид затруднено ввиду отсутствия достаточного количества исходного материала для получения фракции высокоповторяющейся ДНК.

В рамках исследования выполнена апробация созданного метода на изображениях хромосом *M. xanthosomus* и *O. felineus*, полученных после проведения FISH хромосомоспецифичных ДНК-проб (Богомоллов и др., 2012). Как видно из рисунков 3.12 и 3.13, компьютерная обработка позволила успешно идентифицировать материалы соответствующих хромосом.

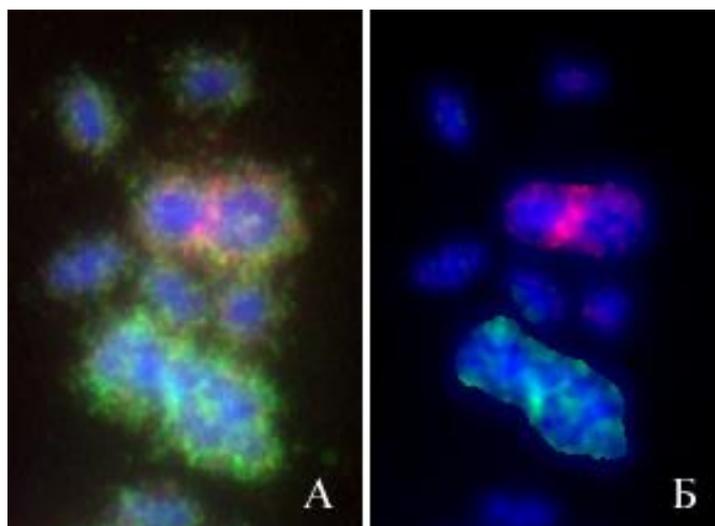


Рисунок 3.12. Пахитенные хромосомы. FISH ДНК-проб из хромосом 1 (зеленый) и 2 (красный) вида *O. felineus*. **А** – Необработанное исходное изображение. **Б** – Результат компьютерной обработки с помощью метода VISSIS.

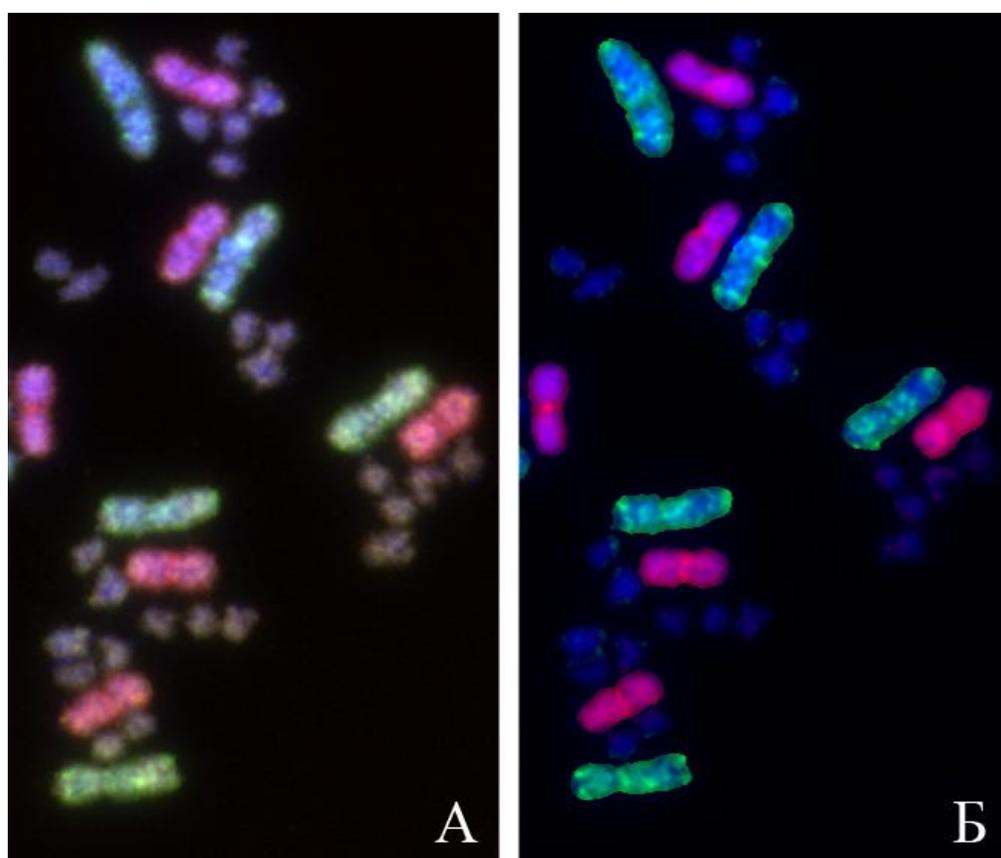


Рисунок 3.13. Пахитенные хромосомы. FISH ДНК-проб из хромосомы 1 (зеленый) и хромосомы 2 (красный) вида *M. xanthosomus*. **А** – Необработанное исходное изображение. **Б** – Результат компьютерной обработки с помощью метода VISSIS.

Было обработано 62 изображения хромосом *M. xanthosomus* и 9 изображений хромосом *O. felineus*. Результаты классификации объектов и эффективность метода VISSIS приведены в таблицах 3.5 и 3.6.

Таблица 3.5 – Результаты классификации объектов на изображениях FISH ДНК-проб из хромосом *M. xanthosomus* и *O. felineus*

	Ошибка первого рода (E_1)	Ошибка второго рода (E_2)	Количество объектов класса <i>Ca</i> и <i>Cb</i> на изображениях	Общее количество хромосом на изображениях
FISH ДНК-проб из хромосом <i>M. xanthosomus</i> 1 и 2	0	0.0378	430	1384
FISH ДНК-проб из хромосом <i>O. felineus</i> 1 и 2	0	0.14	19	76
Примечание. К классам <i>Ca</i> и <i>Cb</i> относятся объекты, содержащие ДНК, специфичную хромосомам, из которых получены ДНК-пробы.				

Таблица 3.6 – Среднее значение и стандартное отклонение отношения сигнал/шум (SNR) после компьютерной обработки FISH ДНК-проб из хромосом *M. xanthosomus* и *O. felineus*.

Название вида	SNR	
	Для результатов FISH ДНК-пробы из хромосомы 1	Для результатов FISH ДНК-пробы из хромосомы 2
<i>M. xanthosomus</i>	18.8±7.3	41.43±15.3
<i>O. felineus</i>	16.8±4	12.97±4.3

Большая доля ошибок классификации объектов на изображениях мейотических хромосом *O. felineus* связана с более интенсивным сигналом ДНК-пробы, полученной из хромосомы 2, в центромерных и прицентромерных районах хромосомы 2. В остальных районах хромосомы сигнал слабее и в некоторых случаях может быть сопоставим по интенсивности с сигналом на других хромосомах. Это приводит к расположению объектов, содержащих ДНК, специфичную хромосоме 2, в пространстве признаков недостаточно далеко от объектов класса *Cns* и, как следствие, к ошибкам классификации. На некоторых обработанных изображениях наблюдается неполное покрытие сигналом всей хромосомы 2, тем не менее она легко идентифицируется после применения метода VISSIS.

Из таблицы 3.5 можно видеть, что среднее значение SNR колеблется от 13 до 41. Такие значения SNR сопоставимы с результатами CISS-гибридизации на хромосомах человека (оценка эффективности CISS-гибридизация приведена в таблице 3.3), а чаще всего превосходят их. Таким образом, полученные результаты подтверждают возможность использования метода в качестве альтернативы CISS-гибридизации в случаях, когда невозможно получить Cot-1 ДНК.

Высокую эффективность метода можно объяснить небольшой долей повторяющихся последовательностей в геноме описторхид. При проведении FISH хромосомоспецифичных ДНК-проб это приводит к более интенсивному сигналу в соответствующих хромосомах по сравнению с сигналами в других районах хромосом. Ранее отмечалось, что результаты компьютерной обработки зависят от типов повторяющихся последовательностей, представленных в ДНК-пробах (см. раздел 3.2). Поэтому большие значения отношения сигнал/шум могут указывать на близость хромосом, из которых получены ДНК-пробы, по составу повторяющихся последовательностей.

3.5 Применение метода VISSIS для идентификации хромосом и крупных хромосомных районов у саранчовых

Саранчовые являются надсемейством с самыми большими размерами геномов среди насекомых. Размер геномов саранчовых варьирует от 6000 до 20000 миллионов пар нуклеотидов (Bensasson et al., 2001). Хромосомная организация крупных геномов, в том числе у саранчовых, остается слабо изученной. В геномах таких видов количество повторяющихся последовательностей слишком велико, что не позволяет провести полную супрессию их гибридизации *in situ*.

С помощью метода VISSIS впервые удалось установить гомеологию крупных С-негативных районов в половых хромосомах у 19 видов саранчовых семейства *Pamphagidae* (Рисунок 3.14) (Jetybayev et al., 2017). FISH-изображения были получены Ильясом Еркиновичем Джетыбаевым в ходе изучения молекулярной эволюции нео-Y-хромосомы при формировании нео-XX/нео-XY гетерогаметного определения пола из исходного для саранчовых – XX/X0. У семейства *Pamphagidae* нео-Y-хромосома, исходно гомологичная аутосоме, вступившей в транслокацию с X-хромосомой, становится гоносомой, резко отличающейся от своего гомолога размерами и присутствием значительных гетерохроматиновых районов.

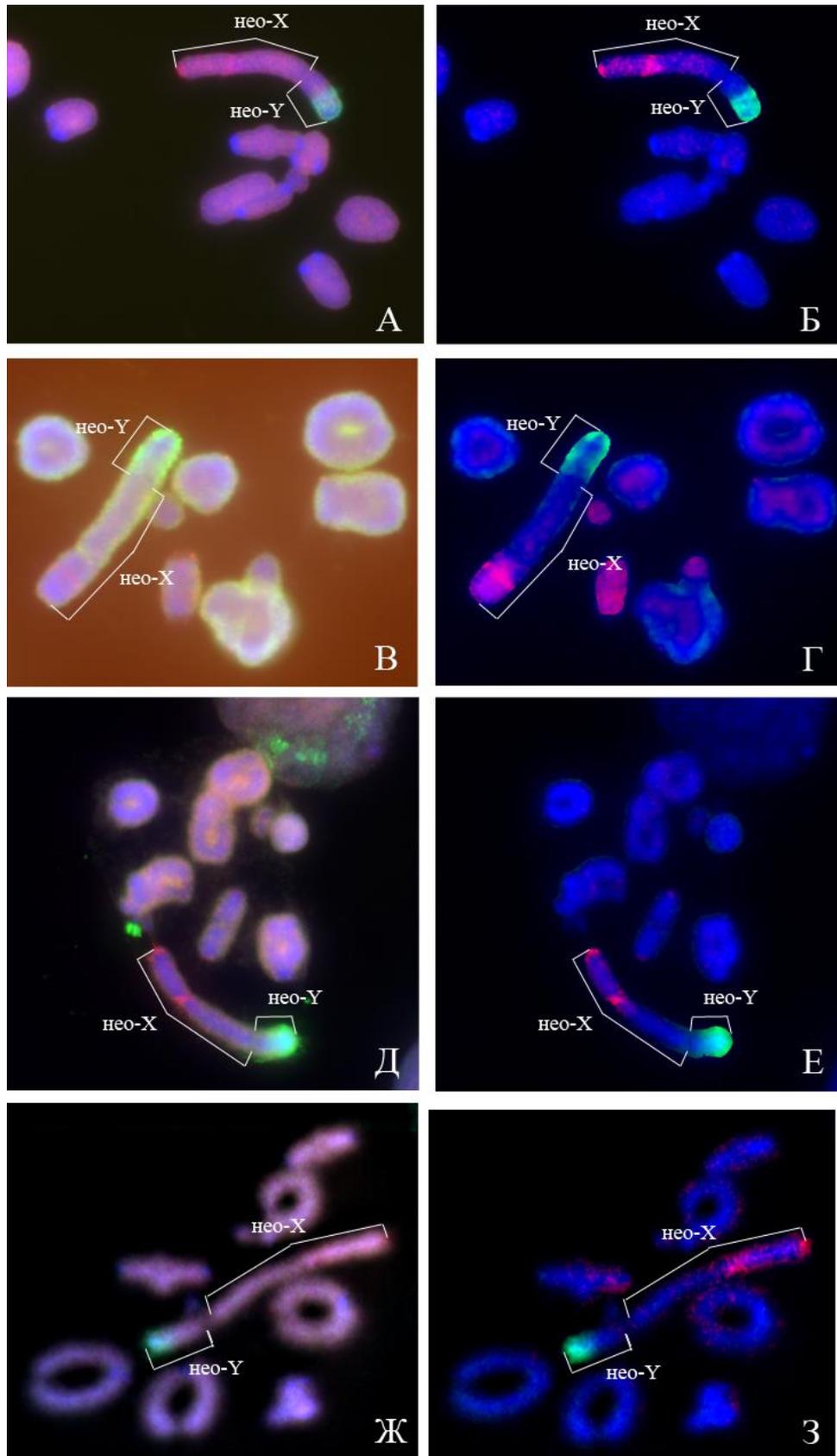


Рисунок 3.14. FISH микродиссекционных ДНК-проб из короткого плеча нео-X хромосомы (красный) и нео-Y хромосомы (зеленый) с хромосомами исходных видов семейства *Pamphagidae*. А, Б – *Nocaracris rubripe*. В, Г – *Paranocarodes tolunayi tolunayi*. Д, Е – *Nocaracris tardus*. Ж, З – *Nocaracris cyanipes*. А, В, Д, Ж – Необработанное исходное изображение. Б, Г, Е, З – Результат компьютерной обработки с помощью метода VISSIS.

Анализ результатов FISH ДНК-проб, полученных из нео-половых хромосом видов семейства *Pamphagidae* (без использования Cot-1 ДНК и компьютерной обработки), позволил лишь охарактеризовать преимущественно кластеризованные повторяющиеся последовательности, локализуемые в районах гетерохроматина. Как видно из рисунка 3.14, сигнал диспергированных повторяющихся последовательностей достаточно интенсивный, что не позволяет распознать районы локализации хромосомспецифичных уникальных последовательностей. Для эффективного распознавания хромосом был применен метод VISSIS, который успешно справился с задачей выявления сигнала хромосомспецифичных последовательностей.

Оценка эффективности метода и результаты классификации объектов на примере обработки изображений хромосом *Nocaracris cyanipes*, *Nocaracris rubripes*, *Paranocarodes tolunayi tolunayi* и *Nocaracris tardus* приведены в таблицах 3.7 и 3.8.

Таблица 3.7 – Результаты классификации объектов на изображениях FISH ДНК-проб из хромосом *Nocaracris cyanipes*, *Nocaracris rubripes*, *Paranocarodes tolunayi tolunayi* и *Nocaracris tardus*

	Ошибка первого рода (E₁)	Ошибка второго рода (E₂)	Количество объектов класса <i>Ca</i> и <i>Cb</i> на изображениях	Общее количество объектов на изображениях
Применение метода на изображениях хромосом <i>Nocaracris rubripes</i>	0	0.1914	18	113
Применение метода на изображениях хромосом <i>Paranocarodes tolunayi tolunayi</i>	0	0.2308	14	79
Применение метода на изображениях хромосом <i>Nocaracris tardus</i>	0	0.136	28	175
Применение метода на изображениях хромосом <i>Nocaracris cyanipes</i>	0	0.1674	50	259
Примечание – К классам <i>Ca</i> и <i>Cb</i> относятся объекты, содержащие ДНК, специфичную хромосомам, из которых получены ДНК-пробы. Плечи нео-Х хромосомы рассматривались как два отдельных объекта.				

Таблица 3.8 – Среднее значение и стандартное отклонение отношения сигнал/шум (SNR) после компьютерной обработки FISH ДНК-проб из хромосом *Nocaracris cyanipes*, *Nocaracris rubripes*, *Paranocarodes tolunayi* и *Nocaracris tardus*

Название вида	Район хромосомы <i>a</i> , из которого получена ДНК-проба	Район хромосомы <i>b</i> , из которого получена ДНК-проба	SNR		Количество проанализированных изображений
			Для результатов FISH ДНК-пробы из хромосомы <i>a</i>	Для результатов FISH ДНК-пробы из хромосомы <i>b</i>	
<i>Nocaracris rubripes</i>	нео-У хромосома	Короткое плечо нео-Х хромосомы	22±7.3	6.34±2.67	6
<i>Paranocarodes tolunayi</i>	нео-У хромосома	Короткое плечо нео-Х хромосомы	12.52±1.52	4.2±0.48	7
<i>Nocaracris tardus</i>	нео-У хромосома	Короткое плечо нео-Х хромосомы	43.57±12.52	13.56±4.8	11
<i>Nocaracris cyanipes</i>	нео-У хромосома	Короткое плечо нео-Х хромосомы	19±7.61	5.4±1.78	25

Большие значения SNR для результатов FISH ДНК-проб из нео-У хромосомы объясняются сохранением интенсивного сигнала в крупном С-позитивном районе после цифровой обработки изображений. Иначе обстоит дело с компьютерной обработкой FISH ДНК-проб из короткого плеча нео-Х. Сигнал выявляется не только на коротком плече нео-Х хромосомы, но и на длинном плече, а в некоторых случаях и на аутосомах. По этой причине значения SNR не очень большие. Ошибки классификации также связаны с наличием интенсивных сигналов на других хромосомах. Тем не менее, цифровая обработка позволяет идентифицировать материал соответствующих хромосом, а эффективность метода может быть сопоставима с результатами CISS-гибридизации на хромосомах человека.

В ходе исследования эволюции У-хромосомы метод VISSIS также применили на кроссгибридизационных изображениях (FISH полученных ДНК-проб неополовых хромосом одного вида на хромосомы других исследуемых видов). Кроссгибридизация позволила проследить эволюцию нео-половых хромосом у 19 видов саранчовых семейства *Pamphagidae*: установить гомологию или ее отсутствие

крупных C-негативных районов в половых хромосомах нео-X и нео-Y. 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. Основными механизмом эволюционных изменений является возникновение новых интеркалярных C-позитивных блоков и потеря C-негативных районов, происходящих из предковой аутосомы. Более подробную информацию о результатах исследования эволюции половых хромосом можно найти в статье (Jetybayev et al., 2017).

3.6 Возможности применения метода и его ограничения

В настоящей главе продемонстрированы примеры использования метода визуализации хромосомоспецифичных последовательностей при проведении FISH на хромосомах человека, описторхид (*Metorchis xanthosomus*, *Opisthorchis felinus*) и саранчовых (*Nocaracris cyanipes*, *Nocaracris rubripes*, *Paranocarodes tolunayi*, *Nocaracris tardus*). Перечисленные виды существенно различаются по размеру генома и по доле повторяющихся последовательностей в нем. Метод позволил успешно идентифицировать материал целых хромосом и отдельных хромосомных районов, в том числе у саранчовых, для которых ранее эту задачу решить не удавалось.

На примере ДНК-проб из хромосом человека было показано, что результаты компьютерной обработки зависят от состава повторяющихся последовательностей хромосом, из которых были получены ДНК-пробы. Чем сильнее различия хромосом по соотношению количества SINE/LINE последовательностей, тем больше ошибок классификации и труднее идентифицировать хромосомоспецифичные районы из-за наличия интенсивного неспецифичного сигнала. Однако, эту особенность метода можно использовать для анализа закономерностей распределения в хромосомных районах различных типов диспергированных повторяющихся последовательностей у слабо изученных видов млекопитающих. Такая возможность была продемонстрирована на примере анализа FISH-изображений с ДНК-пробами из 18 и 19 хромосом человека, контрастных по содержанию LINE- и SINE-повторов.

ЗАКЛЮЧЕНИЕ

Диссертация посвящена созданию и апробации метода анализа микроскопических изображений результатов FISH хромосомо- и районспецифичных ДНК-проб для выявления хромосомспецифичных последовательностей ДНК. Основой для разработки этого метода являлось предположение, что различные хромосомы из одного генома содержат диспергированные повторы, общие для разных хромосом. Использование двух различных хромосомо- или районспецифичных ДНК-проб для FISH позволяет выделить специфичный сигнал на анализируемых хромосомах после элиминации сигнала повторяющихся диспергированных последовательностей ДНК посредством компьютерной обработки изображений. В ходе настоящего исследования был разработан метод VISSIS (visualization of specific signal *in silico*) и реализован в виде программного продукта VisualCS (Visualization Chromosome-specific Signals).

Актуальность работы обусловлена тем, что в настоящее время задача идентификации хромосом и хромосомных районов с помощью FISH ДНК-проб, полученных из этих хромосом, остается не до конца решенной. Стоит отметить, что в значительной степени она решена для идентификации хромосом человека при проведении диагностики хромосомных патологий. CISS-гибридизация хромосомо- и районспецифичных ДНК-проб дает удовлетворительные результаты, исключая из гибридации большую часть повторяющихся последовательностей ДНК. При ее использовании, как правило, также отсутствует сигнал в С-позитивном районе анализируемой хромосомы. Таким образом, С-позитивные районы часто полностью исключаются из анализа. По умолчанию считается, что неокрашиваемые районы хромосомы содержат только повторяющиеся последовательности ДНК. В большинстве случаев это соответствует действительности. Однако, теоретически нельзя исключить инсерцию в С-позитивный район небольшого района из другой хромосомы. Во многих случаях эта проблема не может быть решена даже при проведении 24-х цветной FISH. При переводе в псевдочвета комбинации флуорохромов в С-позитивных районах может возникать артефактное окрашивание небольших участков хромосом. При использовании метода VISSIS С-позитивные районы “окрашиваются” за счет повторяющихся последовательностей, специфичных анализируемой хромосоме

Также отметим, что при проведении CISS-гибридизации с недостаточной супрессией повторяющихся последовательностей, дополнительное применение метода VISSIS значительно улучшает результаты анализа. Такая комбинация методов может использоваться для уменьшения используемого объема Cot-1 ДНК в ходе проведения FISH-диагностики и тем самым снизить затраты на проведения исследований.

Апробация метода на изображениях хромосом человека, полученных после проведения FISH хромосомоспецифичных ДНК-проб, выявило ряд особенностей его применения. Некоторые хромосомы человека значительно различаются по составу диспергированных повторов. Так, например, хромосома 19 состоит преимущественно из G-негативных районов, обогащенных SINE, в то время как хромосома 18, - преимущественно из G-позитивных районов, обогащенных LINE. Поскольку G-негативные и G-позитивные районы обогащены разными типами повторяющихся последовательностей, то использование метода VISSIS для анализа результатов FISH такой пары хромосом приводит к различному окрашиванию районов, отличающихся по относительному содержанию разных типов диспергированных повторяющихся последовательностей (сохранению сигнала после обработки). С одной стороны, ограничивает возможность применения метода VISSIS для анализа результатов FISH, полученных при использовании только определенных пар хромосом. С другой стороны, метод может быть использован для выявления хромосом и хромосомных районов, характеризующихся обогащением разными типами диспергированных повторяющихся последовательностей ДНК, у неизученных ранее видов.

Проведенный анализ относительного присутствия SINE/LINE последовательностей в хромосомах человека позволил определить пары хромосом, ДНК-пробы которых дают при использовании FISH дают максимальную эффективность компьютерной обработки методом VISSIS полученных микроскопических изображений. Такие пары хромосомоспецифичных ДНК-проб приведены в настоящей работе.

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. Вероятно, хромосомы этих видов обеднены диспергированными повторяющимися последовательностями. Поэтому такие результаты компьютерной обработки следовало ожидать.

Более впечатляющими были результаты использования метода VISSIS для идентификации гомологичных районов хромосом у саранчовых, геном которых по размеру в несколько раз превышает геном человека, а высокое обогащение С-негативных хромосомных районов диспергированными повторяющимися последовательностями делали невозможным их специфическое окрашивание FISH ДНК-пробами, полученными из индивидуальных хромосом, даже при проведении CISS-гибридизации. Идентификация гомологичных районов хромосом благодаря использованию метода VISSIS была выполнена у 19 видов саранчовых семейства *Pamphagidae*.

Таким образом, цифровая обработка изображений FISH ДНК-проб, полученных из хромосом, имеющих сходный состав повторяющихся последовательностей ДНК, проведенная с помощью метода VISSIS, значительно снизила сигнал с диспергированных повторов и тем самым позволило выделить специфичный сигнал. Отношение сигнал/шум на таких изображениях сопоставимо с результатами CISS-гибридизации, а в большинстве случаев превосходит их. Метод VISSIS является хорошей альтернативой CISS-гибридизации и/или инструментом для улучшения результатов CISS-гибридизации. И что более важно, метод VISSIS позволяет визуализировать хромосомоспецифичные последовательности у организмов, для которых CISS-гибридизация не дает удовлетворительных результатов, или в принципе не может быть выполнена по объективным причинам.

ВЫВОДЫ

1. Разработаны компьютерный метод VISSIS (visualization 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 литературы

1. Богомолов А.Г., Задесенец К.С., Карамышева Т.В., Подколотный 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
2. 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 – Т. 48 – №6. – С. 881–890
3. Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Осоков М.В., Моржин А.В. Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий. – М.: Физматкнига, 2010. – 672 с
4. Гайнер Т.А., Карамышева Т.В., Каримова О.Г., Корень О.Л., Шлома В.В., Шорина А.Р., Богомолов А.Г., Рубцов Н.Б. Комплексная диагностика хромосомной патологии - деривата хромосомы 4 и малой сверхчисленной маркерной хромосомы // Медицинская генетика. – 2017 – Т. 16 – №. 12 – С. 9-17
5. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
6. Евгеньев М.В. Мобильные элементы и эволюция генома // Молекулярная биология. – 2007. – Т. 41. – №. 2. – С. 234-245.
7. Рубцов Н.Б., Карамышева Т.В. Многоцветие современной цитогенетики, или multicolor FISH today // Вавиловский журнал генетики и селекции. – 1999. – Т. 3 – № 11. –С. 11-15.
8. Рубцов Н.Б., Карамышева Т.В., Гайнер Т.А., Шкляева О.А. Анализ маркерных хромосом: ДНК пробы для оценки возможного клинического значения маркерной хромосомы. // AbidМедицинская генетика. – 2003. – Т. 2 – № 12. – Р. 520-527.
9. Сергеева Е. М., Салина Е. А. Мобильные элементы и эволюция генома растений // Вавиловский журнал генетики и селекции. – 2011. – Т. 15. – №. 2. – С. 382-397.
10. Шапиро Л., Стокман Дж. Компьютерное зрение. – М.: БИНОМ, 2006. – 752 с.

11. Abbo S., Miller T.E., Reader S.M., Dunford, R.P., King, I.P. Detection of ribosomal DNA sites in lentil and chickpea by fluorescent in situ hybridization // *Genome*. 1994. V. 37 (4). P. 713-716.
12. Abid F., Hamami L. A survey of neural network based automated systems for human chromosome classification // *Artificial Intelligence Review*. 2016. V 49(1). P. 41-56.
13. Agam G., Dinstein I. Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997. V. 19(11). P. 1212-1222.
14. Arachchige A.S. Human metaphase chromosome analysis using image processing: Doctor of Philosophy Dissertation. London. 2014. – 138 p.
15. Arachchige A.S., Samarabandu J., Knoll J., Khan W., Rogan P. An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes // In: *Image Processing (ICIP), 17th IEEE International Conference*. 2010. P. 3613-3616.
16. Arachchige A.S., Samarabandu J., Knoll J.H., Rogan P.K. Intensity integrated laplacian-based thickness measurement for detecting human metaphase chromosome centromere location // *IEEE Transactions on Biomedical Engineering*. 2013. V. 60(7). P. 2005-2013.
17. Arora T., Dhir R. A review of metaphase chromosome image selection techniques for automatic karyotype generation // *Medical & biological engineering & computing*. 2016. V. 54(8). P. 1147-1157.
18. Badawi A. M., Hasan K.G., Aly E.E., Messiha R.A. Chromosomes classification based on neural networks, fuzzy rule based, and template matching classifiers // In: *Circuits and Systems, IEEE 46th Midwest Symposium*. 2003. V. 1. P. 383-387.
19. Bakker B., van den Bos H., Lansdorp P.M., Foijer F. How to count chromosomes in a cell: An overview of current and novel technologies // *BioEssays*. 2015. V. 37(5). P. 570-577.
20. Bauman J.G.J., Wiegant J., Borst P., van Duijn P. A new method for fluorescence microscopical localization of specific DNA sequences by *in situ* hybridization of fluorochrome-labelled RNA // *Experimental cell research*. 1980. V. 128(2). P. 485-490.

21. Beliveau B.J., Joyce E.F., Apostolopoulos N., Yilmaz F., Fonseka C.Y., McCole R.B., Chang Y., Li J.B., Senaratne T.N., Williams B.R., Rouillard J., Wu C. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes // *Proceedings of the National Academy of Sciences*. 2012. V. 109 (52). P. 21301-21306.
22. Bensasson D., Petrov D.A., De-Xing Zhang Hartl D.L., Hewitt G.M. Genomic gigantism: DNA loss is slow in mountain grasshoppers // *Molecular Biology and Evolution*. 2001. V. 18(2). P. 246-253.
23. Berensmeier S. Magnetic particles for the separation and purification of nucleic acids // *Applied microbiology and biotechnology*. 2006. V. 73(3). P. 495-504.
24. Bolzer A. Craig J.M., Cremer T. and Speicher M.R. A complete set of repeat-depleted, PCR-amplifiable, human chromosome-specific painting probes // *Cytogenetic and Genome Research*. 1999. V. 84(3-4). P. 233-240.
25. Boyle S, Rodesch M.J., Halvensleben H.A., Jeddloh J.A., Bickmore W.A. Fluorescence *in situ* hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis // *Chromosome Research*. 2011 V. 19(7). P. 901-909
26. Britten R.J., Kohne D.E. Repeated Sequences in DNA // *Science*. 1968. V. 161. P. 529-540.
27. Britten R.J., Graham D.E., Neufeld B.R. Analysis of repeating DNA sequences by reassociation // *Methods in enzymology*. 1974. V. 29. P. 363-419.
28. Bugrov A.G., Karamysheva T.V., Pyatkova M.S., Rubtsov D.N., Andrenkova O.V. Comparative FISH analysis of distribution of B chromosome repetitive DNA in A and B chromosomes in two subspecies of *Podisma sapporensis* (Orthoptera, Acrididae). // *Cytogenet. Genome Res*. 2004. V. 106. P. 284-288.
29. Calvard S., Ridler T.W. Picture thresholding using an iterative selection method // *IEEE Transactions on Systems, Man and Cybernetics*. 1978. V. 8(8). P. 630-632.
30. Charters G. C., Graham J. Trainable grey-level models for disentangling overlapping chromosomes // *Pattern Recognition*. 1999. V. 32(8). P. 1335-1349.
31. Cheng S., Fockler C., Barnes W. Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA // *Proceedings of the National Academy of Sciences*. 1994. V. 91(12). P. 5695-5699.

32. Chen-Liu L.W., Huang B.C., Scalzi J.M., Hall B.K., Sims K.R., Davis L.M., Siebert P.D., Hozier J.C. Selection of hybrids by affinity capture (SHAC): a method for the generation of cDNAs enriched in sequences from a specific chromosome region // *Genomics*. 1995. V. 30(2). P. 388-392.
33. Choi H., Castleman K. R., Bovik A. C. Joint segmentation and classification of M-FISH chromosome images // In: *Engineering in Medicine and Biology Society, IEMBS'04. 26th Annual International Conference of the IEEE*. 2004. V. 1. P. 1636-1639.
34. Choi H., Castleman K. R., Bovik A. C. Segmentation and fuzzy-logic classification of M-FISH chromosome images // In: *Image Processing, 2006 IEEE International Conference*. 2006a. P. 69-72.
35. Choi H., Bovik A.C., Castleman K.R. Maximum-likelihood decomposition of overlapping and touching M-FISH chromosomes using geometry, size and color information // In: *Engineering in Medicine and Biology Society, 28th Annual International Conference of the IEEE*. 2006b. P. 3130-3133.
36. Chudoba I., Plesch A., Lörch T., Lemke J., Claussen U., Senger G. High resolution multicolor-banding: a new technique for refined FISH analysis of human chromosomes // *Cytogenetic and Genome Research*. 1999. V. 84(3-4). P. 156-160.
37. Chung I.B., Kramera J.A., Johnsona M.P., Evansa M.I, Krawetz S.A. The use of cloned repetitive sequences as hybridization competitors to detect single copy sequences // *Genetic Analysis: Biomolecular Engineering*. 1997. V. 14(1). P. 13-15.
38. Cohen A.R. Extracting meaning from biological imaging data // *Molecular biology of the cell*. 2014. V. 25(22). P. 3470-3473.
39. Craig J.M., Kraus J., Cremer T. Removal of repetitive sequences from FISH probes using PCR-assisted affinity chromatography // *Human genetics*. 1997. V. 100(3). P. 472-476.
40. Cremer T, Lichter P, Borden J, Ward DC and Manuelidis L. Detection of chromosome aberrations in metaphase and interphase tumor cells by *in situ* hybridization using chromosome-specific library probes // *Human genetics*. 1988. V. 80(3). P. 235-246.
41. de Faria E. R., Guliato D., de Sousa Santos J. C. Segmentation and centromere locating methods applied to FISH chromosomes images // *Lecture Notes in Computer Science*. 2005. V. 3594. P. 181-189.

42. de Koning A.P., Gu W., Castoe T.A., Batzer M.A., Pollock D.D. Repetitive elements may comprise over two-thirds of the human genome // PLoS genetics. 2011. V. 7 (12). P e1002384.
43. Deleye L., Tilleman L., Vander Plaetsen A.S., Cornelis S., Deforce D., Van Nieuwerburgh F. Performance of four modern whole genome amplification methods for copy number variant detection in single cells // Scientific Reports. 2017. V. 7(1). P. 3422.
44. Devi J., Ko J.M., Seo B.B. FISH and GISH: Modern cytogenetic techniques // Indian Journal of Biotechnology. 2005. V. 4(3). P. 307-315.
45. Dorman S.N., Shirley B.C., Knoll J.H.M., Rogan P. K. Expanding probe repertoire and improving reproducibility in human genomic hybridization // Nucleic acids research. 2013. V. 41(7). P. 1-13.
46. Dugan L.C., Pattee M.S., Williams J.M., Sorensen K., Bedford J.S., Christian A.T. Polymerase chain reaction-based suppression of repetitive sequences in whole chromosome painting probes for FISH // Chromosome Research. 2005. V. 13(1). P. 27-32.
47. Durmaz A.A., Karaca E., Demkow U., Toruner G., Schoumans J., Cogulu O. Evolution of Genetic Techniques: Past, Present, and Beyond // BioMed research international. 2015. V. 2015. Article ID 461524.
48. Fluorescence In Situ Hybridization (FISH). Application Guide / Liehr T. – Jena: Springer, 2017. - P. 588.
49. Fuchs J., Houben A., Brandes A., Schubert I. Chromosome ‘painting’ in plants – a feasible technique? // Chromosoma. 1996. V. 104(5). P. 315-320.
50. Garini Y. Macville M., du Manoir S., Buckwald R.A., Lavi M., Katzir N., Wine D., Bar-Am I., Schrock E., Cabib D., Ried T. Spectral karyotyping // Bioimaging. 1996. V. 4(2). P. 65-72.
51. Geurts R., de Jong H. Chapter 2 Fluorescent *In Situ* Hybridization (FISH) on pachytene chromosomes as a tool for genome characterization // Legume Genomics. Methods and Protocols / R.J. Rose – Callaghan: Humana Press, 2013. – P. 15-24.
52. Grisan E., Poletti E., Ruggeri A. Automatic segmentation and disentangling of chromosomes in Q-band prometaphase images // IEEE Transactions on Information Technology in Biomedicine. 2009. V. 13(4). P. 575-581.

53. Groen F.C.A., Ton K., Smeulders A.W., Young I.T. Human chromosome classification based on local band descriptors // *Pattern Recognition Letters*. 1989. V. 9(3). P. 211-222.
54. Grover D., Majumder P.P., Rao C., Brahmashari S.K., Mukerji M. Nonrandom distribution of alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22 // *Molecular biology and evolution*. 2003. V. 20 (9). P. 1420-1424.
55. Halder A., Halder S., Fauzdar A., Kumar A. Molecular approaches of chromosome analysis: an overview // *Proc. Indian Nat. Sci. Acad.* 2004. V. 70(2). P. 153-221.
56. Han Y. Zhang T., Thammaphichai P., Weng Y., Jiang J. Chromosome-Specific Painting in Cucumis Species Using Bulked Oligonucleotides // *Genetics*. 2015. V. 200(3). P. 771-779.
57. Henegariu O., Heerema N., Lowe Wright L., Bray-Ward P., Ward D. C., Vance G. H. Improvements in cytogenetic slide preparation: controlled chromosome spreading, chemical aging and gradual denaturing // *Cytometry*. 2001. V. 43(2). P. 101-109.
58. Hliscs R., Mühlig P., Claussen U. The spreading of metaphases is a slow process which leads to a stretching of chromosomes // *Cytogenetic and Genome Research*. 1997. V. 76(3-4). P. 167-171.
59. Houben A., Field B.L., Saunders V.A. Microdissection and chromosome painting of plant B chromosomes // *Methods in cell science*. 2001. V. 23(1-3). P. 115-124.
60. Hozier J., Graham R., Westfall T., Siebert P., Davis L. Preparative *in situ* hybridization: Selection of chromosome region-specific libraries on mitotic chromosomes. // *Genomics*. 1994. V. 19(3). P. 441-447.
61. Hozier J.C., Scalzi J.M., Clase A.C. Davis L.M., Liechty M.C. Differential destabilization of repetitive sequence hybrids in fluorescence *in situ* hybridization // *Cytogenetic and Genome Research*. 1998. V. 83(1-2). P. 60-63.
62. Ijdo J.W., Wells R.A., Baldini A., Reeders S.T. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR // *Nucleic acids research*. 1991. V. 19(17). P. 4780.

63. Irwin J.O. On a criterion for the rejection of outlying observations // *Biometrika*. 1925. V. 17(3-4). P. 238-250.
64. ISCN 2013: an international system for human cytogenetic nomenclature / Shaffer L.G., McGowan-Jordan J., Schmid M. – Karger Medical and Scientific Publishers, 2013. - P. 140.
65. Jahani S., Setarehdan S.K. A novel method for centromere and length detection in microscopic images of human chromosomes // In: *Biomedical Engineering (ICBME)*, 18th Iranian Conference. 2011. P. 274-277.
66. Jahani S., Setarehdan S.K. Centromere and length detection in artificially straightened highly curved human chromosomes // *International Journal of Biological Engineering*. 2012. V. 2(5). P. 56-61.
67. Jahani S., Setarehdan S.K., Fatemizadeh E. Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators // In: *Machine Vision and Image Processing (MVIP)*, 7th Iranian. 2011. P. 1-4.
68. Jang H.Y., Kim H.R., Kang M.S., Kim M.H., Zhang B.T. The demand for quantitative techniques in biomedical image informatics // *Biomedical Engineering Letters*. 2014. V. 4(4). P. 319-327.
69. Jensen E. Technical review: *In situ* hybridization // *The Anatomical Record*. 2014. V. 297(8). P. 1349-1353.
70. 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
71. Ji L. Fully automatic chromosome segmentation // *Cytometry Part A*. 1994. V. 17(3). P. 196-208.
72. John H.A., Brinstiel M.L. and Jones K.W. RNA-DNA hybrid at the cytological level// *Nature*. 1969. V. 223. P. 582-587.
73. Kallioniemi A., Kallioniemi O.P., Sudar D., Rutovitz D., Gray J.W., Waldman F., Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors // *Science*. 1992. V. 258 (5083). P. 818-821.

74. Kao J., Chuang J., Wang T. Chromosome classification based on the band profile similarity along approximate medial axis //Pattern Recognition. 2008. V. 41(1) P. 77-89.
75. Kapur J.N., Sahoo P.K., Wong A.K.C. A new method for gray-level picture thresholding using the entropy of the histogram //Computer vision, graphics, and image processing. 1985. V. 29(3). P. 273-285.
76. Karvelis P. S., Fotiadis D.I., Syrrou M., Georgiou I. Segmentation of chromosome images based on a recursive watershed transform //IFMBE Proc. 2005. V. 11(1). P. 1727-1983.
77. Karvelis P. S., Tzallas A.T., Fotiadis D. I., Georgiou I. A multichannel watershed-based segmentation method for multispectral chromosome classification //IEEE transactions on medical imaging. 2008. V. 27(5). P. 697-708.
78. Karvelis P., Likas A., Fotiadis D.I. Identifying touching and overlapping chromosomes using the watershed transform and gradient paths //Pattern Recognition Letters. 2010. V. 31(16). P. 2474-2488.
79. Kazazian H.H. Mobile elements: drivers of genome evolution // Science. 2004. V. 303. P. 1626-1632.
80. Kjeldsen E., Kolvraa S. Chapter 1 FISH Techniques, FISH Probes and Their Applications in Medicine and Biology—An Overview // FISH Technology. – Berlin, Springer Berlin Heidelberg, 2002. – P. 3-50.
81. Knoll J.H.M., Rogan P.K. Sequence-Based, *in situ* detection of chromosomal abnormalities at high resolution // American Journal of Medical Genetics. 2003. V. 121A(3). P. 245-257.
82. Kou Z., Ji L., Zhang X. Karyotyping of comparative genomic hybridization human metaphases by using support vector machines // Cytometry Part A. 2002. V. 47(1). P. 17-23.
83. Landegent J.E., Dirks, R.W., van der Ploeg M. Use of whole cosmid cloned genomic sequences for chromosomal localization by non-radioactive *in situ* hybridization // Human genetics. 1987. V. 77(4). P. 366-370.
84. Langer P.R., Waldrop A.A., Ward D.C. Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes // Proceedings of the National Academy of Sciences. 1981. V. 78(11). P. 6633-6637.

85. Langer S, Kraus J, Jentsch I, Speicher M.R. Multicolor chromosome painting in diagnostic and research applications // *Chromosome Research*. 2004. V. 12(1). P. 15-23.
86. Legrand B., Chang C.S., Ong S.H, Neo S.Y., Palanisamy N. Chromosome classification using dynamic time warping // *Pattern Recognition Letters*. 2008. V. 29(3). P. 215-222.
87. Lerner B. Toward a completely automatic neural-network-based human chromosome analysis // *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 1998. V. 28(4). P. 544-552.
88. Lerner B., Guterman H., Dinstein I., Romem Y. Medial axis transform-based features and a neural network for human chromosome classification // *Pattern Recognition*. 1995. V. 28(11). P. 1673-1683.
89. Levsky J.M., Singer R.H. Fluorescence in situ hybridization: past, present and future // *Journal of cell science*. 2003. V. 116(14). P. 2833-2838.
90. Lichter P., Cremer T., Borden J., Manuelidis L., Ward D.C. Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries // *Human Genetics*. 1988. V. 80(3). P. 224-234.
91. Lichter P., Ledbetter S.A., Ledbetter D.H., Ward D.C. Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines // *Proceedings of the National Academy of Sciences*. 1990. V. 87(17). P. 6634-6638.
92. Liehr T. Classification of FISH Probes // *Fluorescence In Situ Hybridization (FISH) - Application Guide*. / Liehr T. - Jena: Springer Berlin Heidelberg, 2017. - P. 43-47.
93. Liehr T, Starke H., Weise A., Lehrer H., Claussen U. Multicolor FISH probe sets and their applications // *Histology and histopathology*. 2004. V. 19(1). P. 229-237.
94. Liehr T, Glaser A., Kosyakova N. Comparative Genomic Hybridization (CGH) and Microdissection-Based CGH (Micro-CGH) // *Fluorescence In Situ Hybridization (FISH) - Application Guide*. / Liehr T. - Jena: Springer Berlin Heidelberg, 2017. - P. 561-565.

95. Liehr T., Kosyakova N., Weise A. FISH Banding Techniques // Fluorescence In Situ Hybridization (FISH) - Application Guide. / Liehr T. - Jena: Springer Berlin Heidelberg, 2017. - P. 561-565.
96. Ljosa V., Carpenter A.E. Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening // PLoS comput biol. – 2009. V. 5(12). P. e1000603.
97. Ludecke H.J., Senger G., Claussen U., Horsthemke B. Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification // Nature. 1989. V. 338(6213). P. 348-350.
98. Masabanda J.S., Griffin D.K. Generation of chromosome paints: approach for increasing specificity and intensity of signals // Biotechniques. 2003. V. 34. P. 530-536.
99. Markou C, Maramis C, Delopoulos A., Daiou C., Lambropoulos A. Automatic chromosome classification using support vector machines // iConceptPress – Hong Kong. 2012. – P. 1-24.
100. Markova M., Vyskot B. New horizons of genomic in situ hybridization // Cytogenetic and genome research. 2009. V. 126(4). P. 368-375.
101. Minaee S., Fotouhi M., Khalaj B.H. A geometric approach to fully automatic chromosome segmentation // In: Signal Processing in Medicine and Biology Symposium (SPMB), 2014 IEEE. 2014. P. 1-6.
102. Ming D., Tian J. Automatic pattern extraction and classification for chromosome images // Journal of Infrared, Millimeter, and Terahertz Waves. 2010. V. 31(7). P. 866-877.
103. Moallem P., Karimizadeh A., Yazdchi M. Using shape information and dark paths for automatic recognition of touching and overlapping chromosomes in G-band images // International Journal of Image, Graphics and Signal Processing. 2013. V. 5 (5). P. 22.
104. Mohammadi M. R. Accurate localization of chromosome centromere based on concave points // Journal of medical signals and sensors. 2012. V. 2(2). P. 88.
105. Moradi M., Setarehdan S. K. New features for automatic classification of human chromosomes: A feasibility study // Pattern recognition letters. 2006. V. 27(1). P. 19-28.

106. Morozkin E.S., Loseva E.M., Karamysheva T.V., Babenko V.N., Laktionov P.P., Vlassov V.V., Rubtsov N. B. A Method for Generating Selective DNA Probes for the Analysis of C-Negative Regions in Human Chromosomes // *Cytogenetic and Genome Research*. 2011. V. 135(1). P. 1-11.
107. Moter A., Gobel U.B. Fluorescence *in situ* hybridization (FISH) for direct visualization of microorganisms // *Journal of Microbiol Methods*. 2000. V. 41(2). P. 85-112.
108. Muller S., O'Brien P.C., Ferguson-Smith M.A., Wienberg J. Cross-species colour segmenting: a novel tool in human karyotype analysis // *Cytometry*. 1998. V. 33(4). P.445-52
109. Munot M. V., Joshi M. A., Mandhawkar P. Semi automated segmentation of chromosomes in metaphase cells // In: *Image Processing (IPR 2012)*, IET Conference. 2012. P. 1-6.
110. Munot M. V., Mukherjee J., Joshi M. A novel approach for efficient extrication of overlapping chromosomes in automated karyotyping // *Medical & biological engineering & computing*. 2013. V. 51(12). P. 1325-1338.
111. Munoz-Lopez M., Garcia-Perez J. L. DNA transposons: nature and applications in genomics // *Current genomics*. 2010. V. 11(2). P. 115-128.
112. Nair R. M., Remya R. S., Sabeena K. Karyotyping techniques of chromosomes: a survey // *International Journal of Computer Trends and Technology*. 2015. V. 22(1). P.30-34.
113. Navin N., Grubor V., Hicks J., Leibu E., Thomas E., Troge J., Riggs M., Lundin P., Maner S., Sebat J., Zetterberg A., Wigler M. PROBER: oligonucleotide FISH probe design software // *Bioinformatics*. 2006. V. 22(19). P. 2437-2438.
114. Nelson D. L. Interspersed repetitive sequence polymerase chain reaction (IRS PCR) for generation of human DNA fragments from complex sources // *Methods*. 1991. V. 2 (1). P. 60-74.
115. Nelson D.L., Ledbetter S.A., Corbo L., Victoria M.F., Ramírez-Solis R., Webster T.D., Ledbetter D.H., Caskey C.T. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources // *Proceedings of the National Academy of Sciences*. 1989. V. 86(17). P. 6686-6690.

116. Netten H., Young I.T., van Vliet L.J., Tanke H.J., Vrolijk H., Sloos W.C. FISH and chips: automation of fluorescent dot counting in interphase cell nuclei // *Cytometry*. 1997. V. 28(1). P. 1-10.
117. Nietzel A., Rocchi M., Starke H., Heller A., Fiedler W., Wlodarska I., Loncarevic I., Beensen V., Claussen U., Liehr T. A new multicolor-FISH approach for the characterization of marker chromosomes: centromere-specific multicolor-FISH (cenM-FISH) // *Human genetics*. 2001. V. 108(3). P. 199-204.
118. Nisson P.E., Watkins P.C., Menninger J.C., Ward D.C. Improved suppression hybridization with human DNA (COT-1 DNA) enriched for repetitive DNA sequences // *Focus*. 1991. V. 13. P. 42-45.
119. Otsu N. A threshold selection method from gray-level histograms // *IEEE transactions on systems, man, and cybernetics*. 1979. V. 9(1). P. 62-66.
120. Overmyer K., Muller H-W., Gimbel W., Göttert E., Meese E. Enrichment of chromosome specific hncDNAs by magnetic bead coupled Alu sequences // *Molecular biology reports*. 1996. V. 22(1). P. 53-57.
121. Padeken J., Zeller P., Gasser S.M. Repeat DNA in genome organization and stability // *Current opinion in genetics & development*. 2015. V. 31. P. 12-19.
122. Pajor G., Kajtar B., Pajor L., Alpar D. State-of-the-art FISHing: Automated analysis of cytogenetic aberrations in interphase nuclei // *Cytometry Part A*. 2012. V. 81(8). P. 649-663.
123. Pardue M.L., Gall J.G. Molecular hybridization of radioactive DNA to the DNA of cytological preparations // *Proceedings of the National Academy of Sciences*. 1969. V. 64(2). P. 600-604.
124. Peng H., Bateman A., Valencia A., Wren J.D. Bioimage informatics: a new category in Bioinformatics // *Bioinformatics*. 2012. V. 28(8). P. 1057-1057.
125. Pidpala O.V., Yatsishina A.P., Lukash L.L. Human mobile genetic elements: structure, distribution and functional role // *Cytology and genetics*. 2008. V. 42(6). P. 420-430.
126. Pinkel D., Straume T., Gray J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization // *Proceedings of the National Academy of Sciences*. – 1986. V. 83. P. 2934–2938.

127. Pinkel D., Landegent, J., Collins C., Fuscoe J., Seagraves R., Lucas J., Gray J. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4 // Proceedings of the National Academy of Sciences. – 1988. V. 85(23). P. 9138-9142.

128. Piper J., Granum E. On fully automatic feature measurement for banded chromosome classification // Cytometry Part A. 1989. V. 10(3). P. 242-255.

129. Pita M. Orellana J., Martínez-Rodríguez P., Martínez-Ramírez A., Fernandez-Calvín B., Bella J. L. Chapter 10 FISH methods in cytogenetic studies // Functional Analysis of DNA and Chromatin / Stockert J.C., Espada J., Blazquez-Castro A. – Humana Press, 2014. – P. 109-135.

130. Poletti E., Zappelli F., Ruggeri A., Grisan E. A review of thresholding strategies applied to human chromosome segmentation // Computer methods and programs in biomedicine. 2012a. V. 108(2). P. 679-688.

131. Poletti E., Grisan E., Ruggeri A. A modular framework for the automatic classification of chromosomes in Q-band images // Computer methods and programs in biomedicine. 2012b. V. 105(2). P. 120-130.

132. Poon S.S., Martens U.M., Ward R.K., Lansdorp P.M. Telomere length measurements using digital fluorescence microscopy // Cytometry Part A. 1999. V. 36(4). P. 267-278.

133. Pravina V.A. Survey on techniques used for M-FISH image segmentation for classification of chromosomes // Middle-East Journal of Scientific Research. 2015. V. 23(8). P. 1772-1779.

134. Raina S. N., Rani V. GISH technology in plant genome research // Methods in cell science. – 2001. V. 23(1-3). P. 83-104.

135. Rampin M., Bi K., Bogart J. P., Collares-Pereira M.J. Identifying parental chromosomes and genomic rearrangements in animal hybrid complexes of species with small genome size using genomic in situ hybridization (GISH) // Comparative cytogenetics. 2012. V. 6(3). P. 287.

136. Rauch J., Wolf D., Craig J.M. et al. Quantitative microscopy after fluorescence in situ hybridization—a comparison between repeat-depleted and non-depleted DNA probes // Journal of Biochemical and Biophysical Methods. 2000. V. 44(1). P. 59-72.

137. Rens W., Moderegger K., Skelton H., Clarke O., Trifonov V., Ferguson-Smith M.A. A procedure for image enhancement in chromosome painting // *Chromosome Research*. 2006. V. 14(5). P. 497-503.
138. Restif T. Segmentation and Evaluation of Fluorescence Microscopy Images: Doctor of Philosophy Dissertation. Oxford. 2006 – 184 p.
139. Ried T. Schröck E., Ning Y., Wienberg J. Chromosome painting: a useful art // *Human molecular genetics*. 1998. V. 7(10). P. 1619-1626.
140. Riegel M. Human molecular cytogenetics: From cells to nucleotides // *Genetics and molecular biology*. 2014. V. 37(1). P. 194-209.
141. Rogan P.K., Cazcarro P.M., Knoll J.H.M. Sequence-based design of single-copy genomic DNA probes for fluorescence *in situ* hybridization // *Genome research*. 2001. V. 11(6). P. 1086-1094.
142. Rouquier S., Trask B.J., Taviaux S., Engh G.V.D., Diriong S., Lennon G.G., Giorg, D. Direct selection of cDNAs using whole chromosomes // *Nucleic Acids Research*. 1995. V. 23(21). P. 4415-4420.
143. Roshtkhari M.J., Setarehdan S.K. A novel algorithm for straightening highly curved images of human chromosome // *Pattern recognition letters*. 2008. V. 29(9). P. 1208-1217.
144. Rubtsov N.B., Karamisheva T.V., Astakhova N.M., Liehr T., Claussen U., Zhdanova N.S. Zoo-FISH with region-specific paints for mink chromosome 5q: delineation of inter- and intrachromosomal rearrangements in human, pig, and fox // *Cytogenet Cell Genet*. 2000a. V. 90(3-4). - P. 268-70.
145. Rubtsov N., Riemann I., Trifonov V., Karamysheva T., Liehr T., Claussen U., König K. Chromosome microdissection using NIR femtosecond laserpulses and generation of band specific DNA-libraries with DOP-PCR // *Cell Mol. Biol*. 2000b. V.46.
146. Rungruangbaiyok S., Phukpattaranont P. Chromosome image classification using a two-step probabilistic neural network // *Songklanakarin Journal of Science & Technology*. 2010. V. 32(3). P. 255-262.
147. Sampat M. P., Castleman K. R., Bovik A. C. Pixel-by-pixel classification of MFISH images // In: *Engineering in Medicine and Biology, 24th Annual Conference*. 2002. V. 2. P. 999-1000.

148. Sampat M.P., Bovik A.C., Aggarwal J. K., Castleman, K.R. Supervised parametric and non-parametric classification of chromosome images // *Pattern Recognition*. 2005. V. 38(8). P. 1209-1223.
149. Safarik I., Safarikova M. Magnetic nano-and microparticles in biotechnology // *Chemical Papers*. 2009. V. 63(5). P. 497-505.
150. Schier J, Kovar B, Kocarek E, Kunes M. Image Processing for Automated Analysis of the Fluorescence In-Situ Hybridization (FISH) Microscopic Images // In: ICHIT'11 Proceedings of the 5th international conference on Convergence and hybrid information technology. 2011. P. 622-633
151. Schubert I., Fransz P.F., Fuchs J. et al. Chromosome painting in plants // *Chromosome Painting / Sharma A.K., Sharma A.* – Dordrecht: Springer Netherlands, 2001. – P. 57-69.
152. Schwarzacher T. DNA, chromosomes, and *in situ* hybridization // *Genome*. 2003. V. 46(6). P. 953-962.
153. Schwarzacher T., Leitch A.R., Bennett M.D., Heslop-Harrison J.S. *In situ* localization of parental genomes in a wide hybrid // *Annals of Botany*. 1989. V. 64(3). P. 315-324.
154. Schwartzkopf W., Evans B. L., Bovik A.C. Minimum entropy segmentation applied to multi-spectral chromosome images // In: *Image Processing, International Conference*. 2001. V. 2. P. 865-868.
155. Schwartzkopf W., Evans B. L., Bovik A.C. Entropy estimation for segmentation of multi-spectral chromosome images // In: *Fifth IEEE Southwest Symposium*. 2002. P. 234-237.
156. Schwartzkopf W. C., Bovik A. C., Evans B. L. Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images // *IEEE transactions on medical imaging*. 2005. V. 24(12). P. 1593-1610.
157. Sealey P.G., Whittaker P.A., Southern E.M. Removal of repeated sequences from hybridisation probes // *Nucleic acids research*. 1985. V. 13(6). P. 1905-1922.
158. Serbanescu M.S. A k-nearest neighbor approach for chromosome shape classification // *Annals of the University of Craiova-Mathematics and Computer Science Series*. 2010. V. 37(3). P. 142-146.

159. Sethakulvichai W., Manitpornsut S., Wiboonrat M. et al. Estimation of band level resolutions of human chromosome images // In: Computer Science and Software Engineering (JCSSE), International Joint Conference. 2012. P. 276-282.
160. Shakoori A.R. Fluorescence *In Situ* Hybridization (FISH) and Its Applications // Chromosome Structure and Aberrations / BhatAijaz T.A., Wani A.A. – New Delhi: Springer, 2017. – P. 343-367.
161. Shirazi Z.S.P., Zamani A., Mortazavi S.M.J., Zakeri F., Dianatpour M., Mosleh-Shirazi M.A. Developing an Automated Cytogenetic Imaging System for Detection of Dicentric Chromosomes in Biological Dosimetry // Journal of Biomedical Physics and Engineering. 2016. DOI: <https://doi.org/10.22086/jbpe.v0i0.571>.
162. Silva G. S., Souza M. M. Genomic in situ hybridization in plants //Genet Mol Res. 2013. V. 12(3). P. 2953-65.
163. Speicher M. R., Ballard S. G., Ward D. C. Computer image analysis of combinatorial multi-fluor FISH // Bioimaging. 1996. V. 4(2). P. 52-64.
164. Sreejini K.S., Lijiya A., Govindan V.K. M-FISH karyotyping—a new approach based on watershed transform // International Journal of Computer Science, Engineering and Information Technology (IJCEIT). 2012. V 2(2). P. 105-117.
165. Srisang W., Jaroensutasinee K., Jaroensutasinee M. Segmentation of overlapping chromosome images using computational geometry // Walailak Journal of Science and Technology (WJST). 2011. V. 3(2). P. 181-194.
166. Stanley R.J., Keller J.M., Gader P., Caldwell C.W. Data-driven homologue matching for chromosome identification // IEEE Transactions on Medical Imaging. 1998. V. 17(3). P. 451-462.
167. Stuurman N., Swedlow J. R. Software tools, data structures, and interfaces for microscope imaging //Cold Spring Harbor Protocols. 2012. V. 2012(1). P. 50-61.
168. Swennenhuis J.F. Foulk B., Coumans F.A., Terstappen L.W. Construction of repeat-free fluorescence in situ hybridization probes // Nucleic acids research. 2012. V. 40(3). P. e20.
169. Swiger R.R., Tucker J.D. Fluorescence in situ hybridization: a brief review // Environmental and molecular mutagenesis. 1996. V. 27(4). P. 245-254.
170. Tanke H.J., Wiegant J., Van Gijlswijk R.P.M., Bezrookove V., Pattenier H., Heetebrij R.J., Talman E.G., Raap A.K., Vrolijk J. New strategy for multi-colour

fluorescence in situ hybridisation: COBRA: COmBined Binary RAtio labelling // *European Journal of Human Genetics*. 1999. V. 7(1). P. 2-11.

171. Telenius H., Carter N.P., Bebb C.E. Id M.N., Ponder B.A.J., Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer // *Genomics*. 1992. V. 13(3). P. 718-725.

172. Trifonov V.A., Vorobieva N.N., Serdyukova N.A., Rens W. FISH with and without COT1 DNA // *Fluorescence In Situ Hybridization (FISH) – Application Guide.* / Liehr T. – Jena: Springer Berlin Heidelberg, 2017. – P. 123-134.

173. Wang X., Li S., Liu H. Wood M., Chen W.R., Zheng B. Automated identification of analyzable metaphase chromosomes depicted on microscopic digital images // *Journal of biomedical informatics*. – 2008a. V. 41(2). P. 264-271.

174. Wang X., Zheng B., Li S., Mulvihill J.J., Liu, H. A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes // *Computer methods and programs in biomedicine*. 2008b. V. 89(1). P. 33-42.

175. Wang X., Zheng B., Li S., Mulvihill J.J., Wood M.C., Liu H. Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme // *Journal of biomedical informatics*. 2009. V. 42(1). P. 22-31.

176. Waters J.C. Accuracy and precision in quantitative fluorescence microscopy // *The Journal of cell biology*. 2009. V. 185(7). P. 1135-1148.

177. Weiss M.M. Hermsen M.A., Meijer G.A., van Grieken N.C., Baak J.P., Kuipers E.J., Van Diest P.J. Comparative genomic hybridisation // *Molecular pathology*. 1999. V. 52(5). P. 243-251

178. Wenzhong Y., Xiaohui F. A watershed based segmentation method for overlapping chromosome images // In: *Education Technology and Computer Science (ETCS), Second International Workshop*. 2010. V. 1. P. 571-573.

179. Wicker T., Sabot F., Hua-van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. A unified classification system for eukaryotic transposable elements. // *Nature Reviews Genetics*. 2007. V. 8(12), P. 973.

180. Wippold F. J., Perry A. Neuropathology for the neuroradiologist: Fluorescence in situ hybridization // *American journal of neuroradiology*. 2007. V. 28 (3). P. 406-410.

181. Yan W. Enhancement methods for chromosome images // In: Electrical and Control Engineering (ICECE), International Conference. 2011. P. 3024-3026.
182. Yan W., Li D. Feature selection of chromosome images // In: Computer Science and Network Technology (ICCSNT), 2nd International Conference. 2012. P. 677-680.
183. Yang F., Trifonov V., Ng Ling Bee, Kosyakova N., Carter N.P. Generation of Paint Probes from Flow-Sorted and Microdissected Chromosomes// Fluorescence In Situ Hybridization (FISH) - Application Guide. / Liehr T. - Jena: Springer Berlin Heidelberg, 2017. - P. 123-134.
184. Zack G. W., Rogers W. E., Latt S. A. Automatic measurement of sister chromatid exchange frequency // Journal of Histochemistry & Cytochemistry. 1977. V. 25(7). P. 741-753.
185. Zadesenets K.S., Karamysheva T.V., Katokhin A.V., Mordvinov V.A., Rubtsov N.B. Distribution of repetitive DNA sequences in chromosomes of five opisthorchid species (Trematoda, Opisthorchiidae) // Parasitology international. 2012. V. 61(1). P. 84-86.
186. Zhao Y., Kong S. G. Automated classification of touching or overlapping M-FISH chromosomes by region fusion and homolog pairing // Pattern Analysis and Applications. 2013. V. 16(1). P. 31-39.
187. Zhao Y. Wu X., Kong S.G., Zhang L. Joint segmentation and pairing of multispectral chromosome images // Pattern Analysis and Applications. 2013. V. 16 (4). P. 497-506.
188. Zimmer C. From microbes to numbers: extracting meaningful quantities from images // Cellular microbiology. 2012. V. 14 (12). P. 1828-1835.
189. Zwick M.S., Hanson R.E., Islam-Faridi M.N., Stelly D.M., Wing R.A., Price H.J., McKnight T.D. A rapid procedure for the isolation of C0t-1 DNA from plants // Genome. 1997. V. 40(1). P. 138-142.