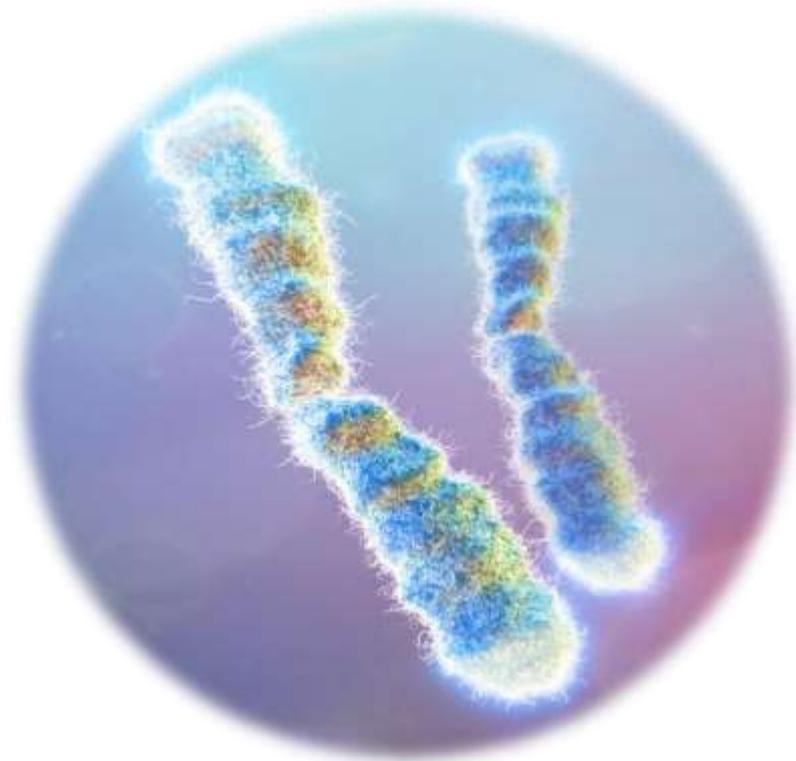


РАЗРАБОТКА МЕТОДА ВИЗУАЛИЗАЦИИ ХРОМОСОМОСПЕЦИФИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК ПРИ ПРОВЕДЕНИИ FISH

03.01.09 Математическая биология, биоинформатика



Богомолов А.Г.
м.н.с., ИЦиГ СО РАН

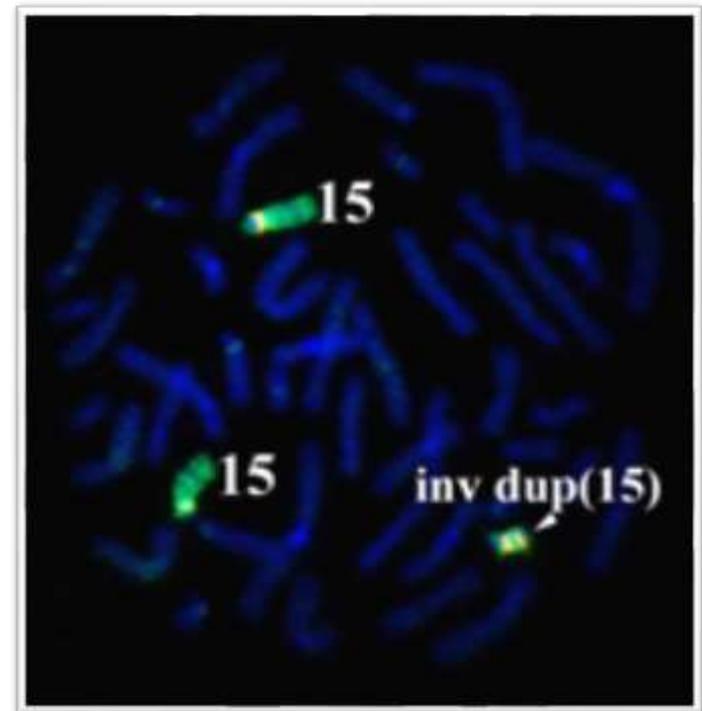
Научный руководитель
д.б.н. Рубцов Н.Б.
г.н.с., ИЦиГ СО РАН

Флуоресцентная гибридизация *in situ* (FISH)

Флуоресцентная гибридизация *in situ*, или метод FISH (англ. *fluorescence in situ hybridization* - FISH) – цитогенетический метод, который применяют для детекции и определения положения специфической последовательности ДНК на метафазных хромосомах или в интерфазных ядрах.

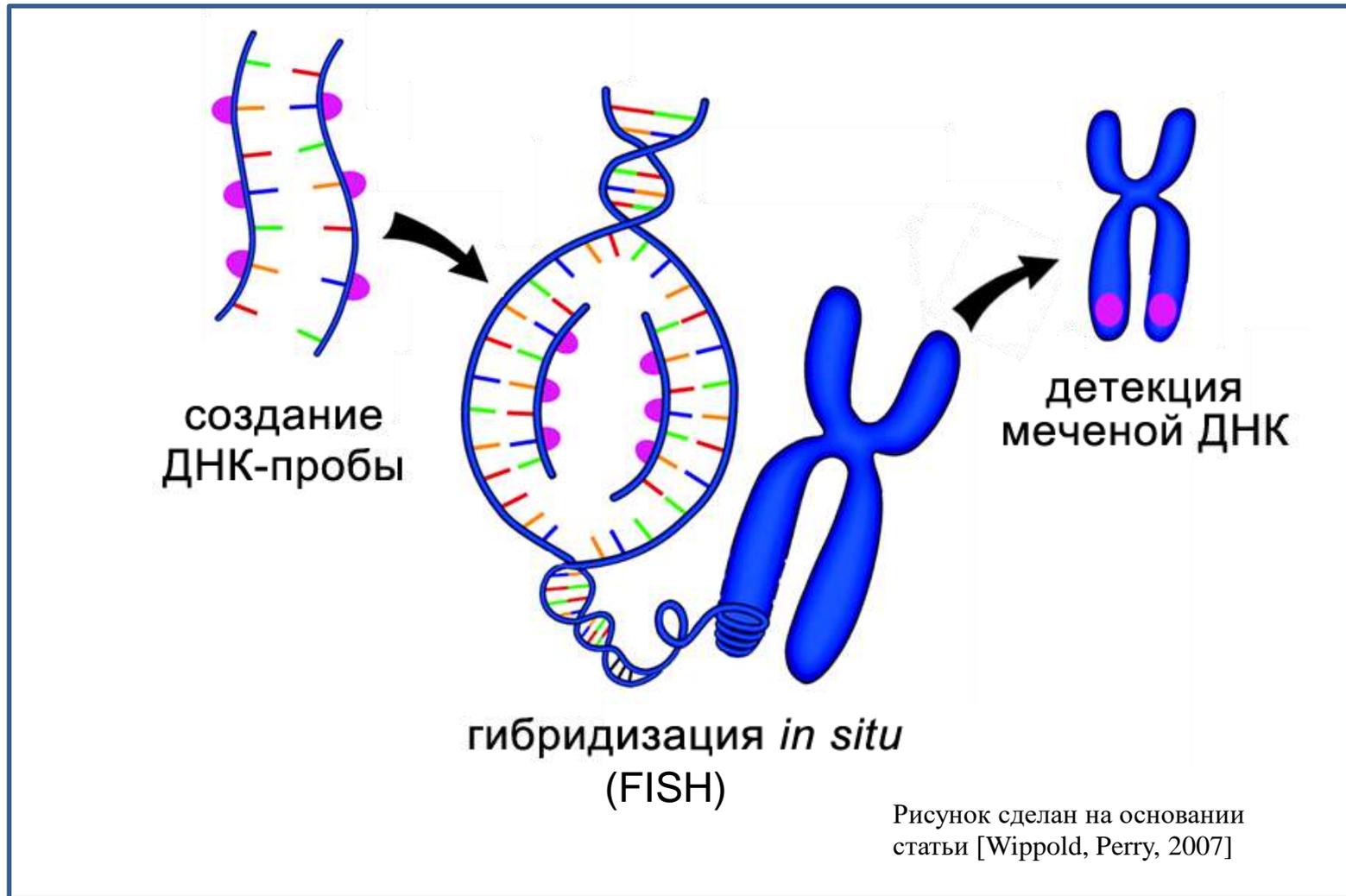


G-окрашивание хромосом
человека



Метод FISH

Принцип флуоресцентной гибридизации *in situ* (FISH)

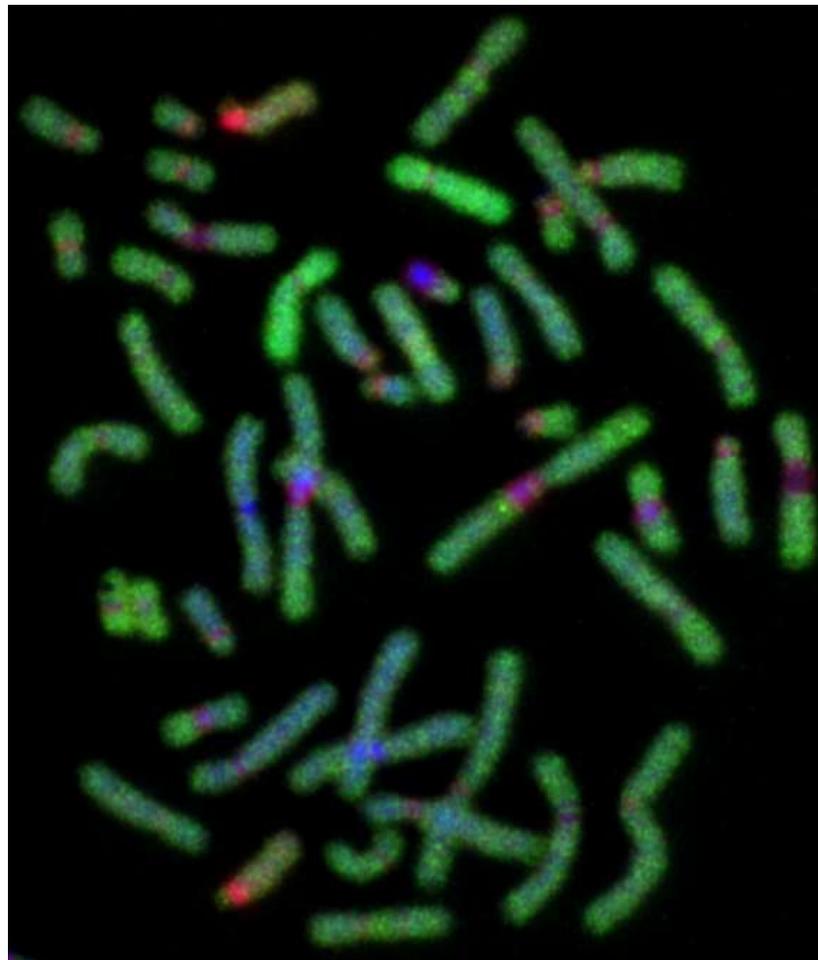


ДНК-пробы - меченные фрагменты ДНК

Проблема детектирования последовательностей ДНК

Проблема: регистрация сигналов повторяющихся последовательностей ДНК на всех хромосомах при проведении FISH протяженных ДНК-проб.

FISH хромосом человека с
ДНК-зондами из хромосом
человека 15  и 10 

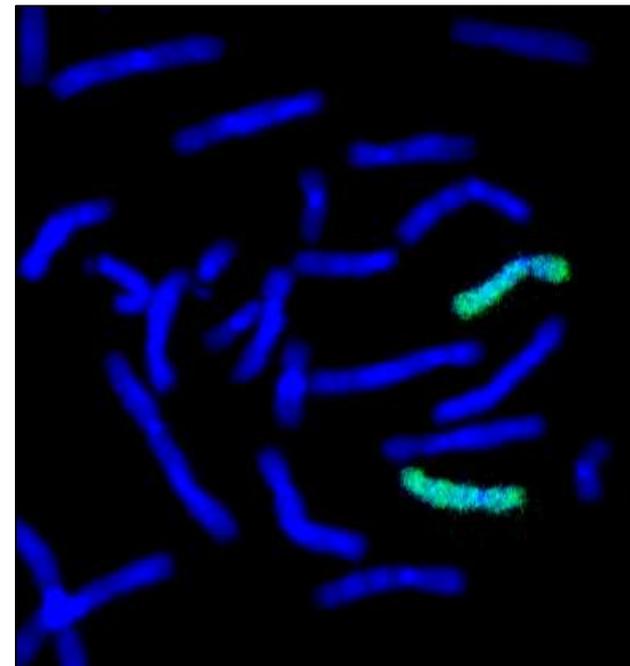
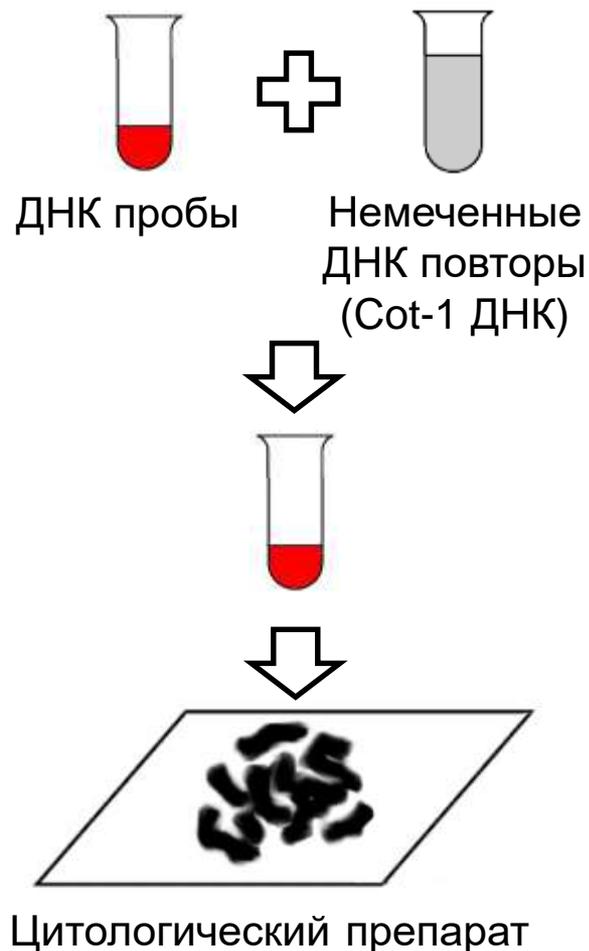


Оригинальное изображение предоставлено
Карамышевой Т.В., с.н.с. ИЦИГ СО РАН

Основное решение: CISS-гибридизация

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

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



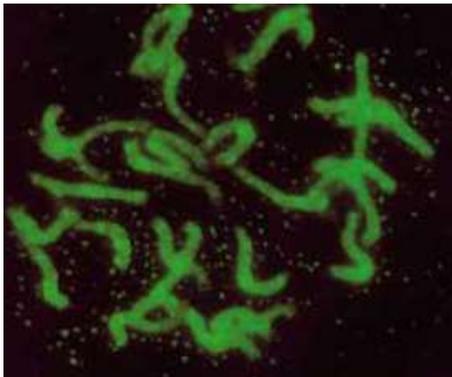
↑
Супрессионная FISH
хромосомспецифичных ДНК
проб из хромосомы 10 человека

Проблема использования CISS-гибридизации



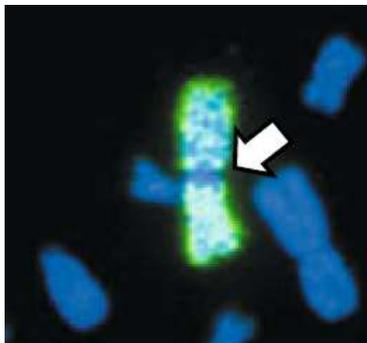
Проблема доступности Cot-1 ДНК

Opisthorchis felinus (Web Atlas of Medical Parasitology)



Неэффективность CISS-гибридизации на организмах в несколько раз превышающих размер генома человека

Неудачная CISS-гибридизации у растений [Schubert et al., 2001]



Подавление всех типов повторяющихся последовательностей

Отсутствие сигнала в С-позитивном районе хромосомы 1 при стандартной CISS-гибридизации

Цели и задачи

Цель: 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.

Задачи:

- 1) Разработать метод и программное обеспечение для визуализации и анализа сигналов хромосомоспецифичных последовательностей ДНК, полученных при проведении двухцветной FISH ДНК-проб из целых хромосом;
- 2) Апробировать метод на изображениях хромосом человека, полученных после проведения двухцветной FISH различных хромосомоспецифичных ДНК-проб с супрессией и без супрессии гибридизации повторяющихся последовательностей;

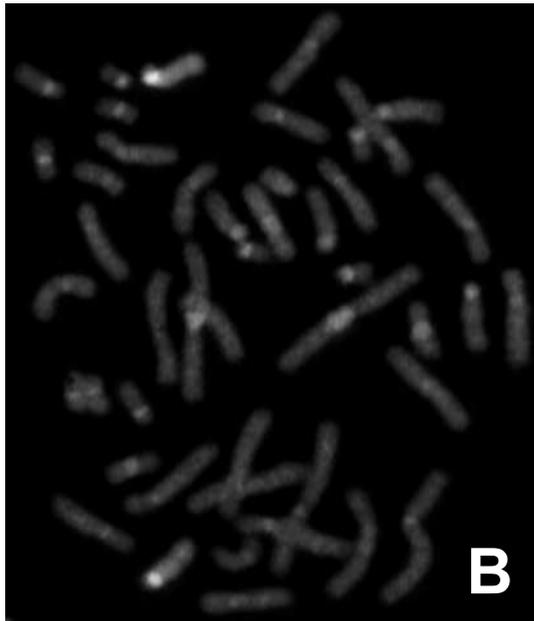
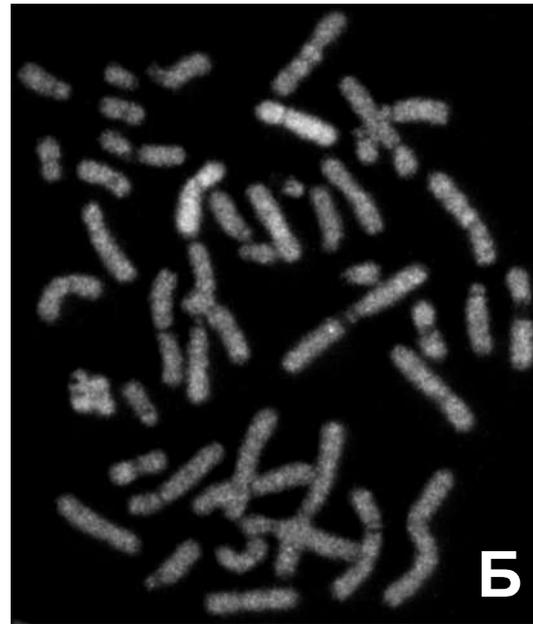
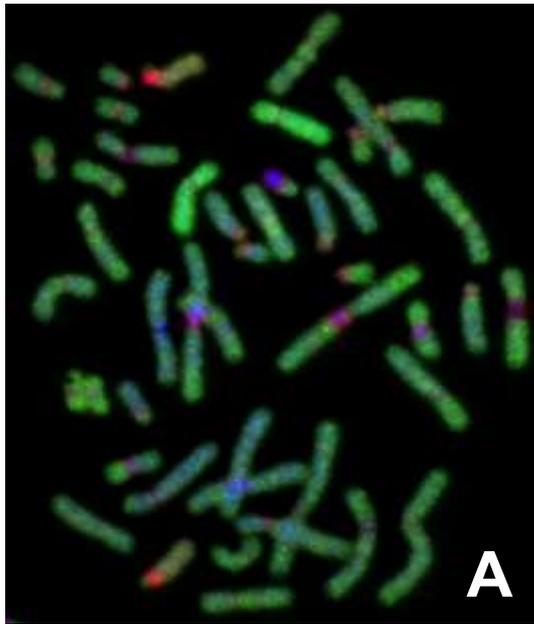
Задачи

- 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.

Положения выносимые на защиту

- 1) Разработан оригинальный метод, который позволяет корректно выделять сигналы хромосомспецифичных последовательностей ДНК в эухроматиновых районах хромосом при проведении FISH ДНК-проб, полученных из целых хромосом или хромосомных районов, без супрессии гибридизации повторяющихся последовательностей.
- 2) Применение данного метода впервые позволило установить гомологию крупных С-негативных районов в хромосомах ряда видов саранчовых семейства *Pamphagidae*, чего ранее не удавалось сделать из-за большого количества повторяющихся последовательностей ДНК в геномах этих видов.

Описание изображений (входные данные)



(А) FISH хромосом человека с ДНК-зондами из хромосом человека 15  и 10 

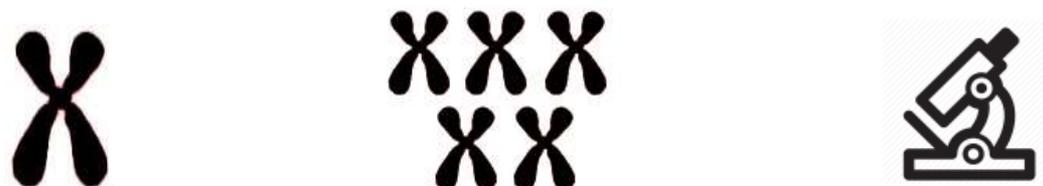
(Б) Результаты FISH с ДНК пробой из 10 хромосомы

(В) Результаты FISH с ДНК пробой из 15 хромосомы

(Г) DAPI бэндинг

Регистрируемый сигнал

Общий сигнал = Специфичный сигнал (**SS**) + Неспецифичный сигнал (**NS**) + Фоновый сигнал (**b**)

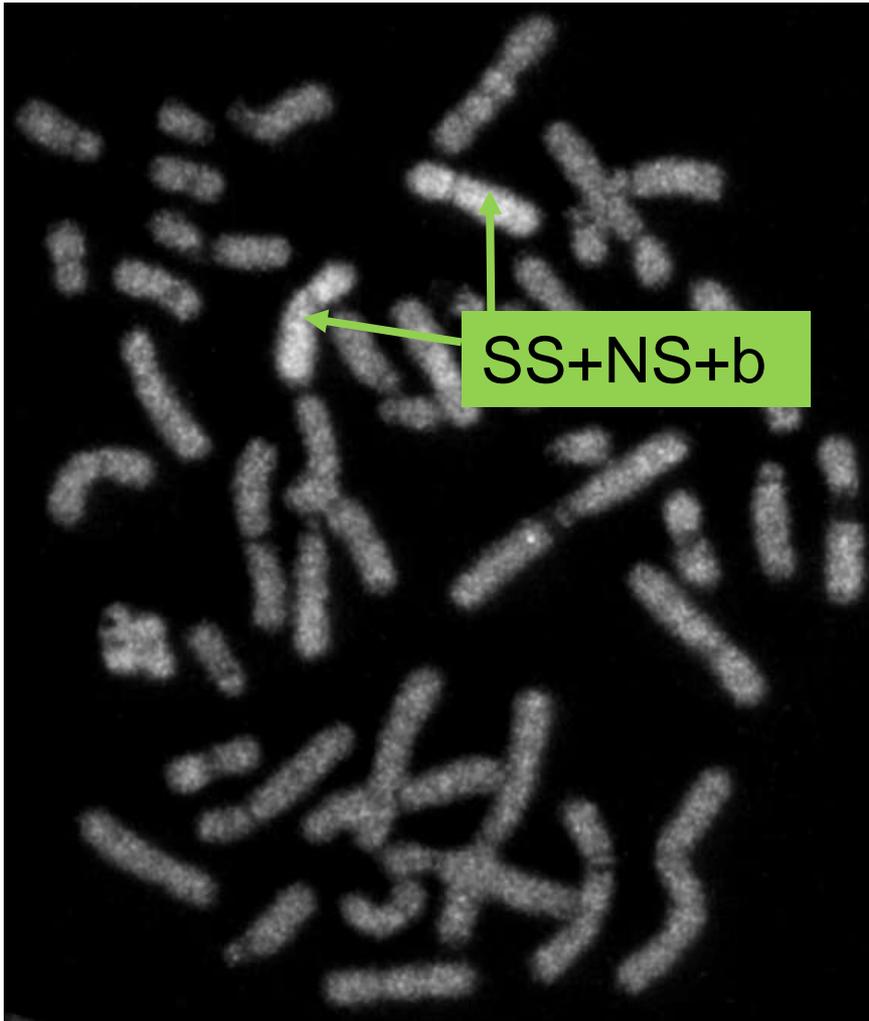


Специфичный сигнал - сигнал с последовательностей, представленных только в хромосоме, из которой была получена ДНК-проба

Неспецифичный сигнал – сигнал с диспергированных повторяющихся последовательностей

Фоновый сигнал – сигнал, обусловлен флуоресцирующими объектами и аппаратными шумами

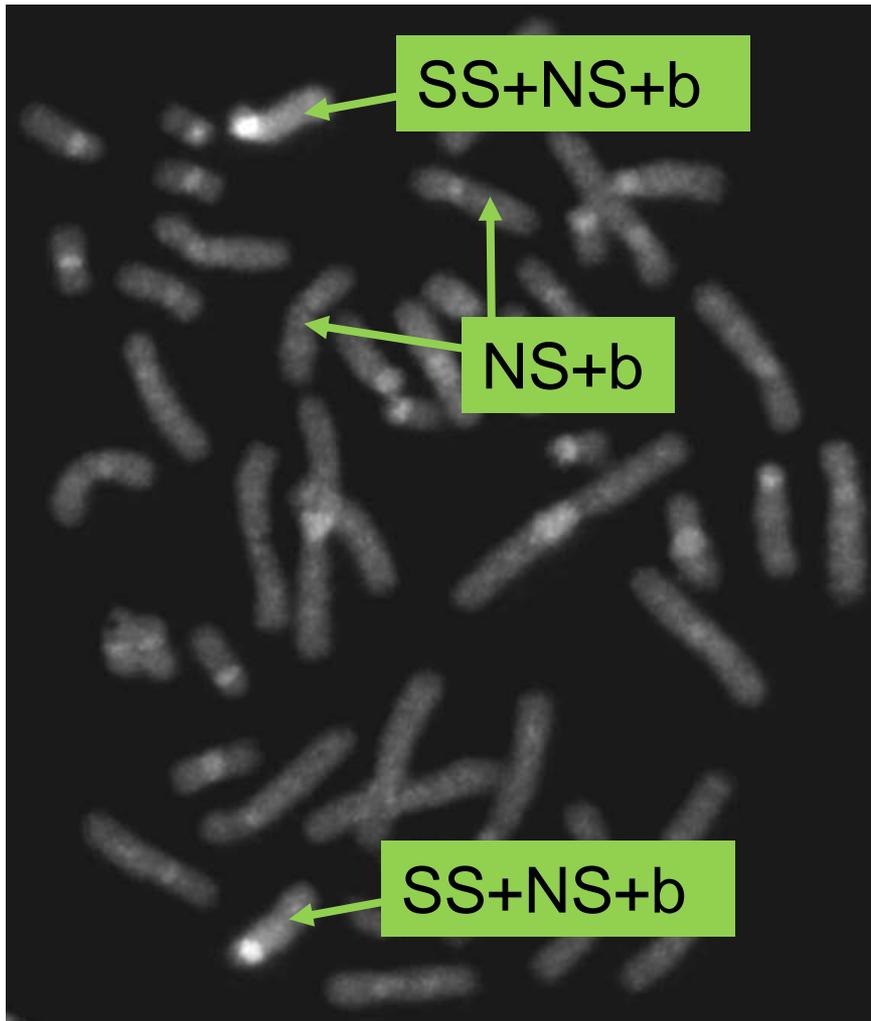
Источники сигналов



← FISH хромосом человека с ДНК-пробой из 15 хромосомы человека

SS- специфичный сигнал
NS- неспецифичный сигнал
b- фоновый сигнал

Источники сигналов



← FISH хромосом человека с ДНК-пробой из 10 хромосомы человека

SS- специфичный сигнал
NS- неспецифичный сигнал
b- фоновый сигнал

Идея метода

Предположение: относительный уровень фонового и неспецифического сигнала приблизительно одинаковый для двух разных ДНК-проб

$$\begin{array}{ccc} i_{ssa} = i_a - i_{ns} - b & \rightarrow & i_{ssa} = i_a - i_b \\ i_{ssb} = i_b - i_{ns} - b & & i_{ssb} = i_b - i_a \end{array}$$

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_{ns} – интенсивность неспецифического сигнала;

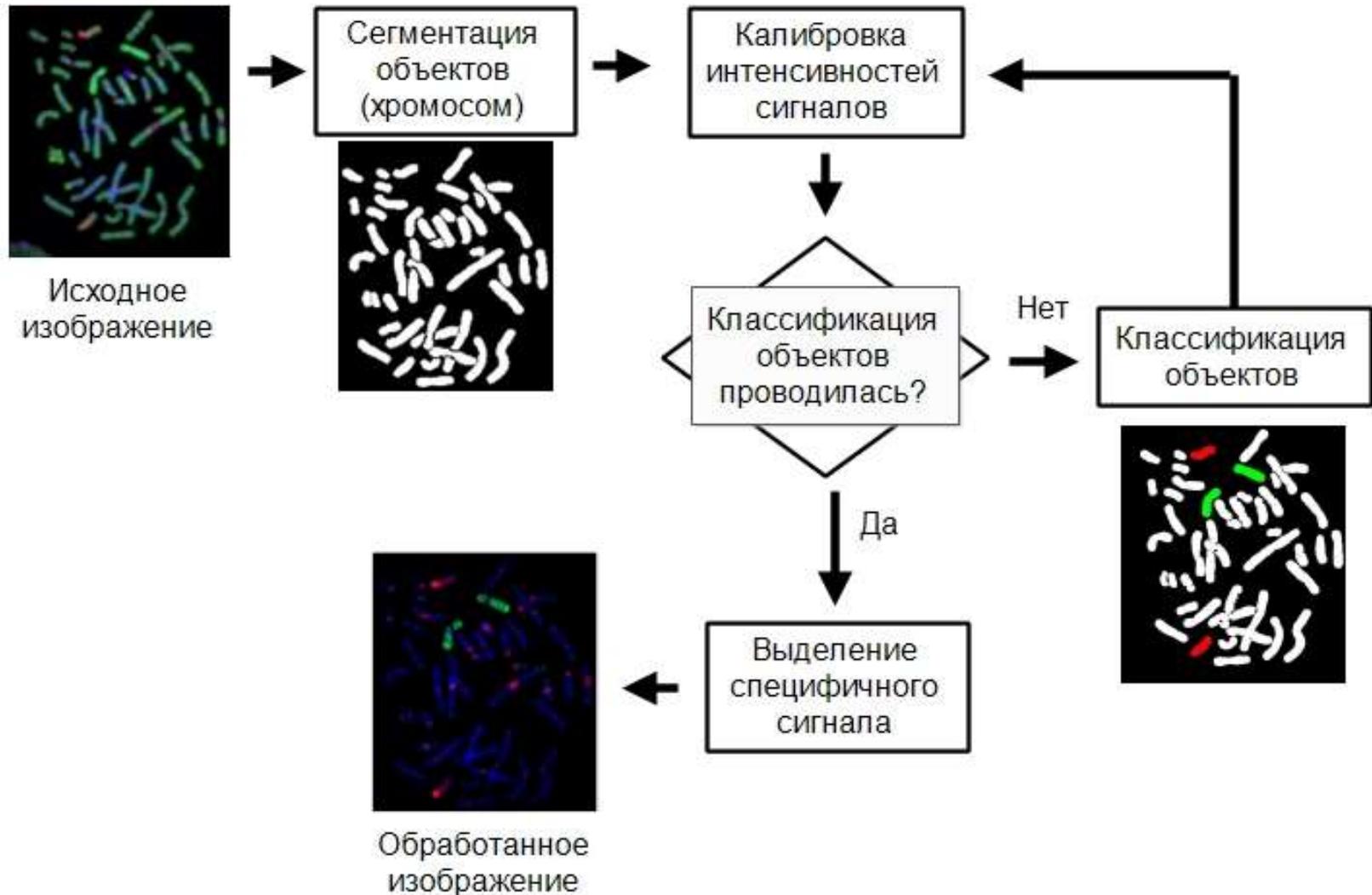
b – интенсивность фонового сигнала;

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 .

Основные этапы метода VISSIS

(visualization specific signal *in silico*)

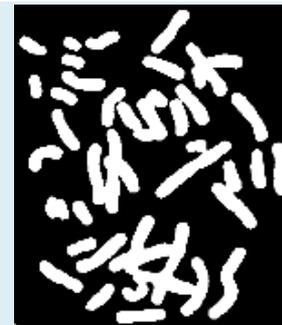


Сегментация объектов

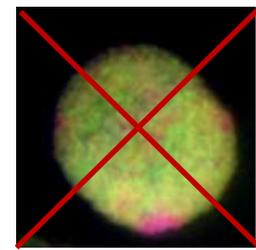
Сегментация - выделение на изображении участков, занятых хромосомами

1) **Удаление шумов** – медианный фильтр (Karvelis et al., 2008; Shirazi et al., 2016)

2) **Пороговая сегментация** – метод Отсу (Otsu, 1979), метод интерактивного выделения (Calvard, Ridler, 1978), алгоритм Triangle (Zack et al., 1977), метод Капура (Kapura et al., 1985).



3) **Удаление ложных объектов** – обнаруживаются с помощью дескрипторов областей (площадь, эксцентриситет).



4) **Разделение кластеров хромосом на отдельные хромосомы** – фильтр средней разности (Poon et al., 1999), метод водораздела (Wenzhong, Xiaohui, 2010)



Калибровка интенсивностей сигналов

Калибровка интенсивностей сигналов – выравнивание уровней интенсивности неспецифических сигналов в точке изображения.

Поправочный коэффициент:

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

Ω - все множество пикселей, принадлежащих объектам, содержащие ДНК, гомологичную только повторяющимся последовательностям;

$i_a(p)$ - интенсивность сигнала с ДНК-пробы, полученной из хромосомы а, в пикселе р;

$i_b(p)$ - интенсивность сигнала с ДНК-пробы, полученной из хромосомы b, в пикселе р.

Классы объектов и характеристики объектов

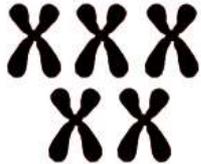
Каждый объект на изображении относился к одному из трех классов:



Ca – объекты, содержащие ДНК, специфичную хромосоме *a*;



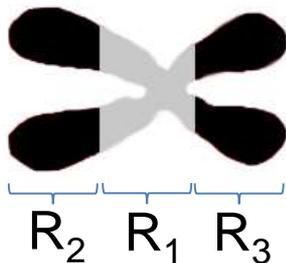
Cb – объекты, содержащие ДНК, специфичную хромосоме *b*;



Cns – объекты, которые не содержат ДНК, специфичную хромосомам *a* и *b*.

Под хромосомами *a* и *b* подразумеваются хромосомы или их районы, из которых получены ДНК-пробы

$$R_{ip} = \log_2 \left(\frac{\mu_a(ip)}{\mu_b(ip)} \right)$$



ip – номер одной из областей, на которые был разделен объект;
 μ_a – средняя интенсивность сигналов с микродиссекционной ДНК-пробы, полученной из хромосомы *a*, в пикселях области *ip*;
 μ_b – средняя интенсивность сигналов с микродиссекционной ДНК-пробы, полученной из хромосомы *b*, в пикселях области *ip*.

Классификация объектов

Классификация объектов – метод ближайшего соседа (Sampat et al., 2004)

$$d(Obj, Clas) = \sqrt{w_1 (R_{1,obj} - R_{1,Clas})^2 + w_2 (R_{2,obj} - R_{2,Clas})^2 + w_3 (R_{3,obj} - R_{3,Clas})^2}$$

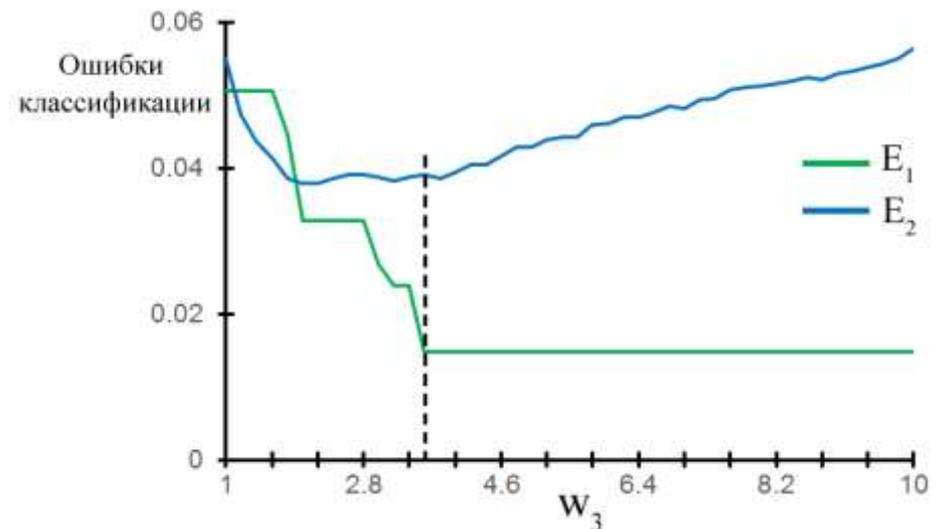
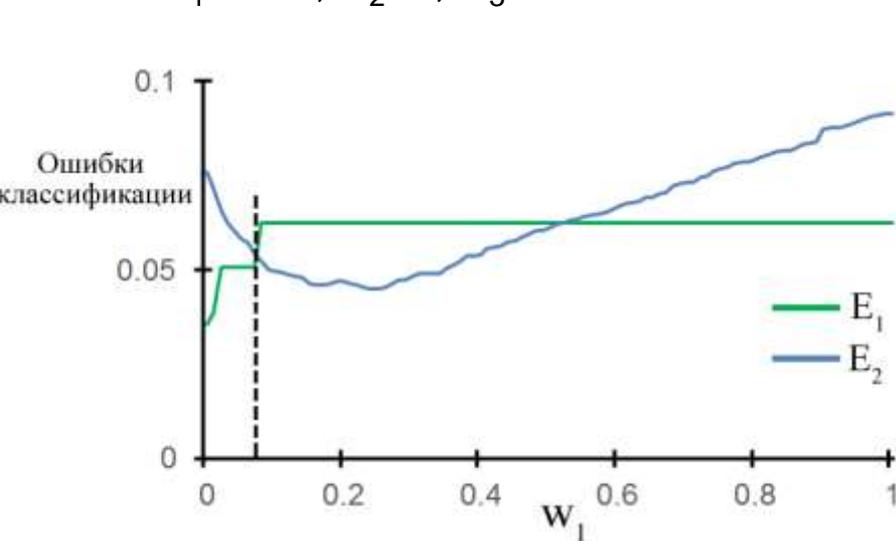
где $\{R_{1,Obj}; R_{2,Obj}; R_{3,Obj}\}$ – характеристики объекта;

$\{R_{1,Clas}; R_{2,Clas}; R_{3,Clas}\}$ – характеристики прототипа класса

Прототип — это вектор отражающий класс объектов.

Подбор коэффициентов на основе анализа результатов классификации

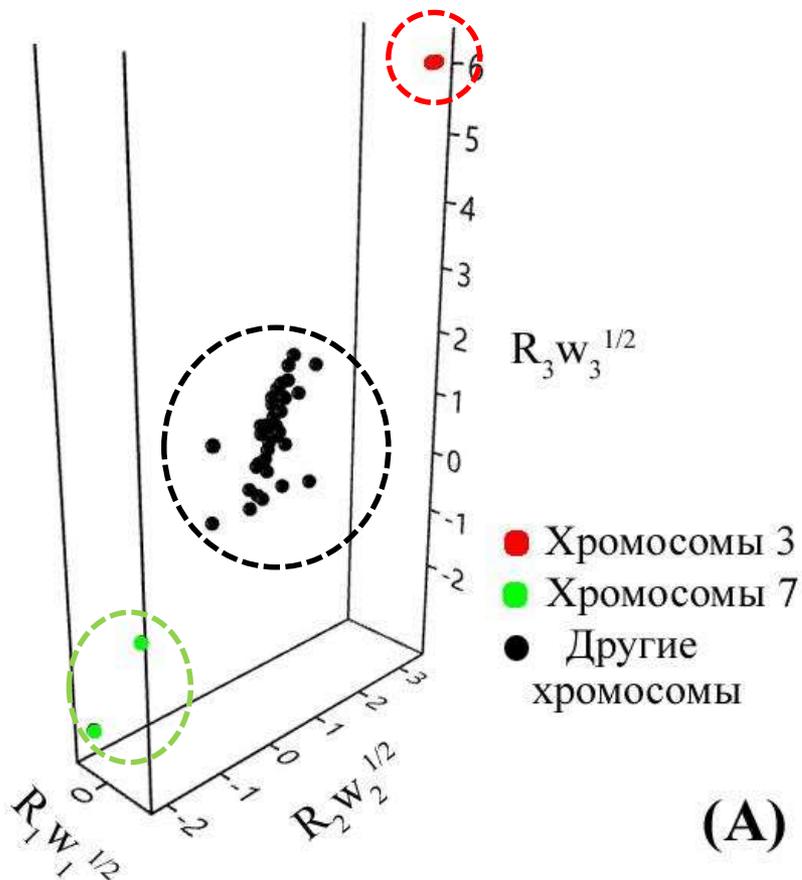
$w_1=0.07$; $w_2=1$; $w_3=3.6$



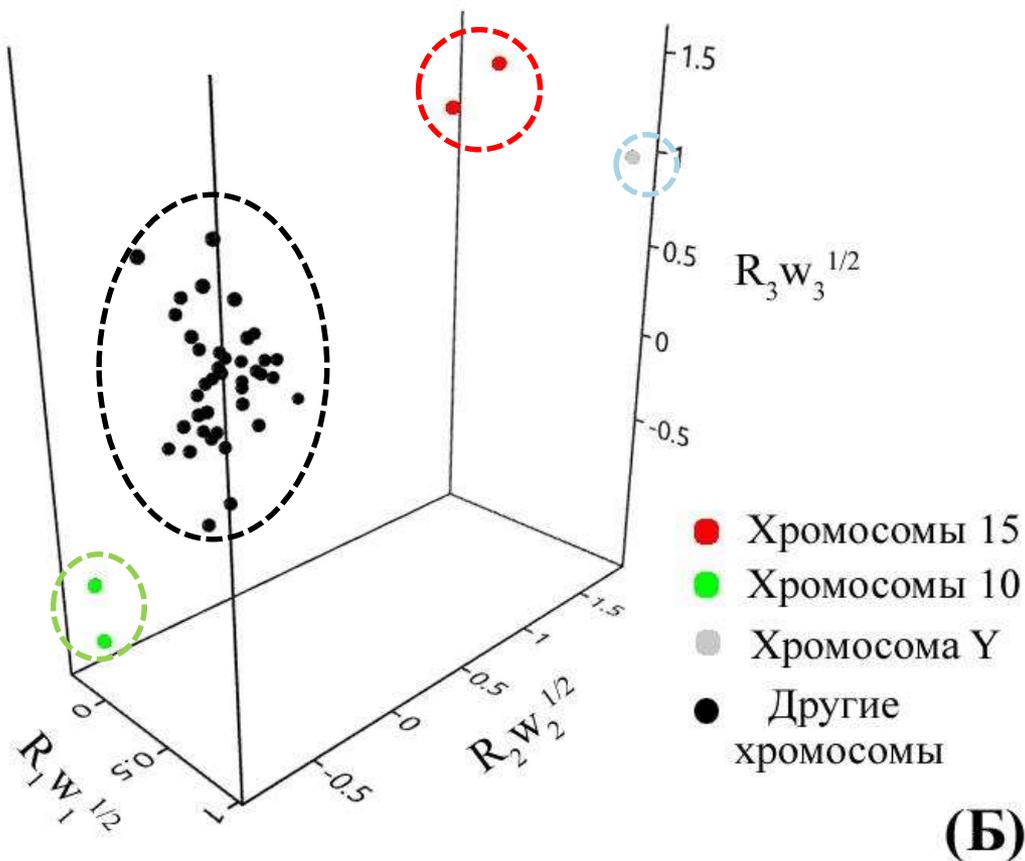
E_1 и E_2 – ошибки классификации первого и второго рода

Объекты в пространстве признаков

CISS-гибридизация



FISH без супрессии повторяющихся последовательностей



Визуализация специфического сигнала

Оценка различий интенсивности сигналов ДНК-проб:

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

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

где $i_a(x, y)$ – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы a , в пиксели (x, y) ;

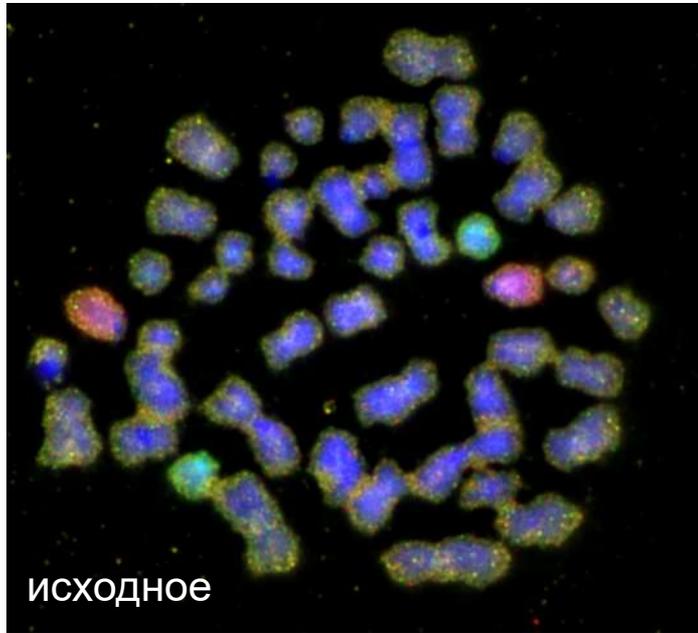
$i_b(x, y)$ – интенсивность сигнала с микродиссекционной ДНК-пробы, полученной из хромосомы b , в пиксели (x, y) .

Новые значения интенсивности сигналов в пиксели:

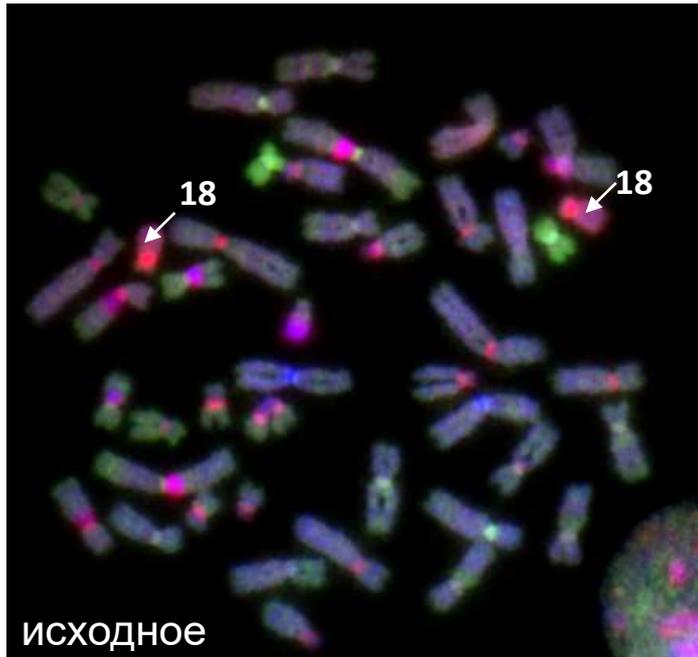
$$i'_a(x, y) = \begin{cases} R(x, y), & \text{если } R(x, y) \geq 0 \\ 0, & \text{если } R(x, y) < 0 \end{cases}$$

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

Применение метода на хромосомах человека



FISH ДНК-проб из
хромосом 13 
и 18 



FISH ДНК-проб из
хромосом 18 
и 19 

Оригинальные
изображения
предоставлены
Карамышевой Т.В.,
с.н.с. ИЦиГ СО РАН

Оценка различия хромосом по соотношению количества коротких (SINE - *Short Interspersed Repeated Element*) и длинных (LINE – *Long Interspersed Repeated Element*) диспергированных повторов:

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

где a и b – номера хромосом, которые сравнивали;

s_a – доля SINE последовательностей в хромосоме a ;

s_b – доля SINE последовательностей в хромосоме b ;

l_a – доля LINE последовательностей в хромосоме a ;

l_b – доля LINE последовательностей в хромосоме b .

Отношение сигнал/Шум (SNR – *Signal/Noise Ratio*):

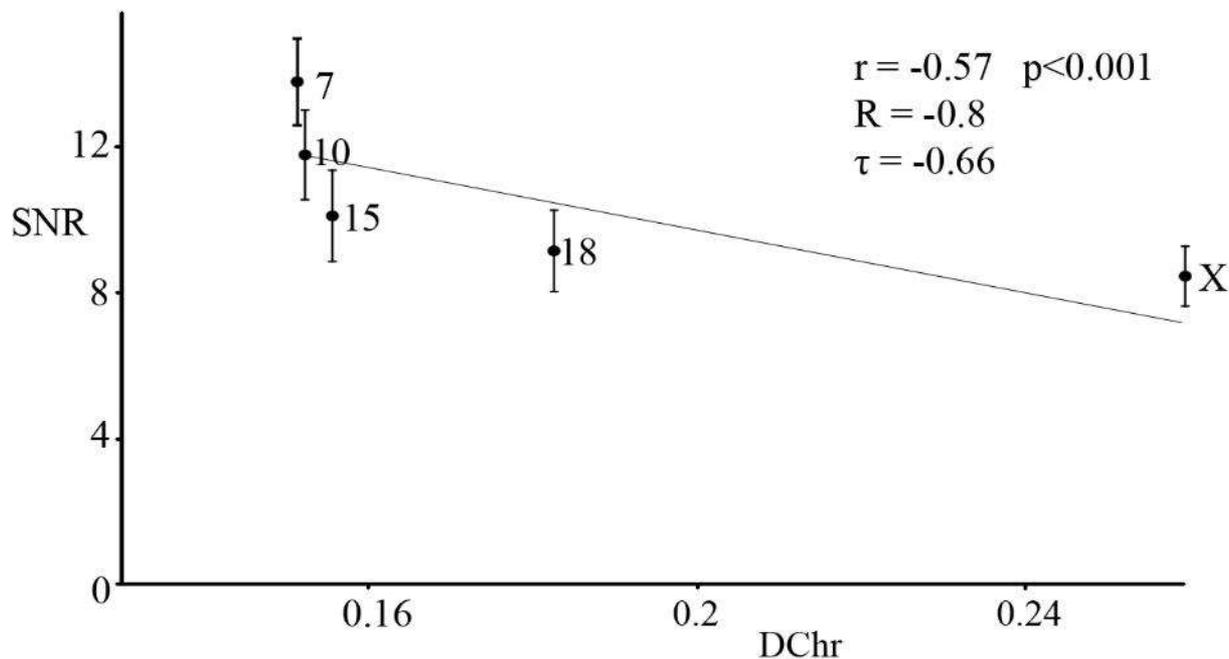
$$SNR = \frac{\mu_s}{\mu_{ns} + 1}$$

μ_s – среднее значение интенсивности специфического сигнала (сигнал на хромосомах, содержавшие ДНК, гомологичную хромосоме, из которой была получена ДНК-проба);

μ_{ns} – среднее значение интенсивности неспецифического сигнала (сигнал на хромосомах, содержащие ДНК, гомологичную только повторяющимися последовательностям).

Анализ эффективности метода

Наблюдается отрицательная корреляция между отношением сигнал/шум и оценкой различий хромосом *DChr* ($p < 0.01$).



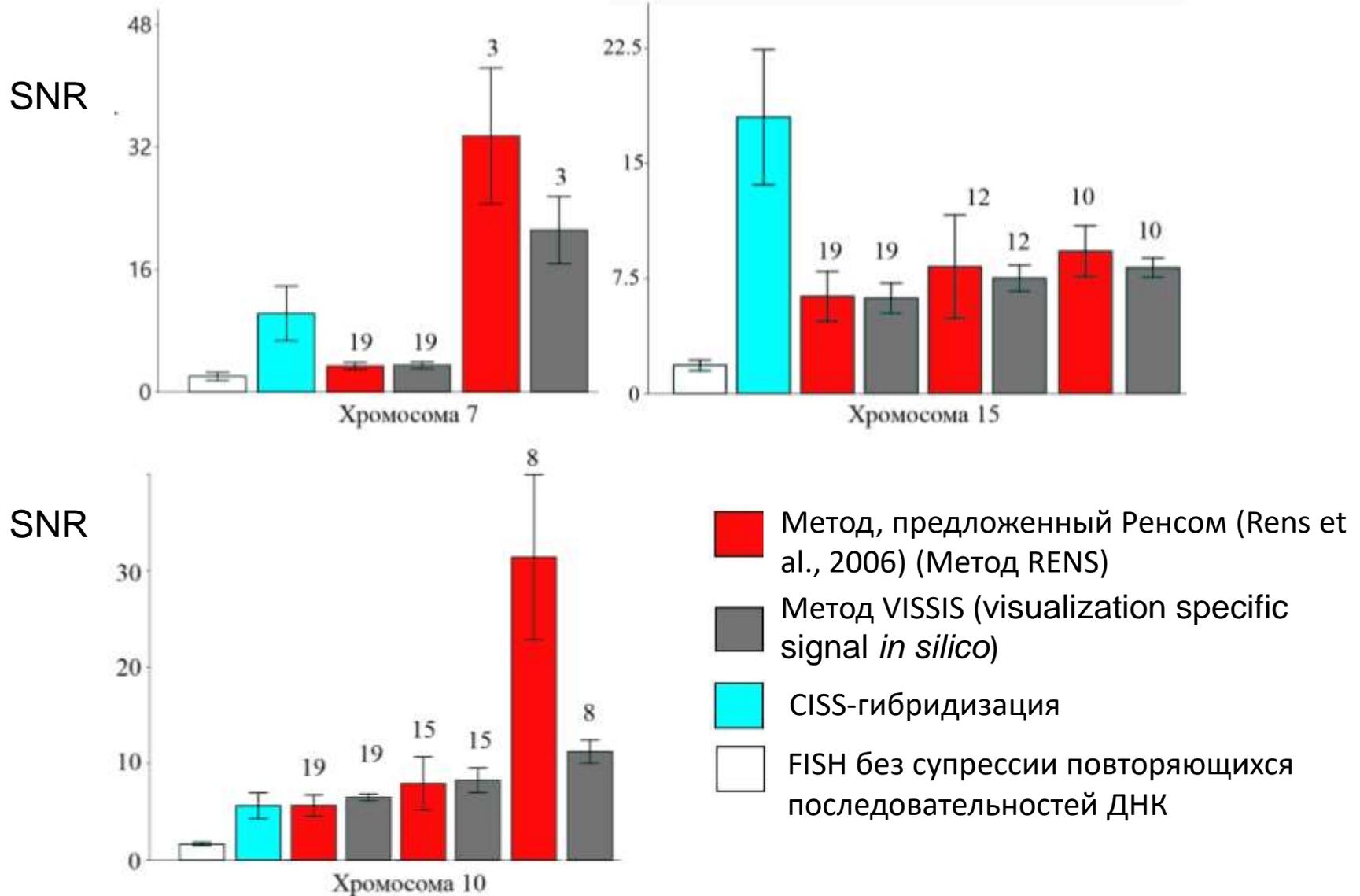
SNR на обработанных изображениях FISH ДНК-проб из 19 хромосомы человека при различных вариантах второй ДНК-пробы. На графике указано среднее значение и стандартное отклонение SNR. Номера хромосом, из которых изготовлена вторая ДНК-проба, указаны справа от значений SNR на графике. r , R и τ – значения коэффициентов корреляции Пирсона, Спирмена и Кендалла, соответственно

Данные по распределению LINE и SINE последовательностям в конкретной хромосоме были взяты из базы данных Genome Browser USCS (<https://genome-euro.ucsc.edu/index.html>)

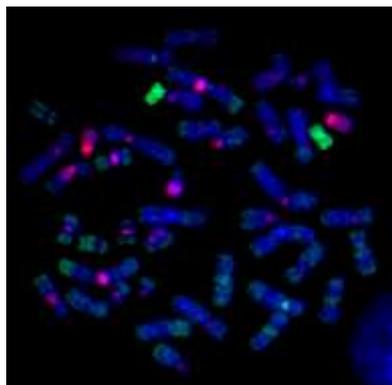
Рекомендованные пары хромосом человека для проведения двухцветной FISH без супрессии гибридизации повторяющихся последовательностей для достижения максимальной эффективности компьютерной обработки методом VISSIS

Номер хромосомы а, из которой была получена первая ДНК-проба	Рекомендованные хромосомы, из которых следует получить вторую ДНК-пробу
1	7, 9, 10, 11, 12, 15, 20
2	3, 4, 5, 6, 7, 8, 9, 10,11, 18
3	2, 4, 5, 6, 7, 8, 11
4	2, 3, 5, 6, 8
5	2, 3, 4, 6, 8, 11
6	2, 3, 4, 5, 7, 8, 11, 18
7	1, 2, 3, 6, 8, 9, 10, 11, 12
8	2, 3, 4, 5, 6, 7, 10, 11, 18
9	1, 2, 7, 10, 14, 15, 18
10	1, 2, 7, 8, 9, 11, 12, 20
11	1, 2, 3, 5, 6, 7, 8, 10, 12
12	1, 7, 10, 11, 20
13	14, 18
14	9, 13, 15, 18
15	1, 9, 14
16	17
17	16, 19
18	2, 6, 8, 9, 13, 14
19	17
20	1, 10, 12
21	13
22	16
X	4
Y	21

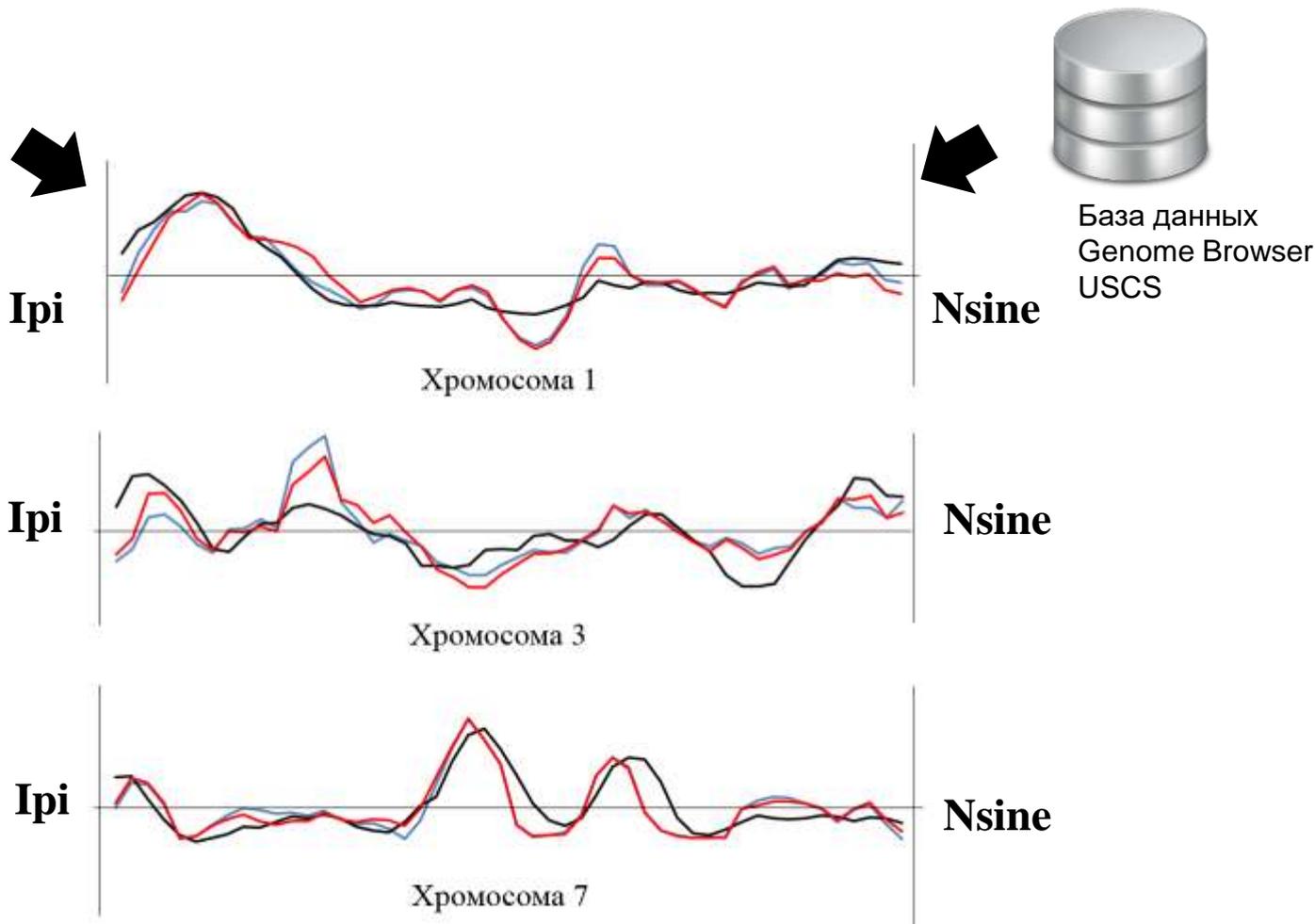
Сравнение метода VISSIS с другими подходами



Сравнение профилей интенсивности с данными по распределению SINE-последовательностей



FISH ДНК-проб из 18 и 19 хромосомы



- Профиль интенсивности
- Распределение SINE последовательностей
- Распределение Alu-повторов

Ipi- Интенсивности сигнала в точке профиля интенсивности
Nsine- количество повторов в хромосомном фрагменте

Сравнение профилей интенсивности с данными по распределению SINE-последовательностей

Оценка различий профилей (*RMSD* - *root-mean-square deviation*):

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

где $P_{img}(j)$ – значение среднего профиля интенсивности в точке j ;

$P_{alu}(j)$ – значение профиля распределения в точке j ;

N – длина профилей (в нашем случае длина профиля составило 50 точек).

$$RMSD_Alu = 0.73 \pm 0.19$$

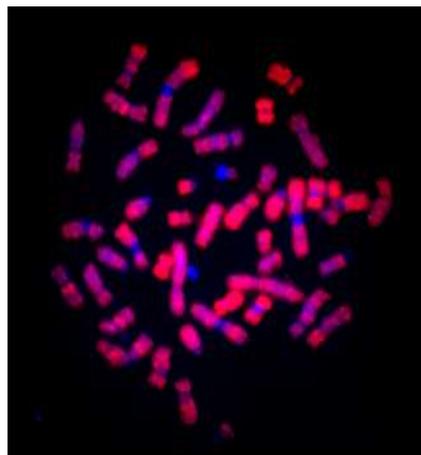
$$RMSD_SINE = 0.73 \pm 0.12$$

RMSD_AlU – количественные оценки различий среднего профиля интенсивности с распределением Alu-повторов

RMSD_SINE количественные оценки различий среднего профиля интенсивности с распределением SINE последовательностей, соответственно

Проведено сравнение профилей интенсивности для 1, 3, 7, 12, 16 и 17 хромосом человека с распределением SINE и Alu последовательностями в соответствующих хромосомах. Профили построены по 14 изображениям хромосом

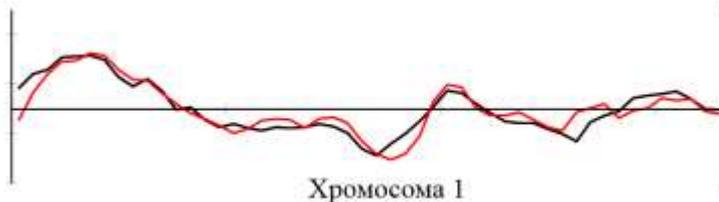
Сравнение профилей интенсивности с данными по распределению SINE-последовательностей



FISH хромосом человека с Alu-последовательностями



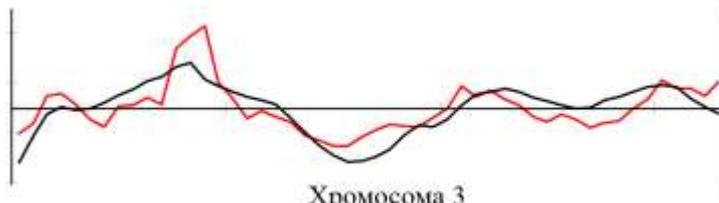
Ipi



Хромосома 1

Nsine

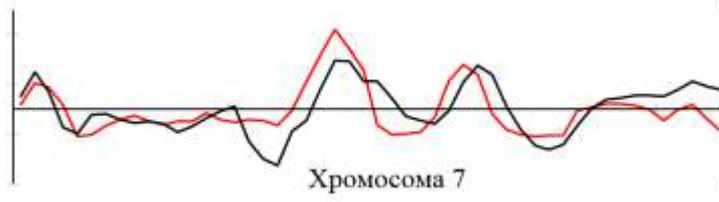
Ipi



Хромосома 3

Nsine

Ipi



Хромосома 7

Nsine

— Профиль интенсивности
— Распределение Alu-повторов



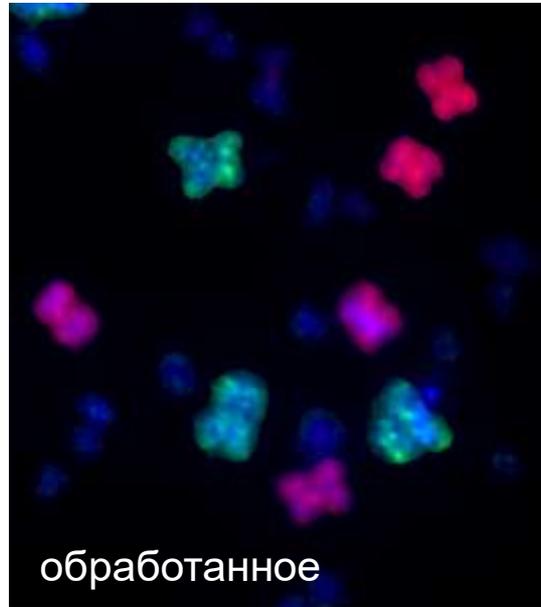
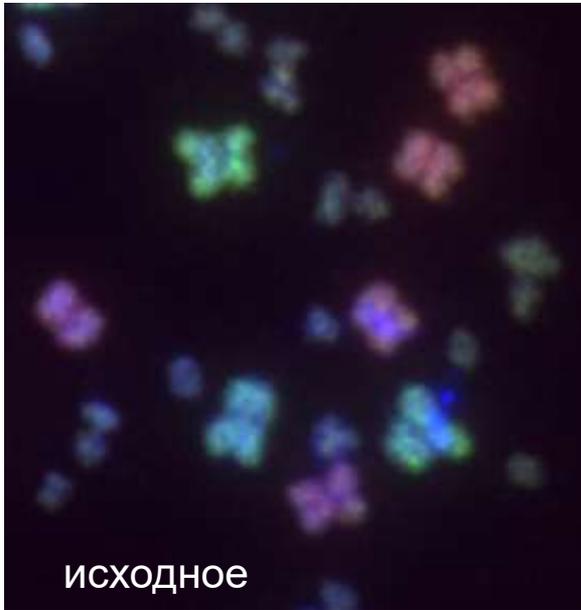
База данных
Genome Browser
UCSC

$$\text{RMSD_AluAlu} = 0.75 \pm 0.2$$

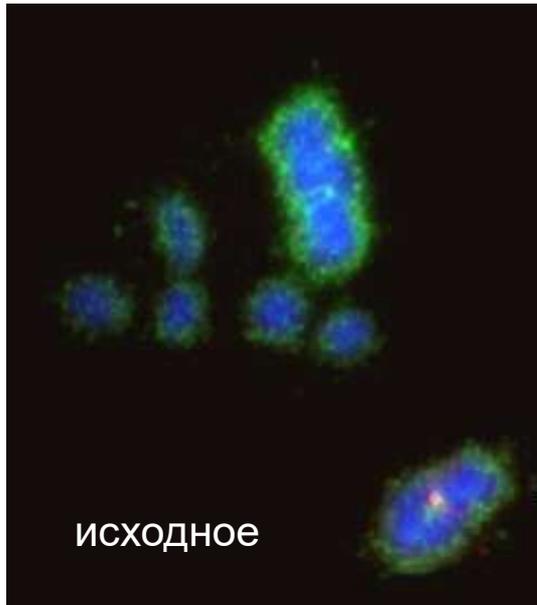
RMSD_AlAlu – количественные оценки различий среднего профиля интенсивности (FISH Alu-последовательностей) с распределением Alu-повторов

Сравнение набора значений оценок RMSD_AlAlu, RMSD_SINE, RMSD_Al не выявило значимых различий медиан (критерий Краскела–Уоллиса, $p > 0.88$).

Применение метода на хромосомах описторхид



Фрагмент метафазной пластинки. FISH ДНК-проб из хромосомы 1  и хромосомы 2  вида *M. xanthosomus*.

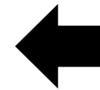
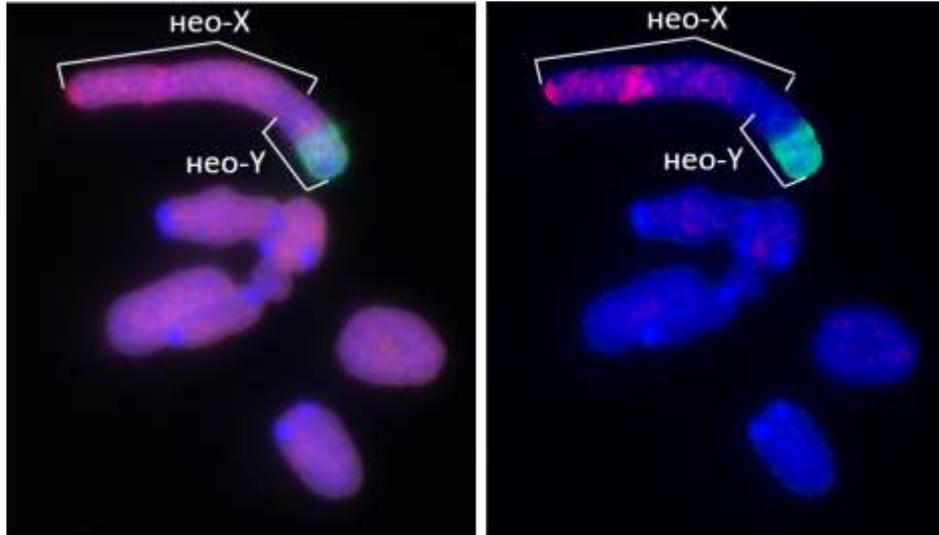


Фрагмент метафазной пластинки. FISH ДНК-проб из хромосомы 1  и хромосомы 2  вида *O. felinus*.

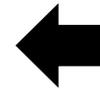
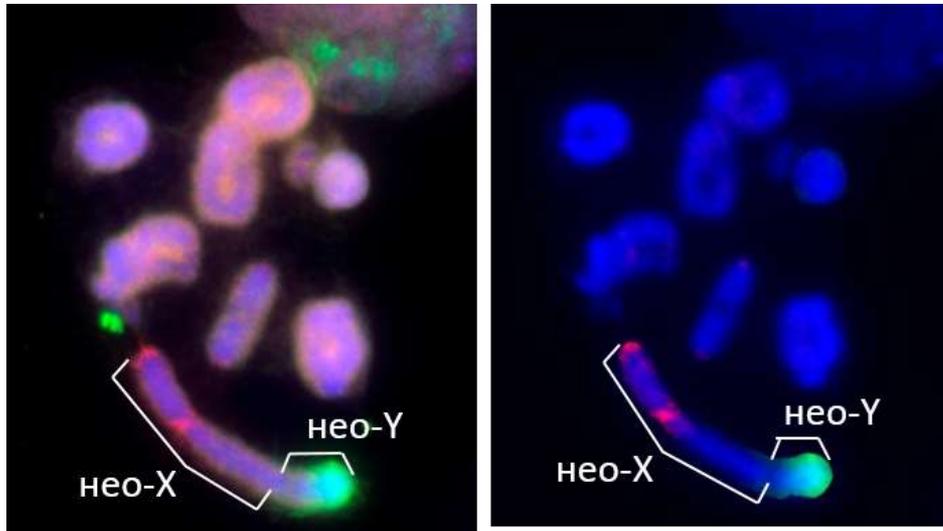
Оригинальные изображения предоставлены
Задесенец К., с.н.с., ИЦиГ СО РАН

Применение метода на хромосомах саранчовых

Применение метода VISSIS впервые позволила установить гомологию крупных С-негативных районов в половых хромосомах нео-Х и нео-У при сравнительном анализе 19 видов саранчовых семейства *Pamphagidae*.



FISH микродиссекционных ДНК-проб из короткого плеча нео-Х хромосомы  и нео-У хромосомы  вида *Nocaracris rubripe* с мейотическими хромосомами этого вида



FISH микродиссекционных ДНК-проб из короткого плеча нео-Х хромосомы  и нео-У хромосомы  вида *Nocaracris rubripe* с мейотическими хромосомами этого вида

Оригинальные изображения предоставлены Джетыбаевым И.Е., н.с., ИЦиГ СО РАН

Выводы

- 1) Разработаны компьютерный метод VISSIS (visualization specific signal *in silico*) и компьютерная программа VisualCS (Visualization Chromosome-specific Signals) позволяющая осуществлять визуализацию сигнала 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.

Выводы

- 4) Обнаружена отрицательная корреляция отношения сигнал/шум на обработанных изображениях с 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 гомологию крупных С-негативных районов в половых хромосомах нео-Х и нео-У при сравнительном анализе 19 видов саранчовых семейства *Pamphagidae*.

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. 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 – V. 16 – №. 12 – С. 9-17 (список ВАК).
4. Karamysheva T.V., Torgasheva A.A., Yefremov Y.R., **Bogomolov A.G.**, Liehr T., Borodin P.M., Rubtsov N.B. Spatial organization of fibroblast and spermatocyte nuclei with different B-chromosome content in Korean field mouse, *Apodemus peninsulae* (Rodentia, Muridae) // *Genome*. – 2017 – V. 60 – N 10 – P. 815-824 (Импакт фактор Web of Science 1.892).
5. **Богомолов А.Г.**, Карамышева Т.В., Рубцов Н.Б. Флуоресцентная гибридизация *in situ* ДНК проб, полученных из индивидуальных хромосом и хромосомных районов // *Молекулярная биология*. – 2014 – V. 48 – N6 – С. 881–890 (Импакт фактор Web of Science 0.977).
6. **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 хромосомоспецифичных 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



Благодарность

Автор благодарит:

д.б.н. Рубцова Н.Б.

к.б.н. Карамышеву Т.В.

к.б.н. Джетыбаева И.Е.

к.б.н. Задесенец К.С.

д.б.н. Орлова Ю.Л.

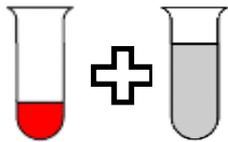
Подколодного Н.Л.

к.б.н. Афонникова Д.А.

к.т.н. Деменкова П.С.

к.б.н. Лаврехе В.В.

Подходы к решению проблемы



ДНК пробы Cot ДНК

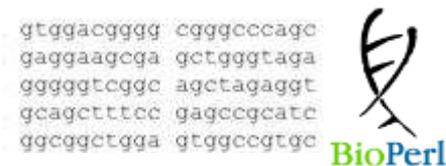
Методы подавления
гибридизации повторенных
последовательностей

(Sealey et al., 1985; Landegent et al., 1987; Lichter et al., 1988; Chung et al., 1997)



Методы элиминации из ДНК
библиотек повторенных
последовательностей

(Chen-Liu et al., 1995; Diriong et al., 1995; Overmyer et al., 1995; Rouquier et al., 1995; Craig et al., 1997; Dugan et al., 2005; Swennenhuis et al., 2012)



Конструирование ДНК проб
без повторенных
последовательностей

(Rogan et al., 2001; Navin et al., 2006; Boyle et al., 2011; Beliveau et al., 2012; Dorman et al., 2013)



Метод дифференциального
удерживания уникальных
последовательностей

(Hozier et al., 1999)



Относительное обогащение
ДНК пробы уникальными
последовательностями

(Nelson et al., 1989, Hozier et al., 1994, Rabbitts et al., 1995, Morozkin et al., 2011)



Компьютерная обработка
изображений

(Rens et al., 2006)

Проверка количества классов

Из гипотезы о компактности классов следует, что если можно разбить множество на два класса, то между ними должен существовать достаточно большой промежуток

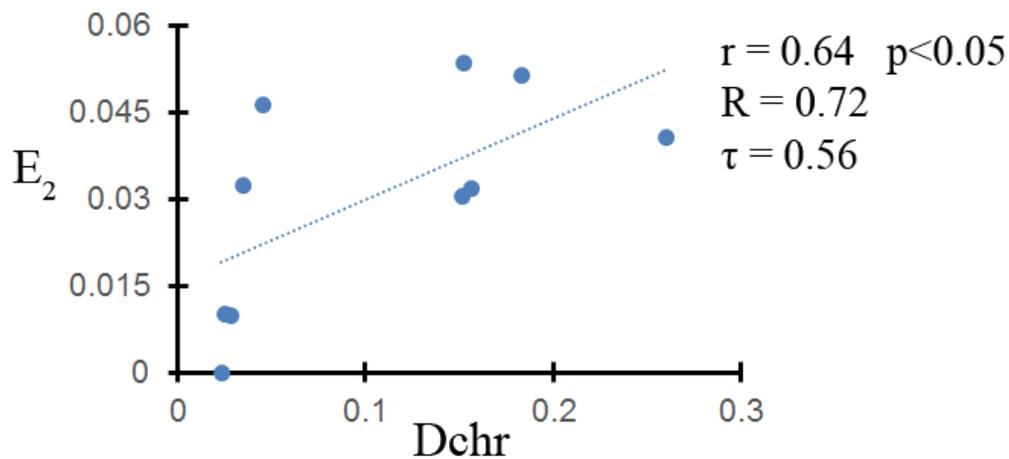
Критерий Ирвина (Irwin, 1925). Для выявления аномальных значений ряда, согласно которому аномальной считается точка Y_t , отстоящая от предыдущей точки Y_{t-1} на величину, большую среднеквадратичного отклонения:

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

где σ - среднеквадратическое отклонение, рассчитываемое по исследуемой выборке с учётом сомнительного значения;

λ_q - критическое (табличное) значение.

Результаты классификации хромосом человека



$$E_1 = \frac{FNns}{Na + Nb}$$

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

$$E_2 = \frac{FNab}{Nns}$$

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