





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

*Колчанов Николай Александрович, проф.
Катохин Алексей Вадимович, к.б.н.
Владимиров Никита Викторович, асп.*

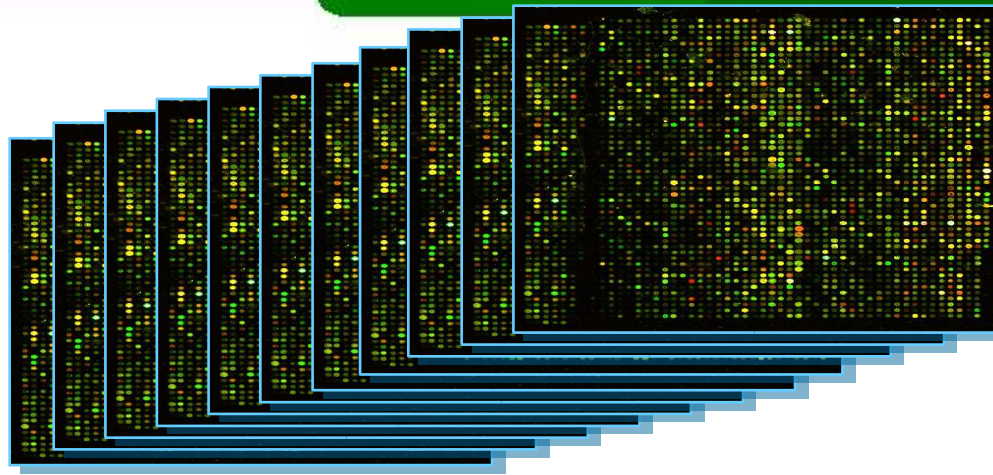
Кафедра информационной биологии ФЕН НГУ
Лаборатория теоретической генетики ИЦГ СО РАН
Сектор функциональной геномики



Компьютерная транскриптомика



Крупномасштабное исследование экспрессионных характеристик геномов с помощью ДНК-чипов



Идентификация генов клеточного цикла на основе статистического анализа профилей экспрессии генов человека

Клетки HeLa S3 были синхронизованы путем ареста клеточного цикла в S-фазе с использованием двойного тимидинового блока или в митозе с использованием тимидин-нокадазолового блока.

Экспрессионные профили для ~ 13 000 генов (42 000 клонов) были оценены с помощью ДНК-чипов и представлены в базе данных SMD (Stanford Microarray Database, <http://genome-www5.stanford.edu/MicroArray/SMD/helpindex.html>.)

Identification of genes periodically expressed in the human cell cycle and their expression in tumors.
Whitfield M. Let al., Mol. Biol. Cell. 2002 Jun;13(6):1977-2000.



Статистический анализ профилей экспрессии генов человека:

монотонное возрастание/уменьшение

Профили экспрессии были исследованы с помощью разработанных нами статистических методов, основанных на регрессионном анализе. Эти методы позволяют группировать гены по сходству профилей экспрессии (цикличность, возрастание, убывание).

Регрессионный анализ монотонности проводился на основе параметрического представления профилей в виде S-образной функции. Перенормировка исходных профилей Z с помощью пробит-преобразования

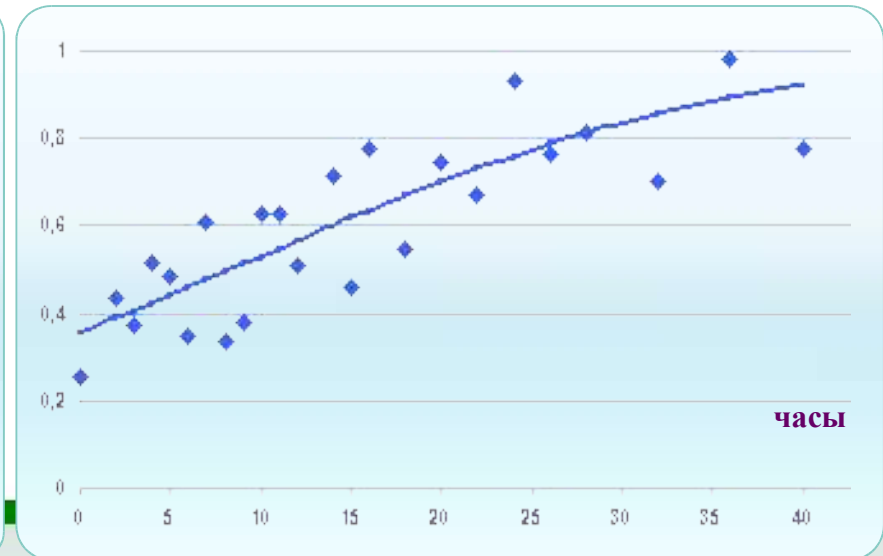
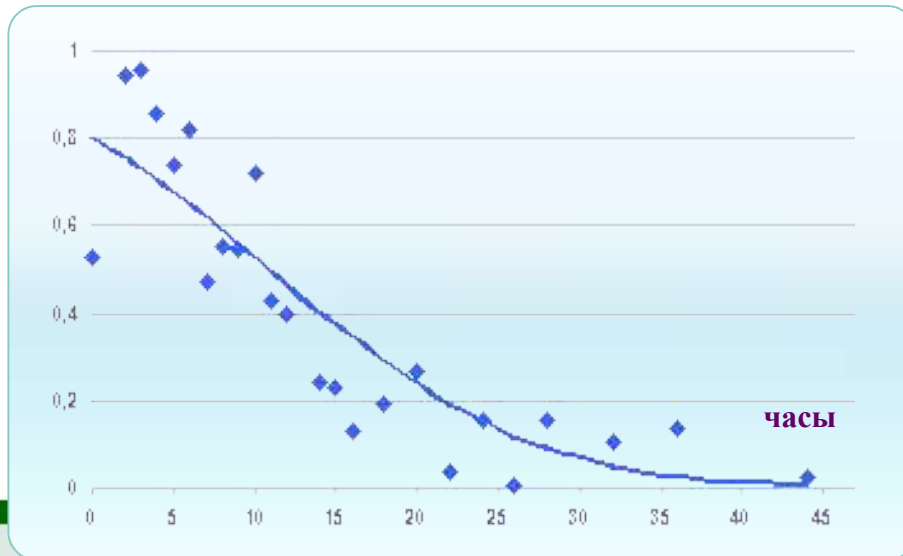
$$\text{PROBIT} = \Phi^{-1}(Z) \quad \text{где} \quad \Phi(x) := (2\pi)^{-1/2} \int_{-\infty}^x \exp(-r^2 / 2) dr,$$

позволила привести их к линейному виду $\text{PROBIT} = a_0 + b_0 \times t + \varepsilon_t$

Среди 42,000 профилей с высоким уровнем значимости ($p\text{-value} < 0.0001$) выявлено **864** монотонных кривых.

Предполагается, что соответствующие гены могут быть вовлечены в процессы созревания клеток, клеточного старения и накопления долгоживущих белковых продуктов.

Пример генов с монотонно изменяющимися профилями экспрессии





Идентификация генов клеточного цикла на основе статистического анализа профилей экспрессии генов человека

Профили экспрессии были исследованы с помощью разработанных нами статистических методов, основанных на регрессионном анализе. Эти методы позволяют группировать гены по сходству профилей экспрессии (цикличность, возрастание, убывание).

$$Y(t) = m_0 + A \times \cos(2\pi \times t / T + \varphi) + \varepsilon t .$$

$$Y(t) = m_0 + A_1 \times \cos(2\pi \times t / T) + A_2 \times \sin(2\pi \times t / T) + \varepsilon t ,$$

где

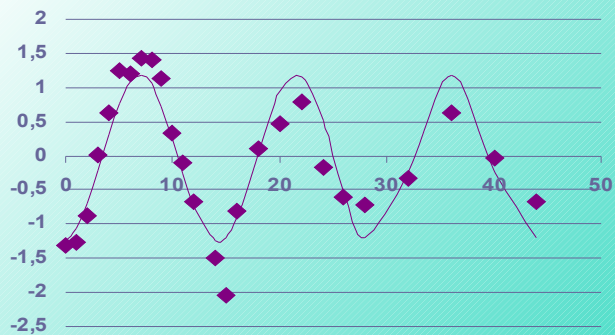
$$A_1 = \cos \varphi, \quad A_2 = -\sin \varphi .;$$

$$A = (A_1^2 + A_2^2)^{-1/2}, \quad \varphi = \arctg (-A_2 / A_1) .$$

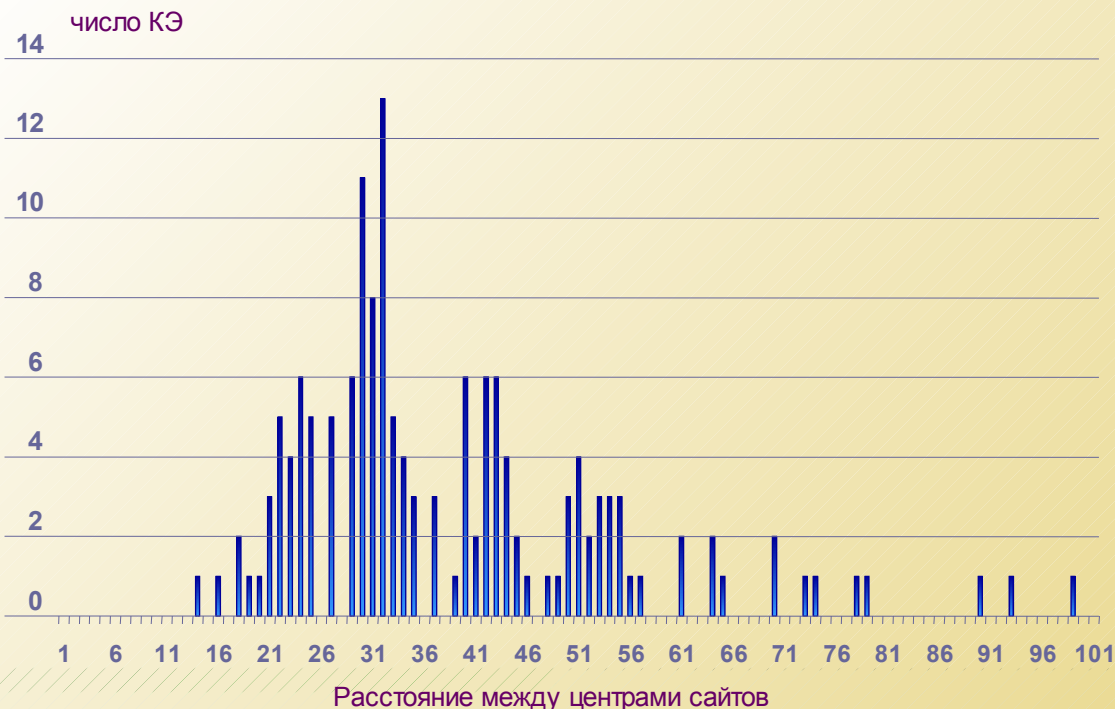
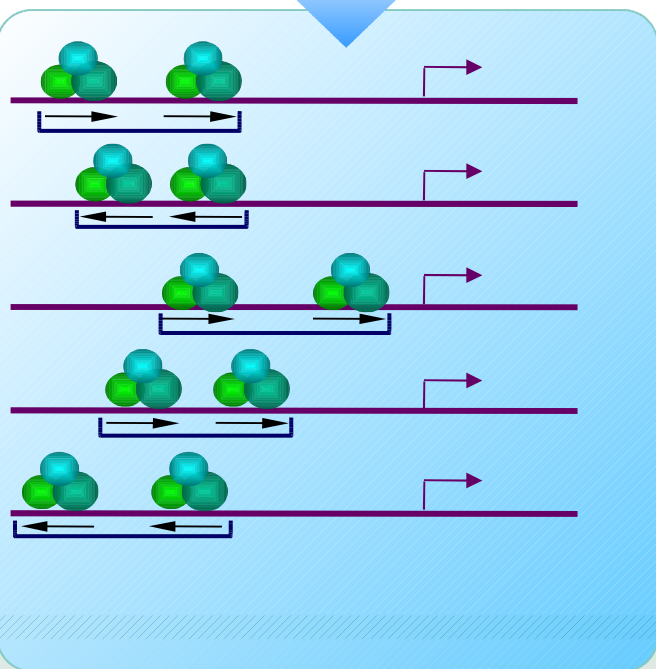
Среди 42,000 профилей с высоким уровнем значимости ($p\text{-value} < 0.0001$) выявлено 4485 периодических кривых, соответствующих ~ **2000** генам.



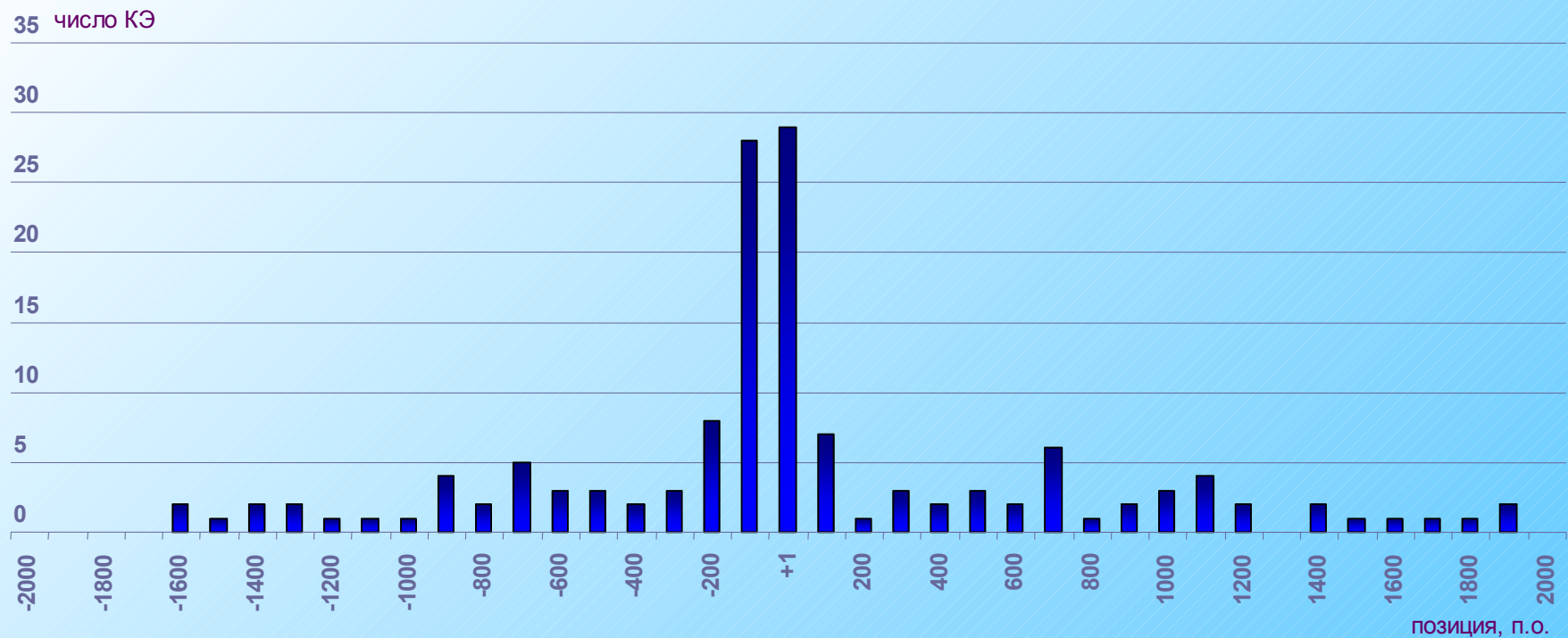
В 5'-районах 96-ти циклически экспрессирующихся генов человека выявлен потенциальный композиционный элемент NF-Y/NF-Y



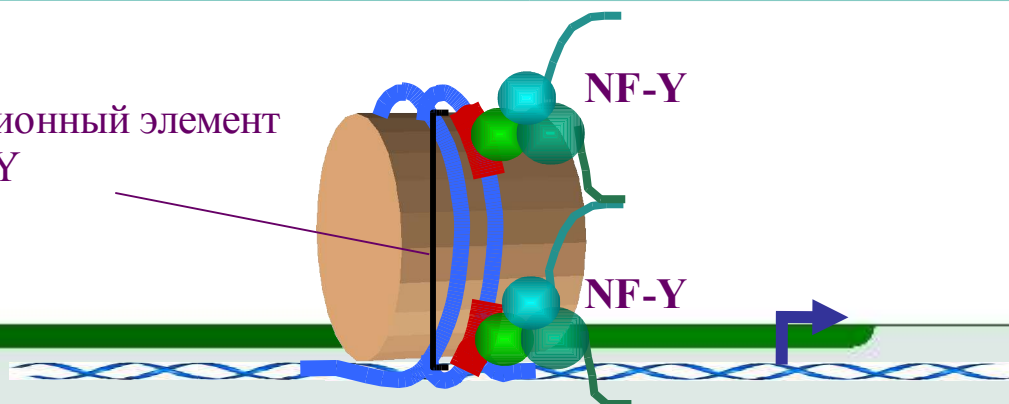
В промоторе гена одного из проанализированных генов циклина В1 экспериментально показано присутствие композиционного элемента, образованного сайтами связывания NF-Y, привлекающего ацетилазу гистонов (p300).
Interactions between p300 and Multiple NF-Y Trimers Govern Cyclin B2 Promoter Function, Salsi et. al. Vol. 278, No. 9, p. 6642–6650, 2003



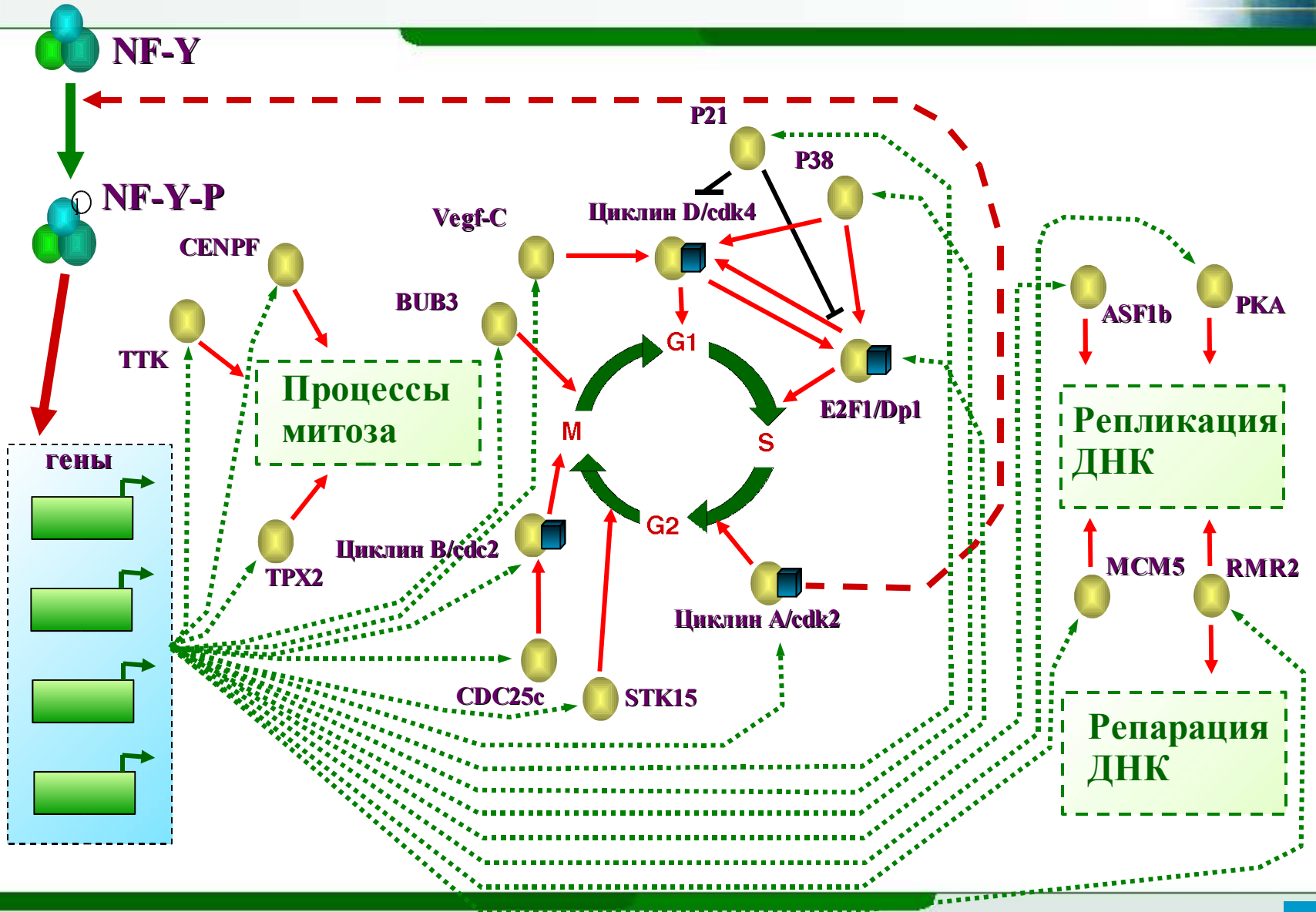
Распределение потенциального композиционного элемента nf-y/nf-y относительно старта транскрипции промоторов циклически экспрессирующихся генов



Композиционный элемент
NF-Y/NF-Y



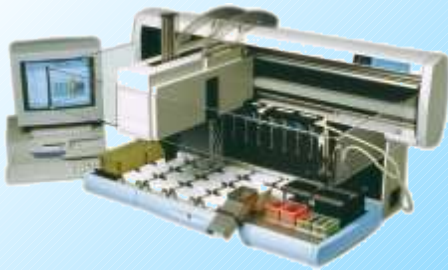
Участие композиционного элемента NF-Y/NF-Y в регуляции клеточного цикла (реконструкция)



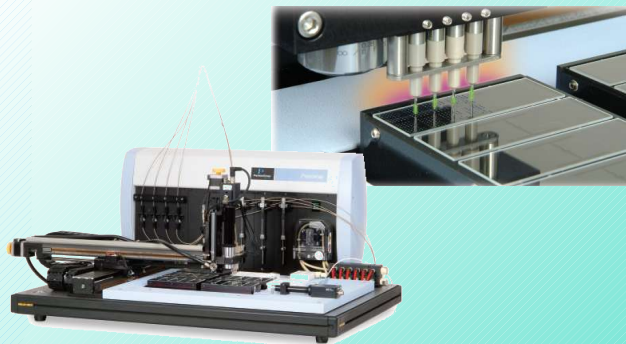


ИЦиГ СО РАН: сектор функциональной геномики (зав. – к.б.н. В.А. Мордвинов)

Оборудование для работы с биочипами



MultiPROBE II HT – роботизированная система для работы с жидкостями: подготовка плат для биочипов, выделение нуклеиновых кислот и т.д.



Piezorray - бесконтактная система для производства биочипов высокой плотности. Прибор является универсальным в отношении плат и субстратов.



ScanArray Lite - конфокальный сканер биочипов: длины волн - 543 и 633 нм, чувствительность – 0.05 флуоресцентных молекул на квадратный микрон.



Оборудование для проведения ПЦР с детекцией в реальном времени



ABI PRISM® 7000 Sequence Detection System. представляет собой многоцветный детектор для мультиплексного количественного анализа, анализа по различению аллелей и качественного плюс/минус анализа. Обладает высокоточной оптикой, совмещенной со сложным многокомпонентным алгоритмом, что обеспечивает получение точных высокопроизводимых результатов.





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ДНК-биочипы – это миниатюризированные матрицы, в которых на подложке в определенном порядке распределены и прикреплены фрагменты ДНК, соответствующие отдельным генам или их частям. Такие организованные микроматрицы позволяют проводить эксперименты по одновременному анализу структуры и экспрессии тысяч генов с помощью параллельной гибридизации.

Развитие методов преобразования результатов этих экспериментов в цифровые данные и методов компьютерной обработки последних обеспечивает возможность анализировать и сопоставлять экспрессию таких массивов генов во множестве экспериментальных условий.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Типы ДНК-биочипов (микроматриц)

На поверхности стекла или полимера

олигонуклеотидные биочипы	кДНКовые биочипы
20-75 н.о.	100-2500 н.о.
Fodor et al., 1993; Lipshutz et al., 1995; Lockhart et al., 1996	DeRisi et al., 1996

В объеме

Фрагменты ДНК, иммобилизованные в микрокаплях геля

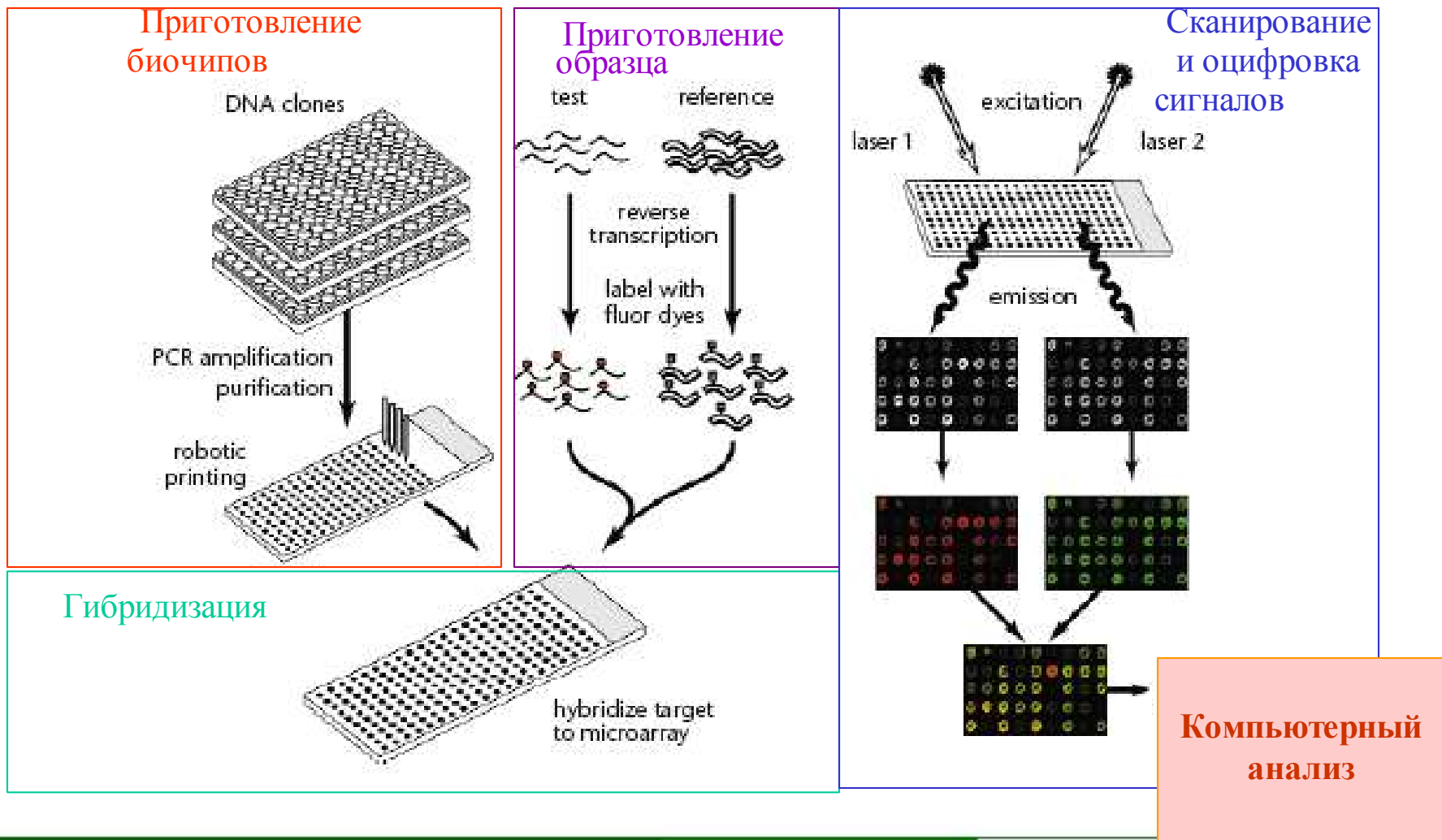
Под руководством акад. А.Д. Мирзабекова в ИМБ РАН
Lysov et al., 1988;
Khrapko et al., 1989

Обычно это микроскопное стекло размером 25 мм x 76 мм,



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

Схема использования кДНК-биочипов (экспериментальная фаза)





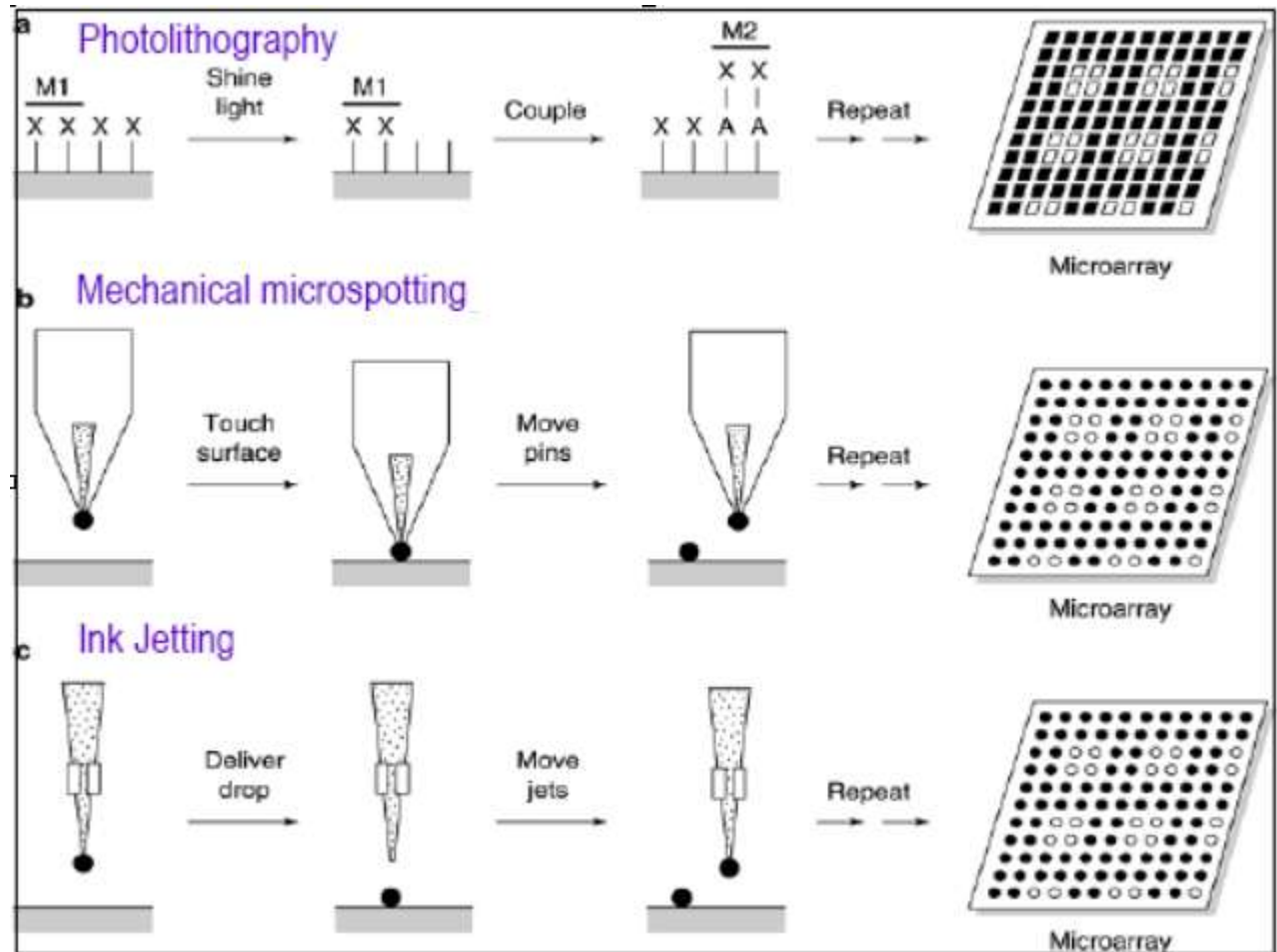
Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

СПОСОБЫ ПРИГОТОВЛЕНИЯ ДНК-БИОЧИПОВ

синтез
олигонуклеотидов
in situ

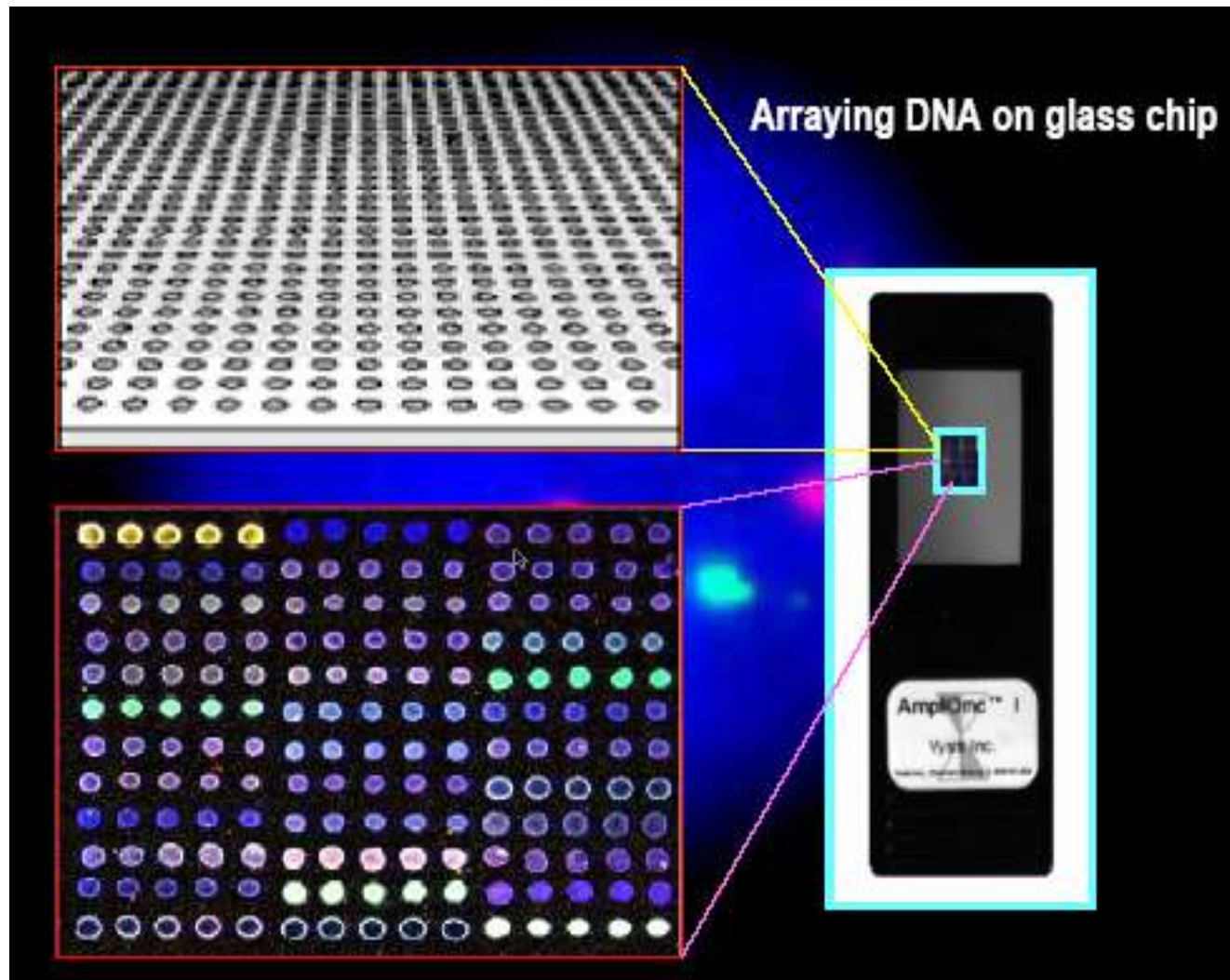
контактная
печать

струйная печать





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Вид
стеклянного
ДНК-биочипа



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

Процесс приготовления ДНК-биочипов





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



При описании ДНК-чипов наиболее часто используют термины:

- проба или зонд (probe), фрагмент ДНК, иммобилизованный на подложке,
- мишень (target) или образец (sample) биологического материала

кДНК-биочипы

Типичный кДНК-биочип состоит из 40000 кДНК-проб длиной 600-2400 н.п.

олигонуклетидные биочипы

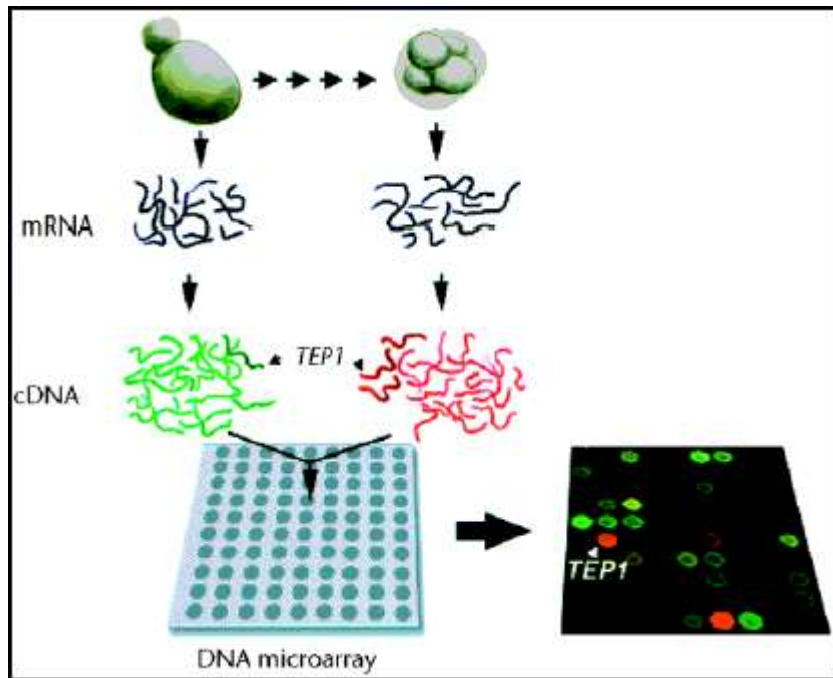
Олигонуклетидные биочипы высокой плотности (high-density oligonucleotide arrays) могут содержать до 500,000 пар проб на одном стекле. Одному гену могут соответствовать 5-15 пар проб.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПРИГОТОВЛЕНИЕ МЕЧЕННОГО ОБРАЗЦА И ГИБРИДИЗАЦИЯ



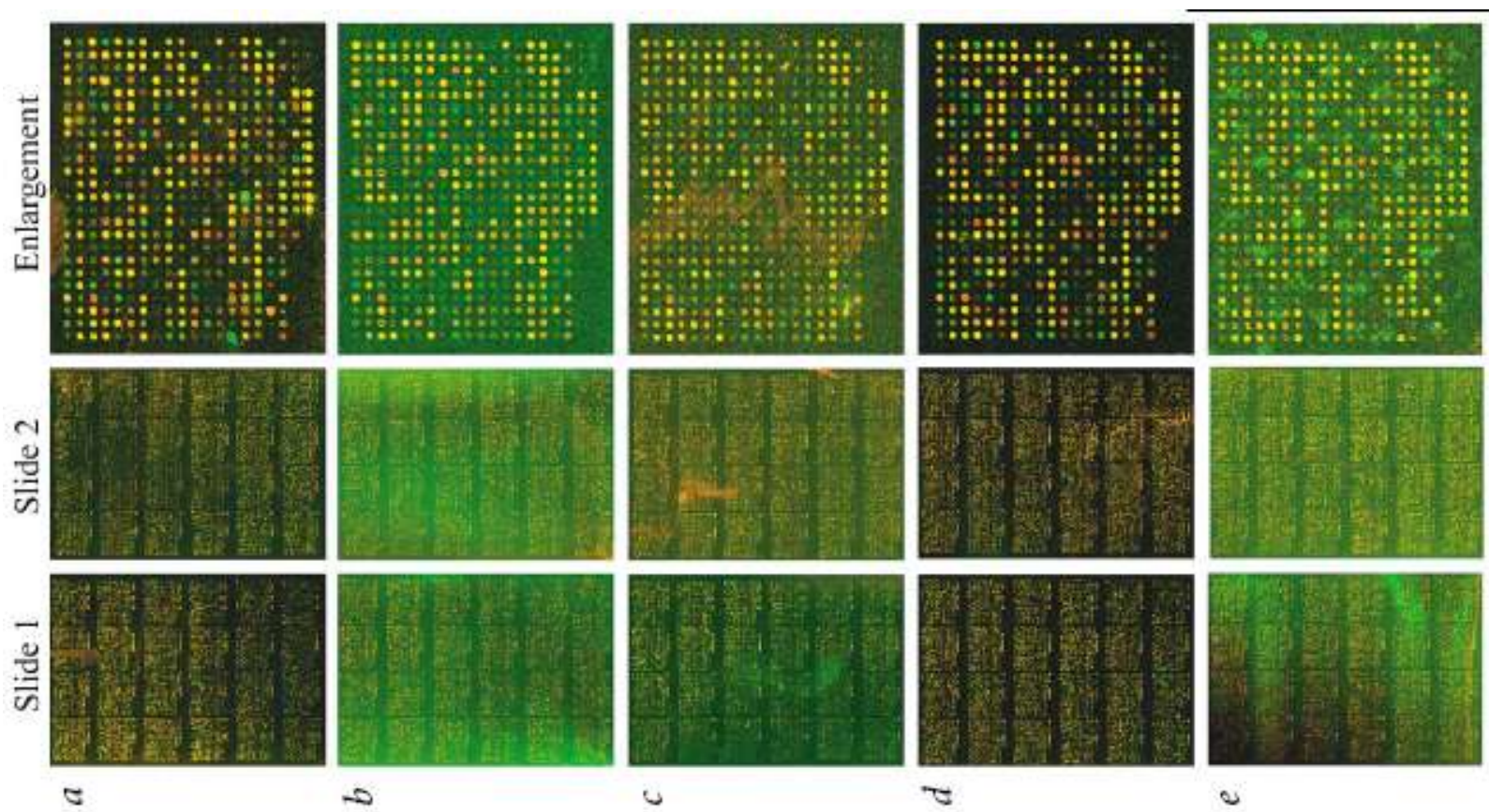
Каждый пул кДНК, приготовленный из двух разных образцов, делится на две равные части молекул, в которые вводятся меченые флуоресцентными группировками нуклеотиды (прямое или не прямое включение). Обычно это зеленый Cyanine3 (Cy3) и красный Cyanine5 (Cy5). Оба меченых образца объединяются для совместной конкурентной гибридизации.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



СКАНИРОВАНИЕ РЕЗУЛЬТАТОВ ГИБРИДИЗАЦИИ И ПОЛУЧЕНИЕ ИЗОБРАЖЕНИЙ





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ИСТОЧНИКИ РАЗБРОСА ХАРАКТЕРИСТИК

СИСТЕМАТИЧЕСКИЕ ОШИБКИ

- Количество образца
- Эффективность:
 - выделения РНК
 - обратной транскрипции
 - введения метки
 - детекции сигнала

Сходный эффект на многие измерения
Коррекция на основе анализа данных



КАЛИБРОВКА

СТОХАСТИЧЕСКИЕ ОШИБКИ

- Успешность ПЦР/Качество ДНК
- Эффективность раскапывания/печати:
 - размер пятен
 - морфология пятен
- Кросс-гибридизация
- Неспецифическая гибридизация

Случайны и не учитываемы, «шум», фон



МОДЕЛИРОВАНИЕ ОШИБКИ



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ОБРАБОТКА ИЗОБРАЖЕНИЙ

Исходные данные, получаемые после сканирования по каждому каналу, – это изображения определенного района поверхности стекла в черно-белой шкале. Обычно это 16-битные TIFF (Tagged Information File Format) изображения.

Они преобразовываются в цифровые данные интенсивности

определения центра пробы (регистрации)

- выделение пикселей картины, относящихся к пробе и не-пробе (сегментация)
- определение значений интенсивности сигнала от пробы (как суммированной величины значений для пикселей каждого сегмента пробы) и определение значений фона (как суммированной величины значений для пикселей каждого сегмента не-пробы) (квантификация)



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ОБРАБОТКА ИЗОБРАЖЕНИЙ

Исследование транскриптома методом ДНК-биочипов основано на предположении, что измеренные в биочипе интенсивности сигнала для каждого гена отражают их относительный уровень экспрессии.

После сопоставления данных об интенсивностях сигналов между пробами на одном биочипе получается статическая информация о дифференциальной экспрессии генов (в какой ткани или типе клеток, на какой стадии, при каком воздействии и т.д.).

После сопоставления данных об интенсивностях сигналов между теми же пробами, полученными в результате отдельных гибридизационных экспериментов, получается динамическая информация об экспрессии генов.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



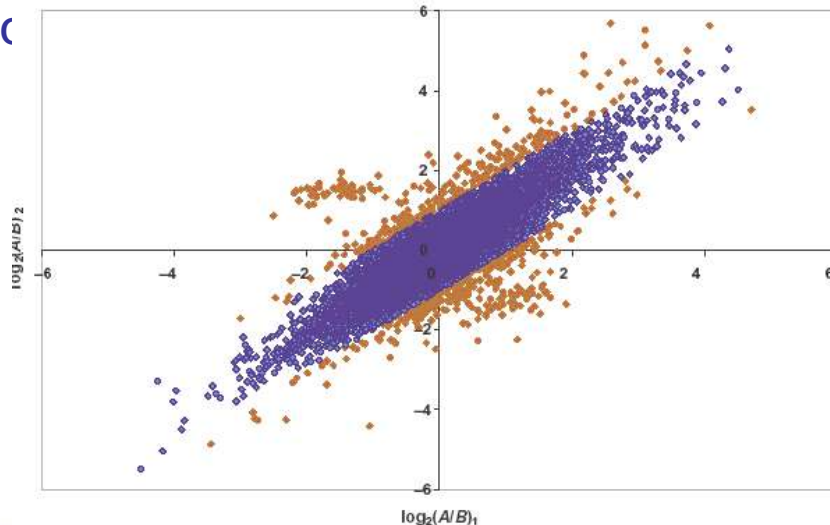
ОБРАБОТКА ИЗОБРАЖЕНИЙ: *Использование реплик*

Биологические реплики – независимо приготовленные меченные образцы, они дают информацию о естественной изменчивости в изучаемой биологической системе, а также случайные различия в процессе приготовления образцов.

Технические реплики – повторы пробы, стёкол, гибридизаций и т.д., но при одном и том же образце, дают информацию о естественных и систематических ошибках методики.

С помощью планирования биочипа и всего эксперимента с помощью технических реплик можно заранее найти способ снижения влияния этих

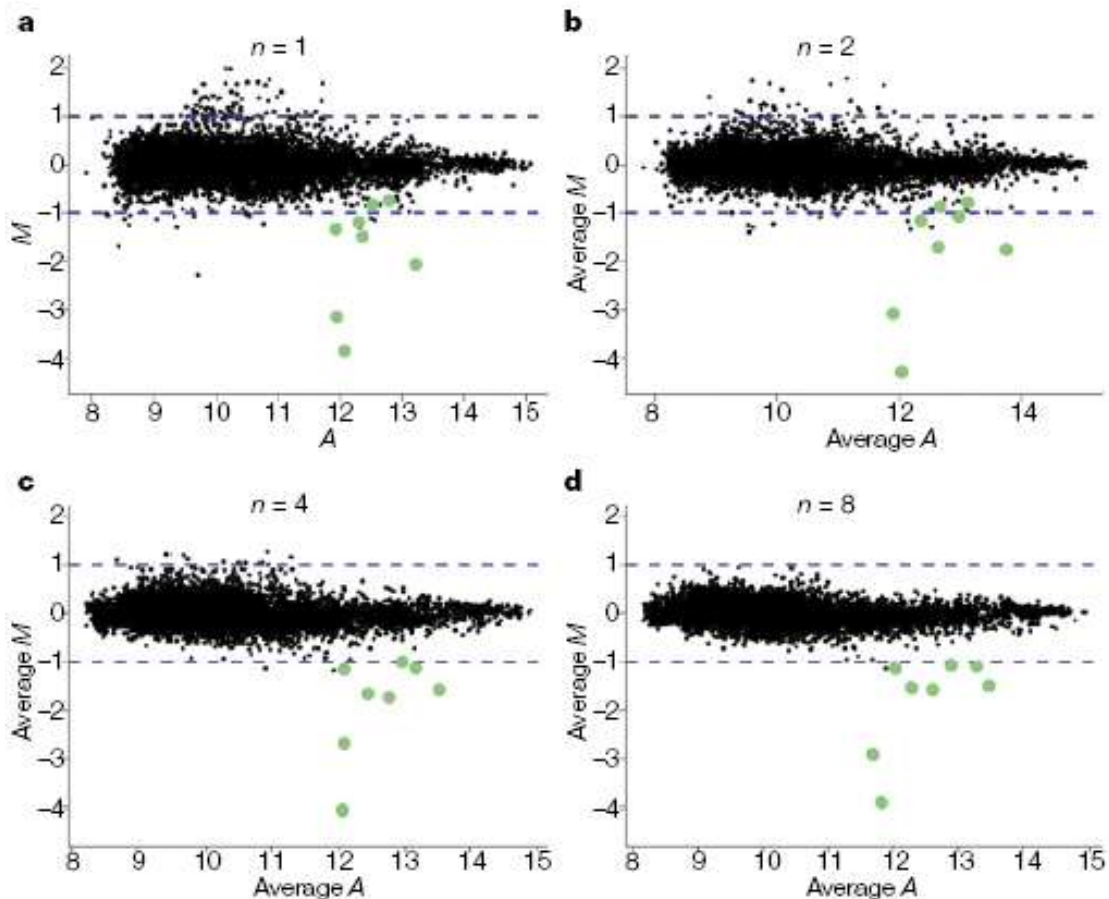
Replicate comparison and trim



Распространенный тип технических реплик – повторная гибридизация с теми же образцами, мечеными наоборот (dye-reversal or flip-dye analysis) и усреднение по репликам.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Биологические реплики.

Гибридизация с образцами РНК от разных особей

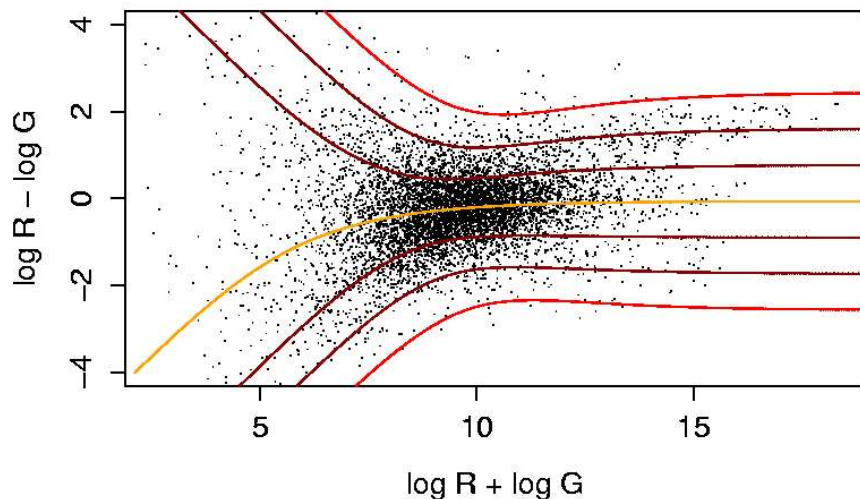


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ОБРАБОТКА ИЗОБРАЖЕНИЙ: *Фильтр по значению дисперсии.*

Используются только те элементы биочипа, интенсивность которых статистически достоверно отличается от фона.



Однако есть опасность потерять биологически значимые различия между генами в области малых интенсивностей.

Вторая опасность – потерять биологически значимые различия между генами в области больших интенсивностей из-за насыщения сигнала (обычно для 16-битного сканнера предел измерения – $2^{16}-1=65,535$ на пиксель).

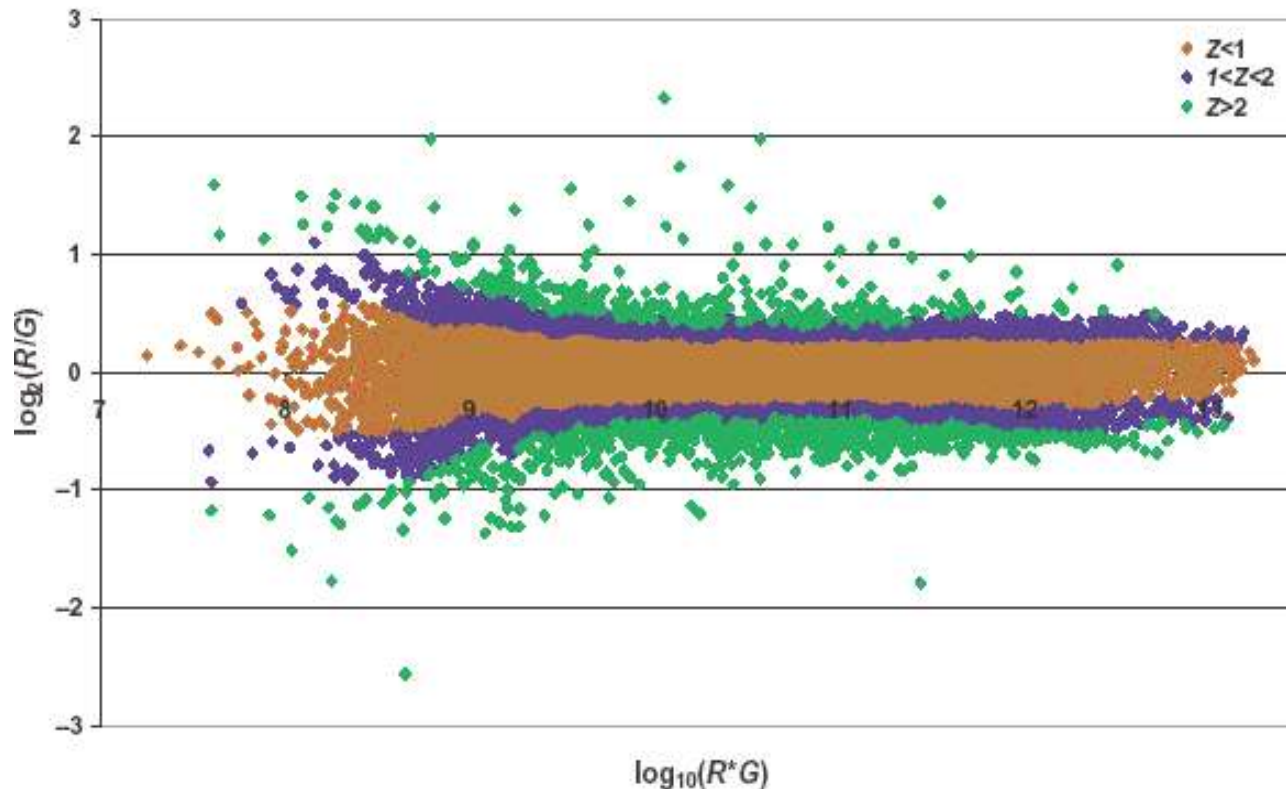


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ОБРАБОТКА ИЗОБРАЖЕНИЙ: *Выявление дифференциально экспрессирующихся генов*

Intensity-dependent Z-scores for identifying differential expression



$$Z_i^{local} = \frac{\log_2(T_i)}{\sigma_{\log_2(T_i)}^{local}}$$



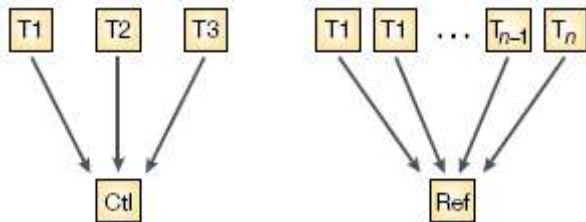
Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПЛАНИРОВАНИЕ (ДИЗАЙН) ЭКСПЕРИМЕНТОВ

Два аспекта планирования:

- (i) Определить какие пробы должны быть, должны ли быть реплики, есть ли возможность для множественных реплик, какие контроли и т.д.
- (ii) Определить расположение проб – дизайн биочипа.



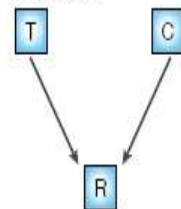
? ??? ?? – ?????? ?? ????????

Обработка T против контроля C.

a Direct



b Indirect



? ?? ?????? ?????????? ?????????? ?????? ?²/2.

? ?? ?????? ?????????? ?????????? ??????

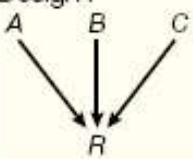
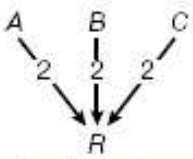
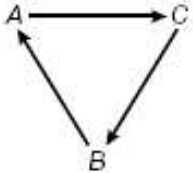
2?²



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПЛАНИРОВАНИЕ (ДИЗАЙН) ЭКСПЕРИМЕНТОВ

Design choices	Number of slides	Units of material (number of samples)	Average variance
Indirect designs			
Design I 	3	$A = B = C = 1$	2.00
Design II 	6	$A = B = C = 2$	1.00
Direct design			
Design III 	3	$A = B = C = 2$	0.67

Однофакторный эксперимент






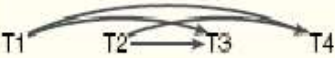
Физические ограничения:
число стекол и количество исходной мРНК



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПЛАНИРОВАНИЕ (ДИЗАЙН) ЭКСПЕРИМЕНТОВ

Design choices	t versus t + 1			Comparisons t versus t + 2		t versus t + 3	Average variance
	t ₁ /t ₂	t ₂ /t ₃	t ₃ /t ₄	t ₁ /t ₃	t ₂ /t ₄	t ₁ /t ₄	
Design I – T1 as common reference 	1.00	2.00	2.00	1.00	2.00	1.00	1.5
Design II – direct: sequential 	1.00	1.00	1.00	2.00	2.00	3.00	1.67
Design III – common reference 	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Design IV – T1 as common reference 	0.67	0.67	1.67	0.67	1.67	1.00	1.06
Design V – direct: loop 	0.75	0.75	0.75	1.00	1.00	0.75	0.83
Design VI – direct: mixed 	1.00	0.75	1.00	0.75	0.75	0.75	0.83

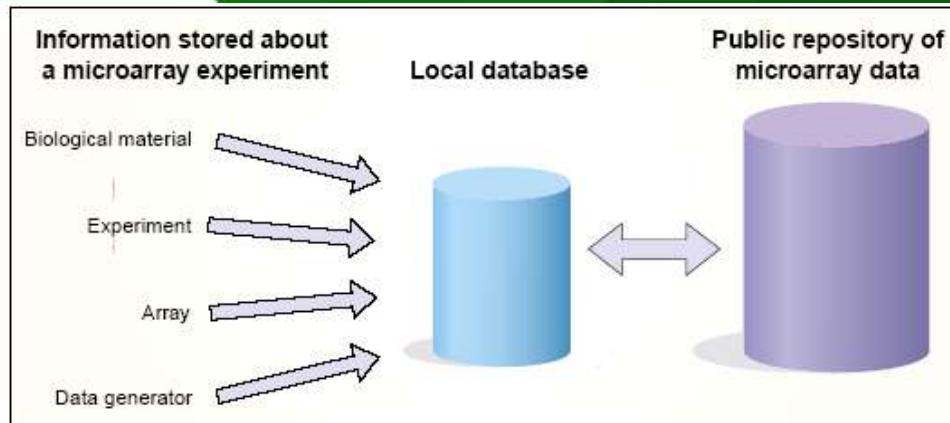
Эксперимент с временными точками

Какова главная цель – выявить различия между всеми состояниями относительно первого

или сравнить развитие экспрессии на всех стадиях



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Общество Microarray Gene Expression Data (MGED) (<http://www.mged.org>), организовано для установления общих стандартов описания данных по биочип-экспериментам, систем обработки, передачи и хранения данных в публичных базах данных.

- MicroArray Gene Expression Markup Language (MAGE-ML) – создан для создания общего формата, чтобы достичь сравнимости результатов
- Minimum Information About a Microarray Experiment (MIAME) – создан для определения типа информации и степени подробности, с которой исследователь обязан ее представить;
- MGED Society Ontology Working Group (<http://www.mged.org/ontology>) – создана для формирования набора контролируемых словарей и онтологий, необходимых



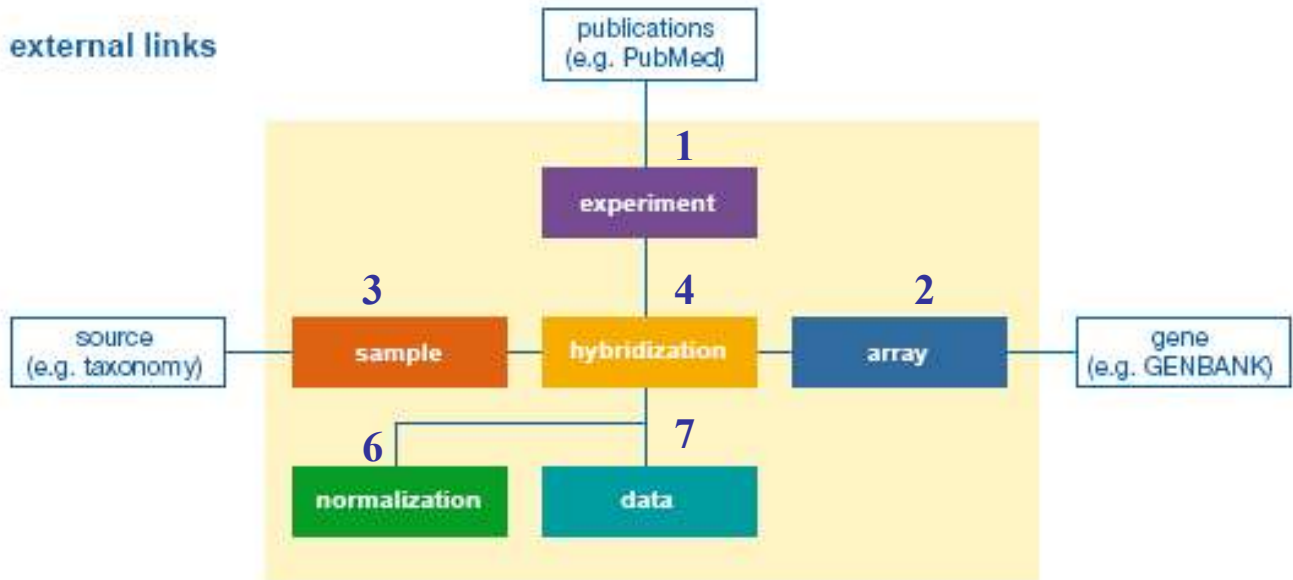
Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Minimum Information About a Microarray Experiment (MIAME)

Минимальная информация об опубликованном эксперименте, основанном на ДНК-биочип-методе, включает шесть типов описаний:

3. План эксперимента – набор отдельных гибридизационных экспериментов
4. План биочипа – содержание пятен/ячеек, компоновка по рядам и т.д.
5. Образцы – источник, приготовление экстрактов, способ мечения
- 6.
- 7.
- 8.



1В



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПРИМЕНЕНИЕ ДНК-БИОЧИПОВ В БИОЛОГИИ

АНАЛИЗ ПОЛИМОРФИЗМА

АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ГЕНОМОВ

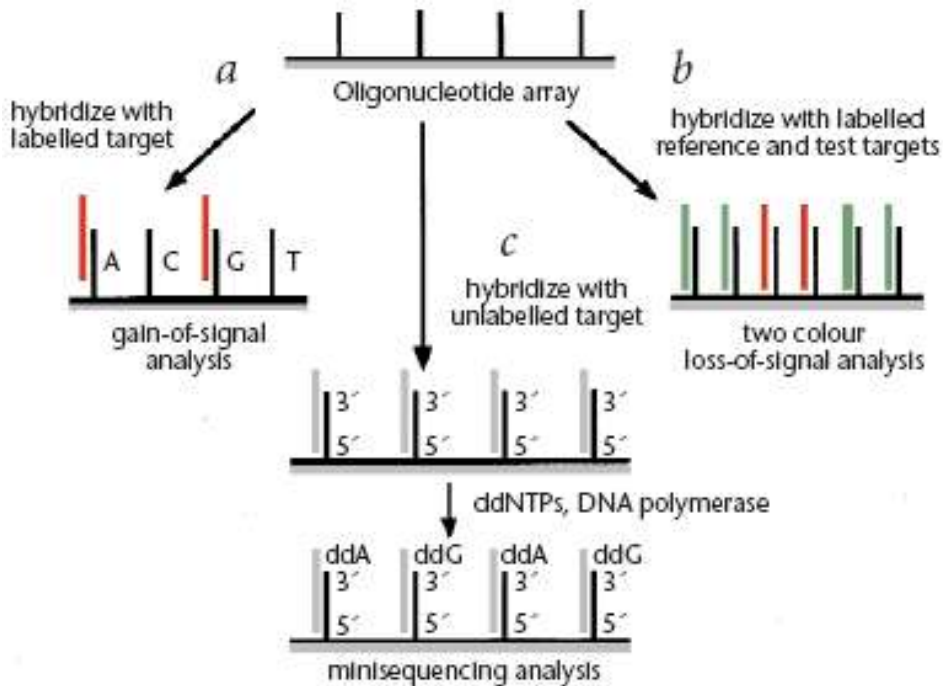


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ПОЛИМОРФИЗМА В ГЕНОМНОЙ ДНК

Мутационный анализ и ресеквенирование с помощью олигонуклеотидных биочипов.



Миллионы нуклеотидных позиций, варьируемых у разных особей, могут быть скринированы с помощью специально разработанных биочипов. SNP-биочипы используются для исследования (1) сцепления между маркерами, (2) неравновесия по сцеплению (linkage disequilibrium), (3) потери гетерозиготности (loss of heterozygosity).

Три экспериментальные стратегии для анализа последовательностей ДНК с помощью биочипов.

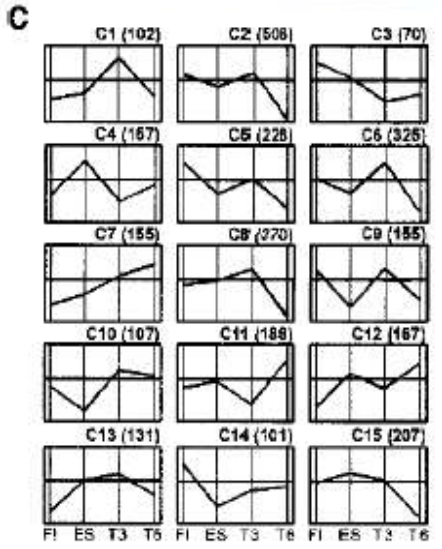
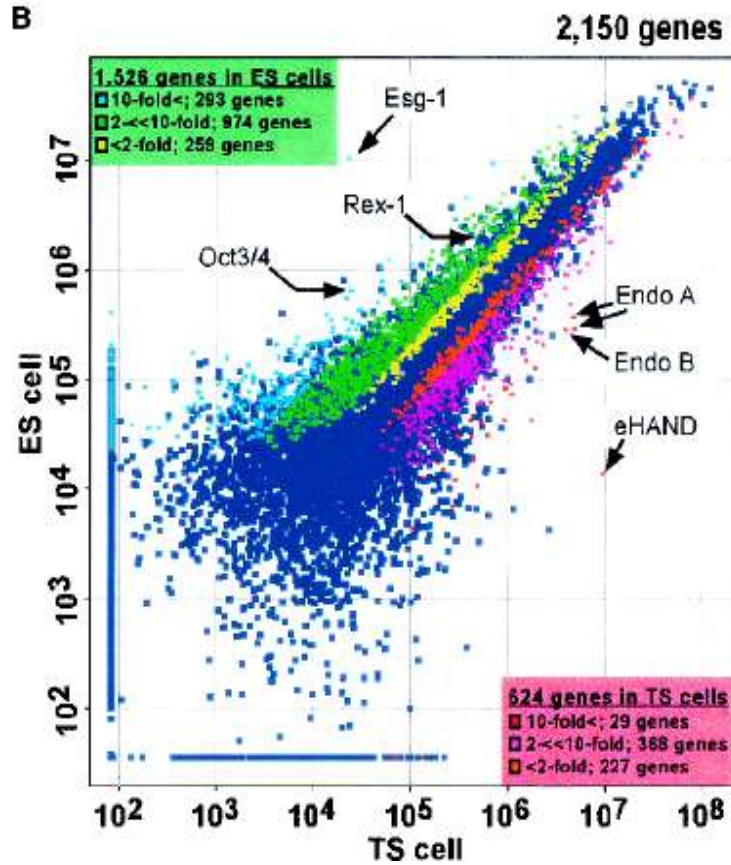
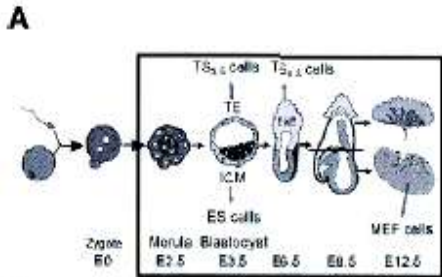


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Молекулярный паспорт стволовых клеток.



Образцы четырёх клеточных линий:
 ES (эмбрионально-стволовые),
 TS_{3,5} и TS_{6,5} (трофобласт-стволовые),
 MEF (мышинные эмбриональные фибробласты).
 Мышиный кДНК-биочип NIA 15К.

Гены, экспрессия которых достоверно ($P < 0.05$)
 ?????? ???? ? ??????
 ?????????, ?????????? ???????.

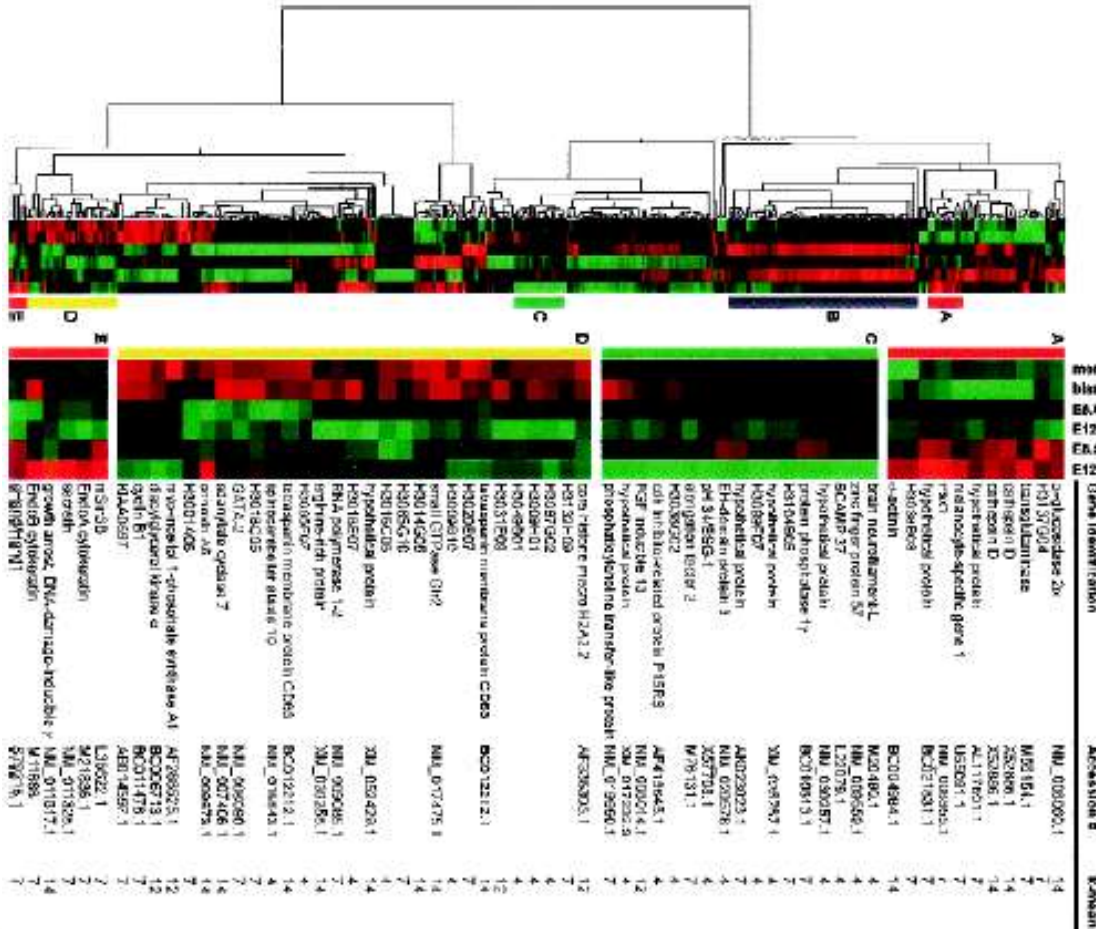
Кластер-анализ методом «к-средних» дифференциально экспрессирующихся генов: 15 кластеров



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ Молекулярный паспорт стволовых клеток.



Иерархическая кластеризация 346 генов, специфично экспрессирующихся в образцах.
 ???????? A ? E - TS ?????; B - MEF; C – ?????? ? ?????? ?????? ??????????; D – ES ??????.

??????? 4 ?????????????? ES
 ?????? ?????? ??????, 7 - TS
 ?????? ??????, 14 - MEF-
 ?????? ??????, 12 – ??? ?
 ??? ES ? TS ??????.



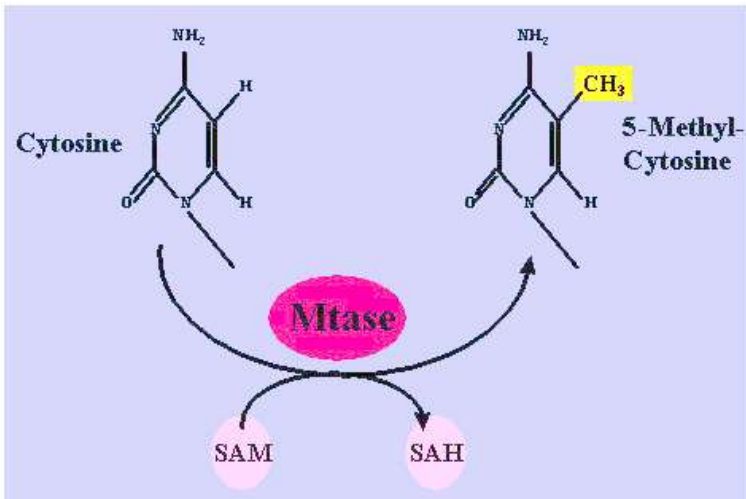


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между экспрессией генов и эпигенетическим состоянием геномных районов вокруг генов.



Метилирование ДНК – один из эпигенетических механизмов. Основной мишенью для метилирования в геноме млекопитающих является цитозин. Чаще всего метилирование происходит в контексте динуклеотидов CpG, хотя CpNG, CC(a/t)GG, CpA и CpT также могут быть метилированы. «CpG-островки» охватывают промоторы и первые экзоны генов.

Метилированное состояние «CpG-островков» часто ведет к подавлению экспрессии генов.

Распределение метилированных сайтов ДНК в нормальных и трансформированных клетках значительно различается – в опухолевых клетках наблюдается гипометилирование одних и гиперметилирование других.

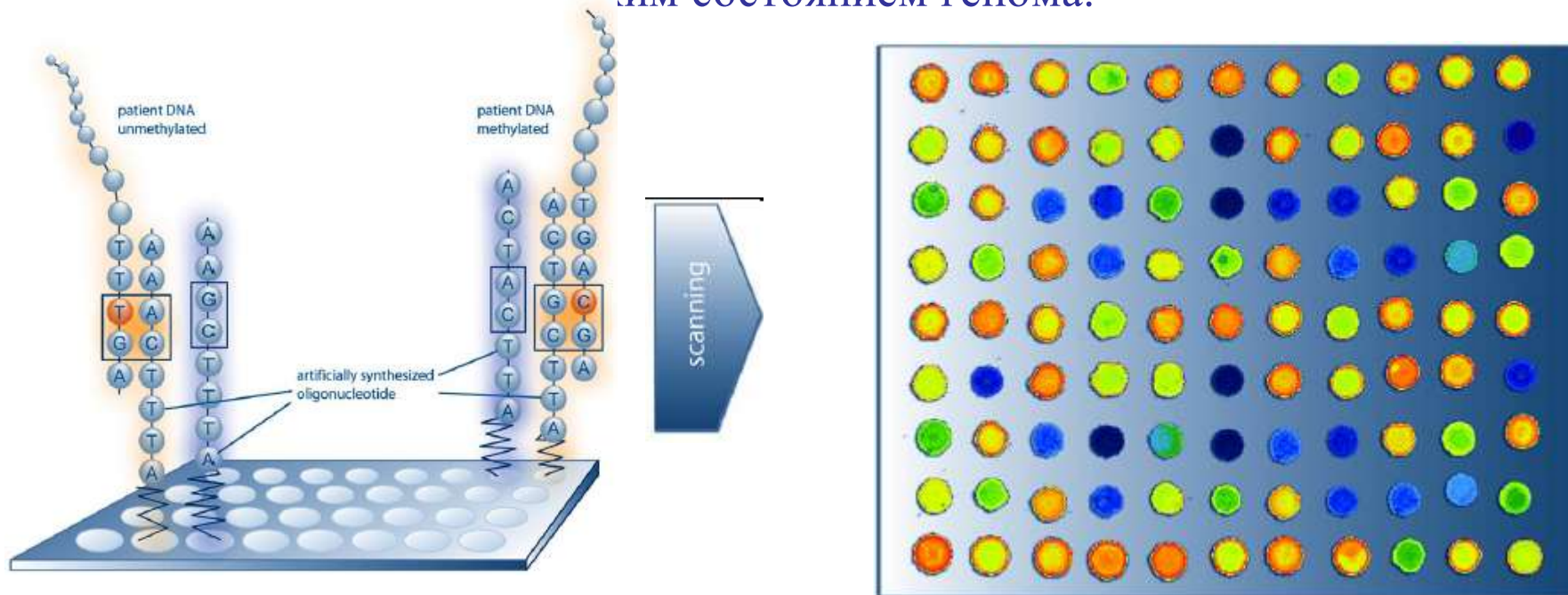


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между экспрессией генов и эпигенетическим состоянием генома.



Образцы из фрагментов геномной ДНК обрабатываются бисульфитом натрия, превращающего неметилированные цитозины в урацилы, затем метятся в процессе амплификации и гибридизуются с олигонуклеотидными биочипами. Для каждого потенциального сайта метилирования разрабатывается пара олигонуклеотидов.

Novik K.L., *et al.*, Epigenomics: genome-wide study of methylation phenomena.

Curr Issues Mol Biol. 2002 4(4):111-128.

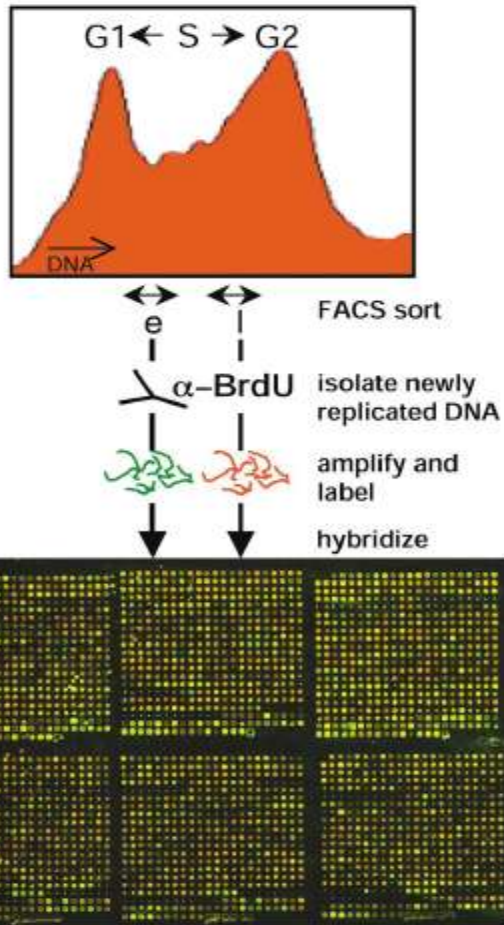


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между экспрессией генов и временем репликации соответствующих районов генома.



- Профиль клеточного цикла, выявляемый после импульсного введения в клетки BrdU, окрашивания клеток пропидиум иодидом (ДНК-специфическим красителем)
- Сортировка клеток по содержанию ДНК с применением FACS (fluorescence-activated cell sorting).
- Иммунопреципитация новосинтезированной ДНК антителами против BrdU, амплификация и введение флюорометки.
- Гибридизация с кДНК-биочипом, содержащим 6500 генов. Три биологические реплики.

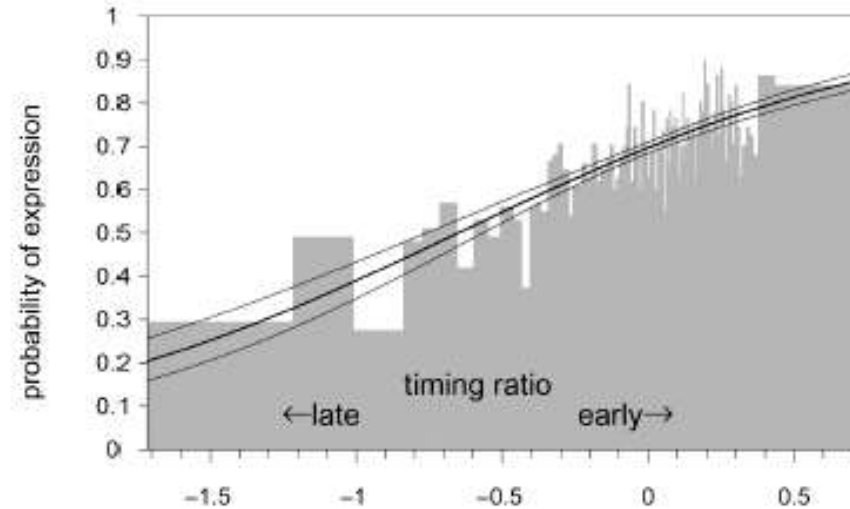
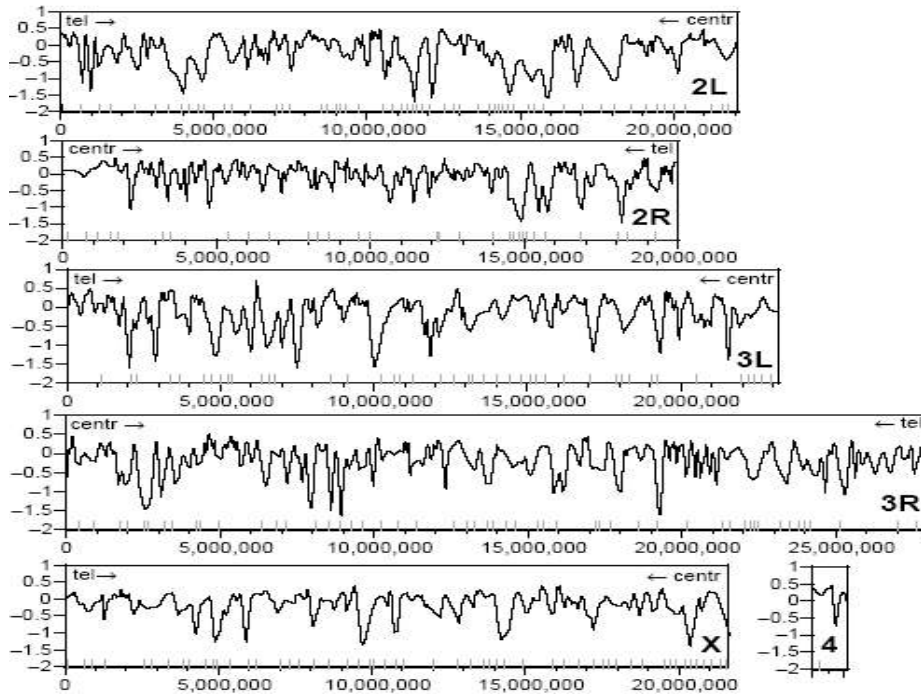


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между экспрессией генов и временем репликации генома.



Репликационный профиль для 6500 генов секвенированной части генома *D. melanogaster*.

Корреляция между транскрипционной активностью гена и временем его репликации.

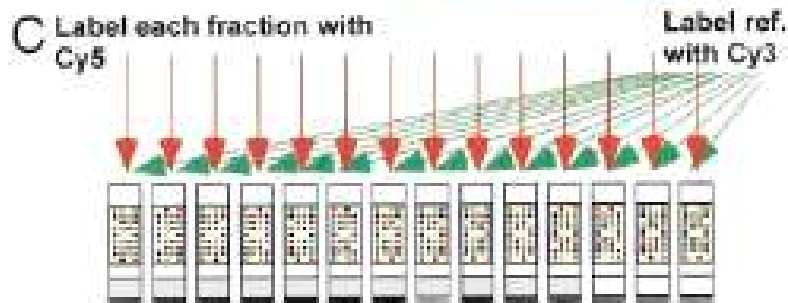
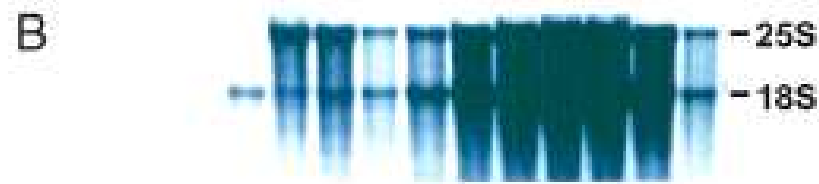
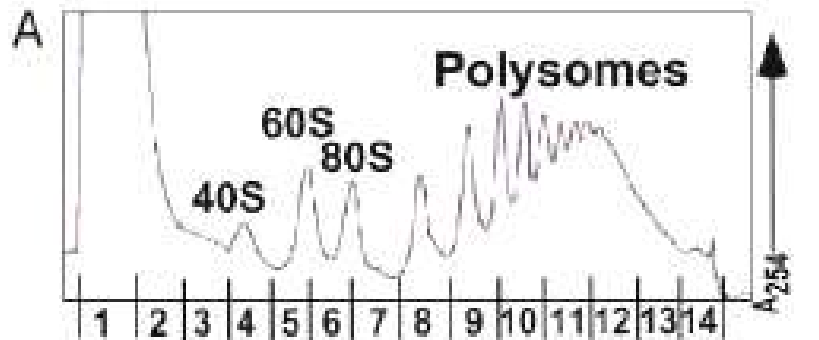


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между транскрипционной и трансляционной активностями генов.



Чем больше уровень трансляции мРНК, тем больше на ней полисом.

- Экспериментальная процедура получения полисомных профилей – седиментация РНК в сахарозном градиенте.
- Отбор образцов из разных фракций.
- Обратная транскрипция и мечение образцов.

- Гибридизация с ДНК-биочипами, содержащими все открытые рамки считывания

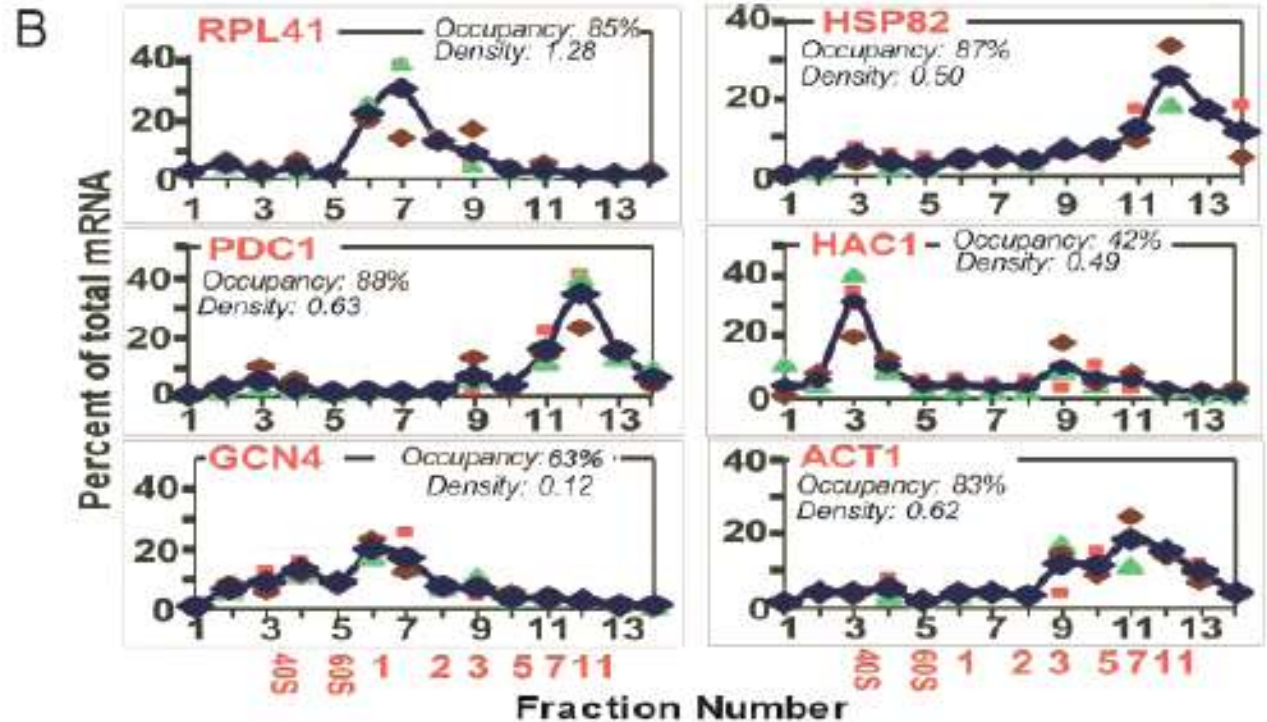
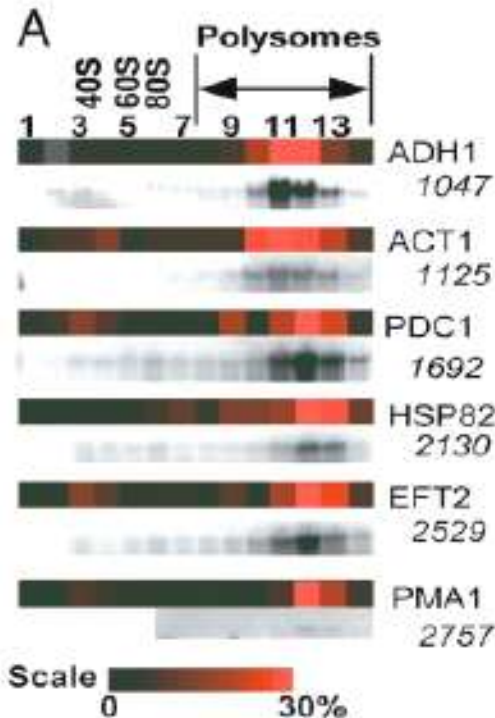
Saccharomyces cerevisiae.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между транскрипционной и трансляционной активностями генов.



Полисомные профили *Saccharomyces cerevisiae* и их проверка методом нозерн-блот-гибридизации.

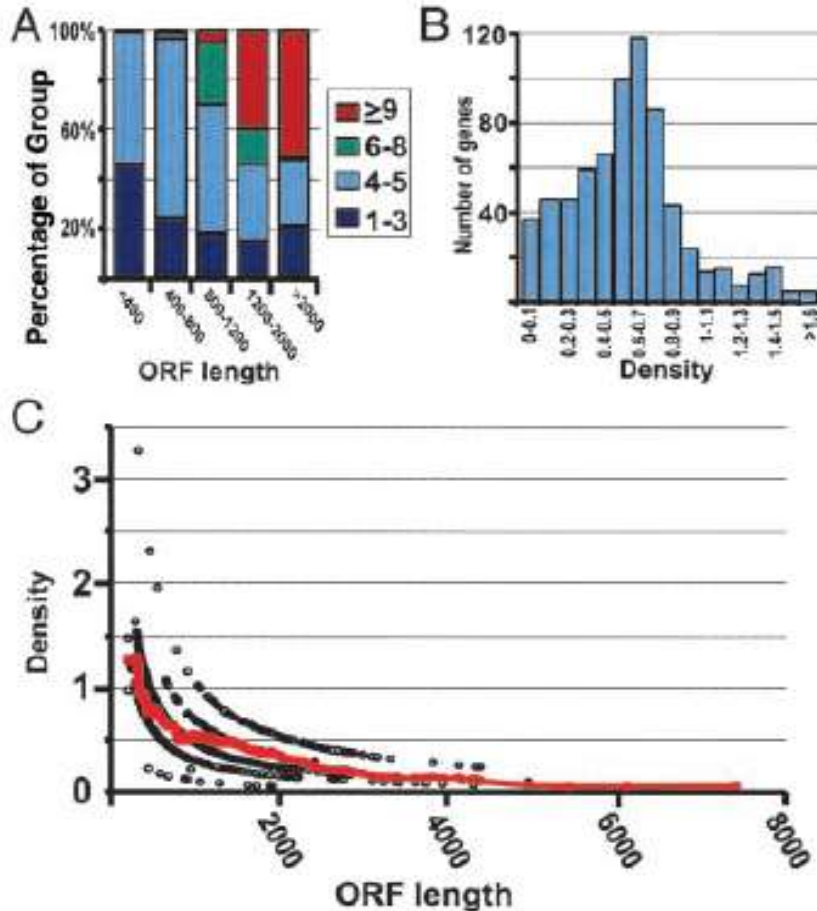


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование связи между транскрипционной и трансляционной активностями генов.



- Число и плотность рибосом, связанных с мРНК:
- А. Гены сгруппированы соответственно длине их ОРС
- В. Число генов как функция плотности рибосом
- С. Плотность рибосом как функция от длины ОРС

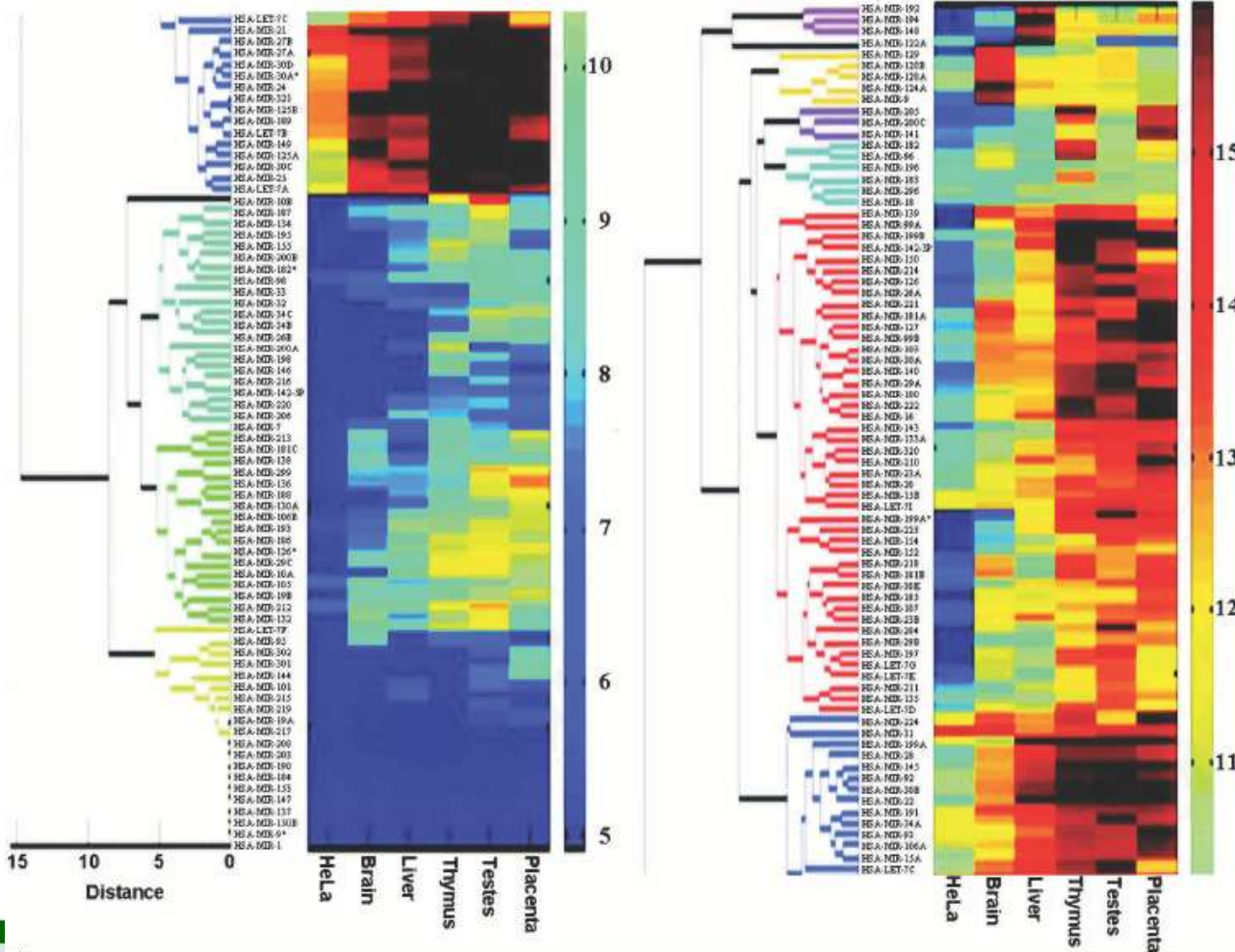


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Исследование профиля экспрессии миРНК.



Профиль экспрессии 150 миРНК человека в пяти органах и клеточной культуре.

Шкала интенсивности представляет \log_2 значений интенсивности сигнала после вычитания фона и нормализации.

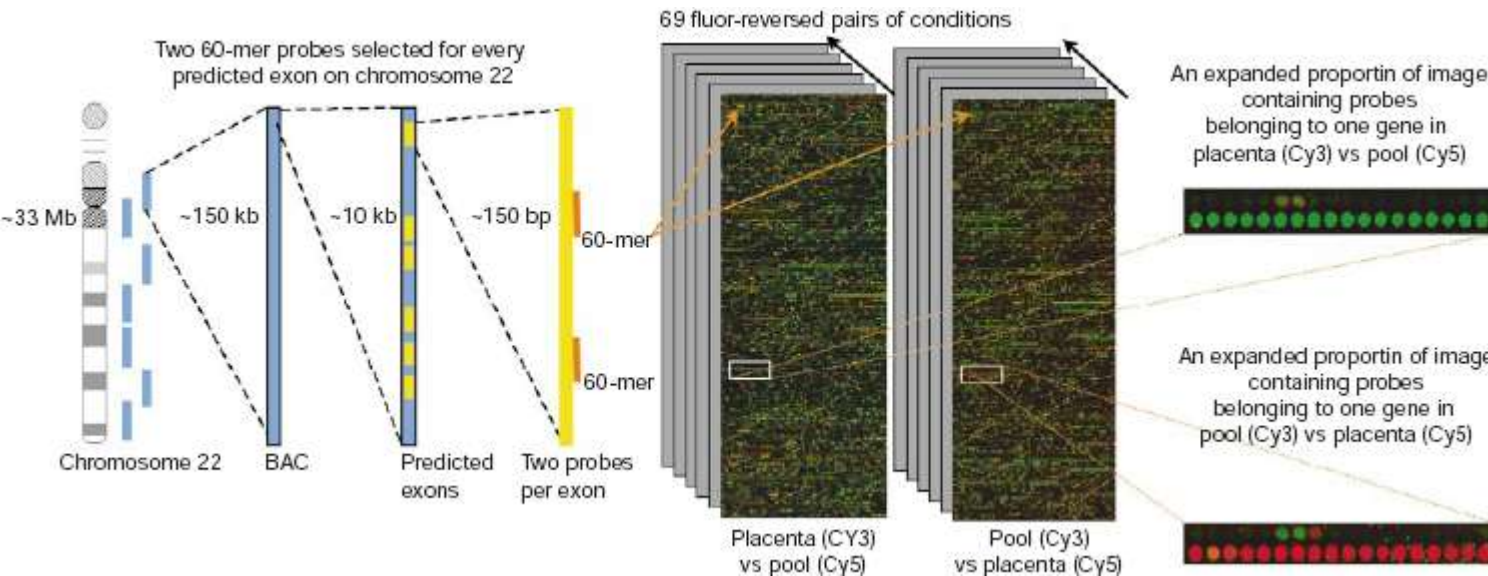


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Экспериментальная аннотация человеческого генома - оценка компьютерных предсказаний наличия генов и определение полноразмерных транскриптов с помощью выявления одновременной экспрессии их экзонов.

Метод применен для анализа экспрессии генов из 22 хромосомы (8,183 экзонов) в 69 парах экспериментальных условий: 572 группы корегулируемых экзонов или 572

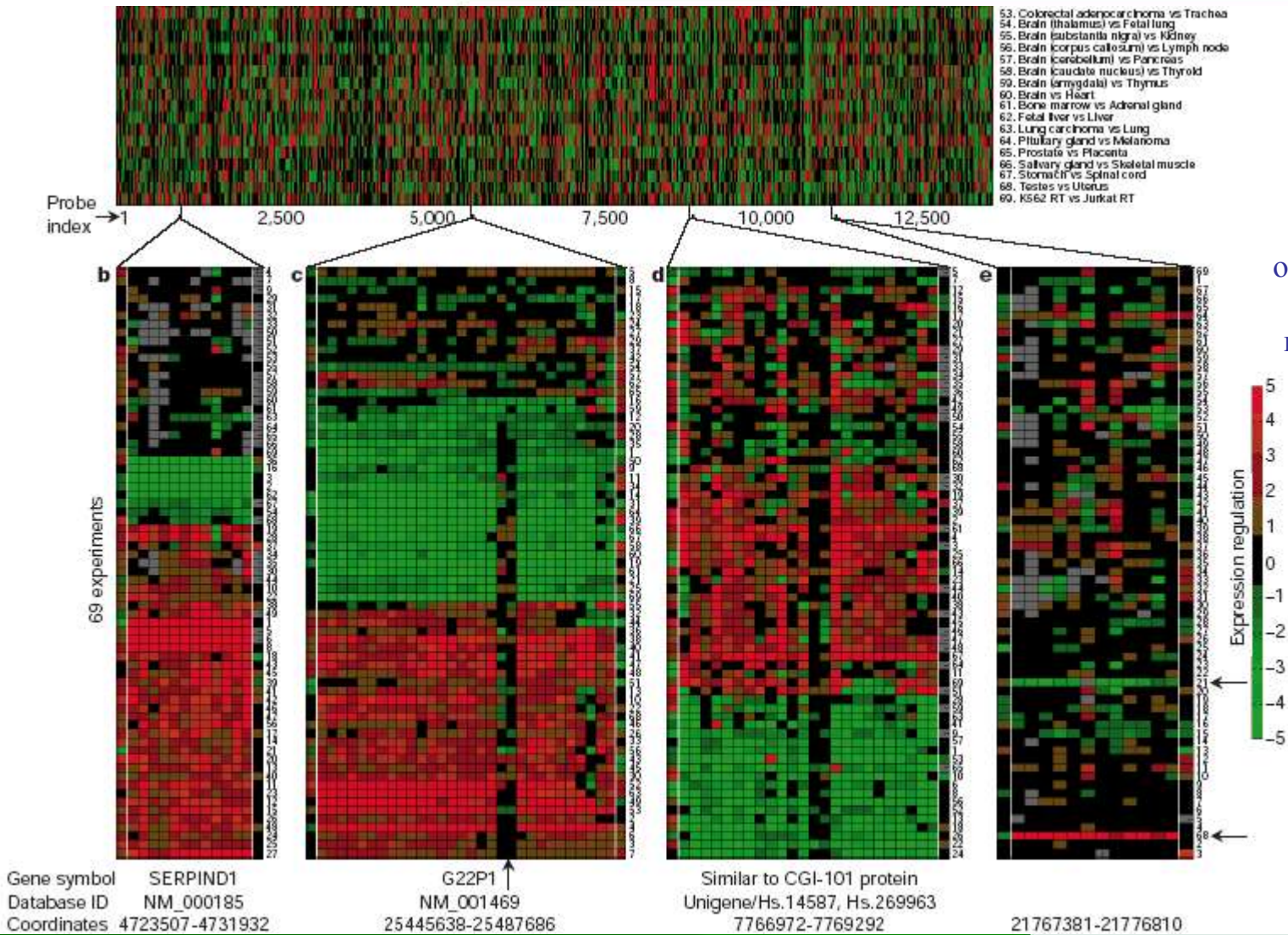


проверенных по экспрессии гена (expression-verified genes - EVGs).

210 (85%) из 247 известных генов и 185 (57%) из 325 предсказанных генов.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

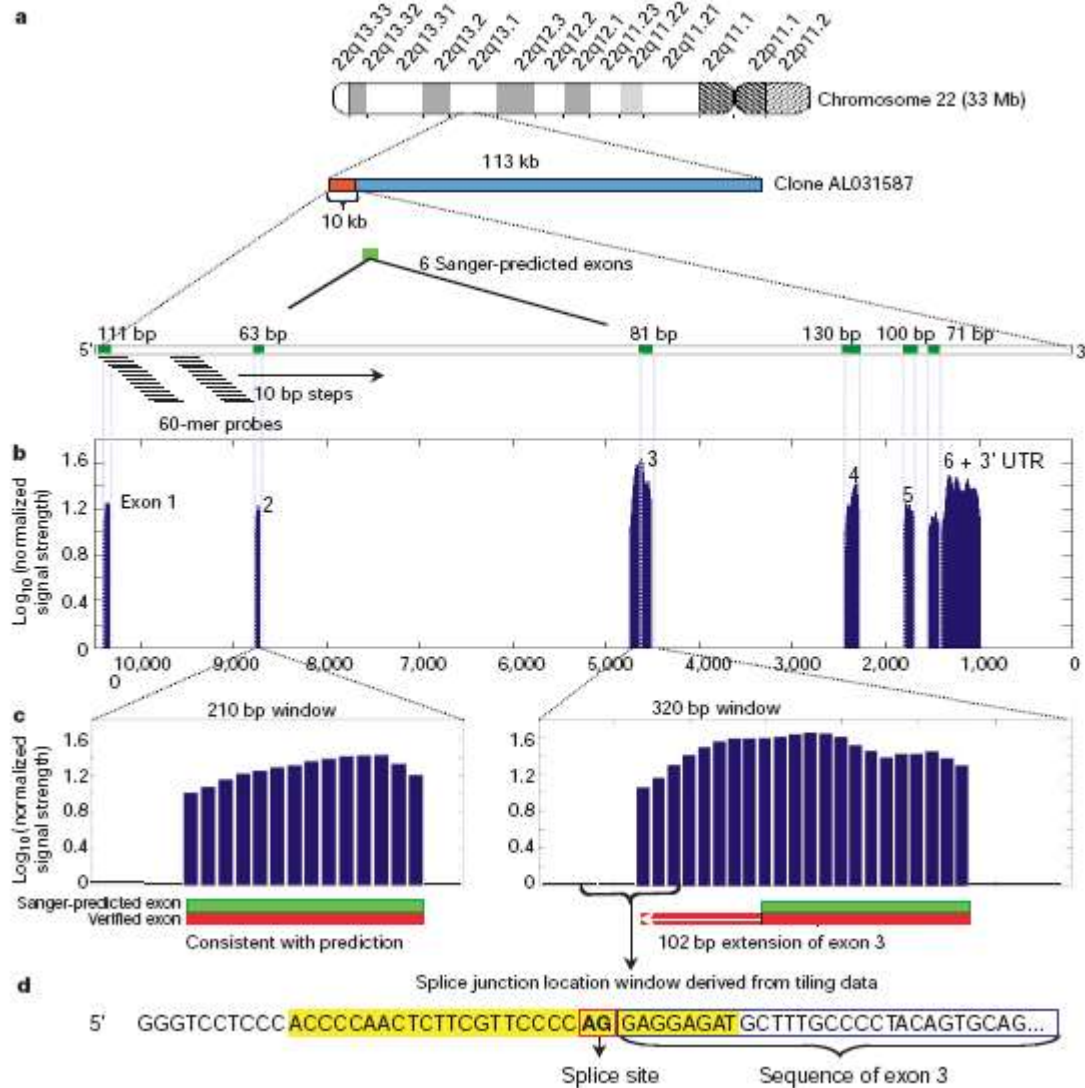


Псевдоцветное
 изображение
 взвешенного по
 ошибкам \log_{10}
 отношения (кр/зел)
 экспрессии
 каждого из ~8000
 экзонов.

Shoemaker DD, *et al.*, Experimental annotation of the human genome using microarray technology. Nature. 2001 409:922-7.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

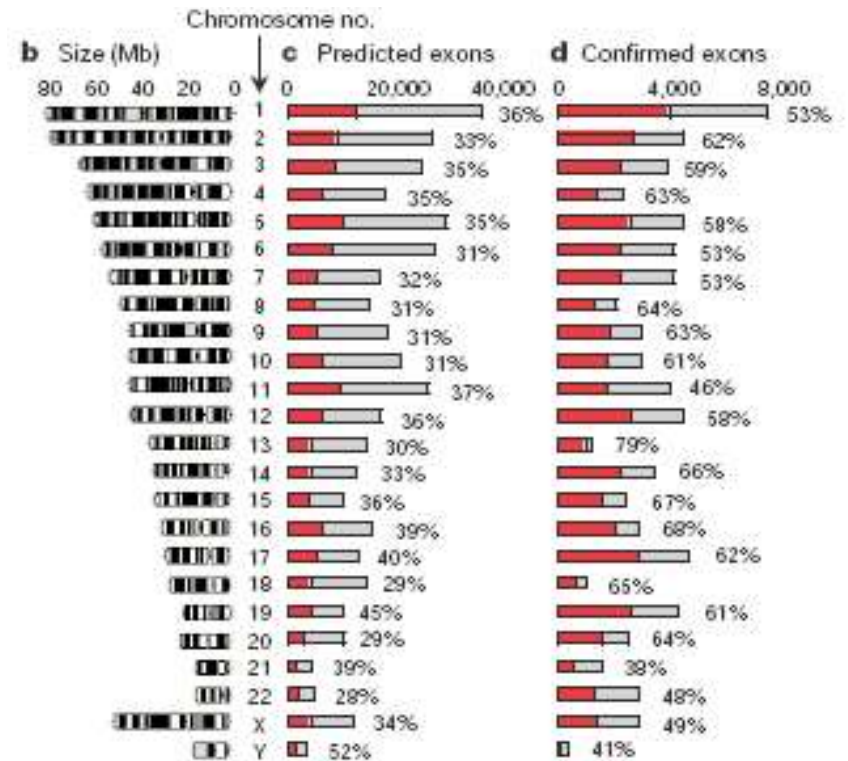
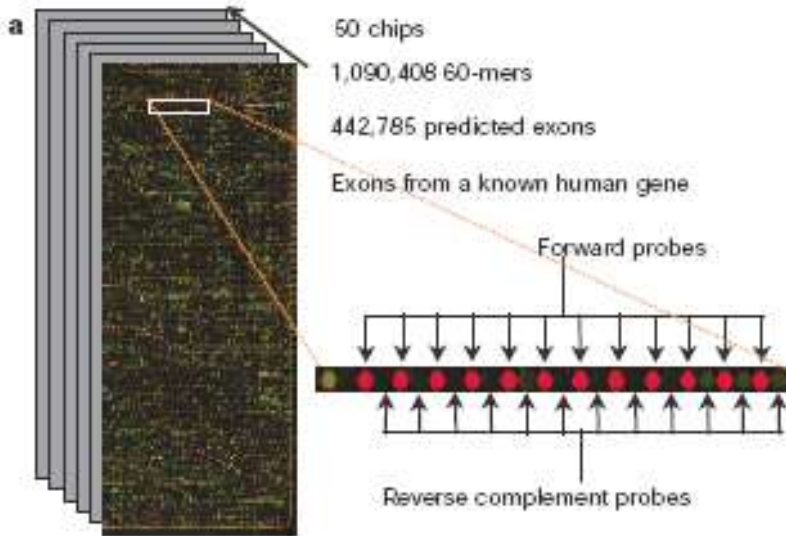


Перекрывающиеся ряды проб, соответствующих обеим цепям отдельных геномных районов хромосомы 22.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов

Сканирование экзонов в масштабах всего генома: 1 090 408 проб, 110 000 обратнo-комплементарных проб, 50 биочипов, два экспериментальных условия (РНК из двух клеточных линий). Красные полосы – доля экспериментально (для этих условий) проверенных экзонов





Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



ПРИМЕНЕНИЕ ДНК-БИОЧИПОВ В МЕДИЦИНЕ

ВХОД

ВЫХОД



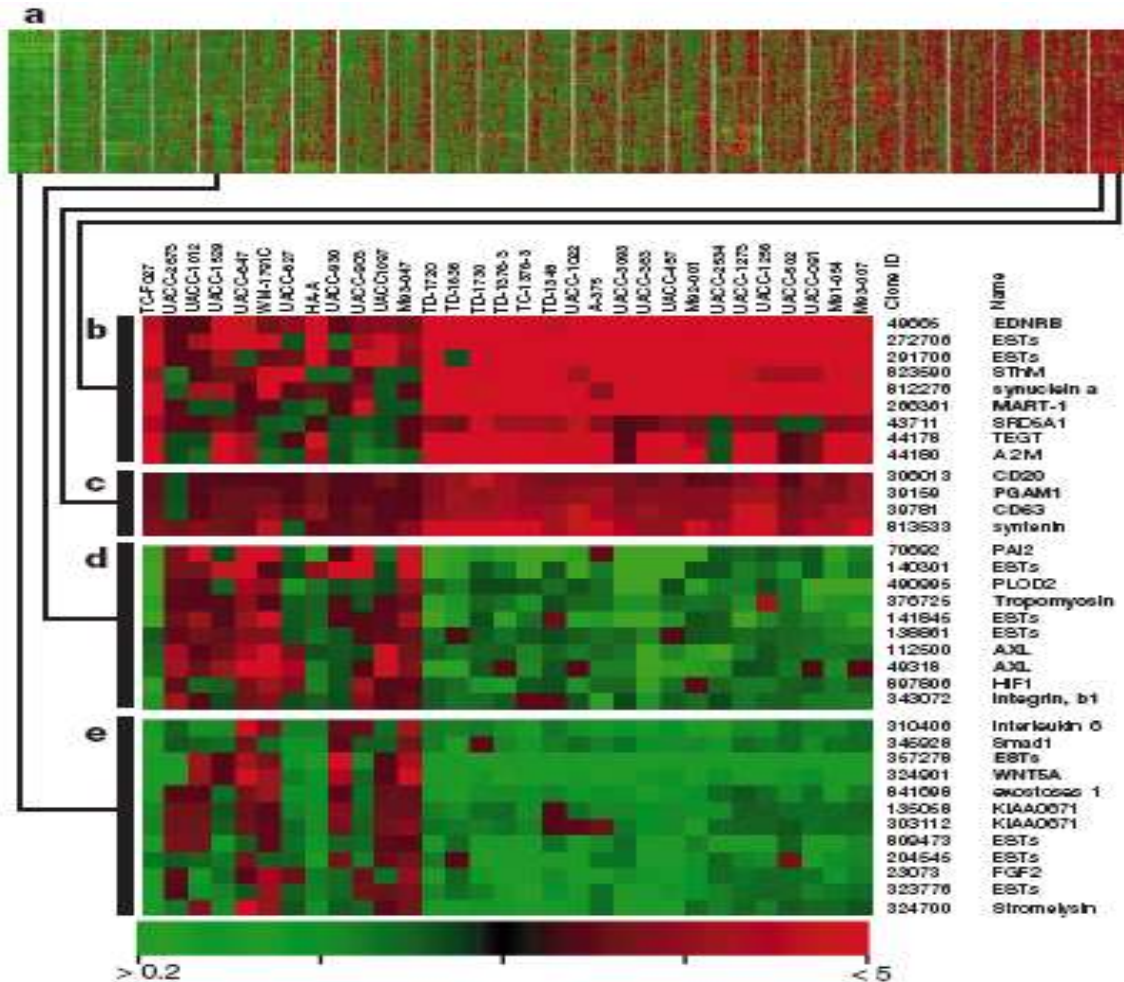


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Молекулярная классификация злокачественных опухолей.



- а. Селекция кластеров меланом
- б. Двумерный кластерный анализ образцов опухолей и генов;
- с. Четыре самых контрастных по экспрессии генов кластера и характеристические гены в них.



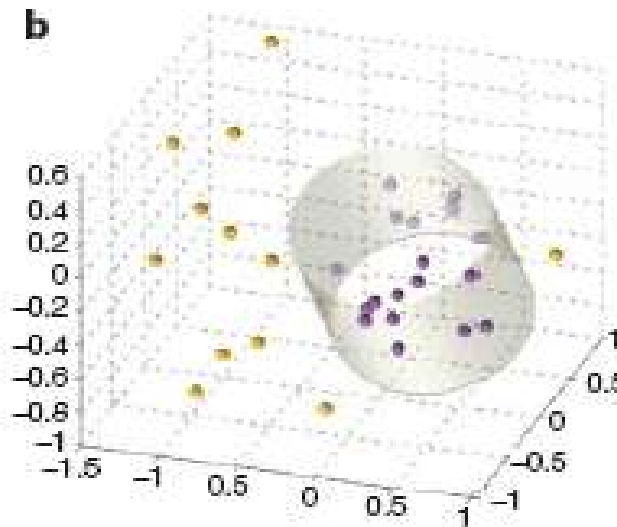
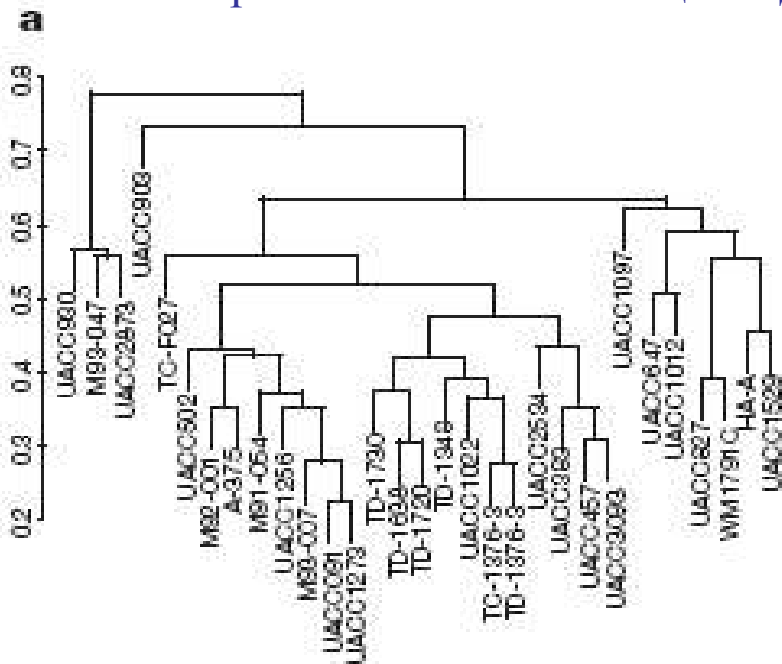
Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Молекулярная классификация злокачественных опухолей.

Таксономия меланом на основе данных профилирования экспрессии генов с помощью кДНК-биочипа (6,971 генов).



а. Дендрограмма иерархической кластеризации 19 меланом по данным экспрессии генов;

б. Трехмерный график значений MDS для 31 меланомы. Показан кластер из 19 образцов.

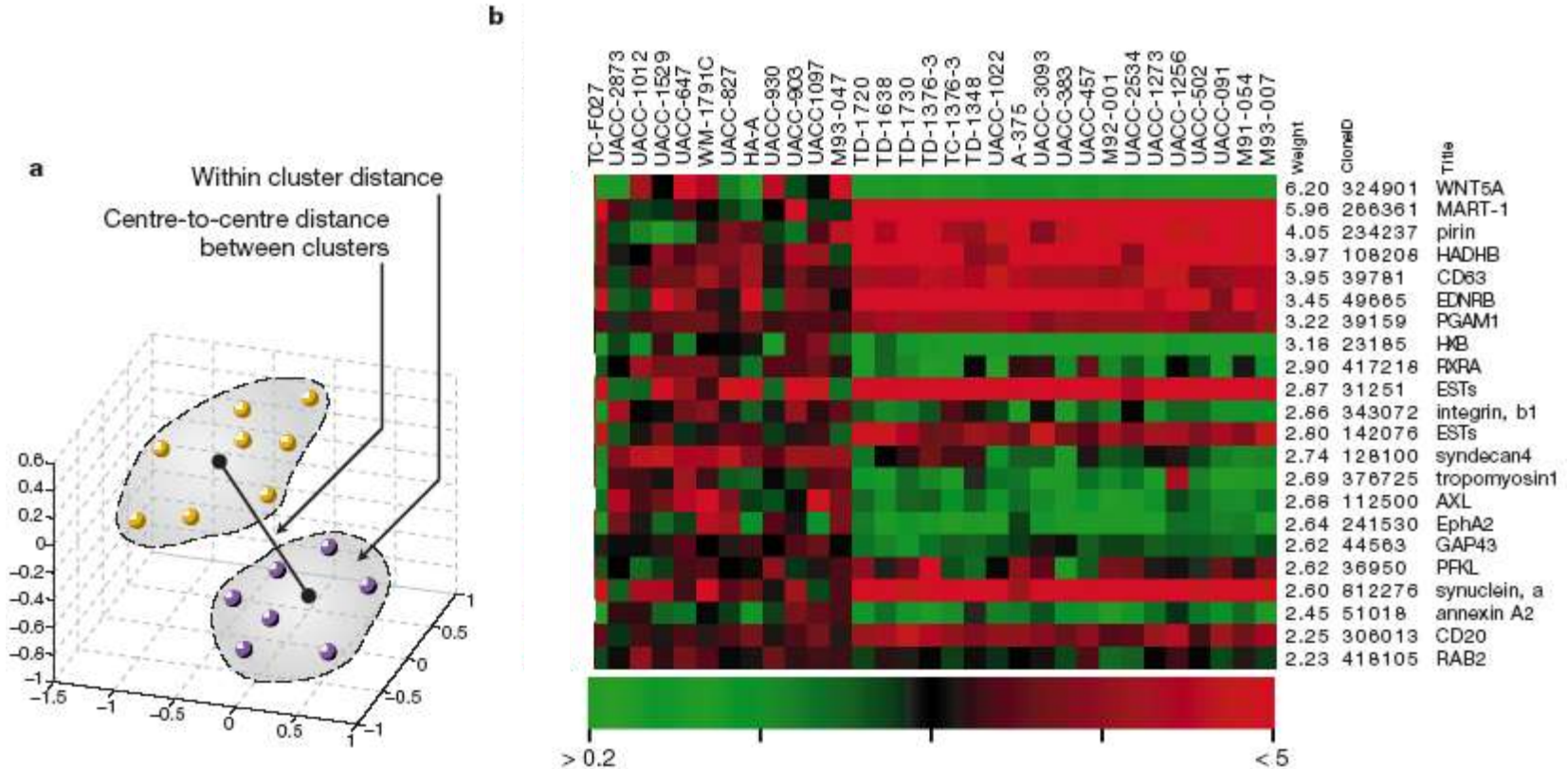


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Молекулярная классификация злокачественных опухолей.



Выявление генов, экспрессия которых позволяет различать кластеры меланом
 а. MDS-анализ распределения генов по их вкладу в минимизацию объема кластера и максимизацию межкластерной дистанции; б. Экспрессия 22 самых заметных генов в кластерах.

Bittner M. *et al.*, Molecular classification of cutaneous malignant melanoma by gene expression profiling.

Nature. 2000 406(6795):536-540.

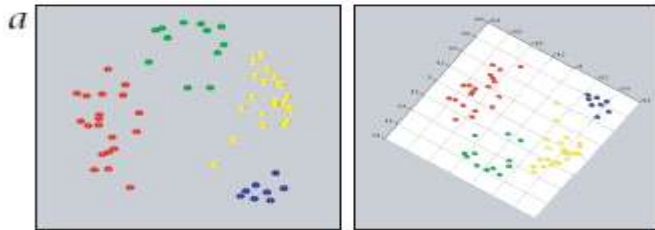


Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



АНАЛИЗ ЭКСПРЕССИИ ГЕНОВ

Молекулярная классификация злокачественных опухолей.



Таксономия нейробластом (NB), рабдомиосарком (RMS), лимфом Ходжкина (NHL) и опухолей семейства Эуинга (EWS) на основе данных профилирования экспрессии генов с помощью кДНК-биочипа (3789 кДНК и 2778 EST).

Анализ экспрессии генов в образцах:
b. две проекции графика значений MDS,
c. иерархическая кластеризация данных по экспрессии генов, выявление характеристических генов.



**Некоторые аспекты
статистического анализа
данных ДНК-чипов**

ОБРАБОТКА ИЗОБРАЖЕНИЙ

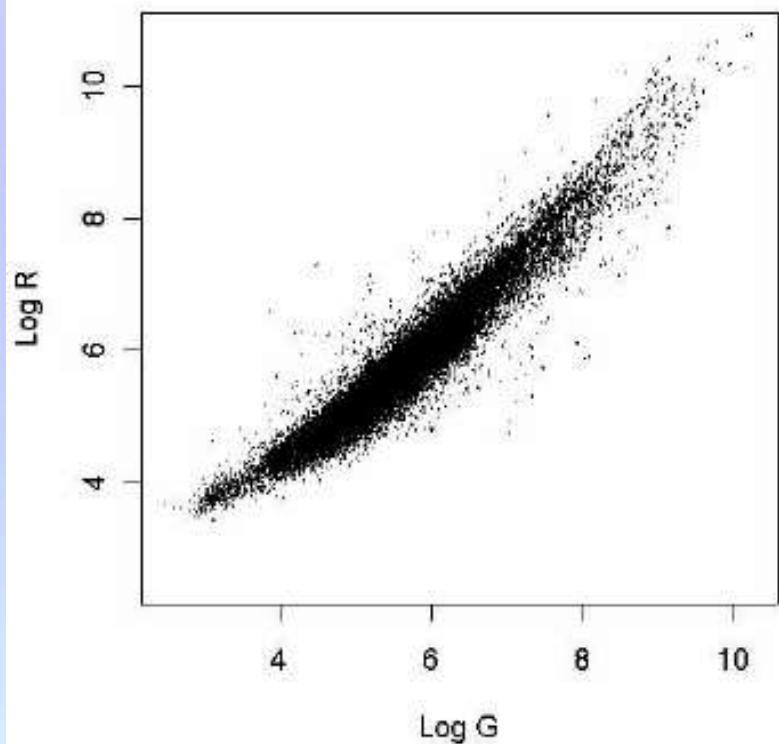
Принято выражать экспрессию гена как отношение между интенсивностью сигнала от исследуемого гена, выявленного, например, Су5-меченным образцом (red), и интенсивностью сигнала от контрольного гена, выявленного, соответственно, Су3-меченным образцом (green): $T_i = R_i / G_i$.

Гены с повышенной в два раза экспрессией будут иметь отношение 2, а с пониженной в два раза – 0.5. Т.е различие будет несимметрично.

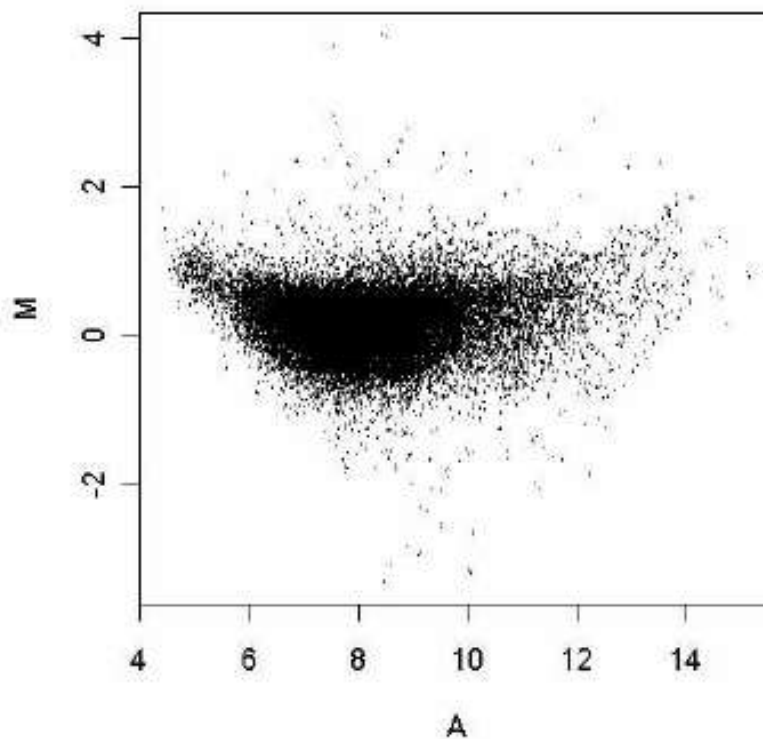
Поэтому применяют преобразование отношения в виде логарифма по основанию 2 : $M = \log_2(R_i / G_i)$. При этом понижение интенсивности в 2 раза и повышение в 2 раза имеют симметричны и равны -1 и +1 соответственно.

Графическое представление СИГНАЛОВ

Способ 1



Способ 2



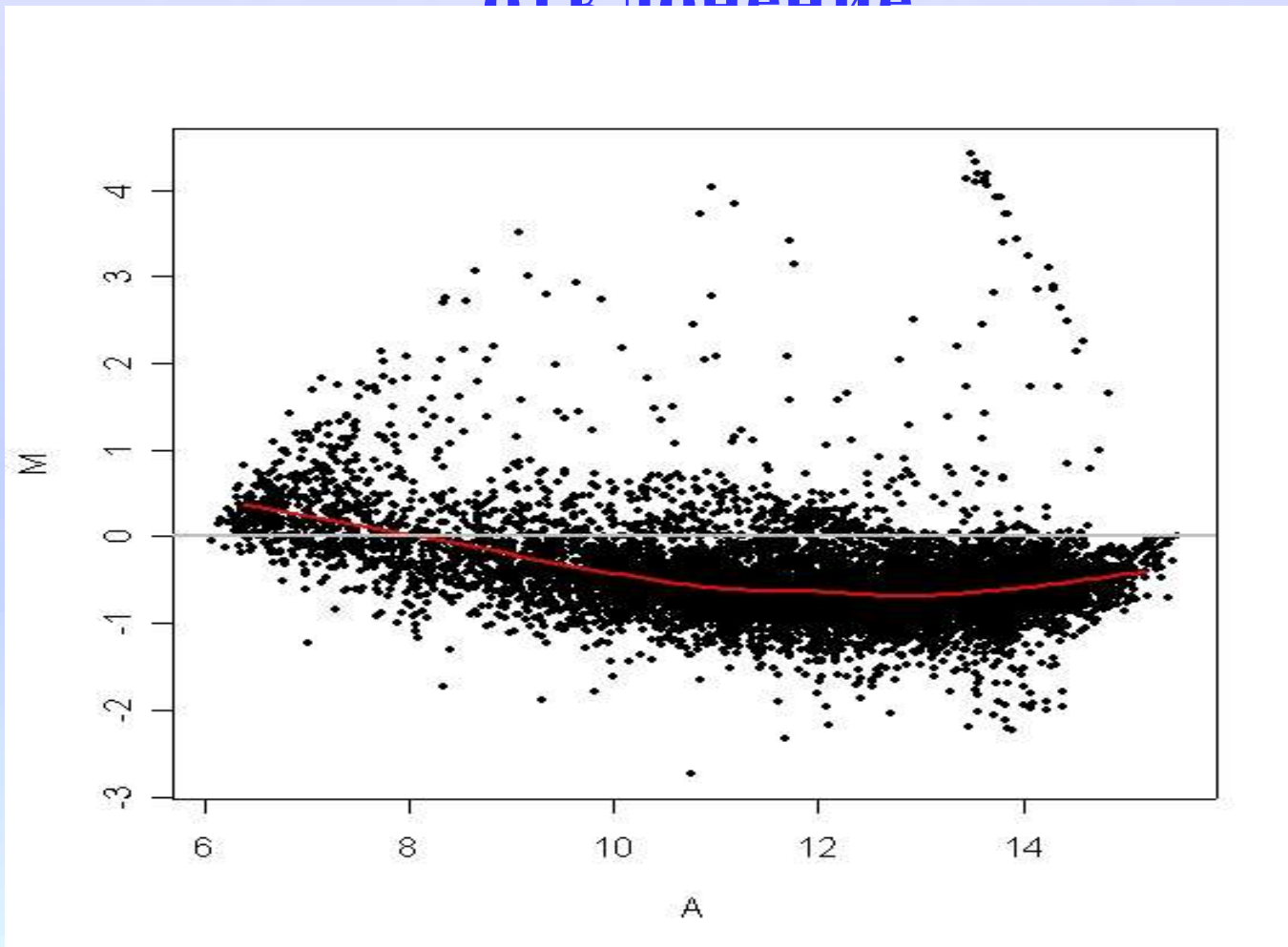
R - значения красных сигналов

G - значения зеленых сигналов

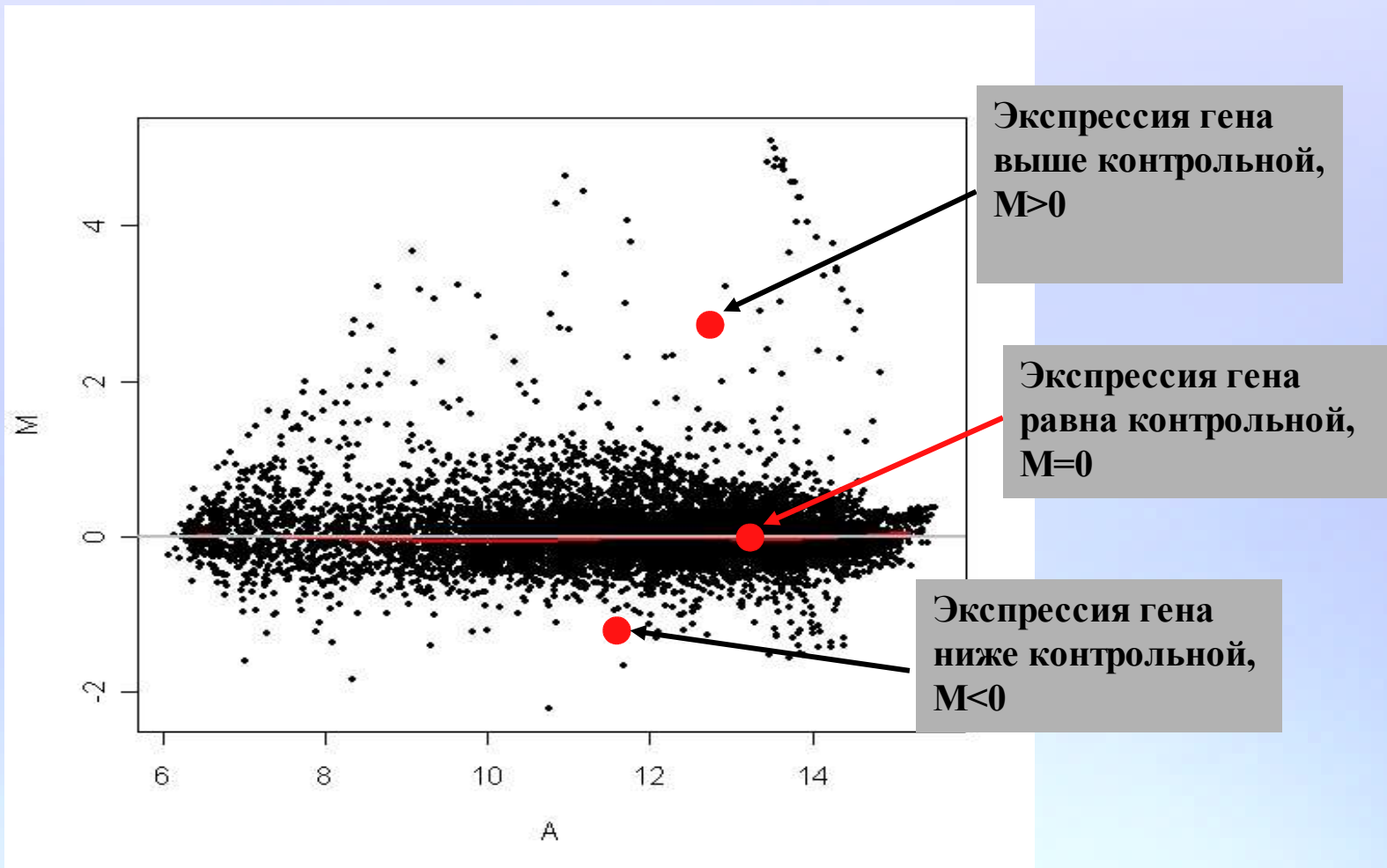
$$M = \log_2 R - \log_2 G,$$

$$A = (\log_2 R + \log_2 G)/2$$

Систематическое отклонение



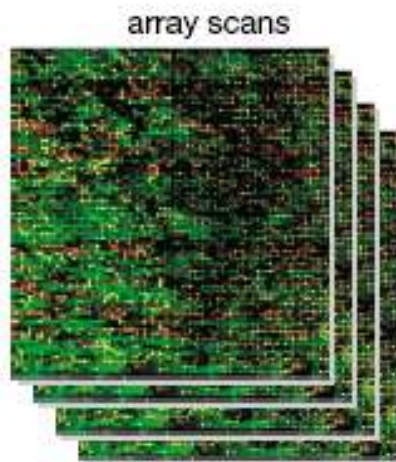
Средняя интенсивность сигнала (красная линия) отклоняется от нулевого значения. Это означает присутствие систематического отклонения значений сигналов.



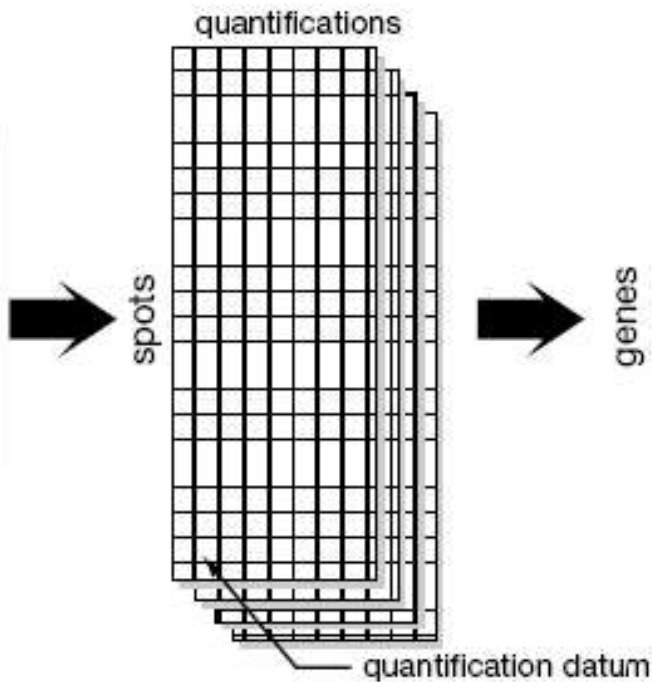
Вид распределения после нормализации. **Красная линия**, обозначающая среднее значение, близка к нулю.

ОБРАБОТКА ДАННЫХ, ПОЛУЧЕННЫХ МЕТОДОМ ДНК-БИОЧИПОВ

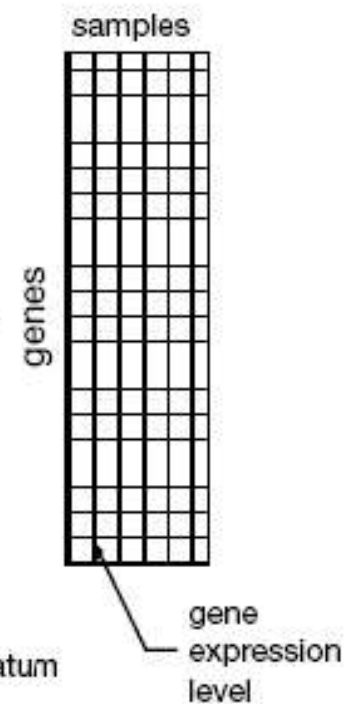
Исходные данные



Матрицы
измерений



Матрицы данных
экспрессии генов



Выделение генов с различающимся уровнем экспрессии

Матрица экспрессии

$\log_2(R/G)$

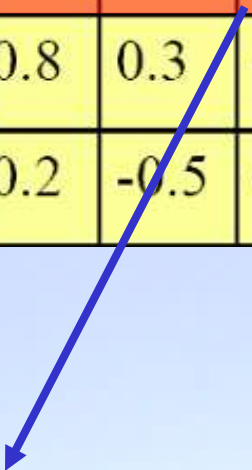


образцы мРНК

0.2	1.1	-1.5	2.0	0.8	0.1	-0.7	1.4
-0.7	-0.1	0.4	1.1	1.3	1.7	0.8	1.5
0.8	0.3	-0.6	-0.4	-0.6	0.1	-0.2	-0.3
0.2	-0.5	0.3	0.9	-1.2	-0.7	0.5	-0.2
⋮				⋮			⋮

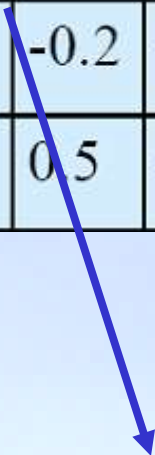
ГЕНЫ

0.2	1.1	-1.5	2.0	0.8	0.1	-0.7	1.4
-0.7	-0.1	0.4	1.1	1.3	1.7	0.8	1.5
0.8	0.3	-0.6	-0.4	-0.6	0.1	-0.2	-0.3
0.2	-0.5	0.3	0.9	-1.2	-0.7	0.5	-0.2



-0.7	-0.1	0.4	1.1
------	------	-----	-----

Выборка 1
среднее m_1



1.3	1.7	0.8	1.5
-----	-----	-----	-----

Выборка 2
среднее m_2

Критерии различия средних в 2-х выборках

Двухвыборочный t-критерий
Стьюдента:

$$T_g = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Применяется для проверки гипотезы о том, что средние m_1 и m_2 в 2-х выборках различны.

Анализ корреляций профилей экспрессии генов

Основные меры сходства профилей экспрессии двух генов

–корреляция Пирсона $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$

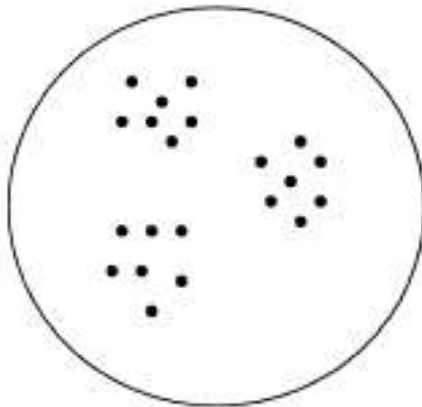
–Евклидово расстояние $d = \sqrt{\sum (x_i - y_i)^2}$

**В зависимости от выбора меры результаты анализа могут
сильно различаться !**

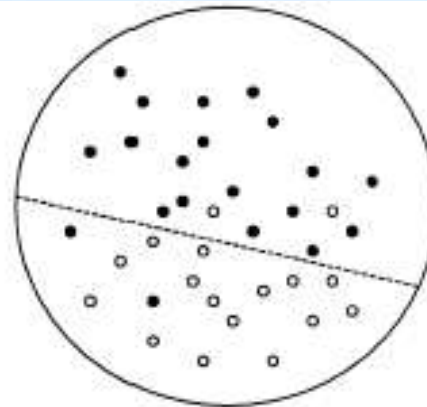
МЕТОДЫ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ

алгоритмы кластеризации (классификации) данных по
экспрессии генов

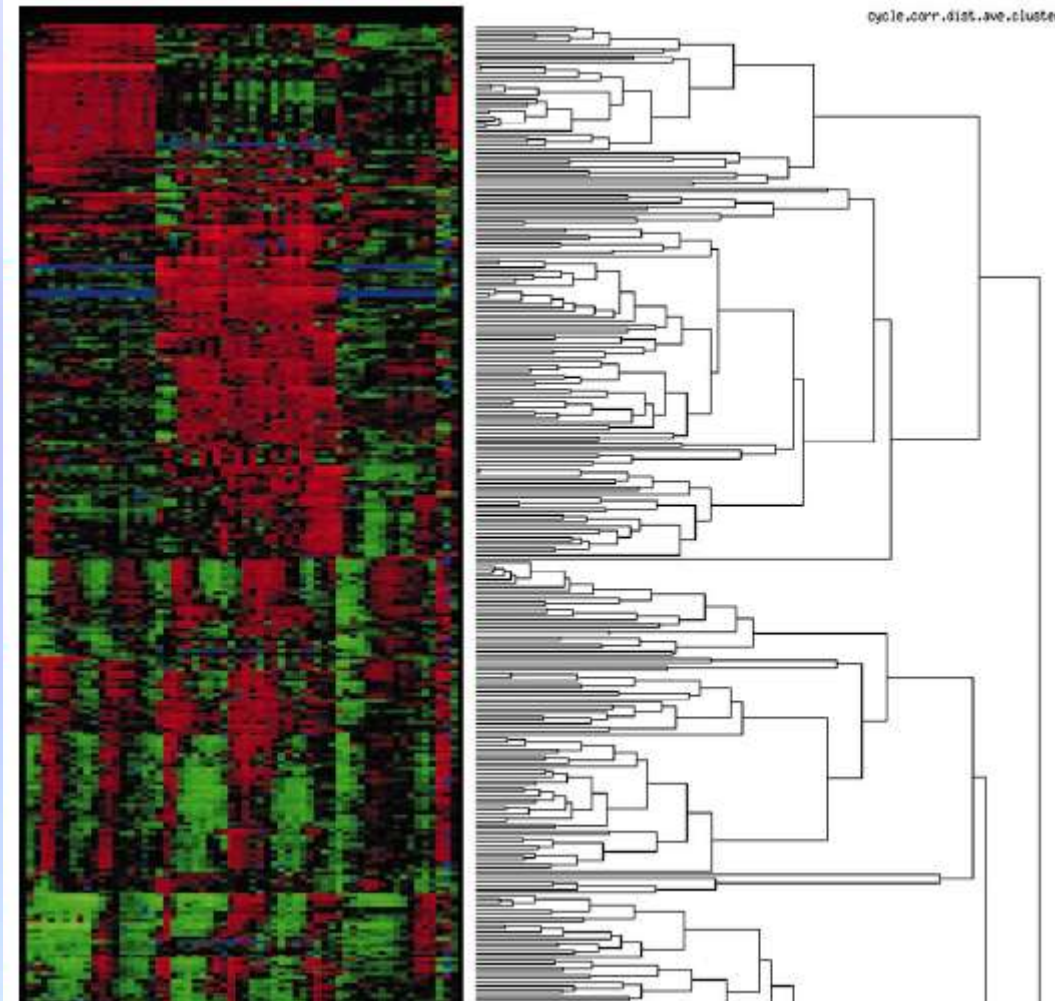
методы безусловной или
неконтролируемой
классификации
(unsupervised)



методы
контролируемой
классификации
(supervised)



Пример кластеризации



Иерархический алгоритм UPGMA (Unweighted Pair Group Method with Arithmetic mean). Кластеризация генов по экспрессии в течение трех стадий клеточного цикла по результатам измерений по 60 разным временным точкам.

Каждый ген представлен вектором – *строкой* из матрицы экспрессии.

Аналогично можно построить дерево из векторов-*столбцов*, и тогда получится разбиение на классы не генов, а *образцов*.

Образцы мРНК (samples)

$$X_{G \times n} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{G1} & x_{G2} & \dots & x_{Gn} \end{bmatrix}$$

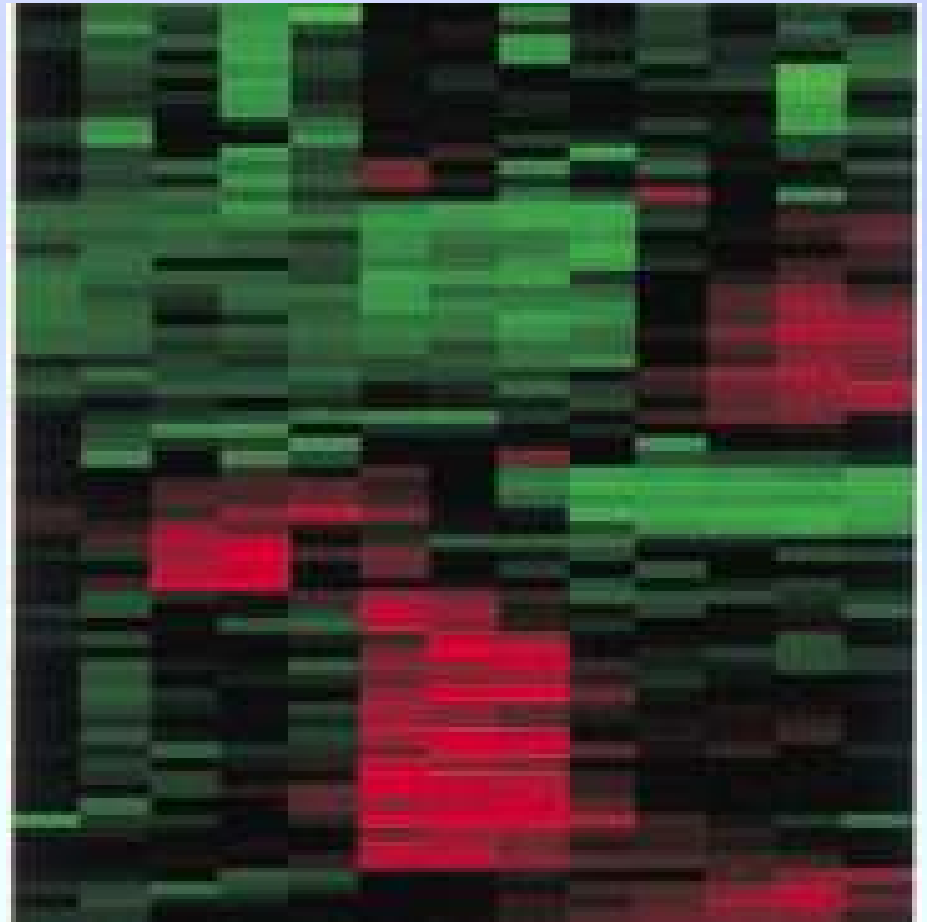
Гены

Визуальный анализ

Позволяет в наглядной форме представить результаты кластеризации, выделить группы генов и образцов со *сходной экспрессией*

Красный – повышение,
Зеленый – понижение
экспрессии относительно
контроля

Образцы



Гены

Полезные WWW-ресурсы

www.r-project.org Свободно распространяемая программная среда для анализа данных. Большие возможности для статистической обработки и графического представления данных, в т.ч. Microarray Analysis.

www.bioconductor.org Проект по созданию библиотек для R для анализа Microarray данных.

<http://rana.lbl.gov> Страничка лаборатории Майкла Айзена, одного из корифеев в области анализа Microarray данных (университет Беркли). Популярны программы, статьи, данные экспериментов.

<http://genome-www5.stanford.edu/> Stanford Microarray Database. В открытом доступе более 200 статей и исходные данные более чем 7700 Microarray экспериментов.

<http://www.ncbi.nlm.nih.gov/geo/> NCBI Gene Expression Omnibus, большая база данных по экспрессии генов, открытый доступ.

www.microarrays.org Протоколы проведения экспериментов, программы для обработки данных.

<http://www.ebi.ac.uk/arrayexpress/> база данных ArrayExpress EBI, открытый доступ.



Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Спасибо за внимание.

Приглашаем студентов-математиков, информатиков, молекулярных биологов проходить практику в секторе функциональной геномики лаборатории теоретической генетики.

