



# **Исследование качественных и количественных характеристик транскриптома (2)**

*Катохин Алексей Вадимович, к.б.н.*

Кафедра информационной биологии ФЕННГУ



# Исследование качественных и количественных характеристик транскриптома



Требования к методам исследования структуры транскриптов и их дифференциального временного и пространственного распределения в клетках и организмах:

- возможность измерения относительного и абсолютного содержания транскриптов определенного гена в клетках разных типов, т.е. возможность сравнения результатов разных экспериментов при разных модификациях методов;
- возможность измерения соотношения транскриптов как можно большего количества генов;
- возможность детекции транскриптов очень слабо или очень специализированно экспрессирующихся генов (чувствительность и достаточно широкий динамический диапазон)
- высокая производительность и эффективность для производства достаточно большого массива данных.



# Исследование качественных и количественных характеристик транскриптома



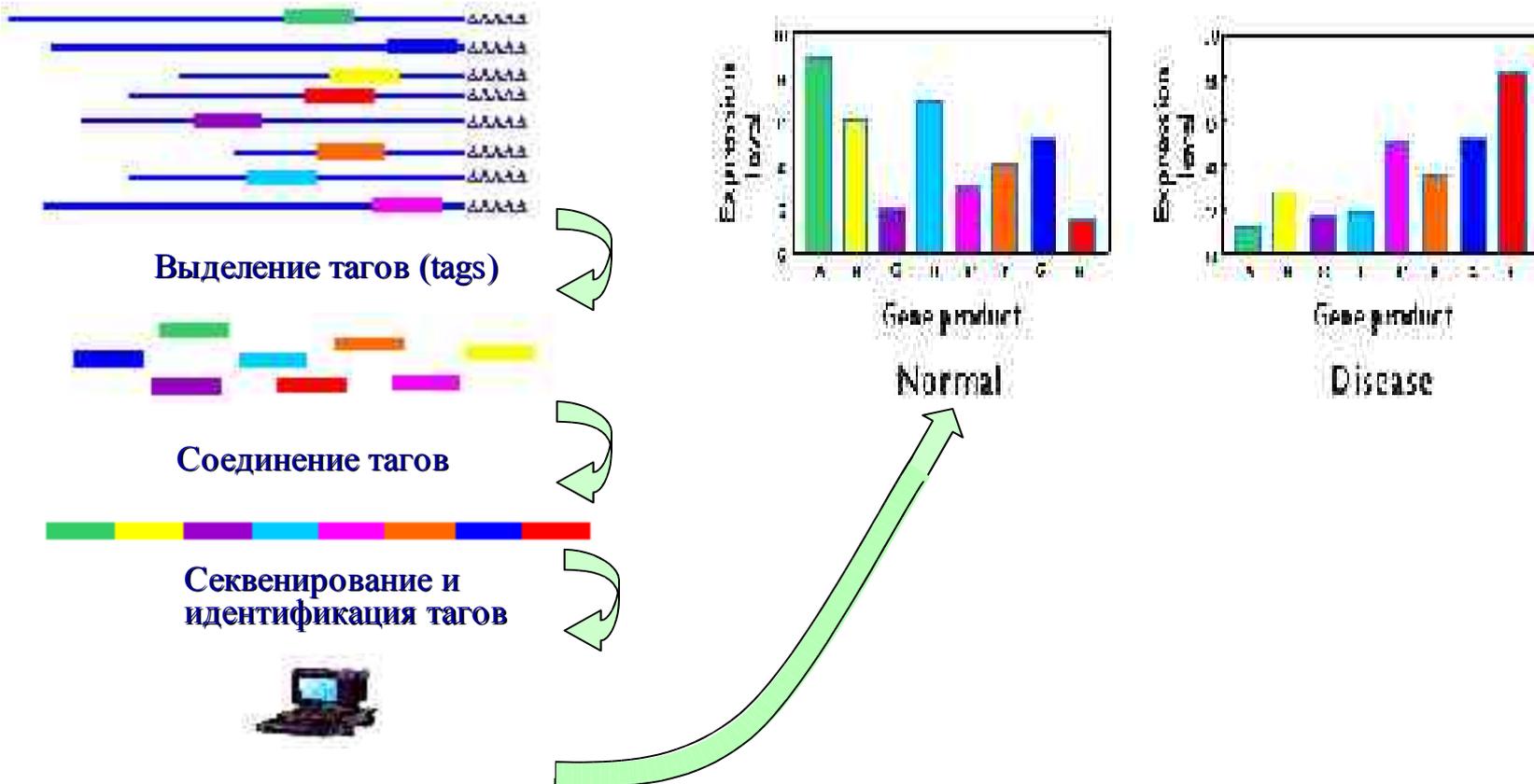
## Характеристика методов детекции и оценки количества

метод	Доверие относит. и абсолют. содержания, сравнимость результатов	количество генов	чувствительность и динамический диапазон	производительность и эффективность
Прямая детекция мРНК гибридизацией транскриптов с одноцеп. ДНК- или РНК-пробами				
Нозерн-блот	+	(1-5) x (5- 20)	--	--
гибридизация Защита от рибонуклеазы	++	(10-15) x (5- 20)	+	--
Детекция кДНК, полученной ОТ-ПЦР из мРНК				
Количественная ОТ-ПЦР	++	десятки	++ (!)	++
Дифференц. дисплей	+	десятки	++ (!)	++
кДНК микрочипы				
Компьютерный анализ сиквенсов кДНК , « <i>in silico</i> гибридизация» транскриптов				
SAGE	+++	тысячи	+++	+++
EST	+	тысячи	+++ (!)	+++



# Исследование качественных и количественных характеристик транскриптома

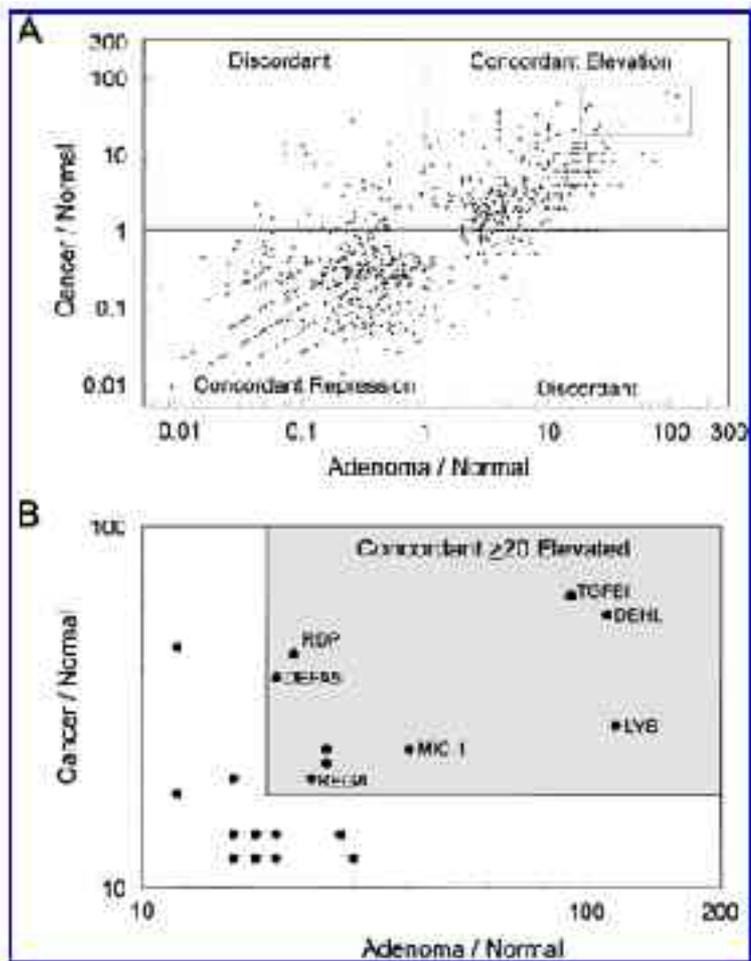
## Серийный анализ экспрессии генов (SAGE)





# Исследование качественных и количественных характеристик транскриптома

## Серийный анализ экспрессии генов (SAGE): приложения



Buckhaults P. *et al.*, 2001  
Secreted and Cell Surface Genes Expressed in Benign and Malignant Colorectal Tumors  
// *Cancer Res*; 61(19):6996-7001

SAGE-анализ предпринят для выявления транскриптов, кодирующих секретлируемые или клеточно-поверхностные белки в добро- и злокачественных опухолях колоректума. Всего было проанализировано 290,394 тагов из разных образцов и выявлено 957 транскриптов, дифференциально экспрессируемых в норме и аденомах или в норме и раке.

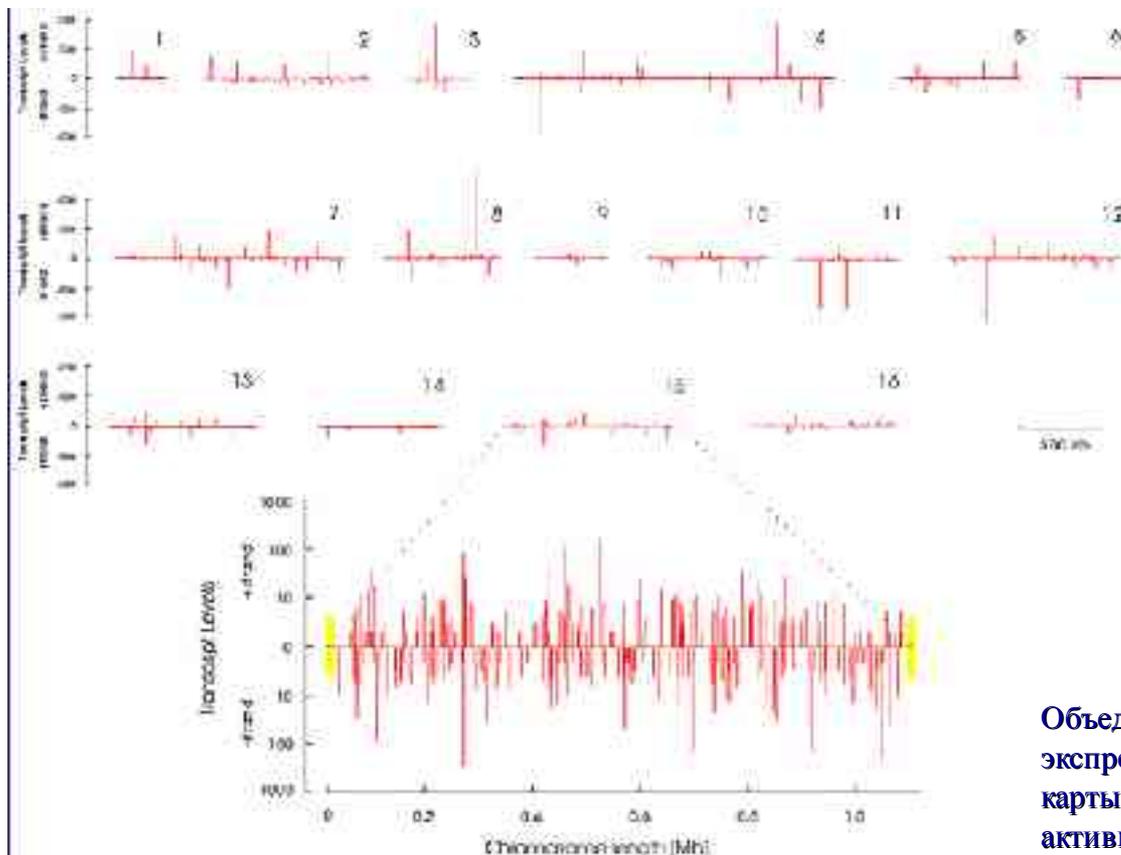
A, распределение значений кратности изменений для достоверно ( $P < 0.05$ ) дифференциально экспрессируемых транскриптов. Отношение выражено в логарифмической шкале. B, увеличенный фрагмент графика A.

Экспрессия транскриптов выявленных генов-кандидатов проверена количественной ОТ-ПЦР.



# Исследование качественных и количественных характеристик транскриптома

## Серийный анализ экспрессии генов (SAGE): приложения



Velculescu V.E. *et al.*, 1997 Analysis of Yeast Transcriptome // Cell 88: 243-251

Анализ транскриптов из дрожжевых клеток трех состояний (лог-фаза, задержанные в S- и G2/M- фазах), всего 60,633 тагов, выявил 4,665 генов с уровнями экспрессии от 0.3 до более 200 транскриптов на клетку.

Число SAGE-тагов определено исходя из требований достижения определенного уровня достоверности выявления редких мРНК. При оценке 15,000 молекул мРНК на клетку секвенирование 20,000 тагов должно дать 72% вероятности обнаружить одну молекулу мРНК на клетку.

Объединение позиционной информации с данными об экспрессии генов позволило построить хромосомные карты экспрессии с районами транскрипционной активности и выявить гены, существование которых было невозможно предсказать, исходя из знания только о последовательности геномного района.



# Исследование качественных и количественных характеристик транскриптома

## Серийный анализ экспрессии генов (SAGE): интернет-ресурсы

<http://www.sagenet.org/>

**SAGE**  
Serial Analysis of Gene Expression

FINDINGS RESOURCES PUBLICATIONS CONFERENCES CONTACT US

**What's New?**

*Serial analysis of gene expression (SAGE) is a powerful tool that allows the analysis of overall gene expression patterns with digital analysis. Because SAGE does not require a preexisting clone, it can be used to identify and quantitate new genes as well as known genes.*

*[Learn more...](#)*



# Исследование качественных и количественных характеристик транскриптома

## Серийный анализ экспрессии генов (SAGE): интернет-ресурсы <http://www.sagenet.org/>

### Вещицы

#### SAGE Data

##### Human Colorectal Tissue

Library Name	Description	Tags Analyzed
H001	Normal colon epithelium	21,227
H002	Normal colon epithelium	22,220
H0003	Colon cancer cells	14,110
H0111	HCT116 colon cancer cells	11,121
SW620	SW620 colon cancer cells	22,228
PKO	PKO colon cancer cells	22,272
TUJ02	primary colon cancer	22,224
TUJ01	primary colon cancer	47,271
HL04	Athymic A	11,111
HL05	Athymic B	11,112

##### Human Pancreatic Cancer

Library Name	Description	Tags Analyzed
PL05	PL05 pancreatic cancer cells	21,226
AS01	AS01 pancreatic cancer cells	22,220
HL01113	Primary pancreatic cancer	11,111
HL0502	Primary pancreatic cancer	11,116

##### Mouse Fibroblast

Library Name	Description	Tags Analyzed
HL01	3T3 cells fibroblast	22,221

##### Yeast

Library Name	Description	Tags Analyzed
Yeast01	Long phase yeast	22,222
Yeast02	Hydroxy-methylated yeast	22,221
Yeast03	Hydroxy-methylated yeast	22,223

- Using the transcriptome to analyze the genome (LongSAGE)



# Исследование качественных и количественных характеристик транскриптома

## Серийный анализ экспрессии генов (SAGE): интернет-ресурсы

Serial Analysis of Gene Expression  
Tag to Gene Mapping **SAGEmap**

GEO PubMed OMIM Entrez UniGene LocusLink

**Public gene expression data** Query

Home | About  
Retrieve  
By ID  
By Accession

In order to support the public use and dissemination of serial analysis of gene expression (SAGE) data, NCBI has recently refurbished the website SAGEmap as a SAGE data resource for the query and retrieval and analysis of SAGE data from any organization. All of the data present on this website has been accessioned in the Gene Expression Omnibus repository.

<http://www.ncbi.nlm.nih.gov/SAGE/>

## SADE

a **SAGE** Adaptation for **D**ownsized **E**xtracts

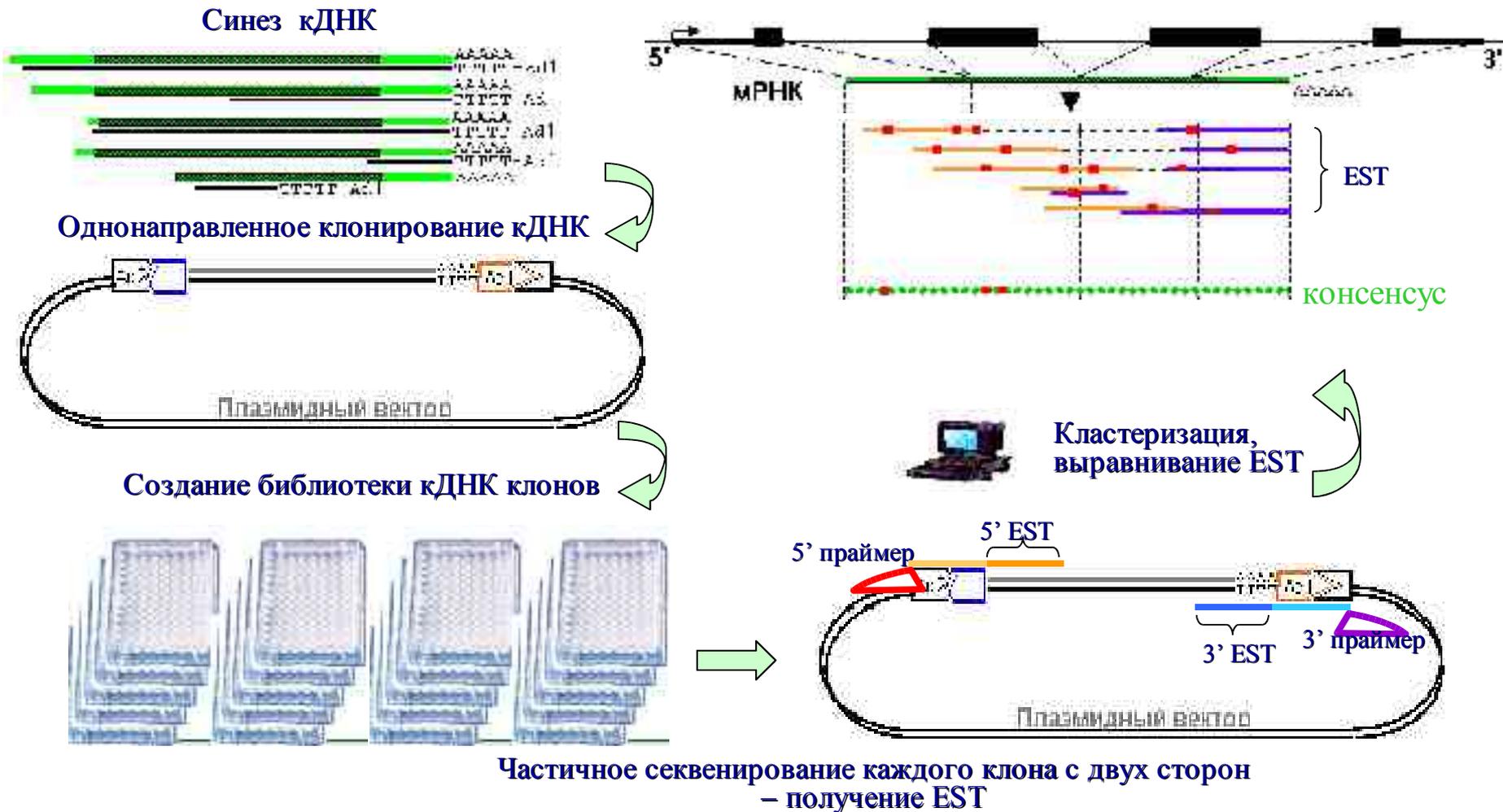
Serial microanalysis (<http://www-dsv.cea.fr/thema/get/sade.html>)



# Исследование качественных и количественных характеристик транскриптома



## Прочитанные фрагменты экспрессированных последовательностей ( ESTs )

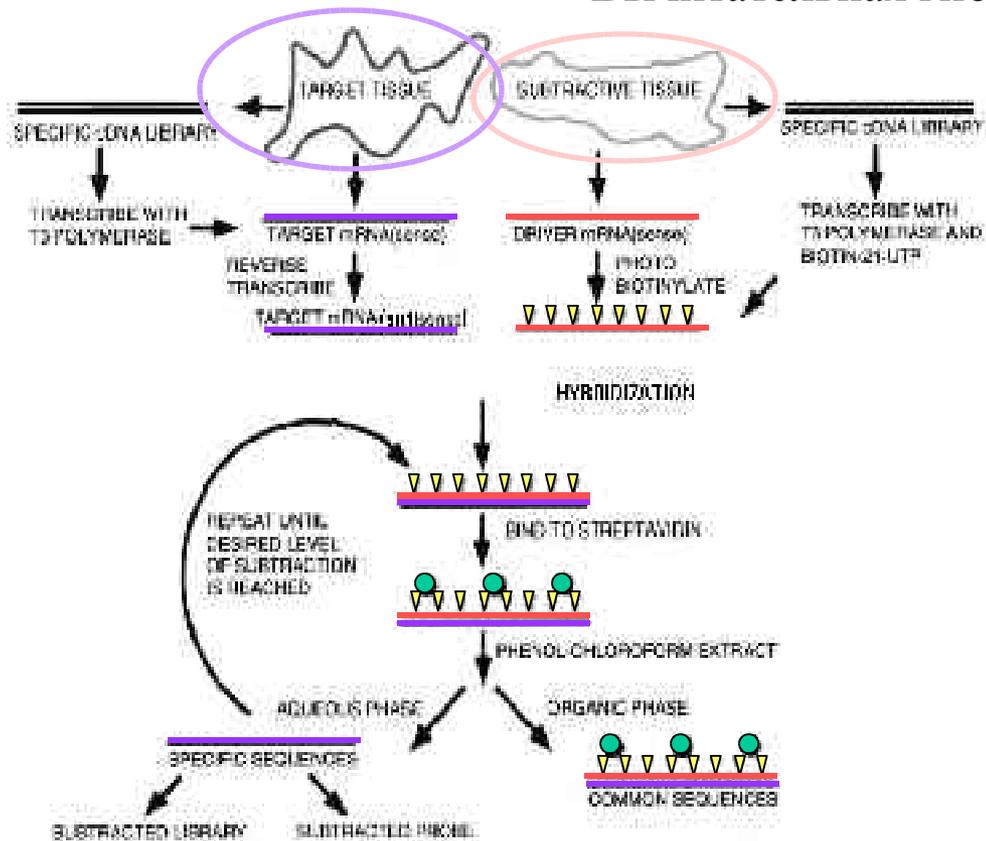




# Исследование качественных и количественных характеристик транскриптома

Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): особенности приготовления библиотек клонов кДНК

## Вычитательная гибридизация



Позволяет выявлять очень редкие ткане- или стадия-специфичные транскрипты,

но изменяет соотношение между транскриптами

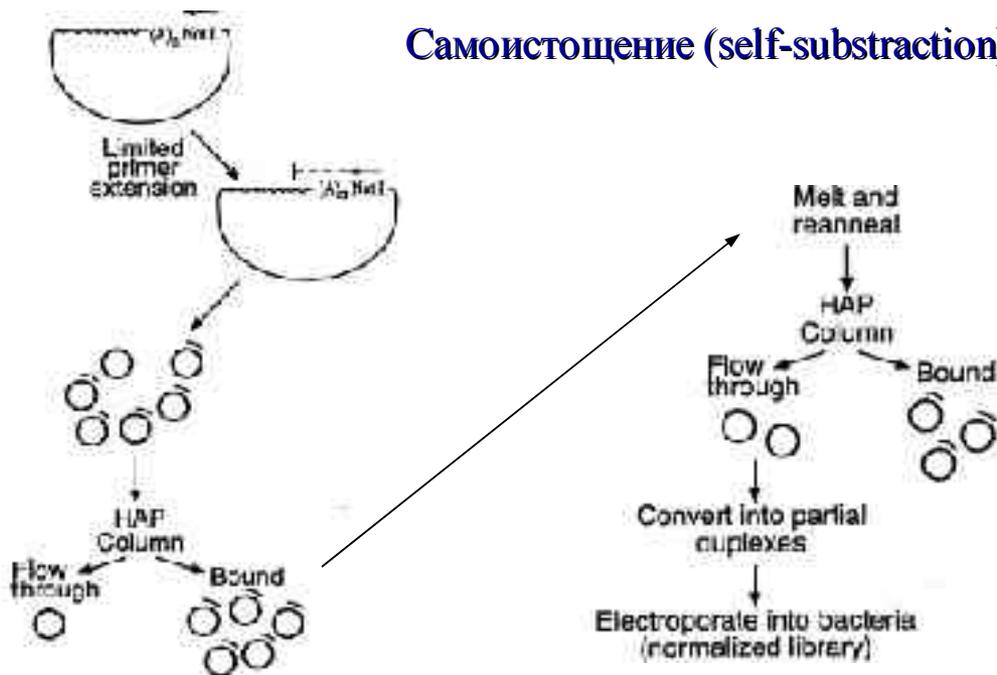


# Исследование качественных и количественных характеристик транскриптома

Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): особенности приготовления библиотек клонов кДНК

## Нормированные библиотеки кДНК клонов

Самоистощение (self-substraction)



Позволяет выявлять очень редкие транскрипты,

**но изменяет соотношение между транскриптами**

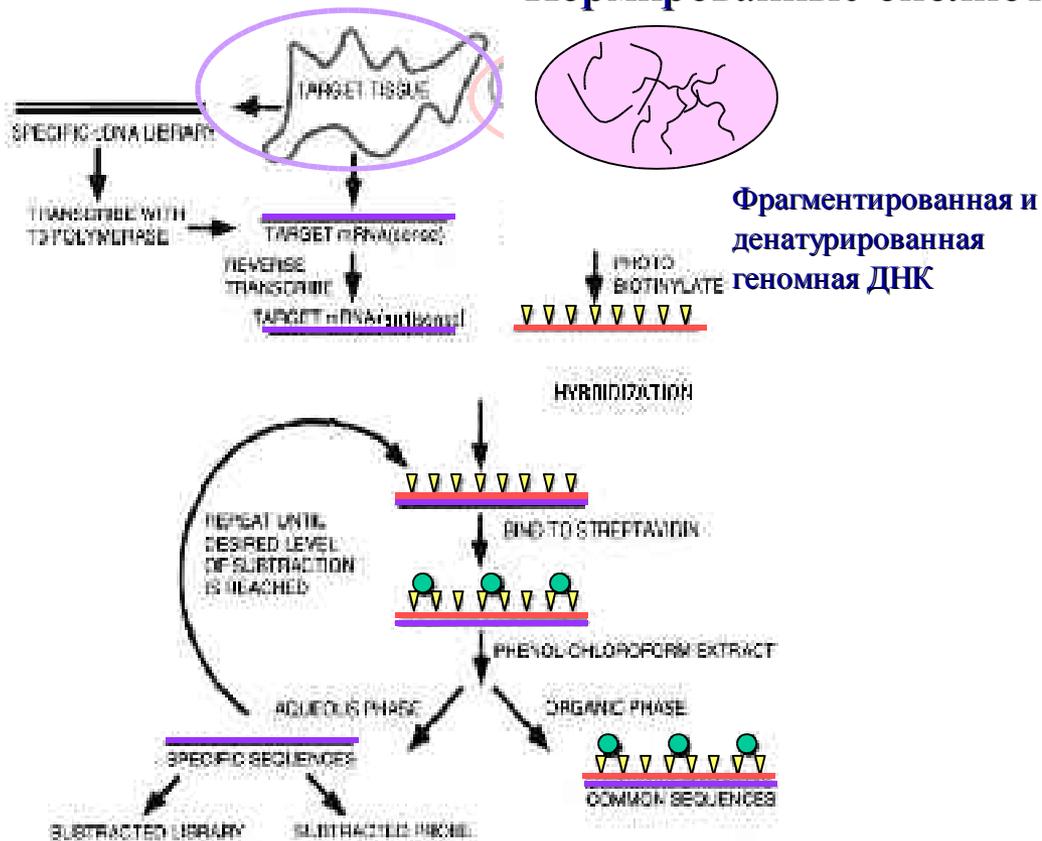
Soares M.B. *et al.*, 1994,  
Construction and characterization of a  
normalized cDNA library. Proc Natl Acad  
Sci U S A.;91(20):9228-32.



# Исследование качественных и количественных характеристик транскриптома

Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): особенности приготовления библиотек клонов кДНК

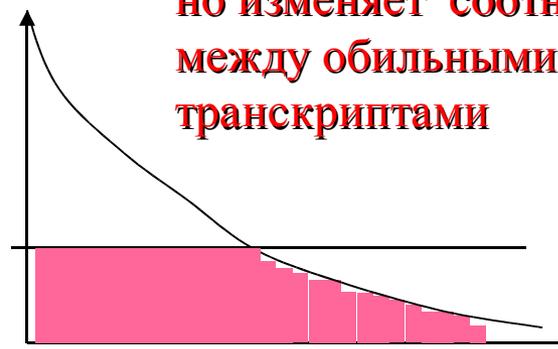
## Нормированные библиотеки кДНК клонов



Истощение геномными последовательностями

Позволяет выявлять очень редкие транскрипты,

но изменяет соотношение между обильными и редкими транскриптами



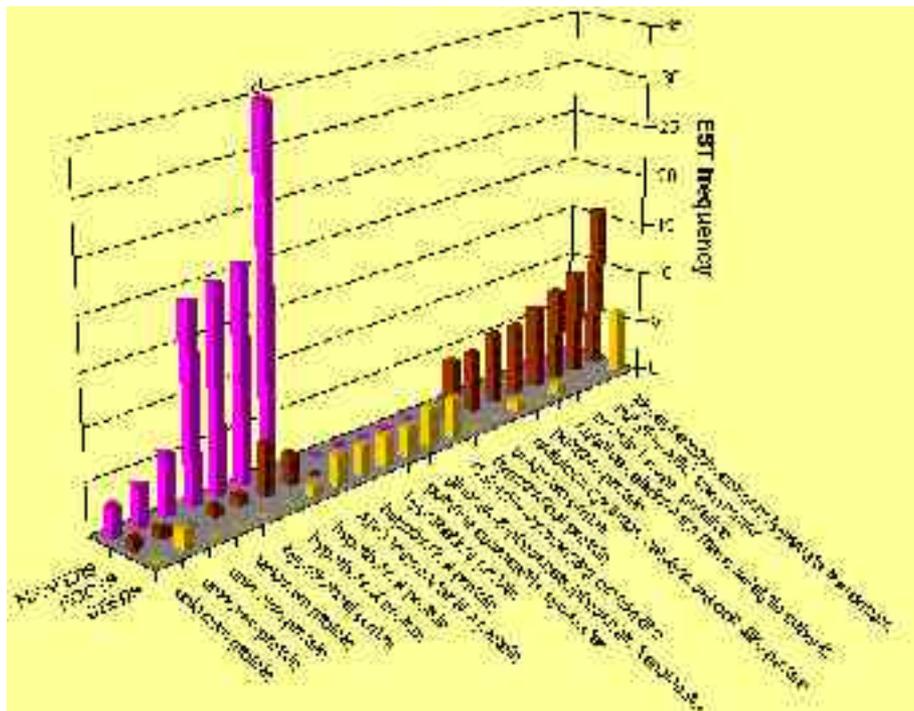


# Исследование качественных и количественных характеристик транскриптома

## Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): приложения

### Цифровой дифференциальный дисплей Digital Differential Display

Audic and Claverie  
(<http://igs-server.cnrs-mrs.fr>)

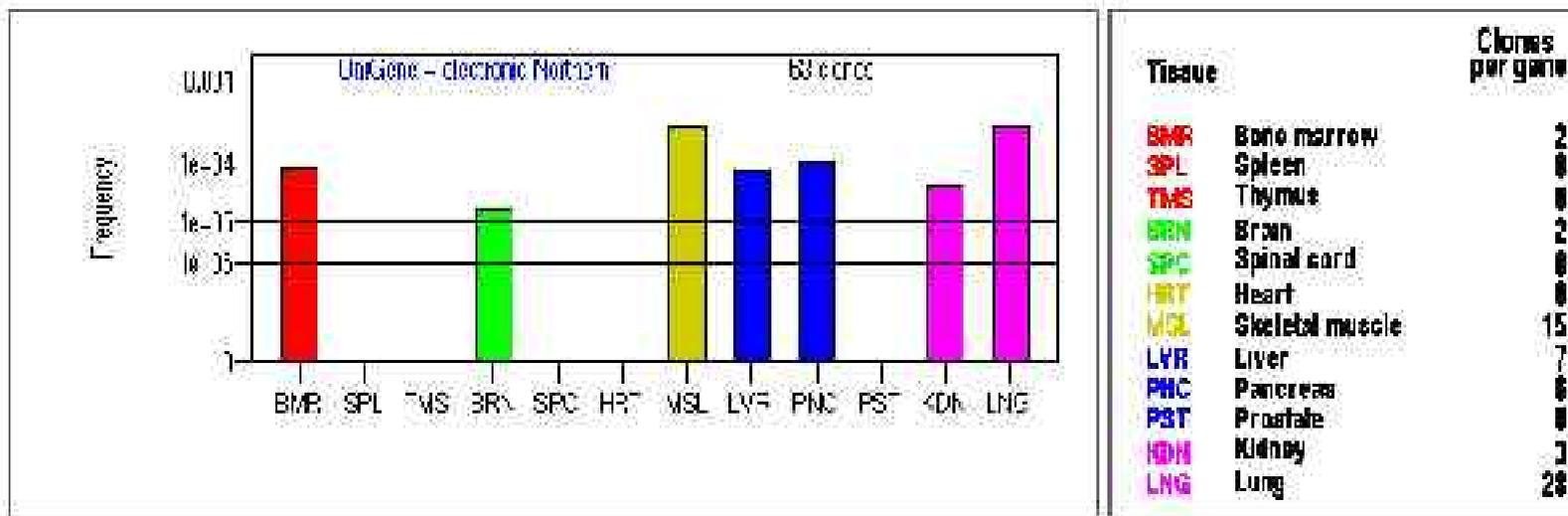


Mégy K, Audic S, Claverie JM. 2002; Heart-specific genes revealed by expressed sequence tag (EST) sampling. *Genome Biol.* 3(12): research0074.1-research0074.11.



# Исследование качественных и количественных характеристик транскриптома

Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): приложения  
Электронный или цифровой Нозерн



The screenshot shows the NCBI UniGene website. At the top, there is the NCBI logo and the UniGene logo. Below the logos, there is a navigation menu with options: Home, Knowledge, Pathway, Science, Methods, Papers, Library. A search bar is visible with the text 'Search UniGene'. Below the search bar, there are links for 'Limits', 'Expression', 'Clones', and 'Details'. The main content area is titled 'Digital Differential Display (DDD)' and includes a 'Page 31 of 36' indicator and a dropdown menu for 'Info Factors'.

<http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi?ORG=Hs>





# Исследование качественных и количественных характеристик транскриптома

## Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): интернет-ресурсы

### The EST Machine

<http://www.tigem.it/ESTmachine.html>

Directory of URLs to selected bioinformatic resources for Expressed Sequence Tags analysis

EST databases	I.M.A.G.E. Consortium / WashU-Merck-IBM Project
<ul style="list-style-type: none"> <li>• ESTs at EMBL (http://www.ebi.ac.uk/EMBL/EST/)</li> <li>• Search of EST using SRS6 of EMBL</li> <li>• UniGene of NCBI (<i>Some papers: How muscular? Same message!</i>, see also the SRS6 implementation of InSilicoBiology, HCOMP, and Tager)</li> <li>• Library browser with EST expression information from UniGene</li> <li>• <i>Sequence, Analysis, Alignment and Consensus Knowledgebase at ZANU</i></li> <li>• TIGR Database (TDB) and TIGR Human Gene Index (see also the TIGR Mouse Gene Index)</li> <li>• <i>GeneID (http://EST.tigr.org)</i></li> <li>• Catalogue EST Database (open)</li> <li>• Mouse cDNA homepage (University of Texas)</li> <li>• <i>Plant EST analysis: community analysis of different plant ESTs</i> (EST sequences at U. Minnesota)</li> </ul>	<ul style="list-style-type: none"> <li>• I.M.A.G.E. Consortium homepage</li> <li>• WashU-Merck-Human EST Project (at BNL Library and WashU Information)</li> <li>• The IMAGE database (I.M.A.G.E. has clustered all three ESTs above for all known human genes)</li> <li>• The Cancer Genome Anatomy Project (CGAP, the National Cancer Institute EST project)</li> <li>• <i>Clonogen and Tumor Suppressor genes: University of Michigan</i></li> <li>• <i>WashU-IBM Mouse EST Project</i></li> <li>• WashU-Ferrel EST Project</li> <li>• <i>Flare Sequence Trace (either for a human or a mouse EST)</i></li> <li>• The LEM (Linking ESTs and their associated Blast Spots) Database</li> <li>• I.M.A.G.E. Consortium Clonal Distributions</li> </ul>



# Исследование качественных и количественных характеристик транскриптома

Прочитанные фрагменты экспрессированных последовательностей ( ESTs ): интернет-ресурсы

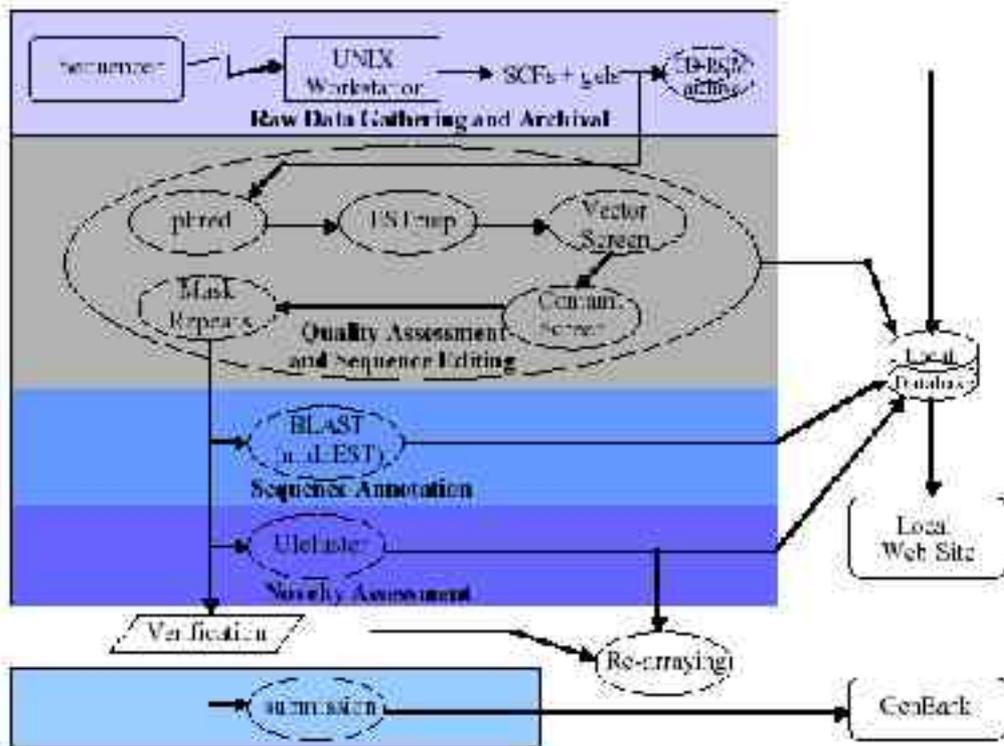


Figure 2. Overview of EST-based Gene Discovery Pipeline.



# Исследование качественных и количественных характеристик транскриптома



Прочитанные фрагменты экспрессированных последовательностей ( ESTs ):  
интернет-ресурсы

<http://www.tigem.it/ESTmachine.html>

<http://cgap.nci.nih.gov/>

[http://industry.ebi.ac.uk/~muilu/EST/EST\\_links.html](http://industry.ebi.ac.uk/~muilu/EST/EST_links.html)

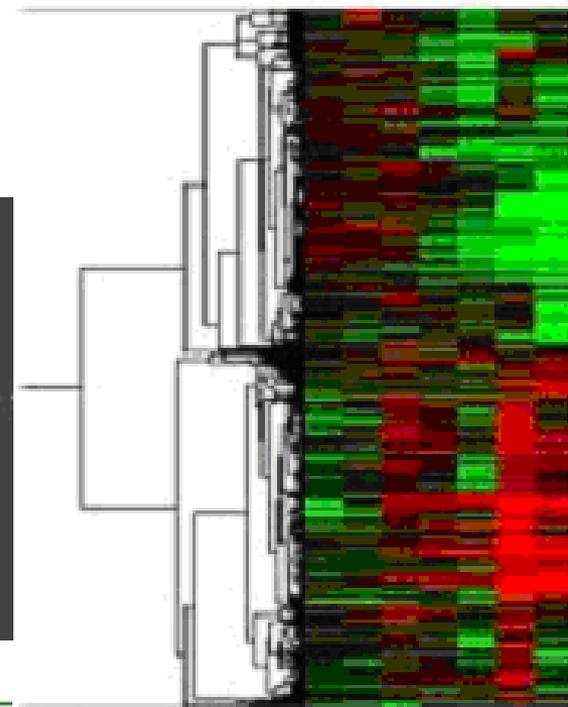
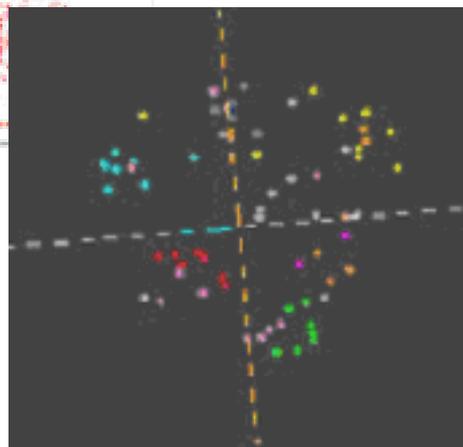
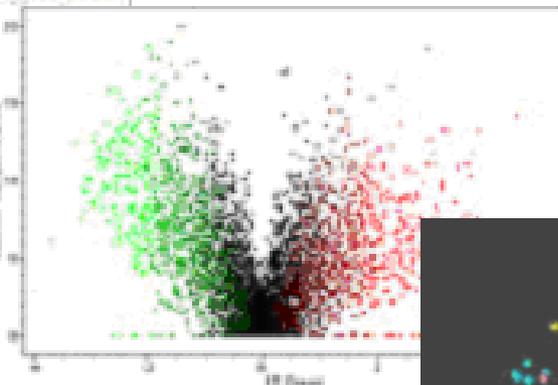
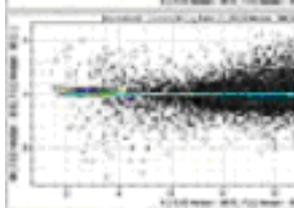
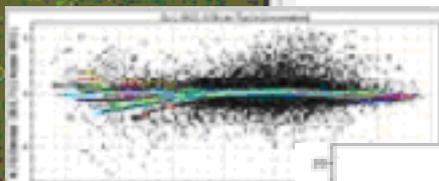
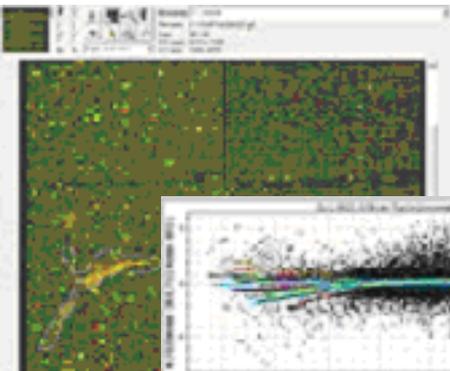
[http://www.ncbi.nlm.nih.gov/dbEST/how\\_to\\_submit.html](http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html)

<http://image.llnl.gov/>

<http://www.cs.jhu.edu/labs/compbio/morgan.html>

<http://hercules.tigem.it/ASSEMBLY/assemble.html>

## ДНК-биочипы (ДНК-микроматрицы)





## **Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)**

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.



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## Типы ДНК-биочипов (микроматриц)

### На поверхности стекла или полимера

олигонуклеотидные  
биочипы

20-75 н.о.

Fodor et al., 1993;  
Lipshutz et al., 1995;  
Lockhart et al., 1996

кДНКовые  
биочипы

100-2500 н.о.

DeRisi et al., 1996

### В объеме

Фрагменты ДНК,  
иммобилизованные в  
микрокаплях геля

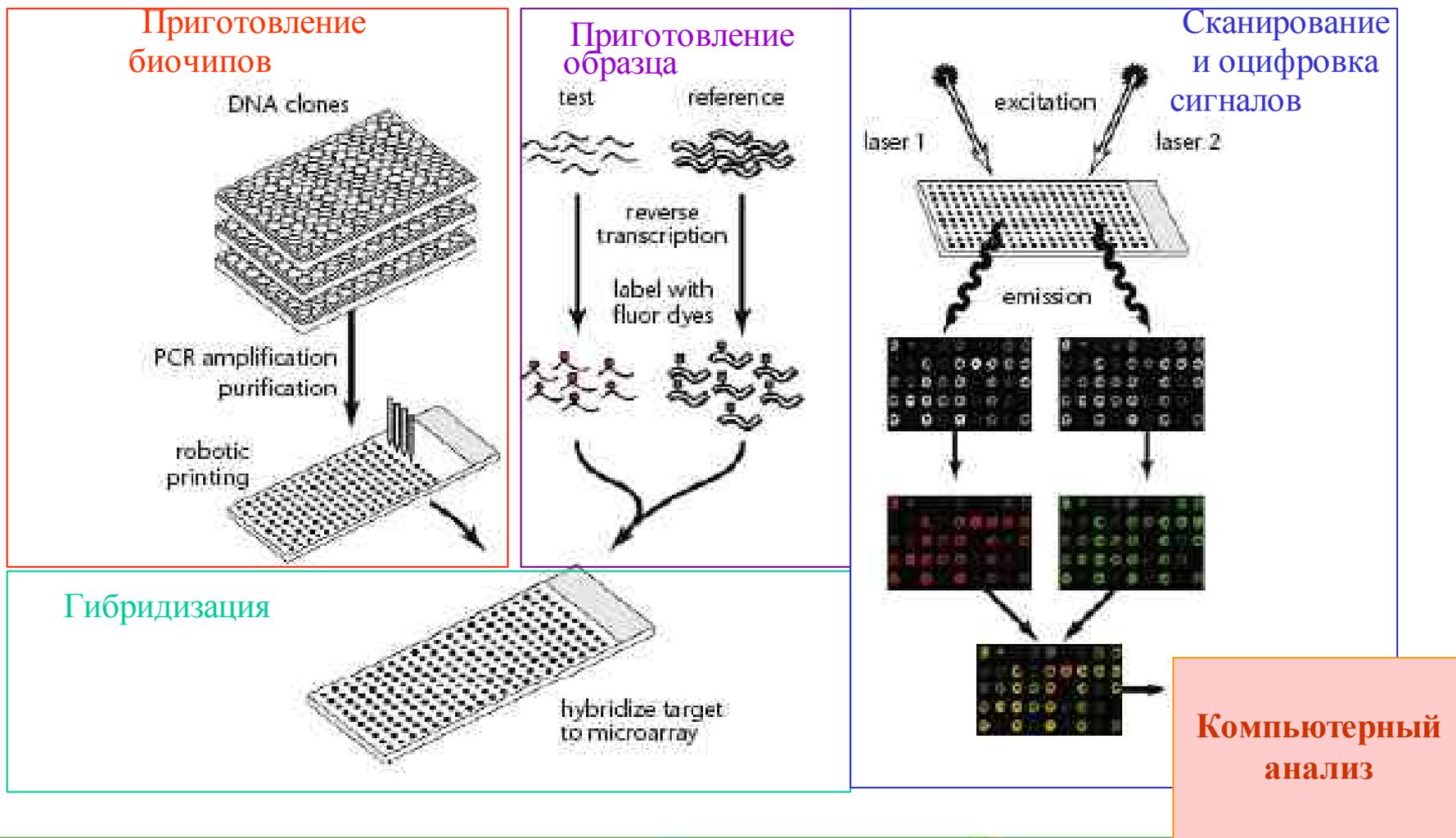
Под руководством акад. А.Д.  
Мирзабекова в ИМБ РАН  
Lysov et al., 1988;  
Khrapko et al., 1989

Обычно это микроскопное стекло размером 25  
мм x 76 мм,



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

Схема использования кДНК-биочипов (экспериментальная фаза)





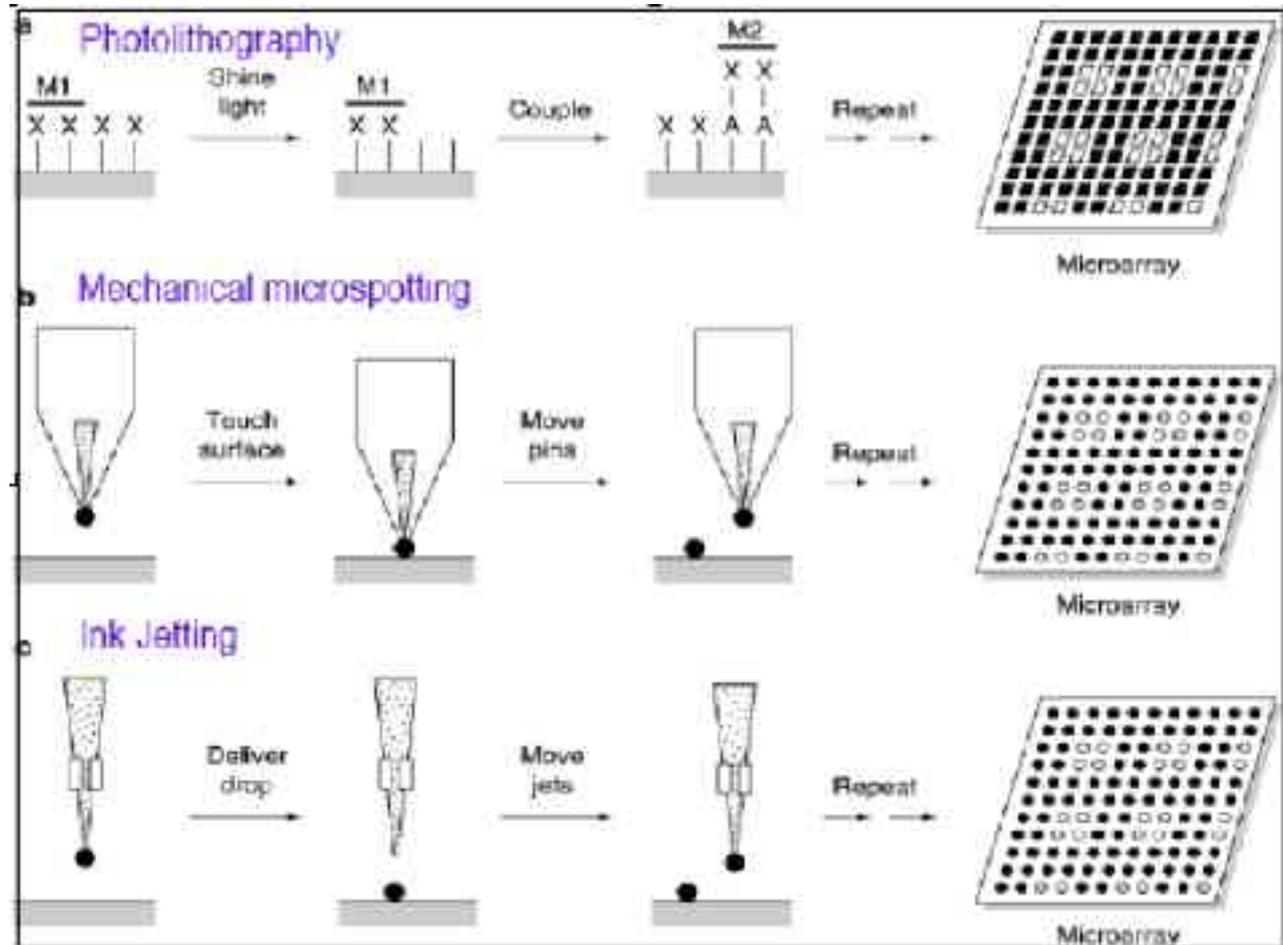
# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## СПОСОБЫ ПРИГОТОВЛЕНИЯ ДНК-БИОЧИПОВ

синтез  
олигонуклеотидов  
*in situ*

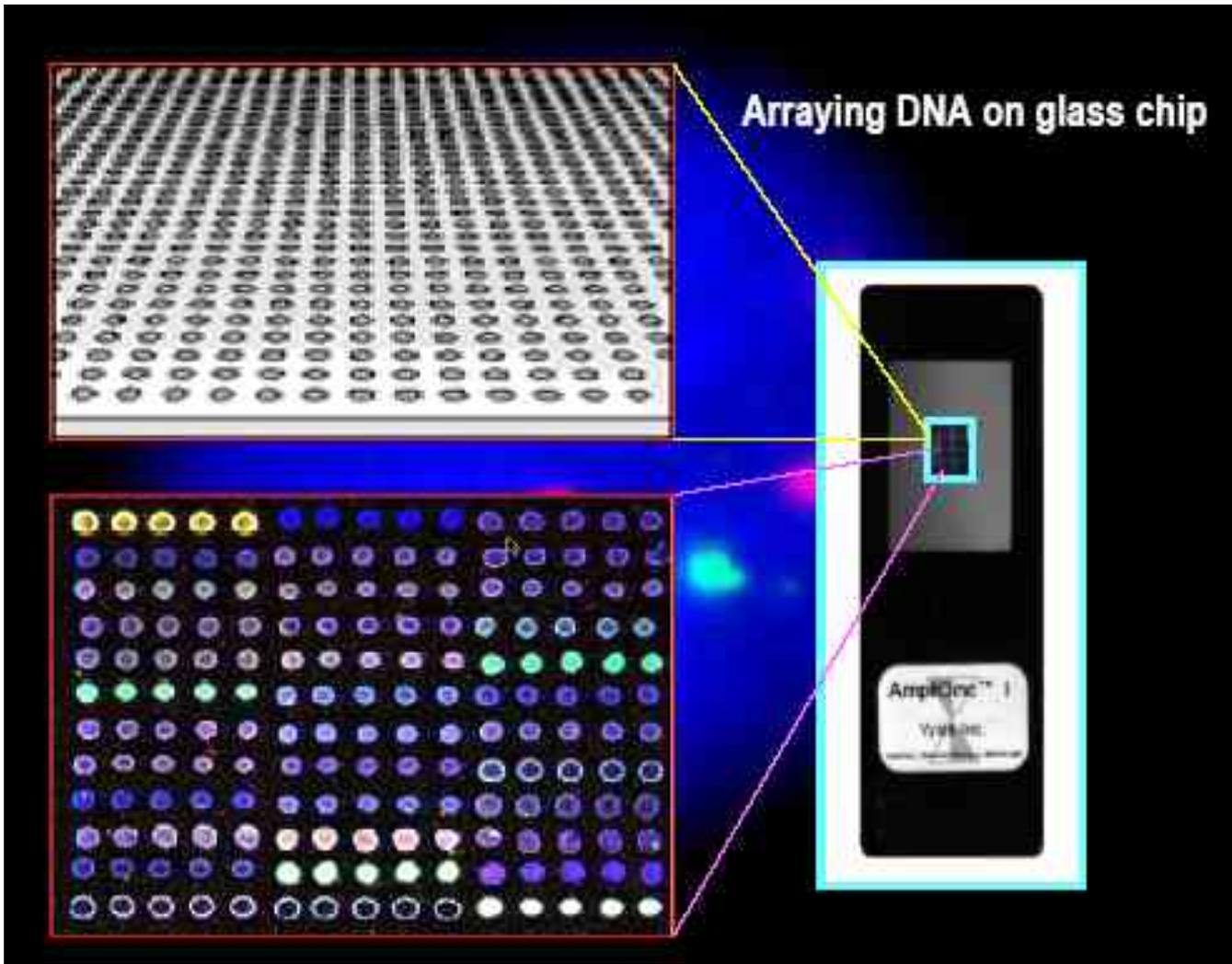
контактная  
печать

струйная печать





# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



Вид  
стеклянного  
ДНК-биочипа



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## Процесс приготовления ДНК-биочипов





# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

При описании ДНК-чипов наиболее часто используют термины:

- «проба» или «зонд» (probe), фрагмент ДНК, иммобилизованный на подложке,
- «мишень» (target) или «образец» (sample), фрагменты ДНК(кДНК) из биологического материала

## кДНК-биочипы

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

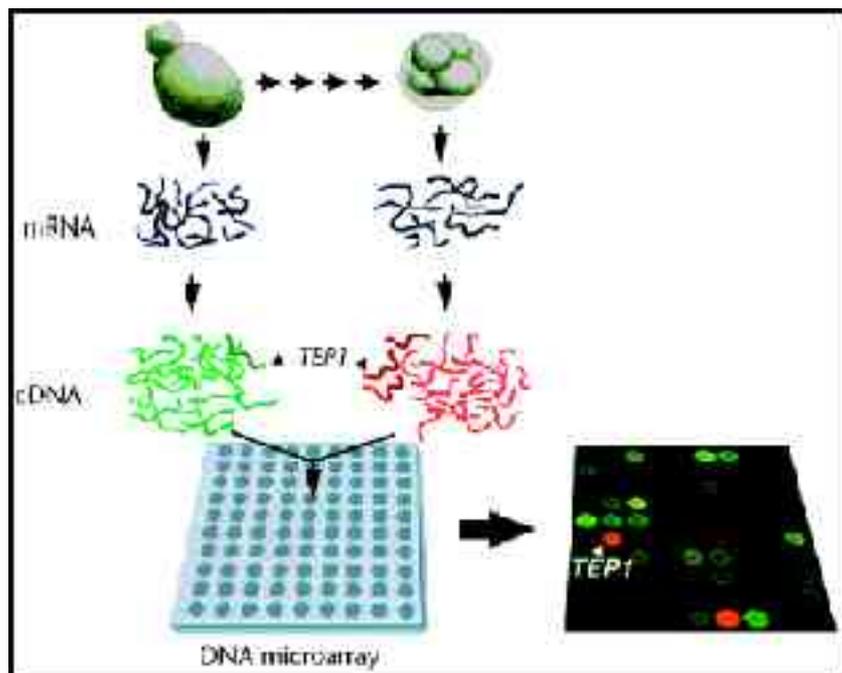
## олигонуклеотидные биочипы

Олигонуклеотидные 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.



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## ПРИГОТОВЛЕНИЕ МЕЧЕННОГО ОБРАЗЦА И ГИБРИДИЗАЦИЯ



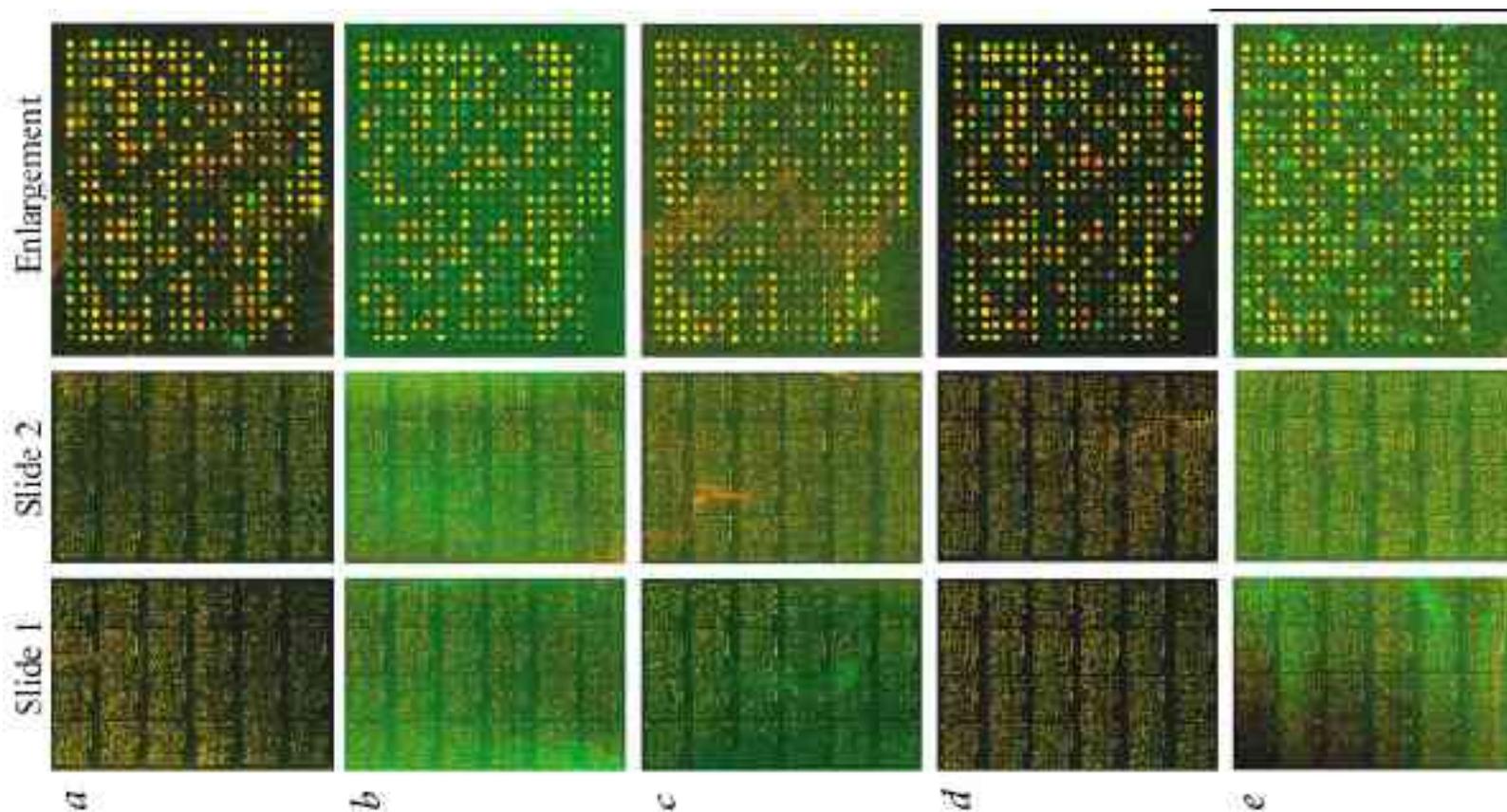
Каждый пул кДНК, приготовленный из двух разных образцов, делится на две равные части молекул, в которые вводятся меченые флуоресцентными группировками нуклеотиды (прямое или не прямое включение). Обычно это зеленый Cyanine3 (Cy3) и красный Cyanine5 (Cy5). Оба меченых образца объединяются для совместной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



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



## СКАНИРОВАНИЕ РЕЗУЛЬТАТОВ ГИБРИДИЗАЦИИ И ПОЛУЧЕНИЕ ИЗОБРАЖЕНИЙ





# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## 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

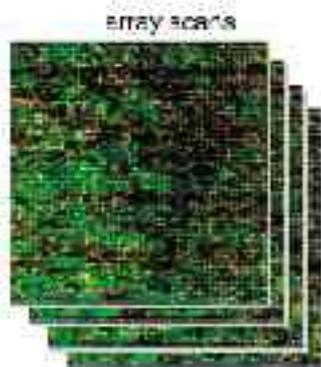


# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

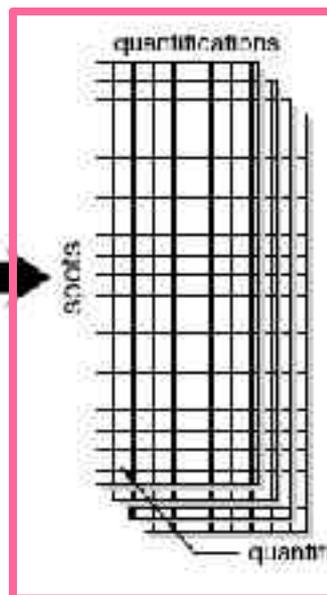


## ОБРАБОТКА ДАННЫХ, ПОЛУЧЕННЫХ МЕТОДОМ ДНК-БИОЧИПОВ

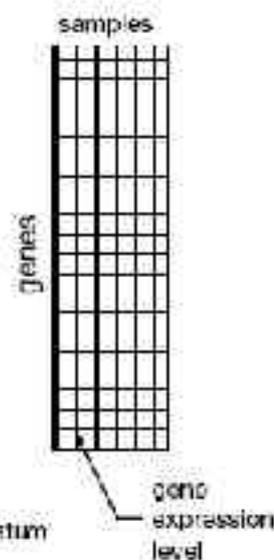
Исходные данные



Матрицы измерений



Матрицы данных по экспрессии генов





# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## 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)  
сигнала гибрида за центром

- 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)



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## 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 биочипов (микроматриц) для исследования транскриптома

## 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 несимметрично.

Поэтому применяют преобразование отношения в виде логарифма по основанию 2 :  $\log_2(R_i / G_i)$  .



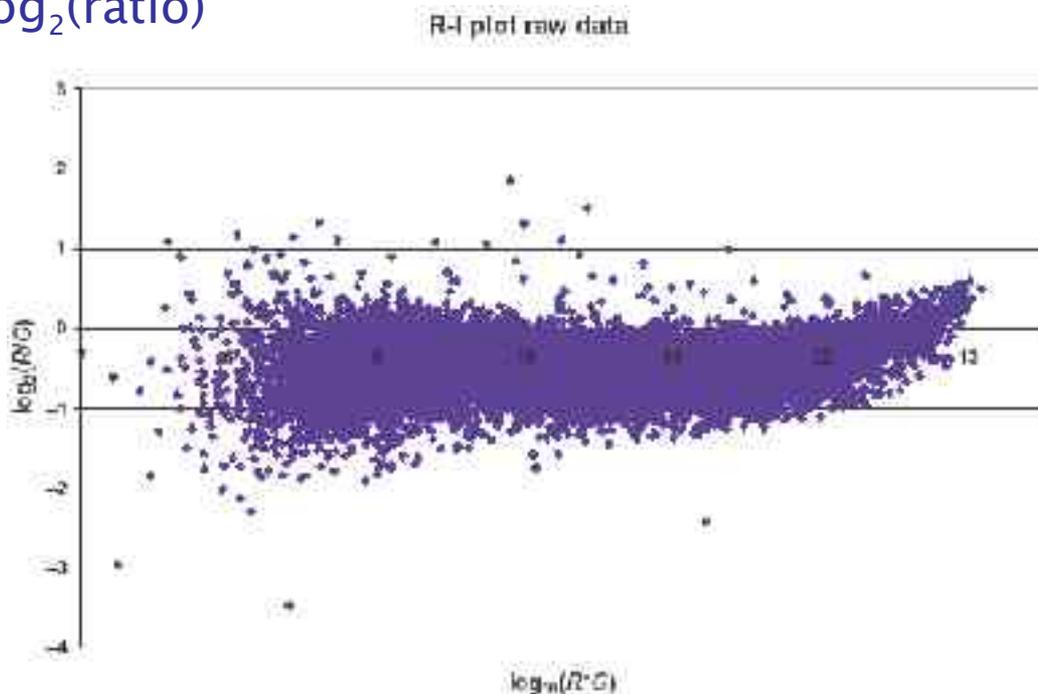
# Применение биочипов (микроматриц) для исследования транскриптома



## ОБРАБОТКА ИЗОБРАЖЕНИЙ *Нормализация.*

Это процесс выявления и исключения систематических ошибок, не связанных с дифференциальной экспрессией

**Генов**  
График 'R-I' (ratio/intensity) может выявить артефакты в измерениях  $\log_2(\text{ratio})$





# Применение биочипов (микроматриц) для исследования транскриптома

## ОБРАБОТКА ИЗОБРАЖЕНИЙ *Нормализация.*

### Нормализация по тотальной интенсивности

Если считать, что примерно одинаковое количество меченых молекул из образцов гибридизуются в пробы в биочипе, то общая интенсивность гибридизации должна быть одинаковой для каждого меченного образца. Нормировочный фактор исчисляется как отношение суммированных интенсивностей по обоим каналам детекции.

$$N_{total} = \frac{\sum_{i=1}^{N_{array}} R_i}{\sum_{i=1}^{N_{array}} G_i}$$

Существуют методы глобальной или локальной нормализации, параметрической и непараметрической нормализации, и т.д.



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

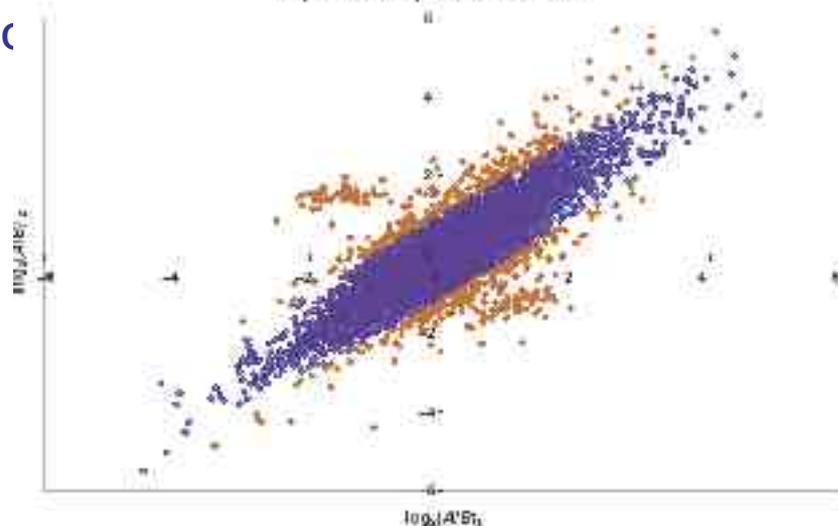
## ОБРАБОТКА ИЗОБРАЖЕНИЙ: *Использование реплик*

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

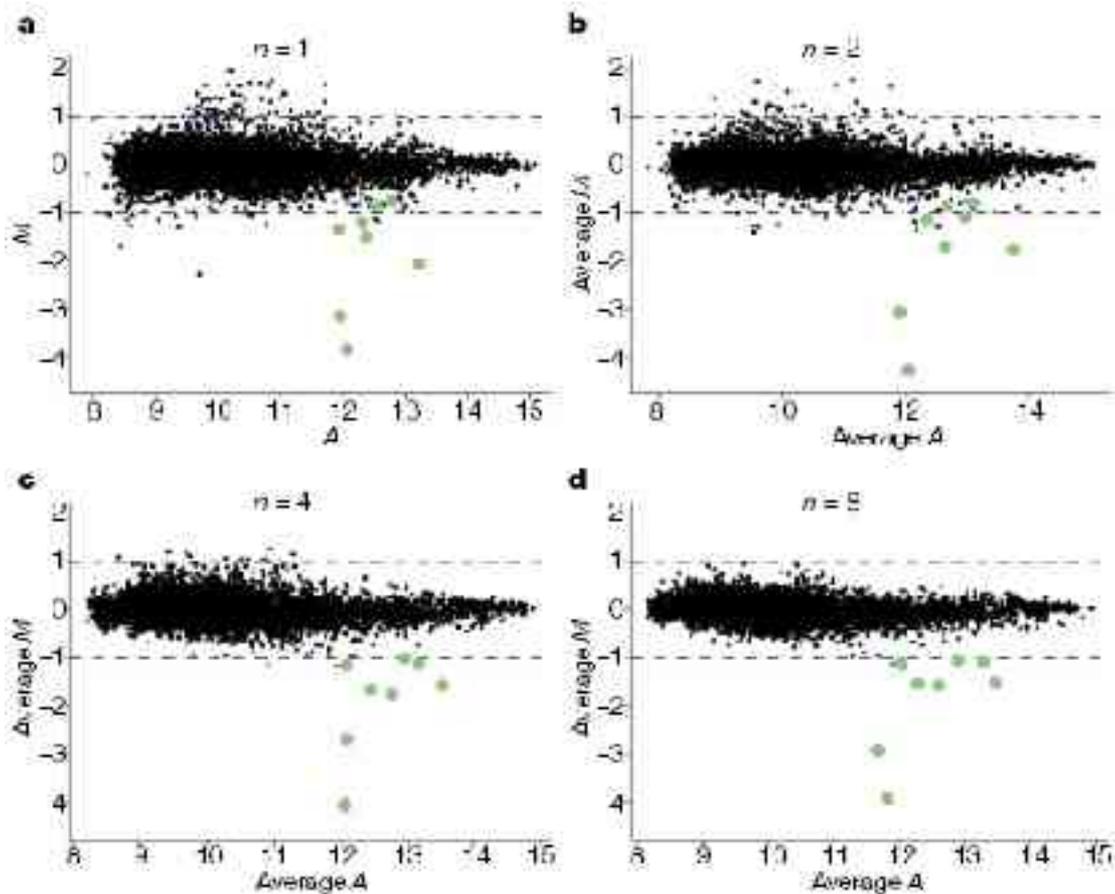
Replicate comparison and trim



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



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



*Биологические реплики.*

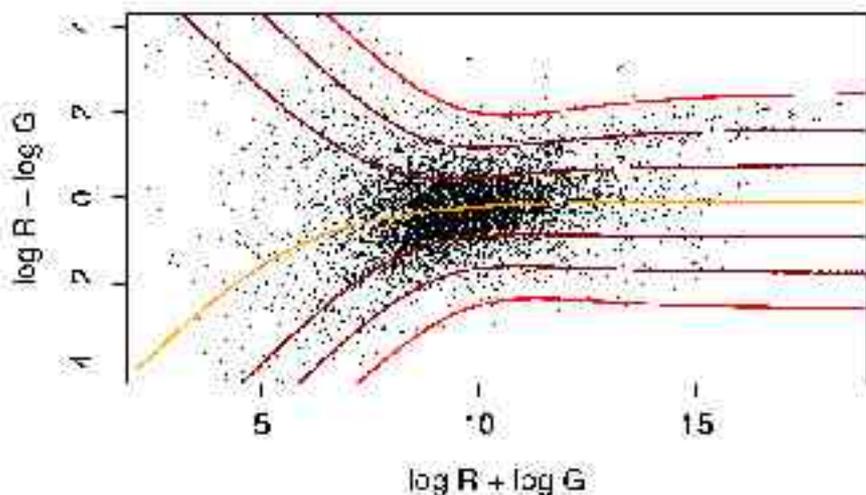
*Гибридизация с образцами РНК от разных особей*



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

**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).

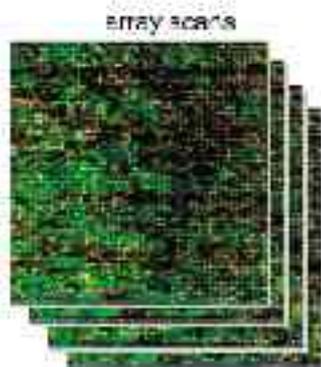


# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

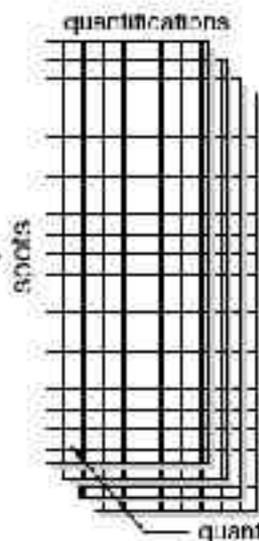


## ОБРАБОТКА ДАННЫХ, ПОЛУЧЕННЫХ МЕТОДОМ ДНК-БИОЧИПОВ

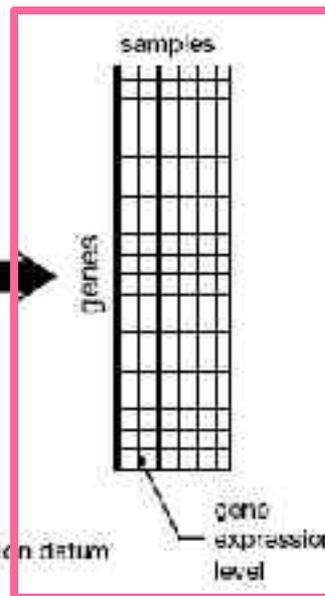
Исходные данные



Матрицы измерений



Матрицы данных по экспрессии генов



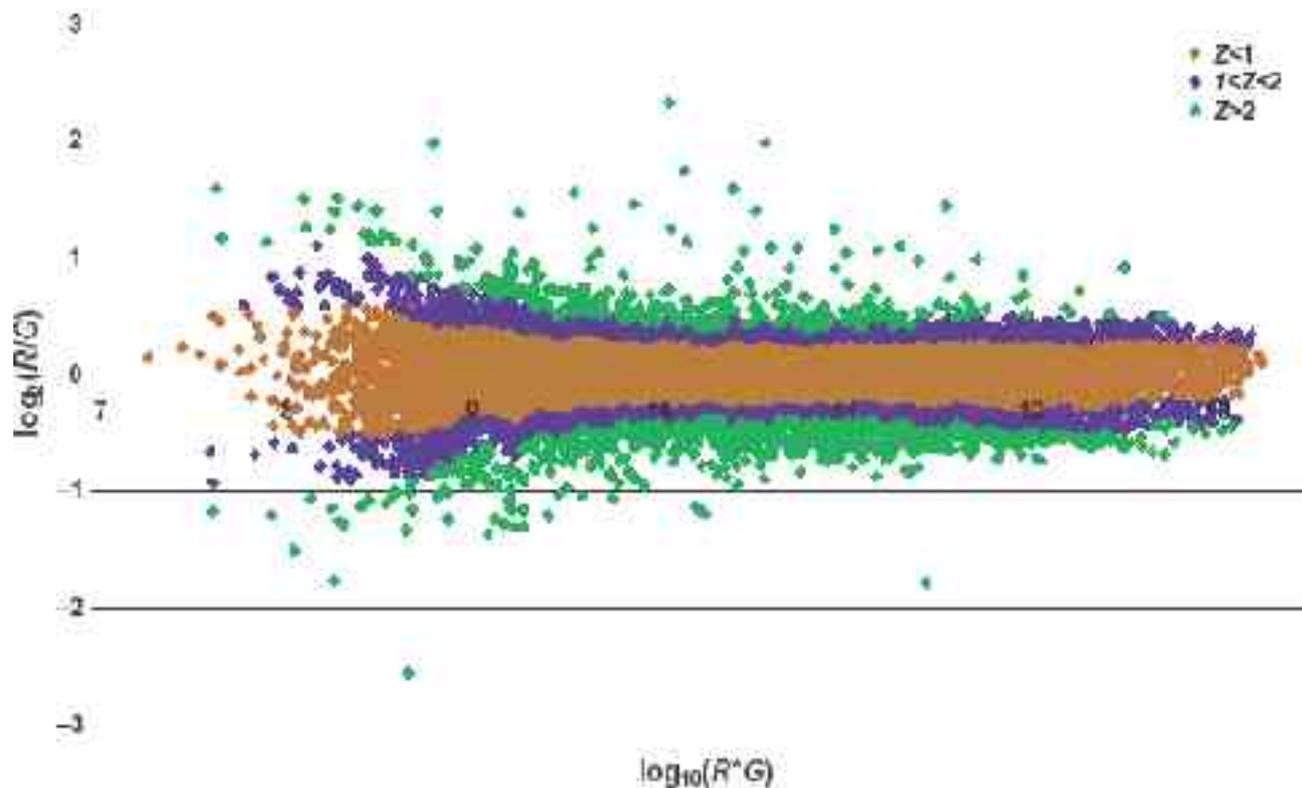


# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## МЕТОДЫ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ

*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 биочипов (микроматриц) для исследования транскриптома

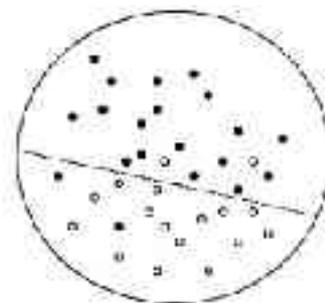
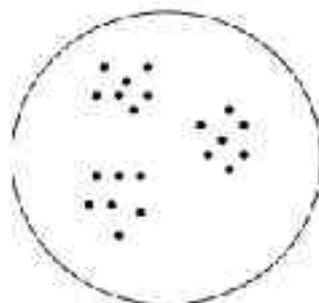


## МЕТОДЫ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ КЛАСТЕРИЗАЦИЯ

алгоритмы кластеризации данных по экспрессии генов

методы безусловной или неконтролируемой классификации

методы контролируемой классификации





# Применение биочипов (микроматриц) для исследования транскриптома



## МЕТОДЫ ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ КЛАСТЕРИЗАЦИЯ

МЕТОДЫ НЕКОНТРОЛИРУЕМОЙ  
КЛАССИФИКАЦИИ

МЕТОДЫ КОНТРОЛИРУЕМОЙ  
КЛАССИФИКАЦИИ

Иерархические алгоритмы  
кластеризации

Неиерархические алгоритмы  
кластеризации

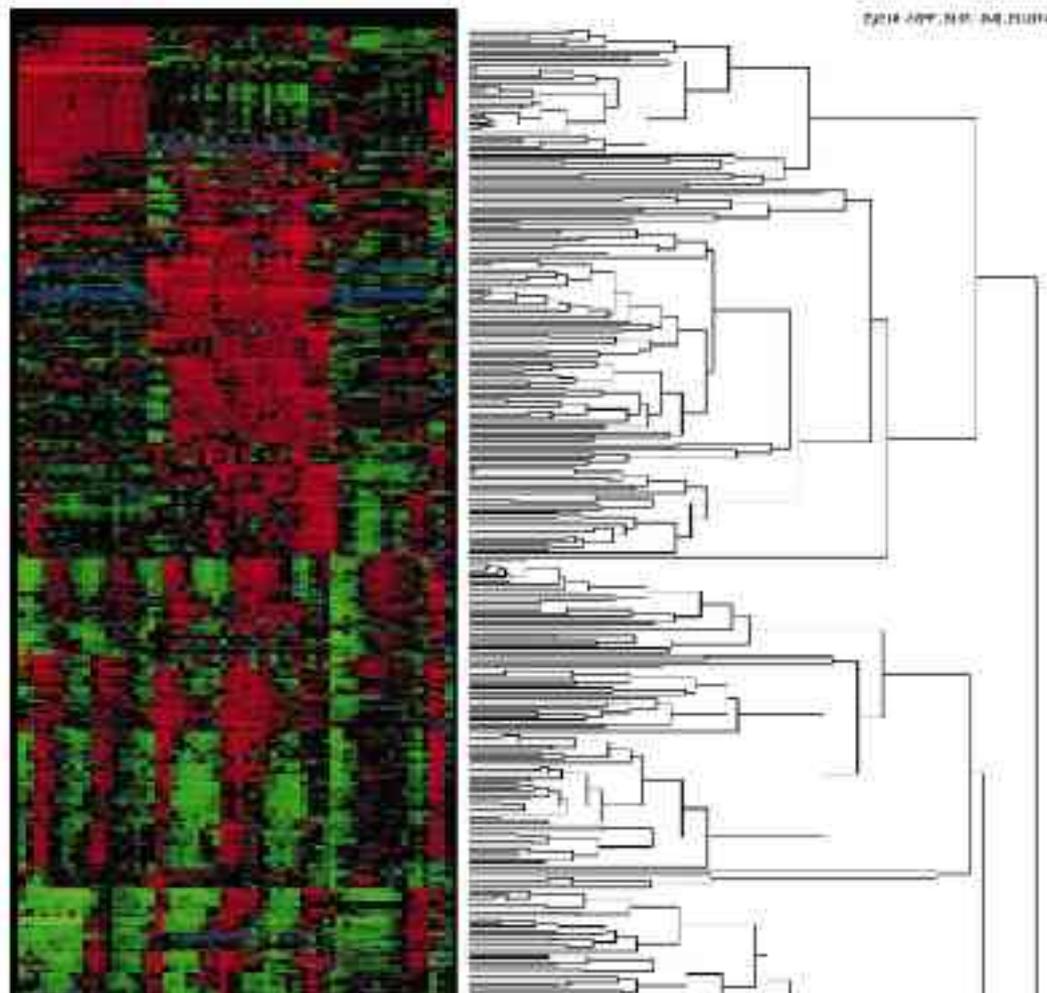
дивизимные      агломеративные

*K*-средних

Самоорганизующиеся  
карты (SOM)



# Применение биочипов (микроматриц) для исследования транскриптома



методы неконтролируемой классификации

Иерархические алгоритмы кластеризации

UPGMA кластеризация генов по экспрессии в течение трех стадий клеточного цикла по результатам измерений по 60 разным временным точкам

GEP (Gene Expression Profiling)  
Профилирование экспрессии генов



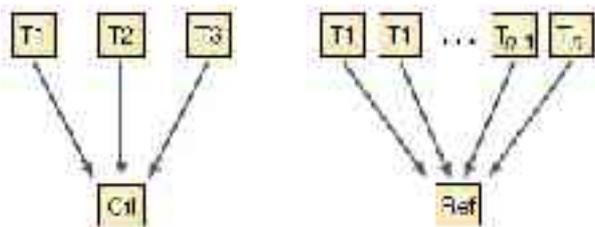
# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



## 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.

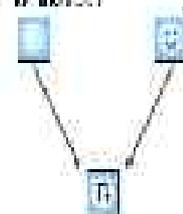


? ??? – ?????? ???????  
Обработка T против контроля C.

a Direct



b Indirect



? ? ? ?????? ?????????? ?????????? ?????? ?<sup>2</sup>/2.

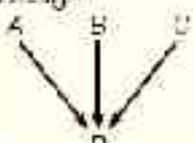
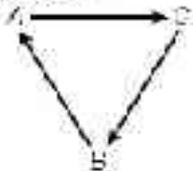
? ? ? ?????? ? ? ?????????? ?????????? ??????  
2<sup>?</sup>



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



## ПЛАНИРОВАНИЕ (ДИЗАЙН) ЭКСПЕРИМЕНТОВ

Design choices	Number of slides	Units of material (number of samples)	Average variance
<i>Indirect designs</i>			
Design 1 	3	$A = B = C = 1$	2.00
Design 2 	3	$A = B = C = 2$	1.00
<i>Direct design</i>			
Design 1 	3	$A - B - C = 2$	0.67

Однофакторный эксперимент

Физические ограничения:  
число стекол и количество исходной мРНК



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)



## ПЛАНИРОВАНИЕ (ДИЗАЙН) ЭКСПЕРИМЕНТОВ

Design choices	t versus t + 1			Comparisons t versus t + 2		t versus t + 3	Average variance
	t <sub>1</sub> /t <sub>2</sub>	t <sub>2</sub> /t <sub>3</sub>	t <sub>1</sub> /t <sub>4</sub>	t <sub>1</sub> /t <sub>3</sub>	t <sub>2</sub> /t <sub>4</sub>	t <sub>1</sub> /t <sub>4</sub>	
Design I – 11 as common reference 	1.00	2.00	2.00	1.00	2.00	1.00	1.0
Design II – direct sequential 	1.00	1.00	1.00	2.00	2.00	5.00	1.67
Design III – common reference 	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Design IV – 11 as common reference 	0.87	0.87	1.57	0.87	1.57	1.00	1.08
Design V – direct hop 	0.75	0.75	0.75	1.00	1.00	0.75	0.83
Design V – direct mbec 	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



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

**MGED Home**

Home Meetings Workgroups Mission MGED Board Site Map

## Microarray Gene Expression Data Society - MGED Society

The Microarray Gene Expression Data (MGED) Society is an international organization of biologists, computer scientists, and data analysts that aims to facilitate the sharing of microarray data generated by functional genomics and proteomics experiments.

The current focus is on establishing standards for microarray data annotation and exchange, facilitating the creation of microarray databases and related software, implementing these standards, and promoting the sharing of high quality, well annotated data within the life sciences community. A long term goal for the future is to extend this mission to other functional genomics and proteomics high-throughput technologies.

Read more about:  
[MGED Sponsors](#) • [Defined MGED Standards](#) • [Historical highlights](#) • [MGED Meetings](#) • [MGED Supported Meetings](#) • [Programming Jamboree](#) • [MGED Goals](#) • [MGED Workgroups](#) • [Relevant Publications](#) • [MGED Board](#)

Latest News:

Publications  
News Archive  
Sponsors

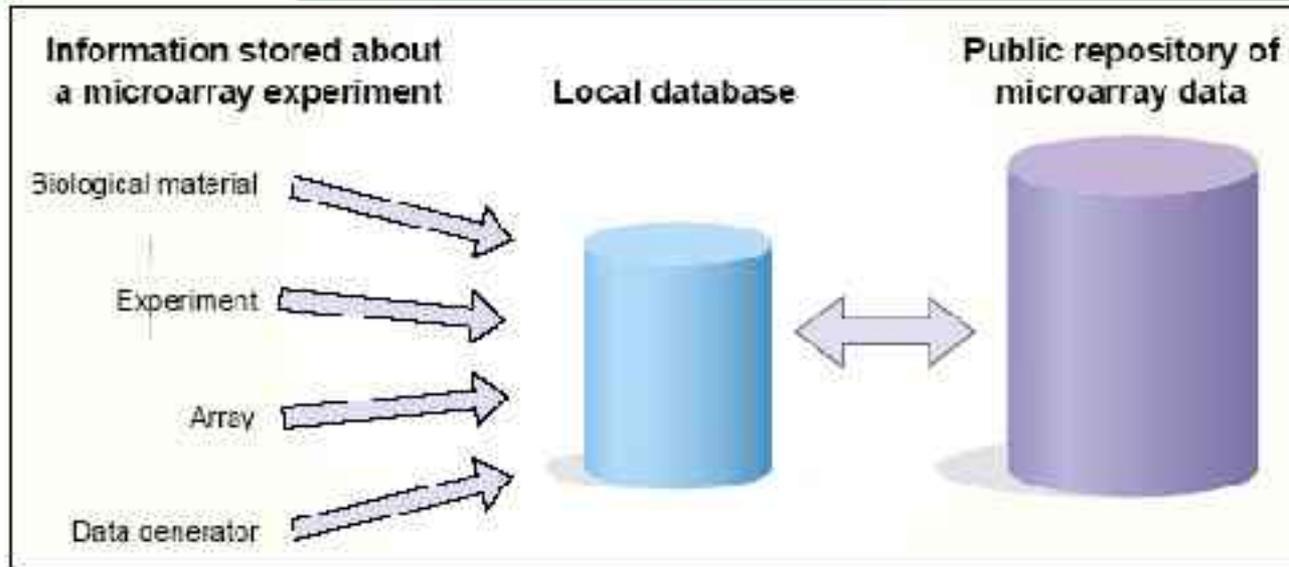
NIAME  
MAGE  
IITN/INGRA (ITWG)  
Transformations  
RSBI WG  
MISFISHIE

OMG  
Sourceforge

MGED 8 Meeting

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.

# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

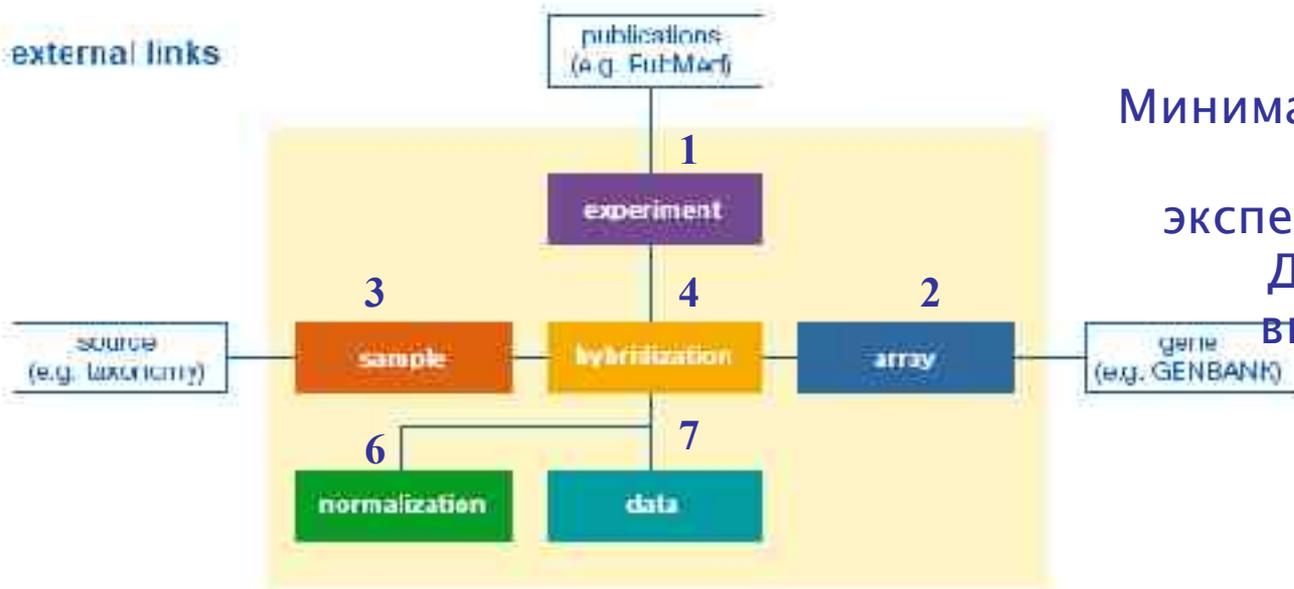


- MicroArray Gene Expression Markup Language (MAGE-ML) – создан для создания общего формата, чтобы достичь сравнимости результатов
- Minimum Information About a Microarray Experiment (MIAME) – создан для определения типа информации и степени подробности, с которой исследователь обязан ее представить;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



# Методы массового исследования транскриптов. ДНК-биочипы (микроматрицы)

## Minimum Information About a Microarray Experiment (MIAME)



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

1. План эксперимента – набор отдельных гибридизационных экспериментов
2. План биочипа – содержание пятен/ячеек, компоновка по рядам и т.д.
3. Образцы – источник, приготовление экстрактов, способ мечения
4. Гибридизация – процедура и параметры
5. Измерение – характеристики изображений и сканнеров
6. Нормализация – алгоритмы и параметры



# Применение ДНК-биочипов (микроматриц) для исследования экспрессии генов



Спасибо за внимание.

