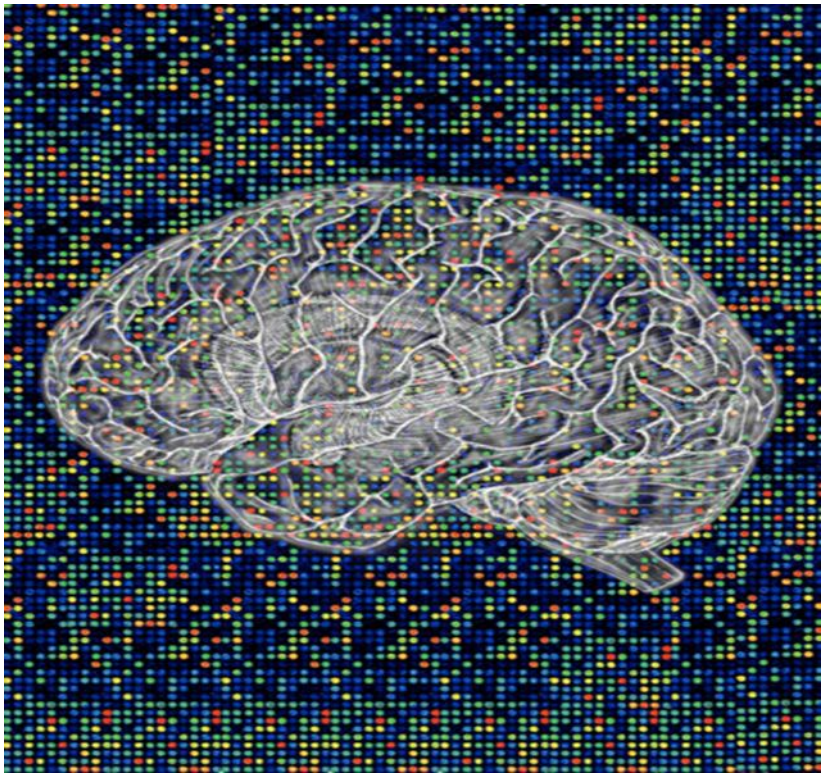
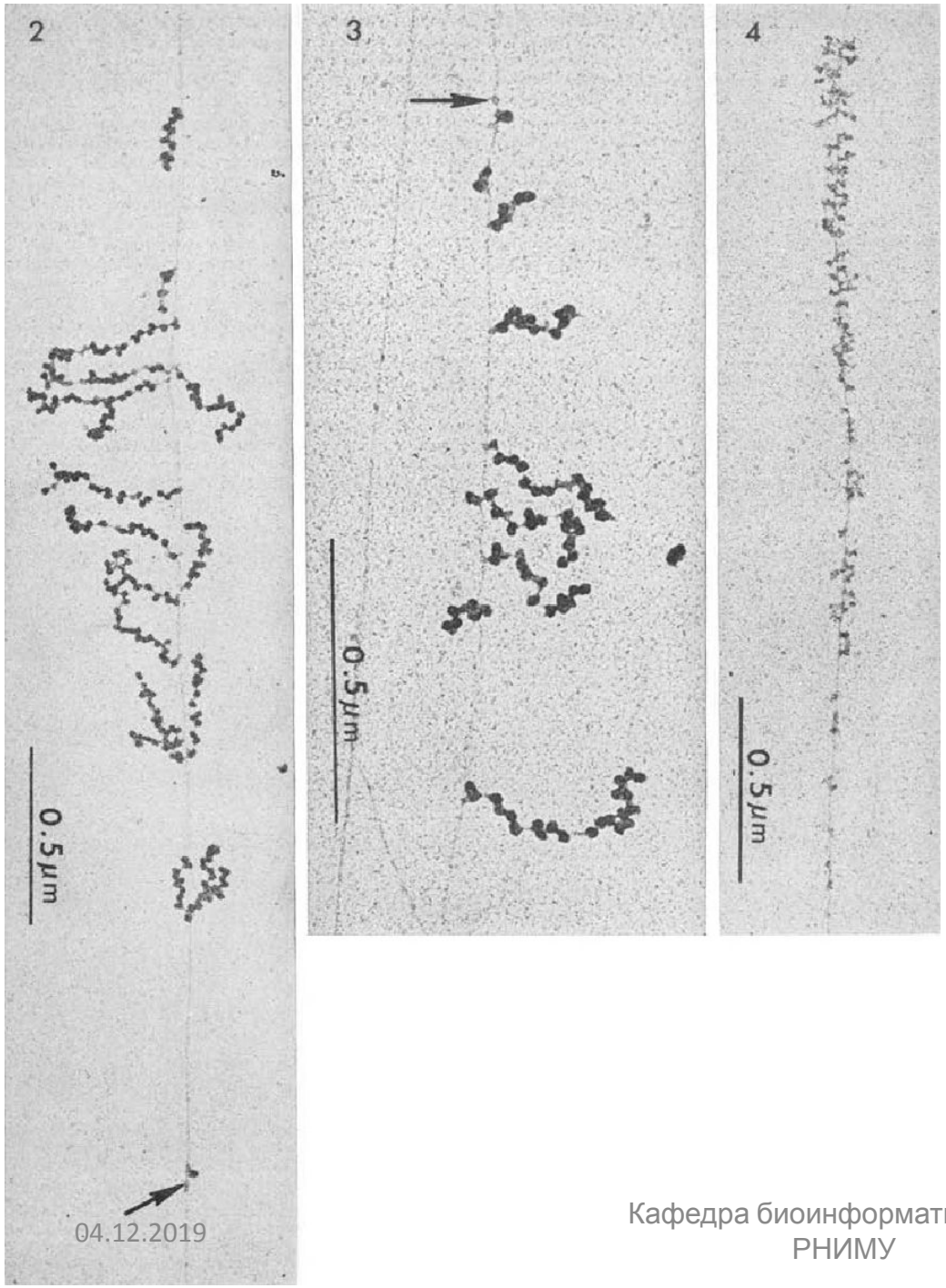
Биоинформатические подходы к анализу РНК. Экспрессия генов: анализ микроэррейных данных

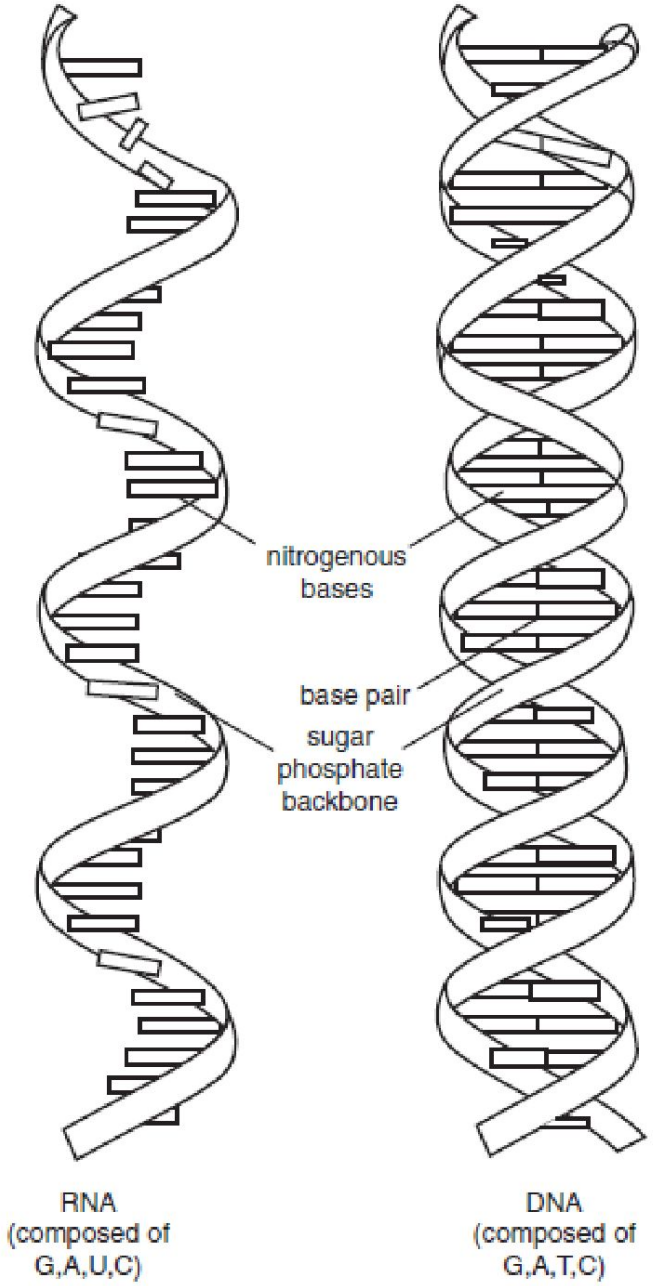


Лекция 8

Многие слайды и материалы используемые в презентации взяты из книги Bioinformatics and Functional Genomics by Jonathan Pevsner Copyright © 2009 by John Wiley & Sons, Inc. и соответствующего курса по биоинформатики Johns Hopkins School of Medicine



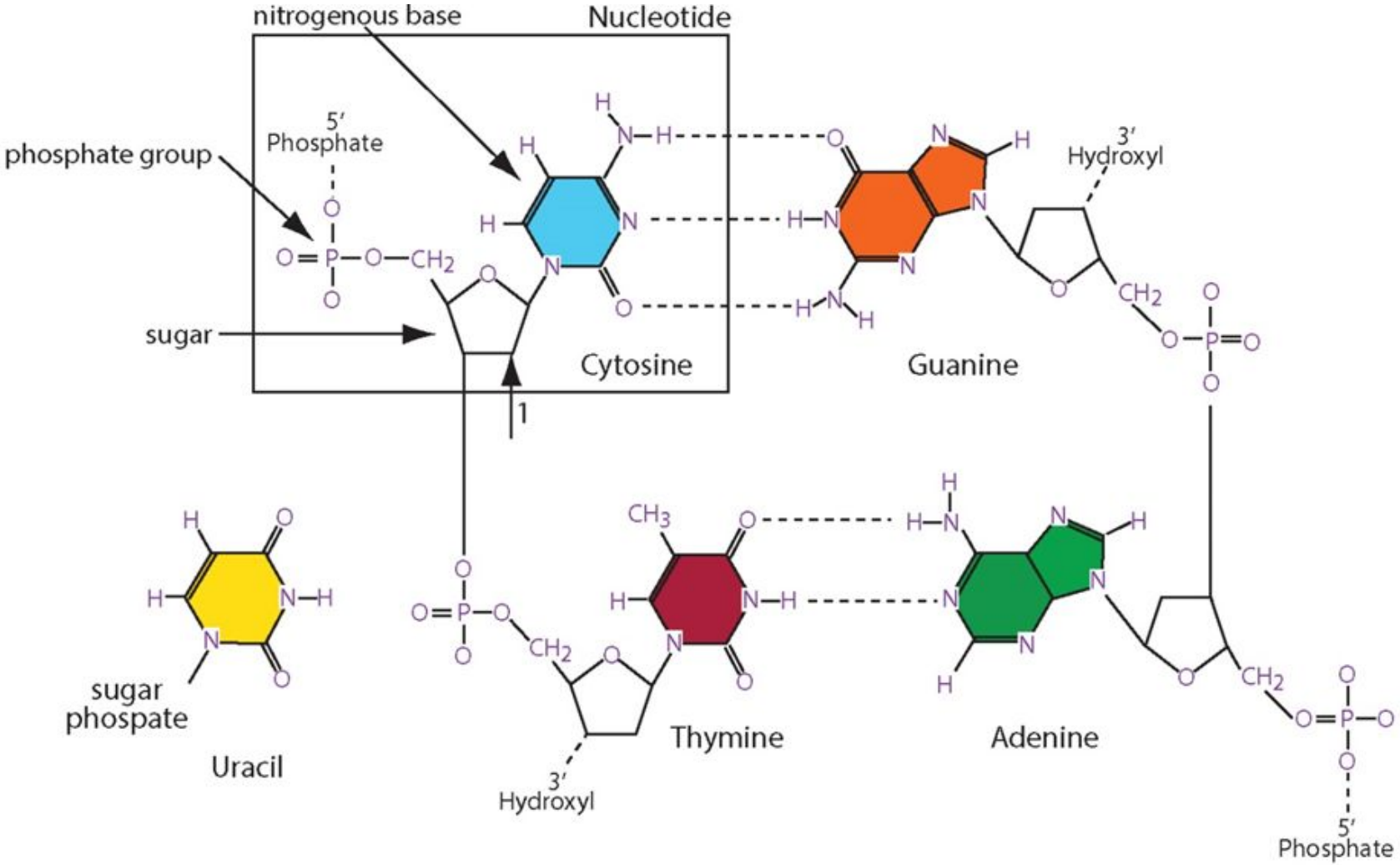
Миллер с коллегами (1970, p. 394) визуализировали генную экспрессию. Они показали хромосомную ДНК Escherichia coli в процессе транскрипции и трансляции. Темные структуры – полирибосомы на мРНК.



Дезоксирибонуклеиновая кислота (ДНК) и рибонуклеиновая кислота (РНК).

В то время как ДНК обычно принимает конформацию двойную спиральную, РНК, как правило, одноцепочечна. Заметным исключением является двухцепочечная структура некодирующих РНК, формирующих структуру в виде шпильки.

Нуклеотидные остатки

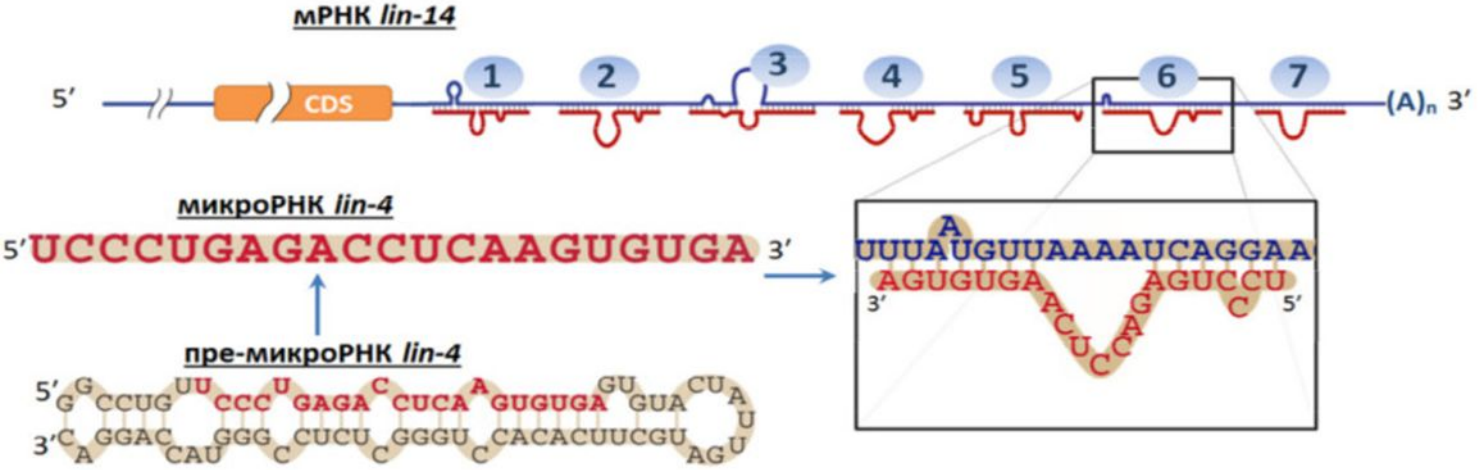


Некодирующие РНК

Несколько ключевых дат:

- 1993 г: демонстрация нового механизма подавления экспрессии:

C.elegans: гены lin-4 и lin-14 (белок LIN-14) антагонисты, но lin-4 не кодирует белок.



- 1999 г: открытие малых интерферирующих РНК (doi: 10.1126/science.286.5441.950)
- 2000 г: открытие второй микроРНК - lin-7 (doi: 10.1038/35002607)
- 2001 г: открытие сотни новых малых РНК, появление термина микроРНК (doi: 10.1126/science.1065329)

Первоначально все некодирующие РНК считали малыми; дальнейшие исследования выявили существенные различия по размерам и по функциям.

Некодирующие РНК (нкРНК)

Большое число и разные функции нкРНК

Классификация по размерам:
малые нкРНК (sncRNA) - размер до 200 нт (в н.в. известны тысячи)
длинные нкРНК (lncRNA, lincRNA) - размер 200-50 тыс. нт (в н.в. известны десятки тысяч)
очень длинные нкРНК (vlincRNA) - размер от 50 тыс.нт до ~700 тыс.нт (в н.в. известны ~2000-3000) (doi: 10.1186/gb-2013-14-7-r73)

Малые нкРНК - высококонсервативны, длинные нкРНК - низкоконсервативные.

Наиболее изучены следующие малые нкРНК:
- малые интерферирующие РНК;
- малые ядерные РНК;
- малые ядрышковые РНК;
- малые РНК, образующие комплексы с piwi-белками;
- малые РНК, образующие шпильки;
- микроРНК.

Малые некодирующие РНК

Наибольшее функциональное значение имеют малые РНК, вовлеченные в процессы генной регуляции - малые интерферирующие РНК (siRNA) и микроРНК (miRNA):

- короткие интерферирующие РНК - длина 20-25 нт
- микроРНК - длина 18-24 нт

- ~60% генов человека регулируются микроРНК (doi:10.1016/j.cell.2004.12.035);
- 2014 г: аннотировано ~1900 микроРНК человека; общее число микроРНК может достичь десятков тысяч;
- огромное разнообразие вариантов регуляции:
  - одна miRNA -> несколько mRNA
  - несколько miRNA -> одну mRNA
  - влияние степени комплементарности

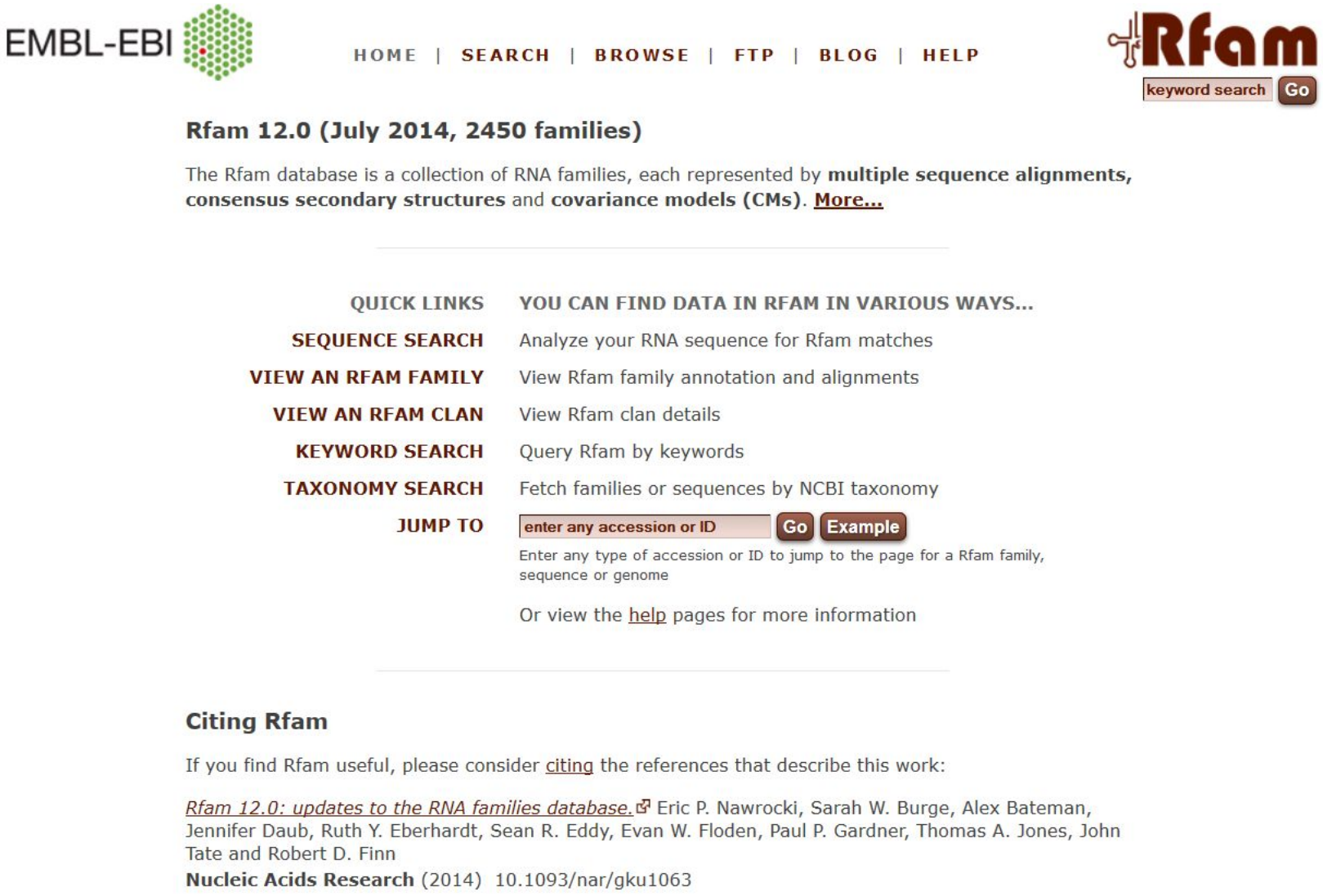
Существует значительное число биоинформационных программ для поиска микроРНК и их генов-мишеней
=> необходимо совершенствование расчетных методов поиска микроРНК, их генов-мишеней и других регуляторных участков

Family	Start	End	Bits score
tRNA	9,734,391	9,734,325	31.22
RSV_RNA	9,990,192	9,989,909	36.74
RSV_RNA	10,142,311	10,142,595	36.83
Metazoa_SRP	10,380,661	10,380,378	122.56
SNORA70	10,385,953	10,386,047	42.25
tRNA	10,492,972	10,492,907	26.05
tRNA	10,493,037	10,492,973	37.46
mir-548	11,052,015	11,051,932	82.85
U6	14,419,904	14,420,010	66.41
U6	14,993,898	14,994,004	76.41
U6	15,340,916	15,340,810	63.69
5S_rRNA	15,443,192	15,443,307	42.69
pRNA	15,448,359	15,448,271	68.52
U6	16,986,602	16,986,708	75.19
U6	17,407,829	17,407,733	41.70
SNORD74	17,657,089	17,657,017	59.88
mir-10	17,911,414	17,911,485	69.08
let-7	17,912,152	17,912,227	62.65
lin-4	17,962,567	17,962,636	76.22
U1	18,091,317	18,091,476	91.09
U6	18,803,865	18,803,965	62.92
tRNA	18,827,177	18,827,107	63.87
Metazoa_SRP	18,878,771	18,879,046	64.51
Y_RNA	18,899,565	18,899,458	41.00
Y_RNA	18,949,116	18,949,224	40.22
RSV_RNA	19,938,102	19,937,818	72.79
U1	20,717,465	20,717,629	93.53
U6	21,728,164	21,728,060	49.35
7SK	21,728,965	21,729,208	75.15
mir-492	21,798,181	21,798,066	40.04
U4	23,577,511	23,577,651	73.90
U2	24,654,231	24,654,058	62.66

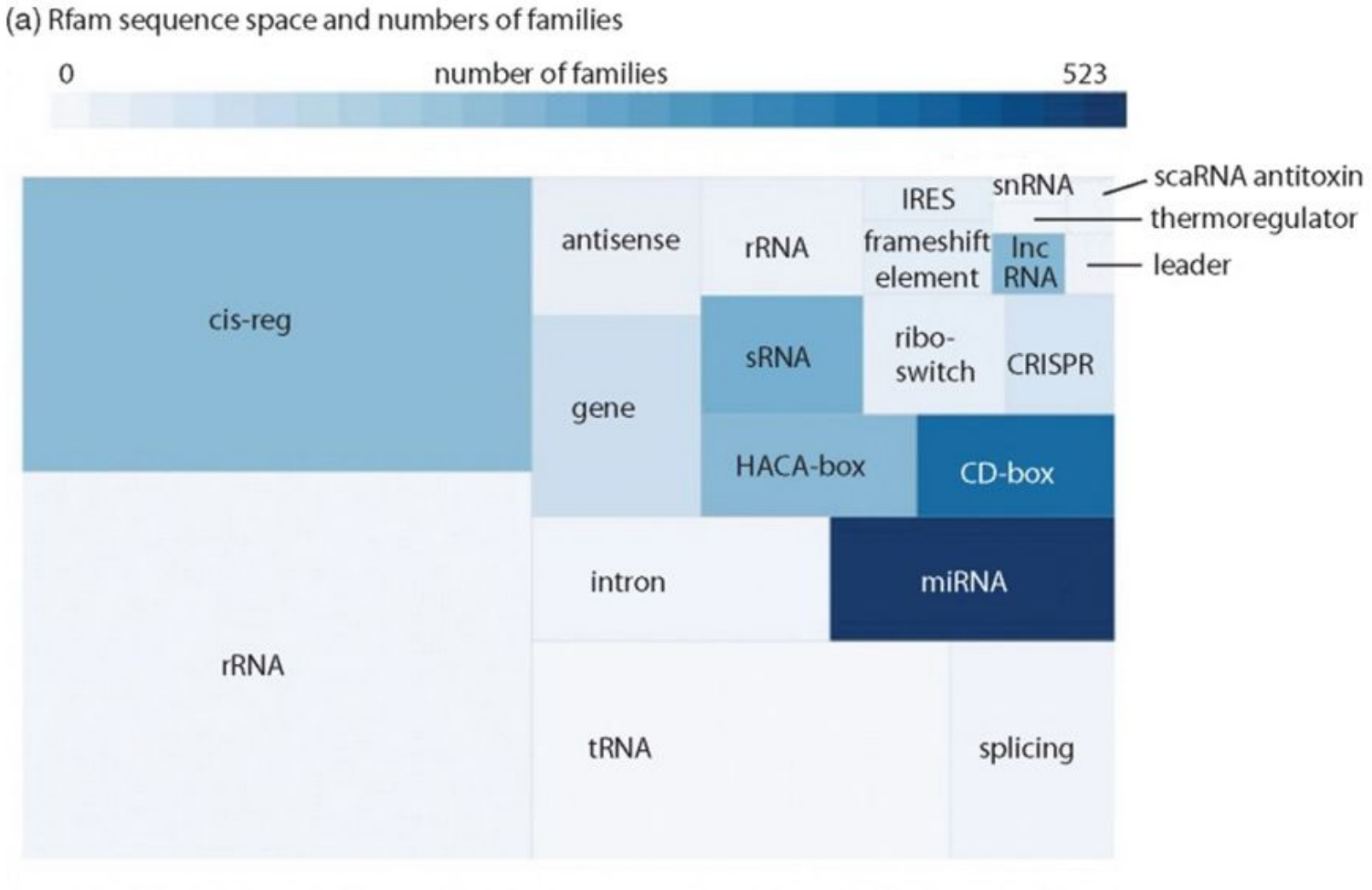
FIGURE 10.4 Noncoding RNA families in the Rfam database that are assigned to human chromosome 21. Only a portion of the entries is shown.

МикроРНК	Кол-во генов-мишеней	Гены-мишени	Функции генов-мишеней
hsa-miR-335-5p	15	NOS3, F13B, CYP19A1, LIF, PLAT, ITGA2, ACE, IL6, ESR2, ASPN, NASE4, SERPINB9, KCNH2, RAB11FIP1, LDLR	Регуляция систем свертывания крови, синтеза эстрогенов и иммунных взаимодействий мать-плод; регуляция хондрогенеза, ангиогенеза, поддерживание разности потенциалов между внешней и внутренней мембраной клеток; формирование, ориентация и слияние внутриклеточных транспортных везикул, синтез холестерина.
hsa-miR-124-3p	13	IL11, IL6, F13B, AGTR1, GTR2, ASPN, ELF3, IGFBP3, CALR, SNAI2, LDLR, CALU, SYPL1	Регуляция клеточного цикла, артериального давления и системы свертывания крови, регуляция хондрогенеза, синтез холестерина кодирование инсулиноподобного фактора роста, регуляция транскрипции генов рецепторов гормонов, фолдинг белков
hsa-miR-26b-5p	11	AGT, AGTR1, LIF, F13B, GSTP1, ASPN, CXCR4, TK1, RAB11FIP1, C1GALT1C1, COL5A1	Регуляция артериального давления, систем свертывания крови, детоксикации и иммунных взаимодействий мать-плод; регуляция хондрогенеза; формирование, ориентация и слияние внутриклеточных транспортных везикул, синтез коллагена

Rfam – база данных (http://rfam.xfam.org/)



Некодирующие РНК



Некодируещие РНК

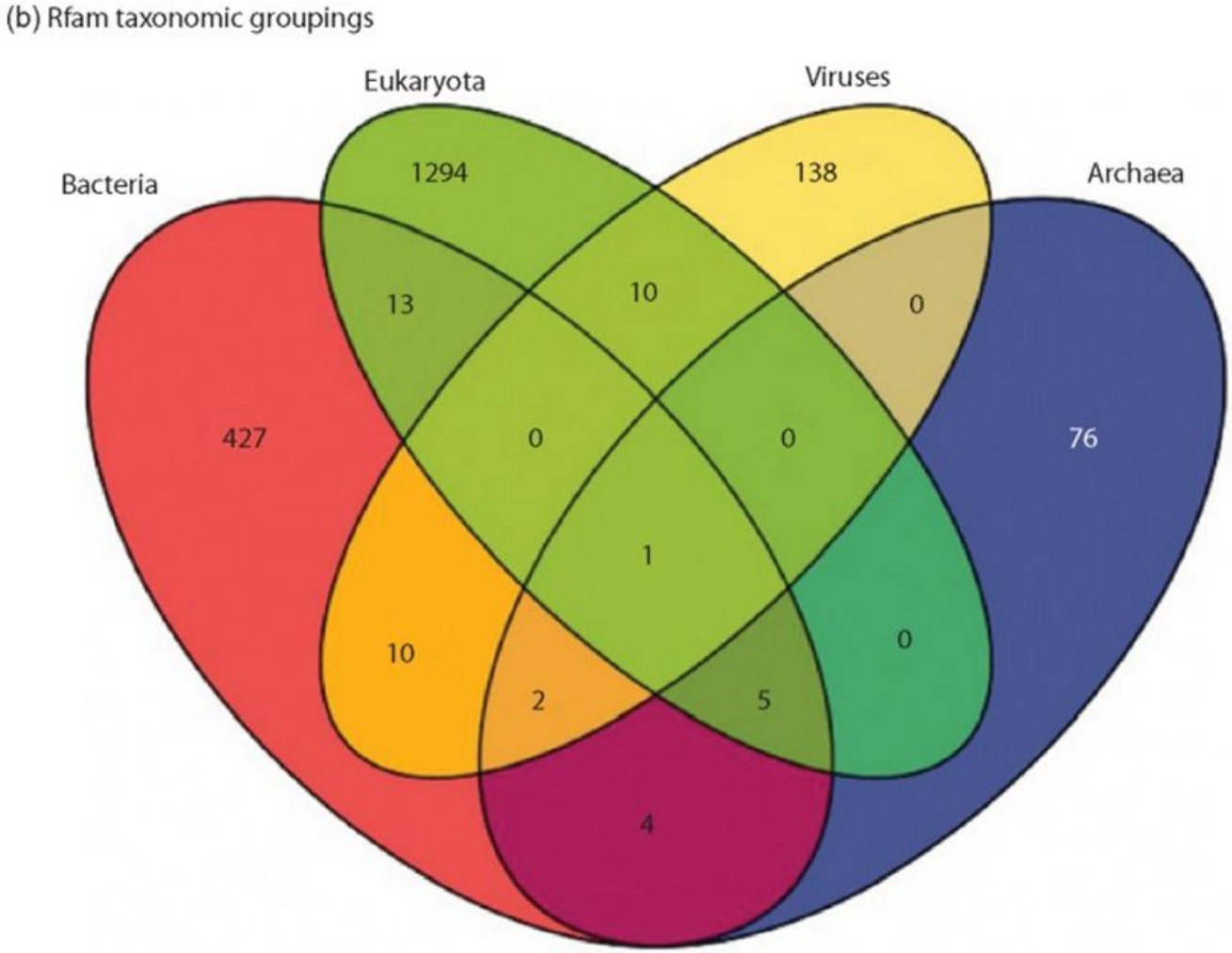


TABLE 10.1 A list of the 13 Rfam entries with the largest number of members. No. full: number of members of the Rfam family (for the full dataset rather than the seed alignment of representative members), rounded to the nearest thousand; Id: the average percent identity of the full alignments.

Name	Accession	No. full	Ave. len. (full)	Id	Type	Description
5_8S_rRNA	RF00002	376,000	152	69	Gene; rRNA	5.8S ribosomal RNA
tRNA	RF00005	298,000	73	46	Gene; tRNA	tRNA
5S_rRNA	RF00001	229,000	116	60	Gene; rRNA	5S ribosomal RNA
UnaL2	RF00436	101,000	54	78	Cis-reg	UnaL2 LINE 3′ element
HIV_POL-1_SL	RF01418	83,000	113	77	Cis-reg	HIV pol-1 stem loop
U6	RF00026	72,000	105	77	Gene; snRNA; splicing	U6 spliceosomal RNA
mtDNA ssA	RF01853	62,000	104	67	Gene; antisense	Mitochondrial DNA control region secondary structure A
Intron_gpI	RF00028	60,000	365	36	Intron	Group I catalytic intron
Intron_gpII	RF00029	51,000	87	54	Intron	Group II catalytic intron
Hammerhead_1	RF00163	49,000	59	70	Gene; ribozyme	Hammerhead ribozyme (type I)
RRE	RF00036	44,000	337	97	Cis-reg	HIV Rev response element
HIV_GSL3	RF00376	39,000	84	82	Cis-reg	HIV gag stem loop 3 (GSL3)
SNORA7	RF00409	26,000	140	79	Gene; snRNA; snoRNA; HACA-box	Small nucleolar RNA SNORA7

Source: Rfam 11.0. Reproduced under the Creative Commons Zero licence, CC0.

TABLE 10.2 Summary of the number of tRNA genes in selected organisms. The "other" category refers to selenocysteine tRNAs (TCA), suppressor tRNAs (CTA,TTA), or tRNAs with undetermined or unknown isotypes. Additionally, some organisms have tRNAs with introns (e.g., human, 32; *P. falciparum*, 1; *Arabidopsis*, 83).

Organism	Common name	No. tRNAs decoding the 20 amino acids	No. predicted pseudogenes	Other	Total
Homo sapiens	Human	506	110	9	625
Pan troglodytes	Chimpanzee	456	0	3	459
Mus musculus	Mouse	432	0	3	435
Canis familiaris	Dog (Canfam1)	898	0	8	906
Drosophila melanogaster	Fruit fly	298	4	2	304
Saccharomyces cerevisiae	Baker's yeast	286	6	3	295
Arabidopsis thaliana	Plant	630	8	1	639
Plasmodium falciparum	Malaria parasite	35	0	0	35
Methanococcus jannaschii	Archaeon	36	0	1	37
Escherichia coli K12	Bacterium	86	1	1	88
Mycobacterium leprae	Bacterium	45	0	0	45

Source: http://genome.ucsc.edu, courtesy of UCSC.

тРНК

(a)

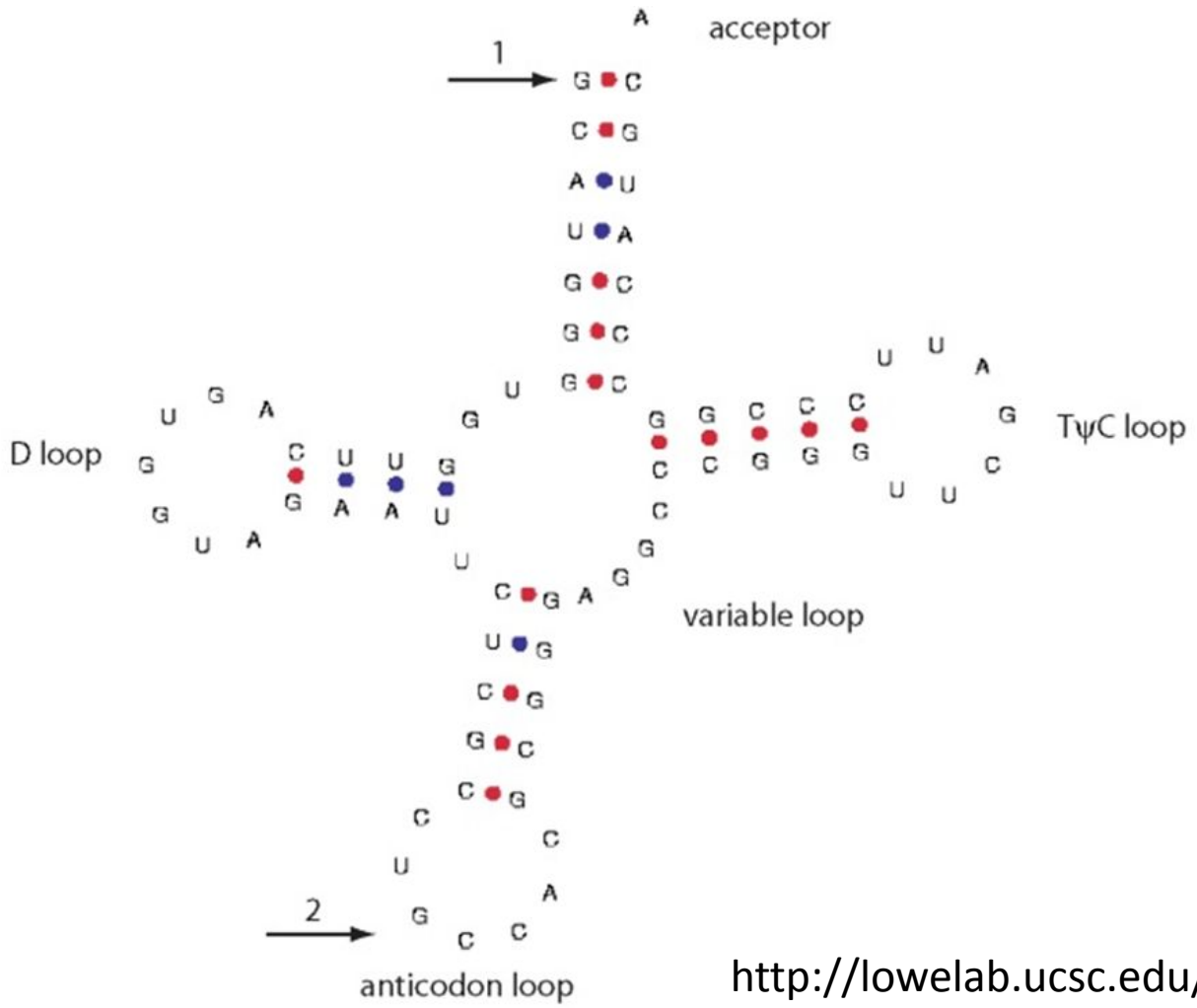
```
Isotype / Anticodon Counts:

Ala    : 0        AGC:        GGC:        CGC:        TGC:
Gly    : 1        ACC:        GCC: 1      CCC:        TCC:
Pro    : 0        AGG:        GGG:        CGG:        TGG:
Thr    : 0        AGT:        GGT:        CGT:        TGT:
Val    : 0        AAC:        GAC:        CAC:        TAC:
Ser    : 0        AGA:        GGA:        CGA:        TGA:        ACT:        GCT:
Arg    : 0        ACG:        GCG:        CCG:        TCG:        CCT:        TCT:
Leu    : 0        AAG:        GAG:        CAG:        TAG:        CAA:        TAA:
Phe    : 0        AAA:        GAA:
Asn    : 0        ATT:        GTT:
Lys    : 0                                CTT:        TTT:
Asp    : 0        ATC:        GTC:
Glu    : 0                                CTC:        TTC:
His    : 0        ATG:        GTG:
Gln    : 0                                CTG:        TTG:
Ile    : 0        AAT:        GAT:                    TAT:
Met    : 0                                CAT:
Tyr    : 0        ATA:        GTA:
Supres: 0                                 CTA:        TTA:
Cys    : 0        ACA:        GCA:
Trp    : 0                                CCA:
SelCys: 0                                             TCA:
```

(b)

```
Your-seq.trna1 (1-71)   Length: 71 bp
Type: Gly       Anticodon: GCC at 33-35 (33-35) Score: 71.03
         *    |    *    |    *    |    *    |    *    |    *    |    *    |
Seq: GCATGGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCGATTCCCGGCCCATGCA
Str: >>>>>>>..>>>>.......<<<<.>>>>>.......<<<<<....>>>>>.......<<<<<<<<<<<.
```

http://lowelab.ucsc.edu/tRNAscan-SE/

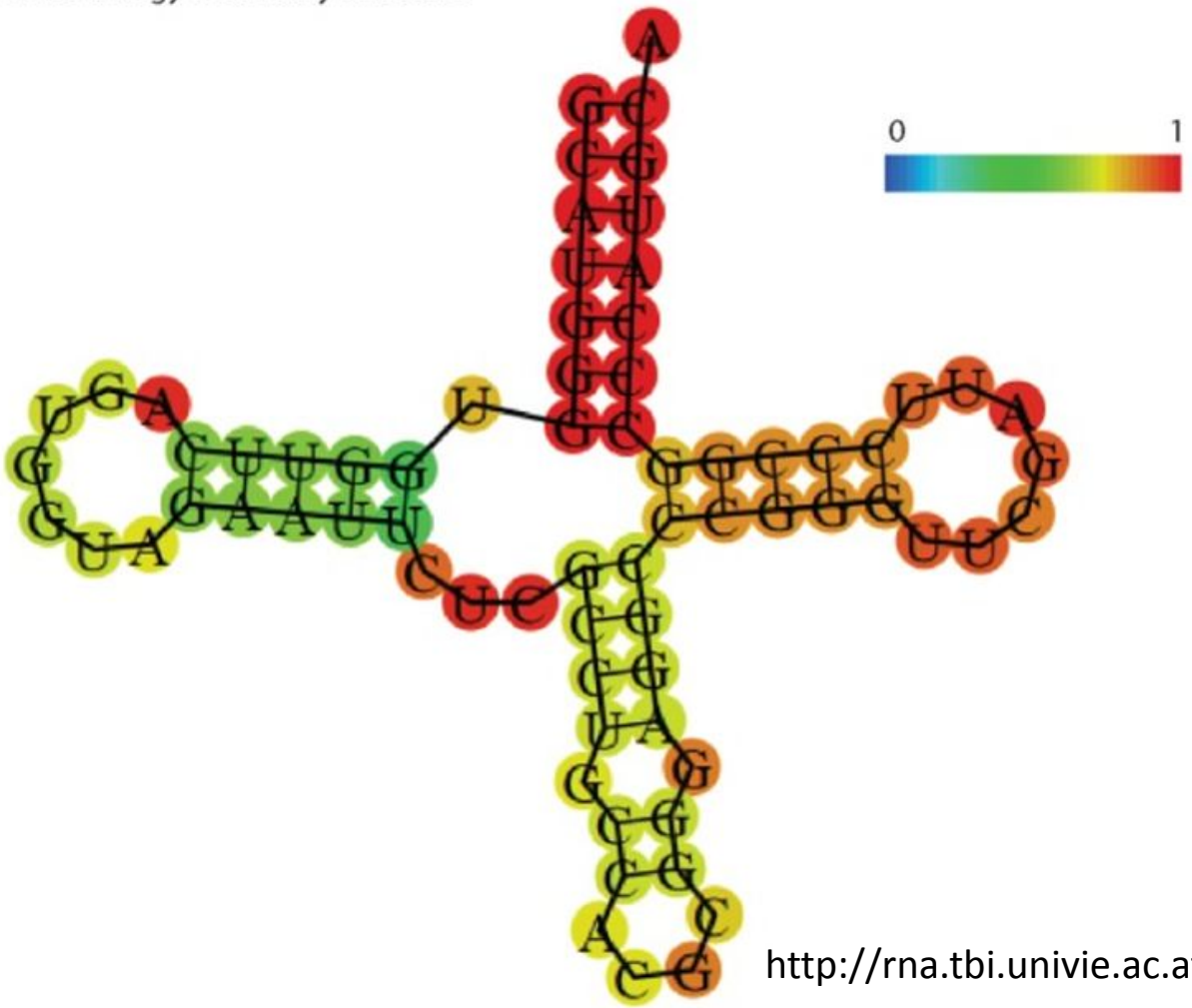
тРНК



http://lowelab.ucsc.edu/tRNAscan-SE/

(a) Minimum free energy prediction (colored by base pairing probability): analyzing a tRNA at with Vienna 2.0

(b) Minimum free energy secondary structure

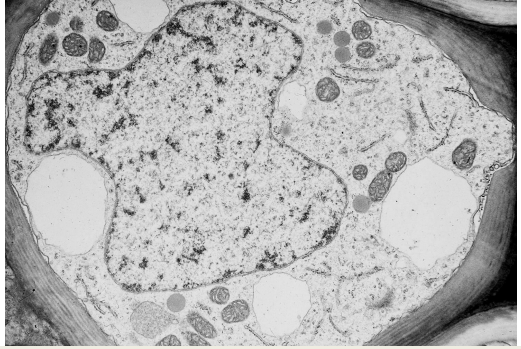


http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

Экспрессия генов: анализ микроэрейных данных

- Экспрессия генов
- Микрочипы (Microarrays)
- Предварительная обработка
  - нормализация
  - диаграммы рассеяния
- Статистический анализ
  - Т-тест
  - ANOVA
  - расстояния
  - кластеризация
  - анализ главных компонентов (PCA)

Сравнение экспрессии генов в этом типе клеток …



…после вирусной инфекции

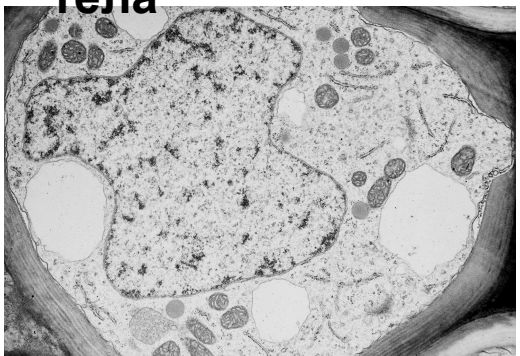
… по отношению к животному с нокаут геном

…с образцами от пациентов

…после применения лекарства

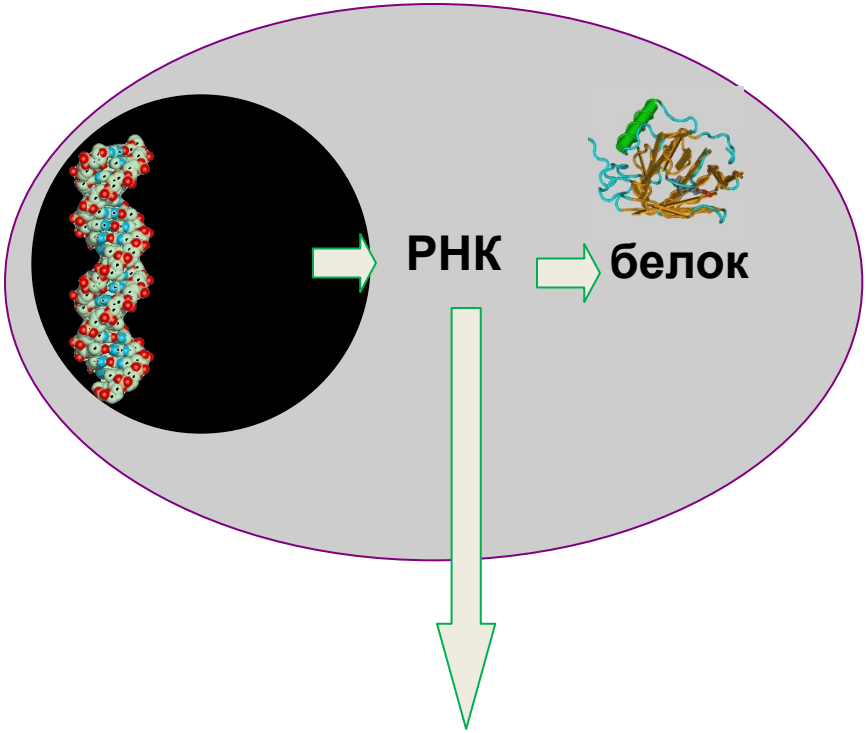
…в разное время жизни

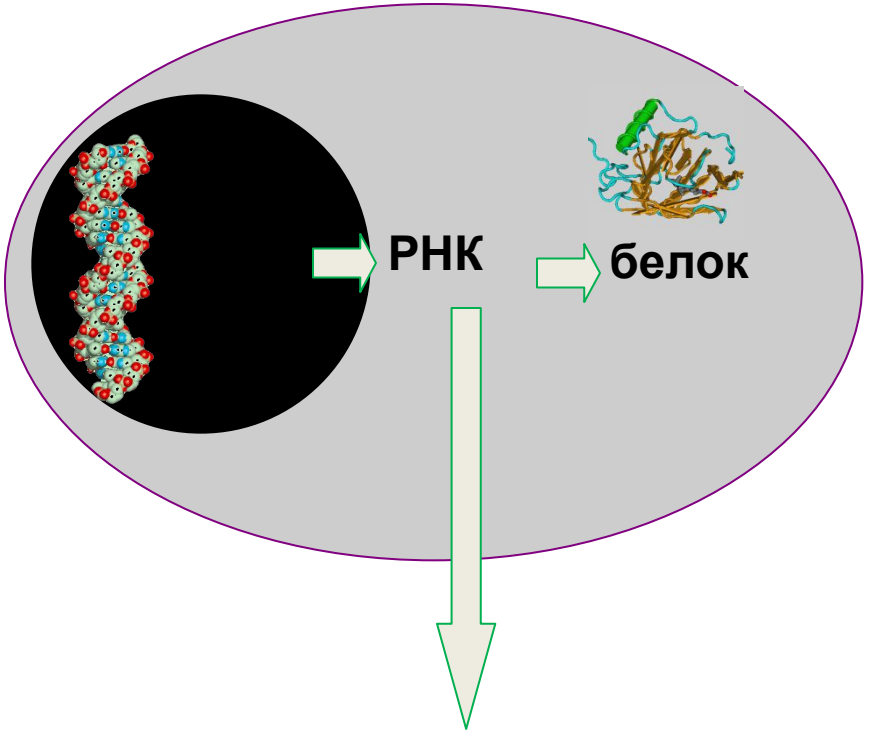
…в различных участках тела



Экспрессия генов зависит от контекста, и регулируется несколькими основными способами

- по тканям и органам (например, мозг по сравнению с почкой)
- в процессе развития (например плода по сравнению с взрослой ткани)
- в динамическом ответ на сигналы внешней среды (например, лекарств)
- при патологических состояниях
- с помощью активности других генов
- эпигенетически





cDNA ⟷ cDNA

UniGene

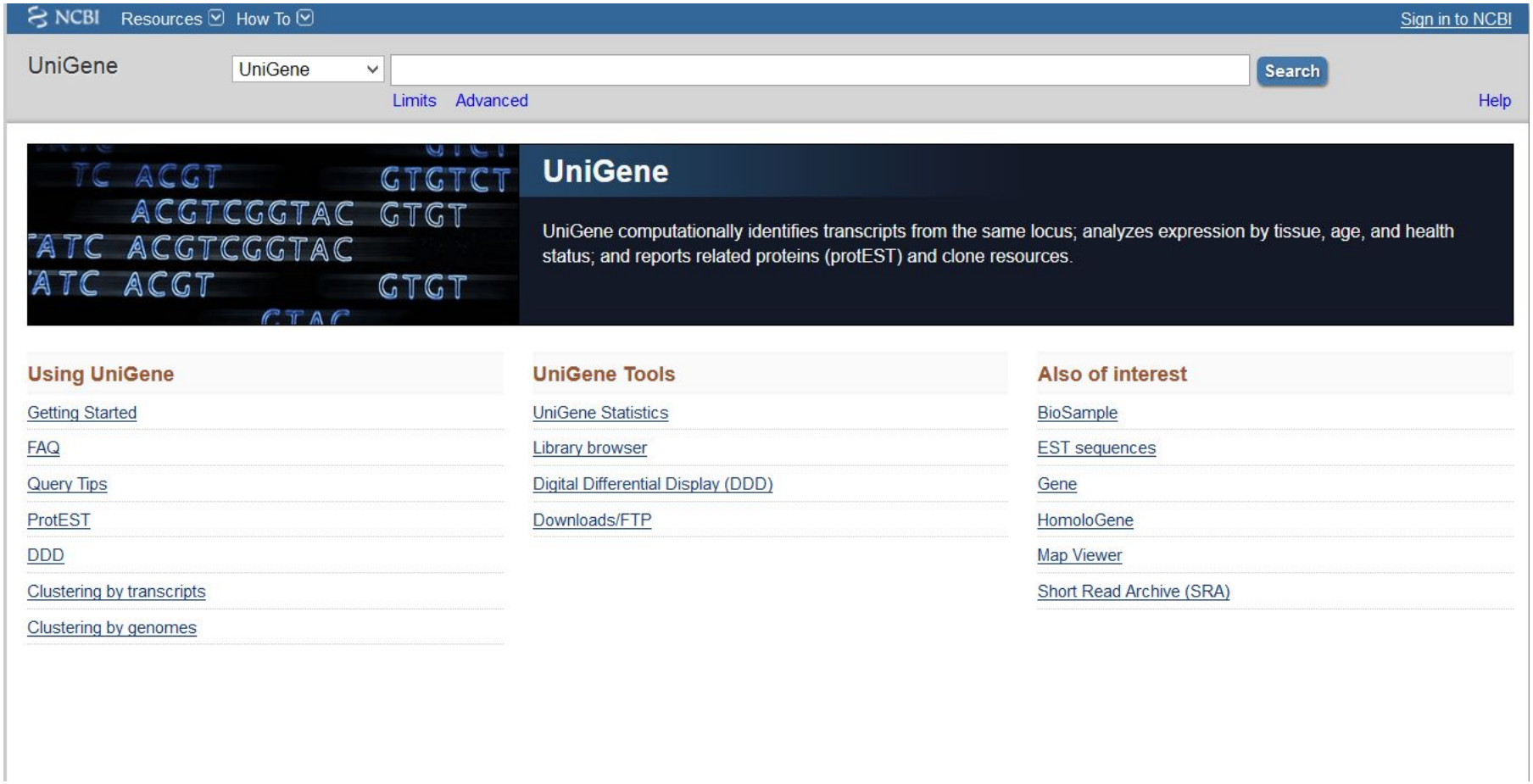
SAGE

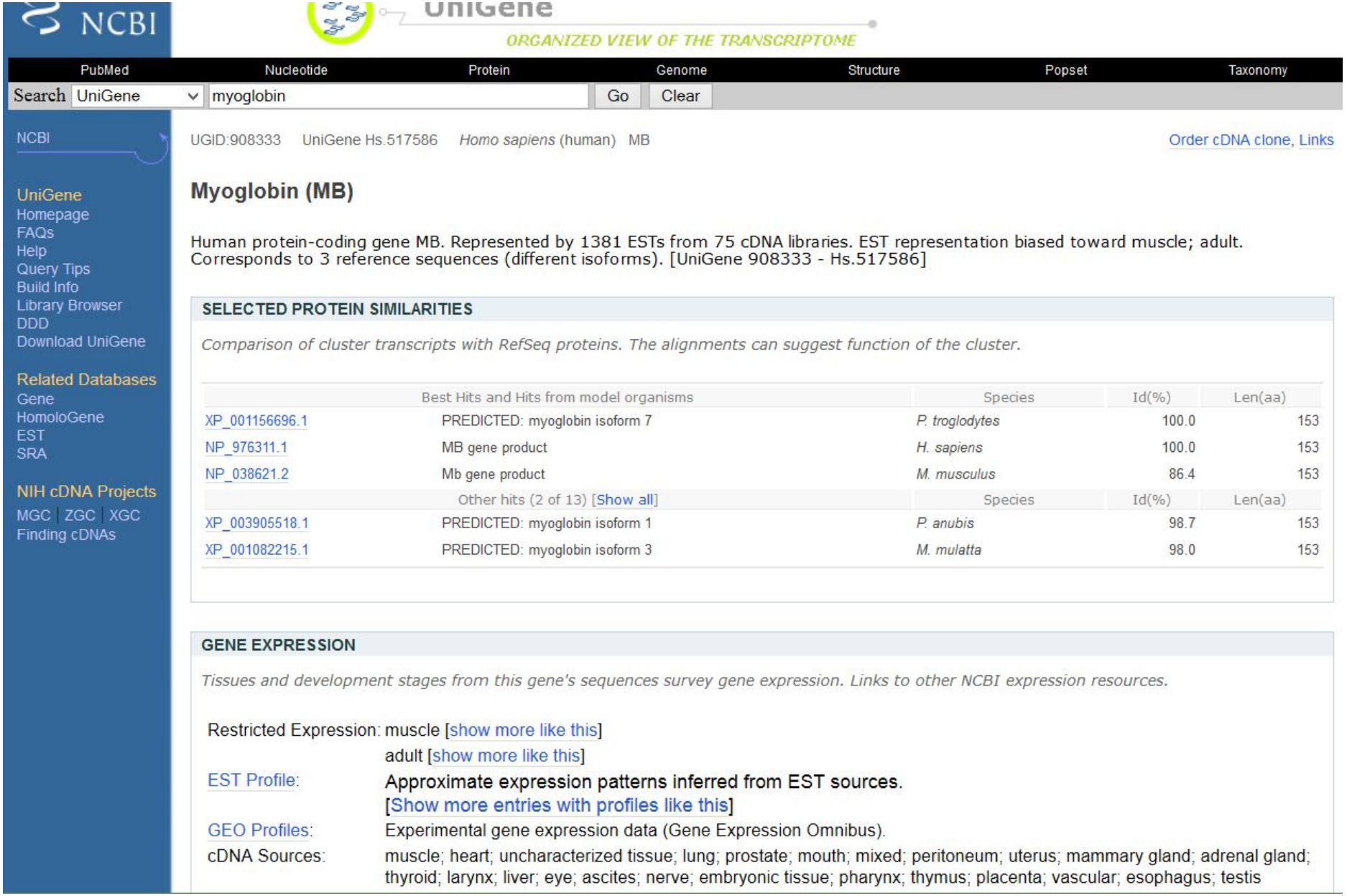
microarray

next-generation sequencing!!!

UniGene: уникальные гены представленные ESTs

- Unigene в NCBI: www.ncbi.nlm.nih.gov/UniGene
- Unigene кластеры содержат много ESTs (Expressed Sequence Tags - короткая подпоследовательность кДНК последовательности)
- Данные Unigene приходят из многих библиотек кДНК.
- Таким образом, когда вы смотрите на ген в Unigene вы получите информацию о величине и месте его экспрессии.


NCBI Resources How To
Sign in to NCBI
UniGene
UniGene
Search
Limits Advanced
Help
UniGene
UniGene computationally identifies transcripts from the same locus; analyzes expression by tissue, age, and health status; and reports related proteins (protEST) and clone resources.
Using UniGene
Getting Started
FAQ
Query Tips
ProtEST
DDD
Clustering by transcripts
Clustering by genomes
UniGene Tools
UniGene Statistics
Library browser
Digital Differential Display (DDD)
Downloads/FTP
Also of interest
BioSample
EST sequences
Gene
HomoloGene
Map Viewer
Short Read Archive (SRA)


NCBI
UniGene
ORGANIZED VIEW OF THE TRANSCRIPTOME
PubMed Nucleotide Protein Genome Structure Popset Taxonomy
Search UniGene myoglobin Go Clear
NCBI
UniGene
Homepage
FAQs
Help
Query Tips
Build Info
Library Browser
DDD
Download UniGene
Related Databases
Gene
HomoloGene
EST
SRA
NIH cDNA Projects
MGC | ZGC | XGC
Finding cDNAs
UGID:908333 UniGene Hs.517586 Homo sapiens (human) MB
Order cDNA clone, Links
Myoglobin (MB)
Human protein-coding gene MB. Represented by 1381 ESTs from 75 cDNA libraries. EST representation biased toward muscle; adult.
Corresponds to 3 reference sequences (different isoforms). [UniGene 908333 - Hs.517586]
SELECTED PROTEIN SIMILARITIES
Comparison of cluster transcripts with RefSeq proteins. The alignments can suggest function of the cluster.
Best Hits and Hits from model organisms Species Id(%) Len(aa)
XP_001156696.1 PREDICTED: myoglobin isoform 7 P. troglodytes 100.0 153
NP_976311.1 MB gene product H. sapiens 100.0 153
NP_038621.2 Mb gene product M. musculus 86.4 153
Other hits (2 of 13) [Show all] Species Id(%) Len(aa)
XP_003905518.1 PREDICTED: myoglobin isoform 1 P. anubis 98.7 153
XP_001082215.1 PREDICTED: myoglobin isoform 3 M. mulatta 98.0 153
GENE EXPRESSION
Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.
Restricted Expression: muscle [show more like this]
adult [show more like this]
EST Profile: Approximate expression patterns inferred from EST sources.
[Show more entries with profiles like this]
GEO Profiles: Experimental gene expression data (Gene Expression Omnibus).
cDNA Sources: muscle; heart; uncharacterized tissue; lung; prostate; mouth; mixed; peritoneum; uterus; mammary gland; adrenal gland; thyroid; larynx; liver; eye; ascites; nerve; embryonic tissue; pharynx; thymus; placenta; vascular; esophagus; testis

EST Profile

Hs.517586 - MB: Myoglobin

Breakdown by Body Sites

	Hs.517586	
adipose tissue	0	0 / 12866
adrenal gland	60	2 / 32940
ascites	75	3 / 39834
bladder	0	0 / 29860
blood	0	0 / 122252
bone	0	0 / 71618
bone marrow	0	0 / 48737
brain	0	0 / 1092688
cervix	0	0 / 48486
connective tissue	0	0 / 149072
ear	0	0 / 16100
embryonic tissue	0	0 / 212896
esophagus	49	1 / 20154
eye	19	4 / 208840
heart	1764	158 / 89524
intestine	0	0 / 231981
kidney	0	0 / 210778
larynx	298	7 / 23466
liver	34	7 / 205291
lung	62	21 / 334815
lymph	0	0 / 44302
lymph node	0	0 / 89748
mammary gland	6	1 / 151230
mouth	120	8 / 66150
muscle	10312	1097 / 106371
nerve	193	3 / 15535
ovary	0	0 / 101488
pancreas	0	0 / 213440

Restricted pools are repesented by orange border

Liver	98	13 / 131488
Lung	0	0 / 282332

Pool name

Transcripts per million(TPM)

Spot intensity based on TPM

Gene EST / Total EST in pool

Breakdown by Health State

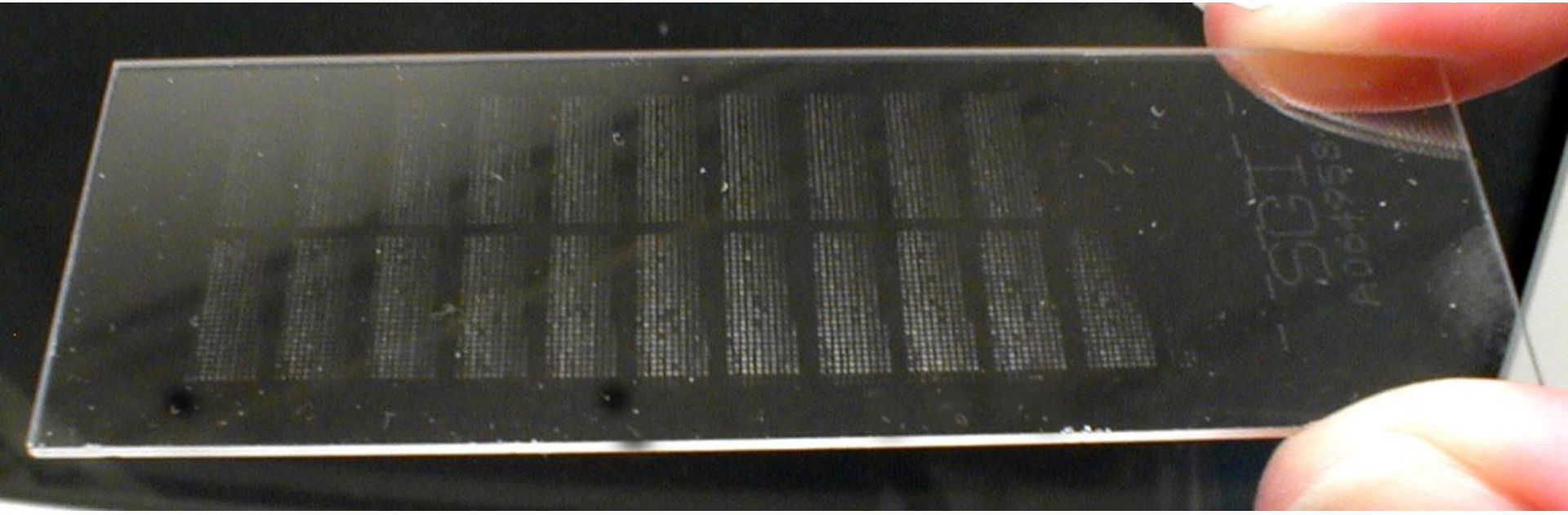
	Hs.517586	
adrenal tumor	158	2 / 12655
bladder carcinoma	0	0 / 17584
breast (mammary gland) tumor	0	0 / 93090
cervical tumor	0	0 / 34484
chondrosarcoma	0	0 / 82838
colorectal tumor	0	0 / 112517
esophageal tumor	57	1 / 17245
gastrointestinal tumor	25	3 / 118498
germ cell tumor	0	0 / 263230
glioma	0	0 / 107194
head and neck tumor	82	11 / 133826
kidney tumor	0	0 / 68872
leukemia	0	0 / 94479
liver tumor	20	2 / 96023
lung tumor	9	1 / 102765
lymphoma	0	0 / 72196
non-neoplasia	0	0 / 96623
normal	344	1147 / 3328811
ovarian tumor	0	0 / 76185
pancreatic tumor	0	0 / 105004
primitive neuroectodermal tumor...	0	0 / 127001
prostate cancer	38	4 / 103844
retinoblastoma	0	0 / 46439
skin tumor	0	0 / 125373
soft tissue/muscle tissue tumor	7	1 / 125265
uterine tumor	11	1 / 90107

Breakdown by Developmental Stage

	Hs.517586	
embryoid body	0	0 / 69969
blastocyst	0	0 / 61448
fetus	80	45 / 556978
neonate	0	0 / 31070
infant	0	0 / 23511
juvenile	0	0 / 55574
adult	311	598 / 1921829

Microarrays: инструмент для измерения экспрессии генов

микрочипом является твердый носитель (такой как мембрана или предметное стекло микроскопа), на котором ДНК известной последовательности нанесена в виде решетки матрицы.



Microarrays: инструмент для измерения экспрессии генов

Наиболее распространенная форма микрочипа используется для измерения экспрессии генов. РНК выделяют из образцов, представляющих интерес. РНК, как правило, превращают в кДНК, помеченную флуоресцентной (или радиоактивной) меткой, а затем гибридизуют на микрочип для того, чтобы измерить уровни экспрессии тысяч генов.

Преимущества микроэррейных экспериментов

- **Скорость**: данные о > 20000 транскриптов несколько дней
- **Всесторонность исследования**: весь геном дрожжей или мыши на чипе
- **Гибкость**: Пользовательские микроэрреи могут быть сделаны, для представляющих интерес генов
- **Легкость создания**: Добавление РНК в чип
- **Стоимость?**: Чип, представляющий 20 000 генов за $ 300

Недостатки микроэррейных экспериментов

- **Цена:** Некоторые исследователи не могут позволить себе сделать статистически значимое количество измерений
- **Значимость РНК**:
  - Отсутствие корреляции между экспрессией генов и количеством белка
  - Полная транскрипция генома плохо понимаема
  - Много некодирующих РНК не представлены в микрочипах
- **Контроль качества:**
  - Артефакты при анализе изображения
  - Артефакты при анализе данных
  - Недостаточно внимание к планированию эксперимента
  - Не хватает правильной статистической обработки

Этап 1: Экспериментальный дизайн

Этап 2: РНК и пробо подготовка

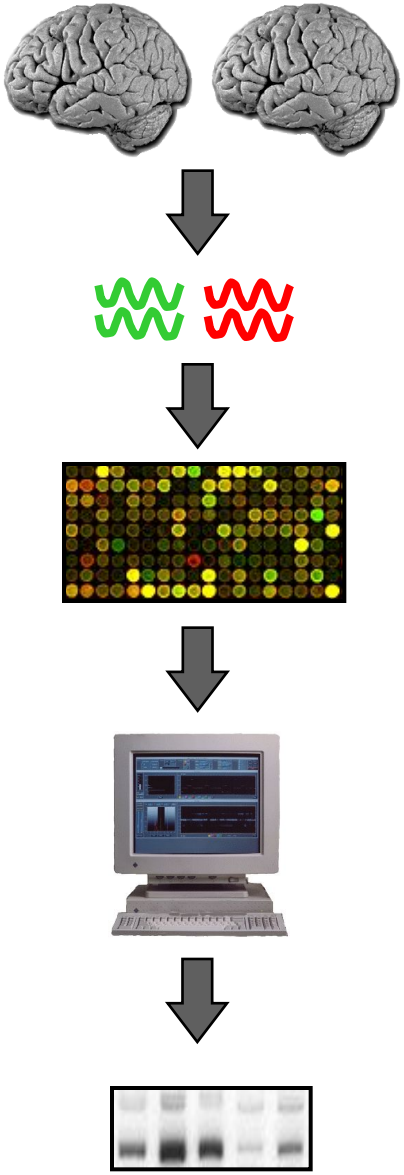
Этап 3: Гибридизация с ДНК чипом

Этап 4: Анализ изображений

Этап 5: Анализ микроэррейных данных

Этап 6: Биологическое подтверждение

Этап 7: Микроэррейные базы данных



Этап 1: Экспериментальный дизайн

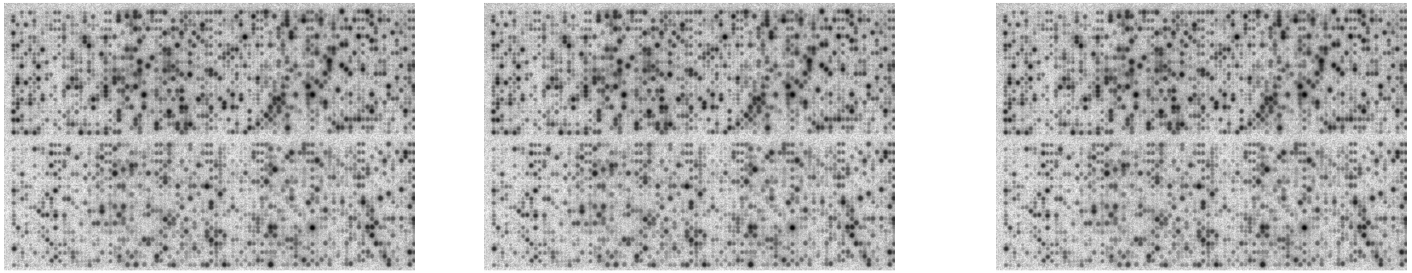
[1] Биологические образцы: технические и биологические повторы: определить подход к анализу данных с самого начала

[2] Выделение РНК, преобразование, маркировка, гибридизация

[3] Расположение элементов массива на поверхности: рандомизации может уменьшить пространственные артефакты

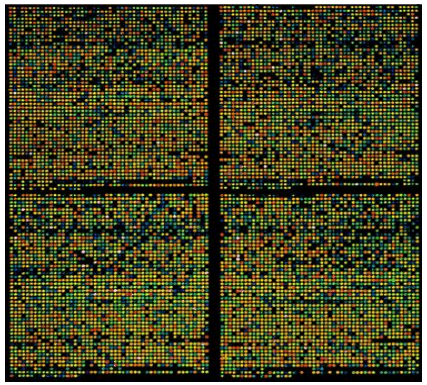
Один образец на массив
(Affymetrix или платформы на радиоактивных метках)

Sample 1 **Sample 2** **Sample 3**

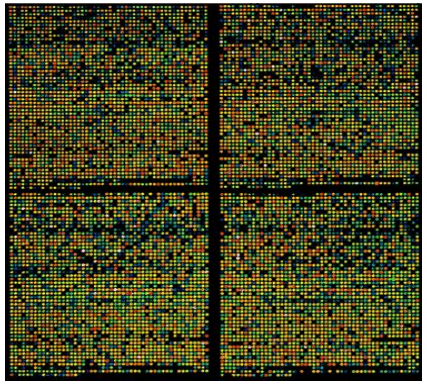


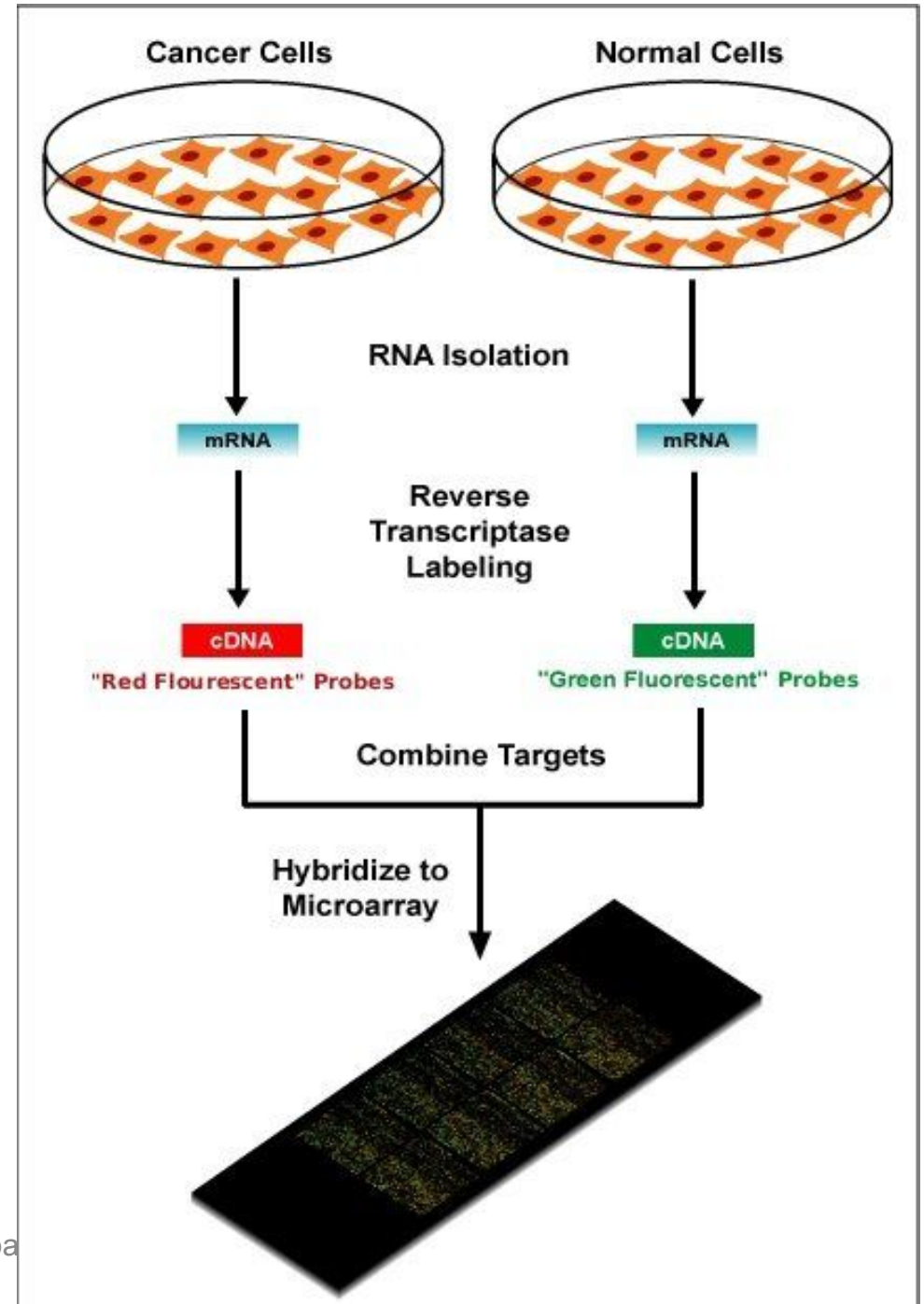
Два образца на массив

Samples 1,2



Samples 1,3





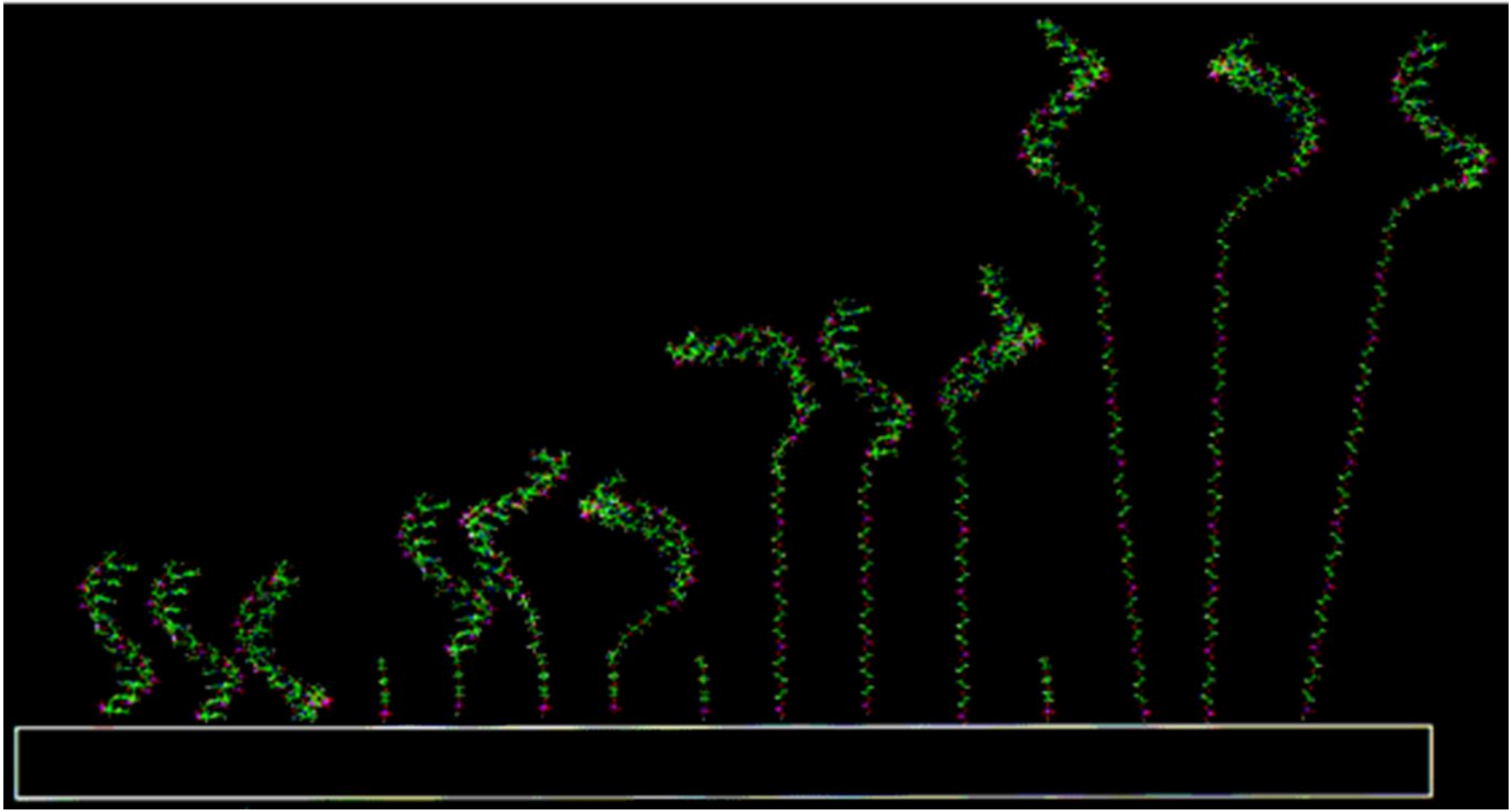
Этап 2: РНК и пробо подготовка

- Для Affymetrix чипов, нужна полная РНК (около 5 мкг)
- Подтвердите чистоту, запустив ее в агарозном геле
- Один из самых больших источников ошибки, связанных с выделения РНК;
- Использование соответствующего сбалансированного, рандомизированого дизайна эксперимента.

Этап 3: Гибридизация с ДНК чипом

- Массив состоит из кДНК или олигонуклеотидов
- Олигонуклеотиды могут быть нанесены с помощью фотолитографии
- Образец преобразуется в кРНК или кДНК

Поверхность чипа

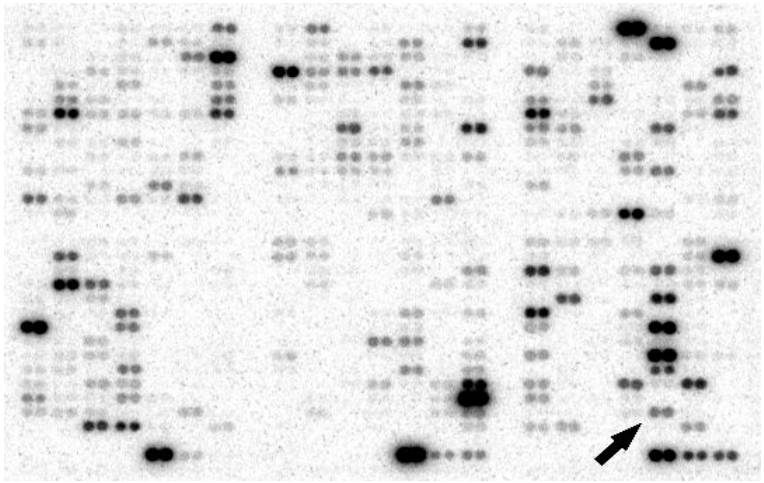


Этап 4: Анализ изображений

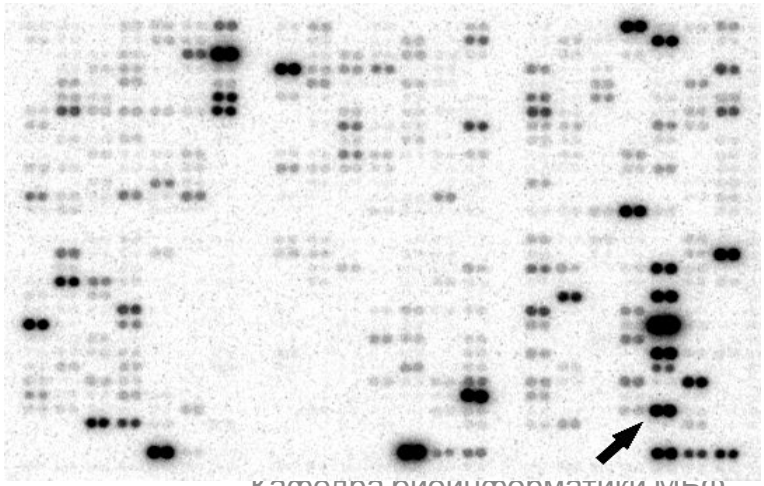
- Уровни РНК-транскриптов являются количественными
- Интенсивность флуоресценции или радиоактивности измеряют с помощью сканера

Дифференциальная генная экспрессия на кДНК микроэррее

Контроль



Синдром Ретта



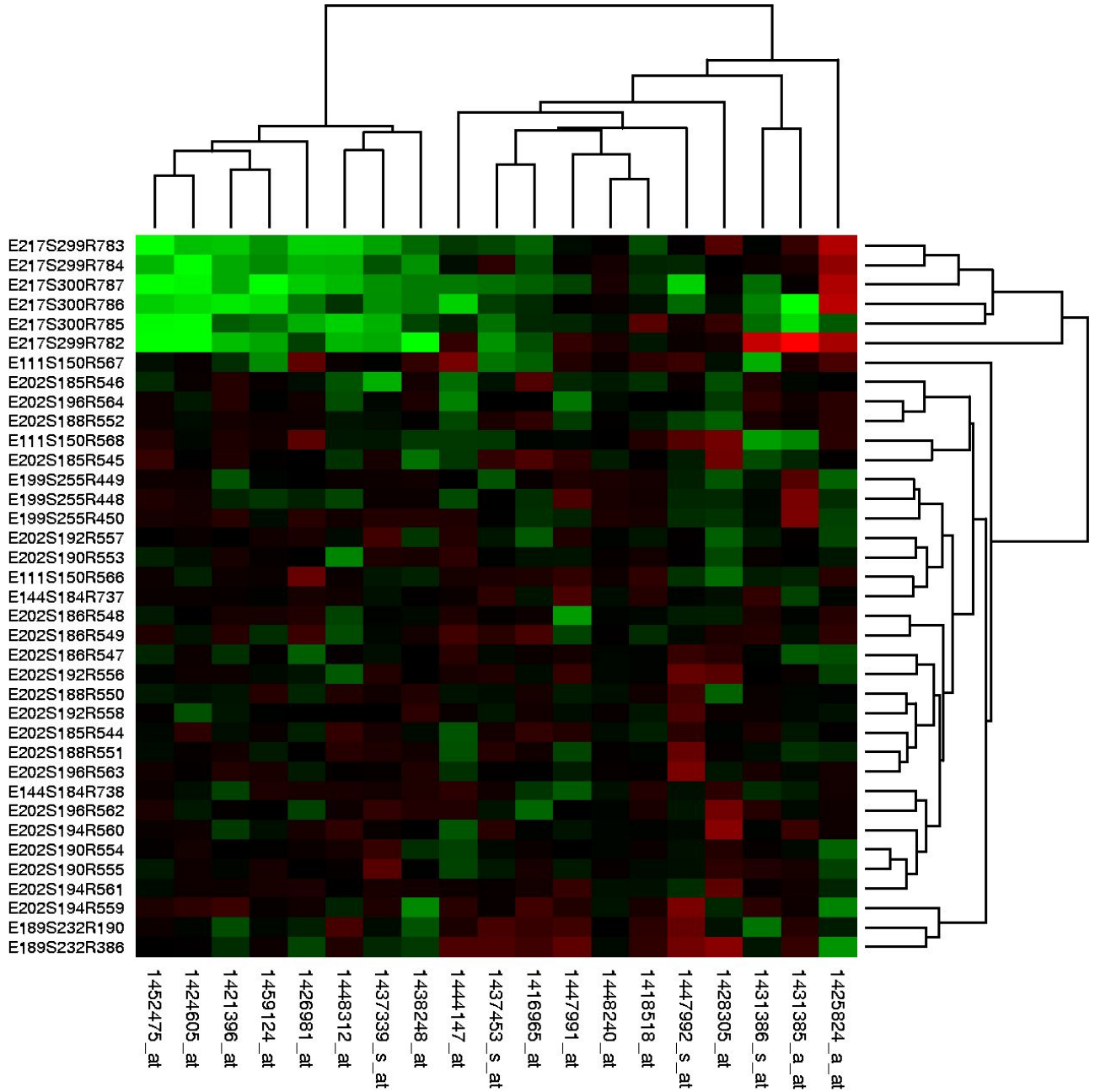
a B Crystallin гиперэкспрессирован при Синдроме Ретта

Этап 5: Анализ микроэррейных данных

- **Проверка гипотезы**
  - Как можно сравнить массивы?
  - Какие РНК-транскрипты (гены) регулируются?
  - Являются ли различия подлинным?
  - Каковы критерии для статистической значимости?

- **Кластеризация**
  - Есть ли значимые закономерности в данных (например, группы)?

- **Классификация**
  - Есть ли у РНК-транскриптов предсказанные заранее группы, такие как подтипы болезней?



Значения экспрессии генов из микроэррейных экспериментов могут быть представлены в виде тепловой карты для визуализации результатов анализа данных

Этап 6: Биологическое подтверждение

- Микроэррейные эксперименты можно рассматривать как «генераторы гипотез».
- Дифференциальное регулирования РНК-транскриптов может быть измерено с помощью независимых анализов, таких как
  - Нозерн-блот
  - Полимеразная цепная реакция (ПЦР)
  - Гибридизация

Этап 7: Микроэррейные базы данных

- Есть две основных базы данных
  - Gene expression omnibus (GEO) в NCBI
  - ArrayExpress в European Bioinformatics Institute (EBI)

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession Search

Getting Started

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- GEO BLAST
- Programmatic Access
- FTP Site

Browse Content

- Repository Browser

DataSets:	3848
Series:	56613
Platforms:	14175
Samples:	1378901

Information for Submitters

- Login to Submit
- Submission Guidelines
- Update Guidelines
- MIAME Standards
- Citing and Linking to GEO
- Guidelines for Reviewers
- GEO Publications

NCBI Resources How To Sign in to NCBI

Search NCBI databases

Help

ubiquitin Search

Results found in 37 databases for "ubiquitin"

Literature

Books	1,009	books and reports
MeSH	333	ontology used for PubMed indexing
NLM Catalog	66	books, journals and more in the NLM Collections
PubMed	43,356	scientific & medical abstracts/citations
PubMed Central	83,307	full-text journal articles

Health

ClinVar	3,356	human variations of clinical significance
dbGaP	46	genotype/phenotype interaction studies
GTR	367	genetic testing registry
MedGen	35	medical genetics literature and links
OMIM	985	online mendelian inheritance in man
PubMed Health	9	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genome assembly information
BioProject	261	biological projects providing data to NCBI
BioSample	15	descriptions of biological source materials
Clone	10,754	genomic and cDNA clones
dbVar	33,020	genome structural variation studies

Genes

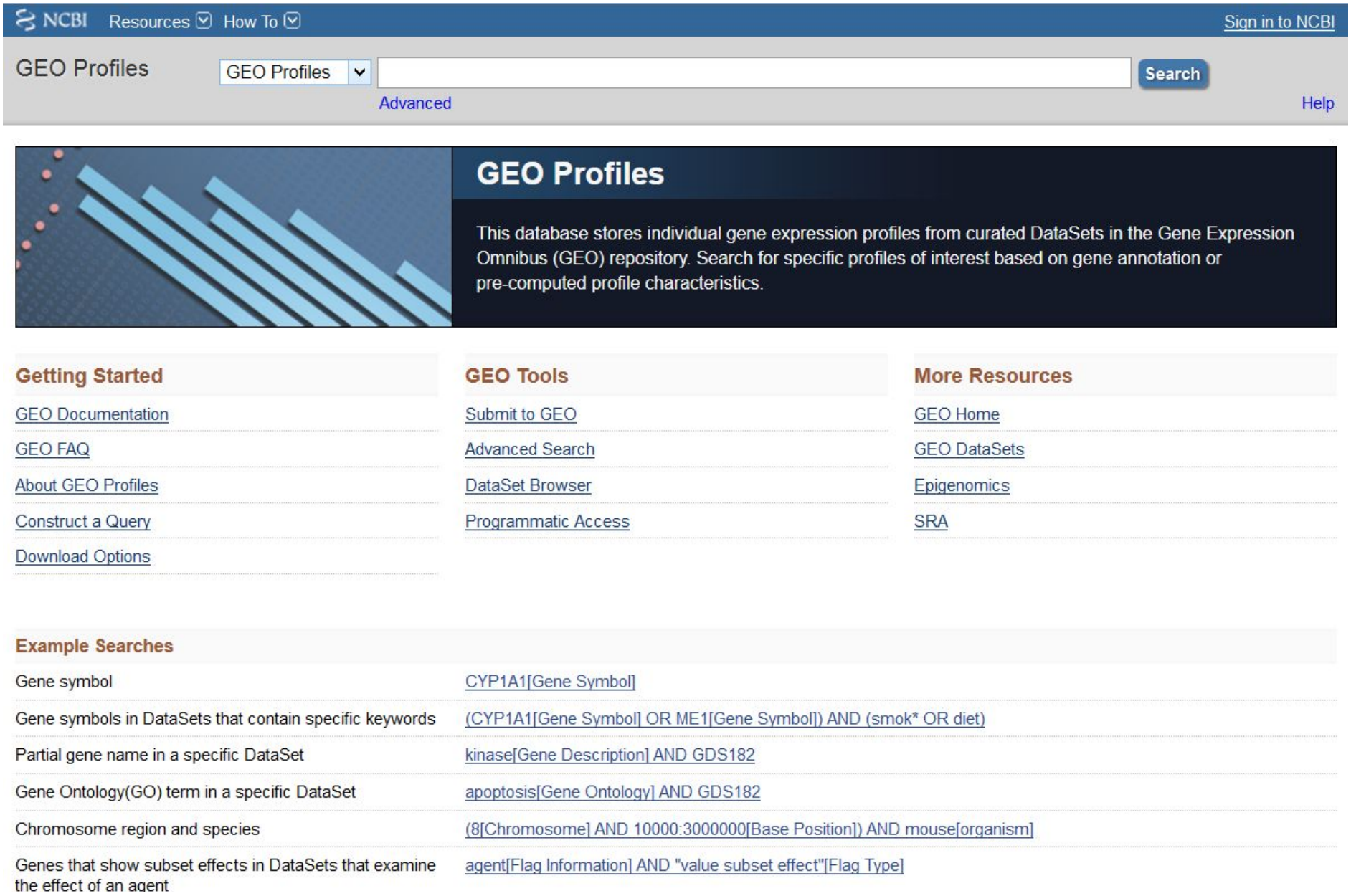
EST	22,145	expressed sequence tag sequences
Gene	159,204	collected information about gene loci
GEO DataSets	615	functional genomics studies
GEO Profiles	2,725,449	gene expression and molecular abundance profiles
HomoloGene	860	homologous gene sets for selected organisms
PopSet	221	sequence sets from phylogenetic and population studies
UniGene	13,193	clusters of expressed transcripts

Proteins

Conserved Domains	777	conserved protein domains
Protein	461,008	protein sequences
Protein Clusters	888	sequence similarity-based protein clusters
Structure	2,585	experimentally-determined biomolecular structures

Chemicals

BioSystems	18,804	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	12,636	bioactivity screening studies
PubChem Compound	9	chemical information with structures, information and links


NCBI Resources How To
Sign in to NCBI
GEO Profiles
GEO Profiles
Search
Advanced
Help
GEO Profiles
This database stores individual gene expression profiles from curated DataSets in the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.
Getting Started
GEO Documentation
GEO FAQ
About GEO Profiles
Construct a Query
Download Options
GEO Tools
Submit to GEO
Advanced Search
DataSet Browser
Programmatic Access
More Resources
GEO Home
GEO DataSets
Epigenomics
SRA
Example Searches
Gene symbol
CYP1A1[Gene Symbol]
Gene symbols in DataSets that contain specific keywords
(CYP1A1[Gene Symbol] OR ME1[Gene Symbol]) AND (smok* OR diet)
Partial gene name in a specific DataSet
kinase[Gene Description] AND GDS182
Gene Ontology(GO) term in a specific DataSet
apoptosis[Gene Ontology] AND GDS182
Chromosome region and species
(8[Chromosome] AND 10000:3000000[Base Position]) AND mouse[organism]
Genes that show subset effects in DataSets that examine the effect of an agent
agent[Flag Information] AND "value subset effect"[Flag Type]

NCBI Resources How To Sign in to NCBI

GEO Profiles

GEO Profiles ubiquitin Search

Save search Advanced Help

Show additional filters

Gene symbol
Select ...

Gene keyword
Select ...

Organism
Select ...

Gene ontology
Select ...

Differential expression
Up/down genes

DataSet keyword
Select ...

GEO accession
Select ...

Clear all

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Subgroup effect **Send to:**

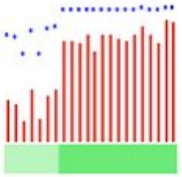
Filters: Manage Filters

Results: 1 to 20 of 2725449

<< First < Prev Page 1 of 136273 Next > Last >>

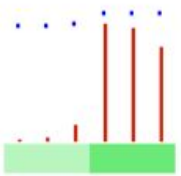
1. UBD - Alcoholic hepatitis

Annotation: UBD, **ubiquitin** D (multiple annotations exist)
Organism: Homo sapiens
Reporter: GPL570, 205890_s_at (ID_REF), GDS4389, 10537 (Gene ID), 2550 (Gene ID), NM_006398
DataSet type: Expression profiling by array, transformed count, 22 samples
ID: 85455638
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors



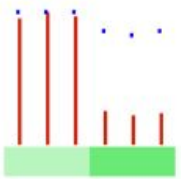
2. HSPA1B - Modeled microgravity effect on activated T lymphocytes

Annotation: HSPA1B, heat shock 70kDa protein 1B (multiple annotations exist)
Organism: Homo sapiens
Reporter: GPL96, 200800_s_at (ID_REF), GDS1806, 3303 (Gene ID), 3304 (Gene ID), NM_005345
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 20229428
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors



3. HERC5 - Extracellular matrix protein cysteine rich 61 (CCN1) effect on LN229 glioma cells

Annotation: HERC5, HECT and RLD domain containing E3 **ubiquitin** protein ligase 5
Organism: Homo sapiens



Profile data

Download profile data

Profile pathways

Find pathways

Find related data

Database: Select

Find items

Search details

ubiquitin[All Fields]

Search See more...

NCBI Resources How To Sign in to NCBI

GEO Profiles

GEO Profiles Search

Advanced Help

Display Settings: Summary Send to:

Ufd1l - 22q11 microdeletion syndrome model: hippocampus

Annotation: Ufd1l, **ubiquitin** fusion degradation 1 like

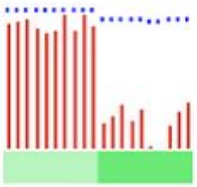
Organism: Mus musculus

Reporter: GPL1261, 1418087_at (ID_REF), GDS3479, 22230 (Gene ID), BC006630

DataSet type: Expression profiling by array, transformed count, 20 samples

ID: 58037418

GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors



Profile data

Download profile data

Profile pathways

Find pathways

Related information

GEO DataSets

Gene

UniGene

Profile neighbors

Chromosome neighbors

Sequence neighbors

Homologene neighbors

Free in PMC

HomoloGene

Pathways + GO

PubMed

Taxonomy

Nucleotide

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | E

NCBI > GEO > **Accession Display**

Not logged in

Scope: Self | Format: HTML | Amount: Quick | GEO accession: GPL1261 | GO

Platform GPL1261 Query DataSets for GPL1261

Status	Public on May 25, 2004
Title	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array
Technology type	in situ oligonucleotide
Distribution	commercial
Organism	Mus musculus
Manufacturer	Affymetrix
Manufacture protocol	see manufacturer's web site

All probe sets represented on the GeneChip Mouse Expression Set 430 are included on the GeneChip Mouse Genome 430 2.0 Array. The sequences from which these probe sets were derived were selected from GenBank«, dbEST, and RefSeq. The sequence clusters were created from the UniGene database (Build 107, June 2002) and then refined by analysis and comparison with the publicly available draft assembly of the mouse genome from the Whitehead Institute for Genome Research (MGSC, April 2002).

Description

Affymetrix submissions are typically submitted to GEO using the GEOarchive method described at http://www.ncbi.nlm.nih.gov/projects/geo/info/geo_affy.html

June 03, 2009: annotation table updated with netaffx build 28
June 07, 2012: annotation table updated with netaffx build 32

NCBI Resources How To Sign in to NCBI

GEO Profiles

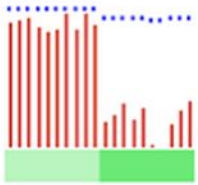
GEO Profiles Search

Advanced Help

Display Settings: Summary Send to:

Ufd1l - 22q11 microdeletion syndrome model: hippocampus

Annotation: Ufd1l, **ubiquitin** fusion degradation 1 like
Organism: Mus musculus
Reporter: GPL1261, 1418087_at (ID_REF), GDS3479, 22230 (Gene ID), BC006630
DataSet type: Expression profiling by array, transformed count, 20 samples
ID: 58037418



GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

Profile data

Download profile data

Profile pathways

Find pathways

Related information

- GEO DataSets
- Gene
- UniGene
- Profile neighbors
- Chromosome neighbors
- Sequence neighbors
- Homologene neighbors
- Free in PMC
- HomoloGene
- Pathways + GO
- PubMed
- Taxonomy
- Nucleotide

NCBI

DATASET BROWSER (CURATED)

GEO Gene Expression Omnibus

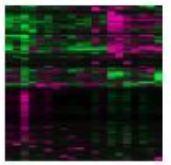
Search for GDS3479[ACCN] | Search | Clear | Show All | Advanced Search

DataSet Record GDS3479: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	22q11 microdeletion syndrome model: hippocampus
Summary:	Analysis of hippocampi of Df(16)A/+ animals. Df(16)A/+ animals carry microdeletions of about 1.3Mb in the locus syntenic to human 22q11.2. Individuals with 22q11.2 microdeletions show behavioral and cognitive deficits and are at high risk of developing schizophrenia.
Organism:	*Mus musculus*
Platform:	GPL1261: [Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array
Citations:	Stark KL, Xu B, Bagchi A, Lai WS et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat Genet* 2008 Jun;40(6):751-60. PMID: 18469815 Xu B, Hsu PK, Stark KL, Karayiorgou M et al. Derepression of a neuronal inhibitor due to miRNA dysregulation in a schizophrenia-related microdeletion. *Cell* 2013 Jan 17;152(1-2):262-75. PMID: 23332760

Reference Series:	GSE10784	**Sample count:**	20
Value type:	transformed count	**Series published:**	2008/04/10

Cluster Analysis



Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

Data Analysis Tools

Find genes

- Compare 2 sets of samples
- Cluster heatmaps
- Experiment design and value distribution

Find gene name or symbol: Go

Find genes that are up/down for this condition(s): ☑ strain Go

NCBI Resources How To Sign in to NCBI

GEO Profiles

GEO Profiles Search

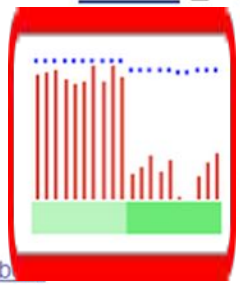
Advanced Help

Display Settings: Summary Send to:

Ufd1l - 22q11 microdeletion syndrome model: hippocampus

Annotation: Ufd1l, **ubiquitin** fusion degradation 1 like
Organism: Mus musculus
Reporter: GPL1261, 1418087_at (ID_REF), GDS3479, 22230 (Gene ID), BC006630
DataSet type: Expression profiling by array, transformed count, 20 samples
ID: 58037418

GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighb
Homologene neighbors



Profile data

Download profile data

Profile pathways

Find pathways

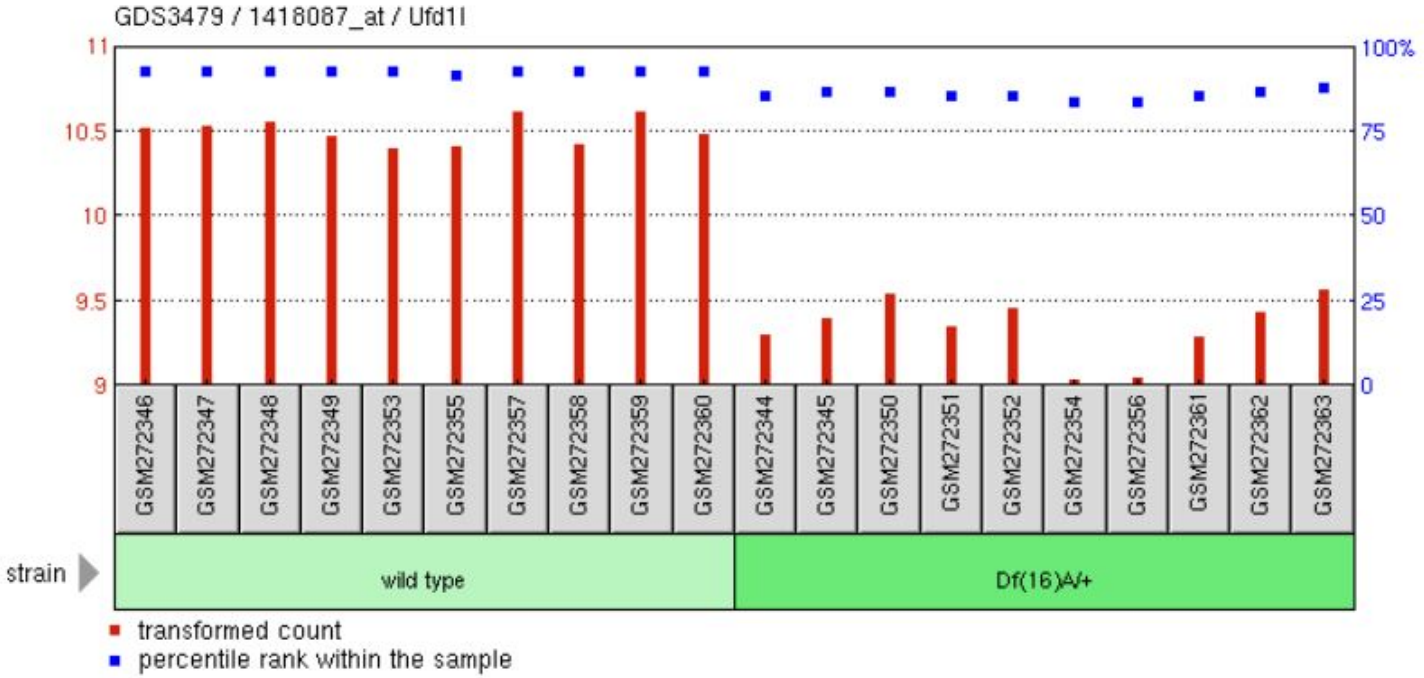
Related information

GEO DataSets
Gene
UniGene
Profile neighbors
Chromosome neighbors
Sequence neighbors
Homologene neighbors
Free in PMC
HomoloGene
Pathways + GO
PubMed
Taxonomy
Nucleotide

Profile GDS3479 / 1418087_at / Ufd1l

Title 22q11 microdeletion syndrome model: hippocampus

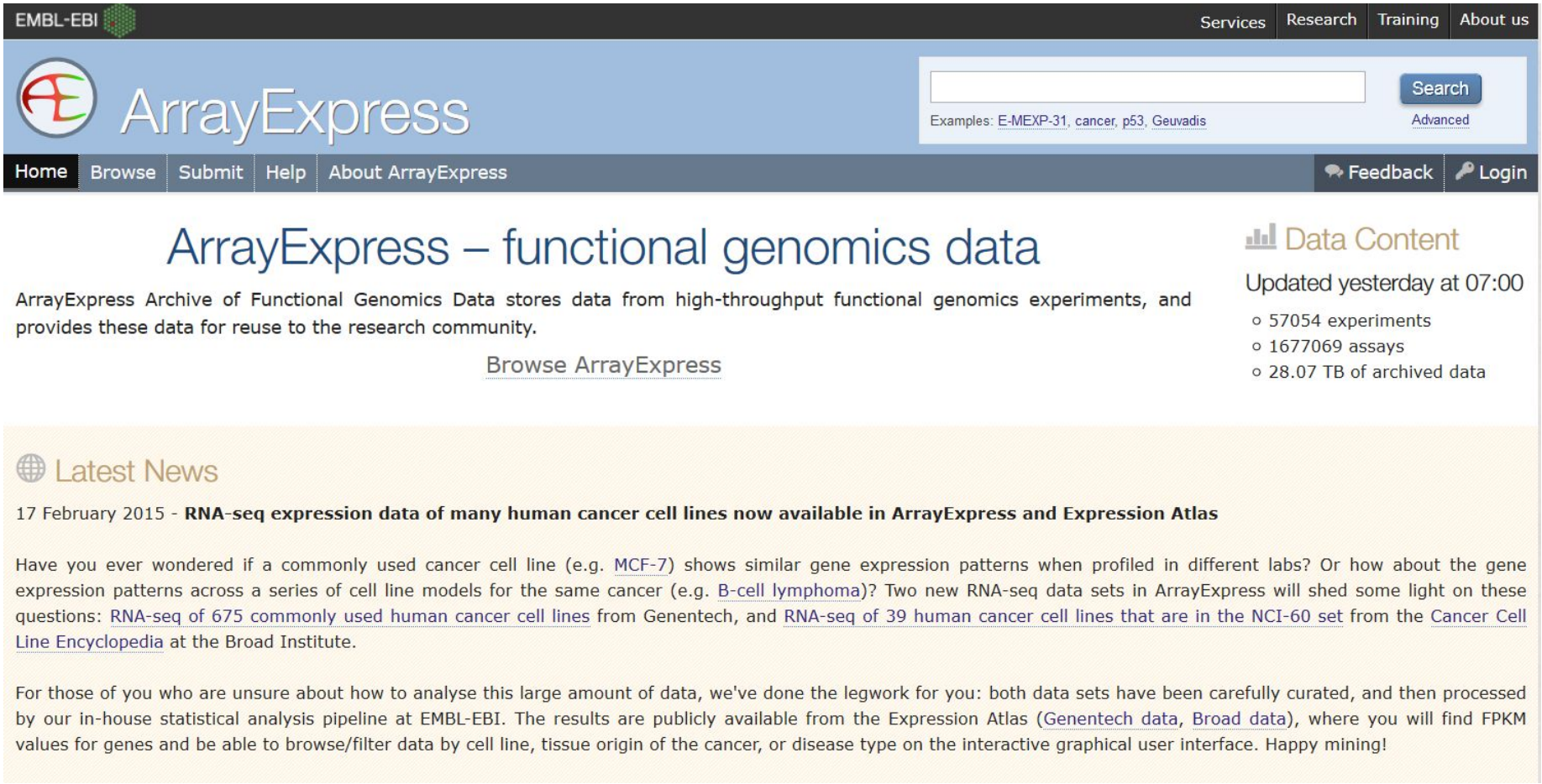
Organism Mus musculus

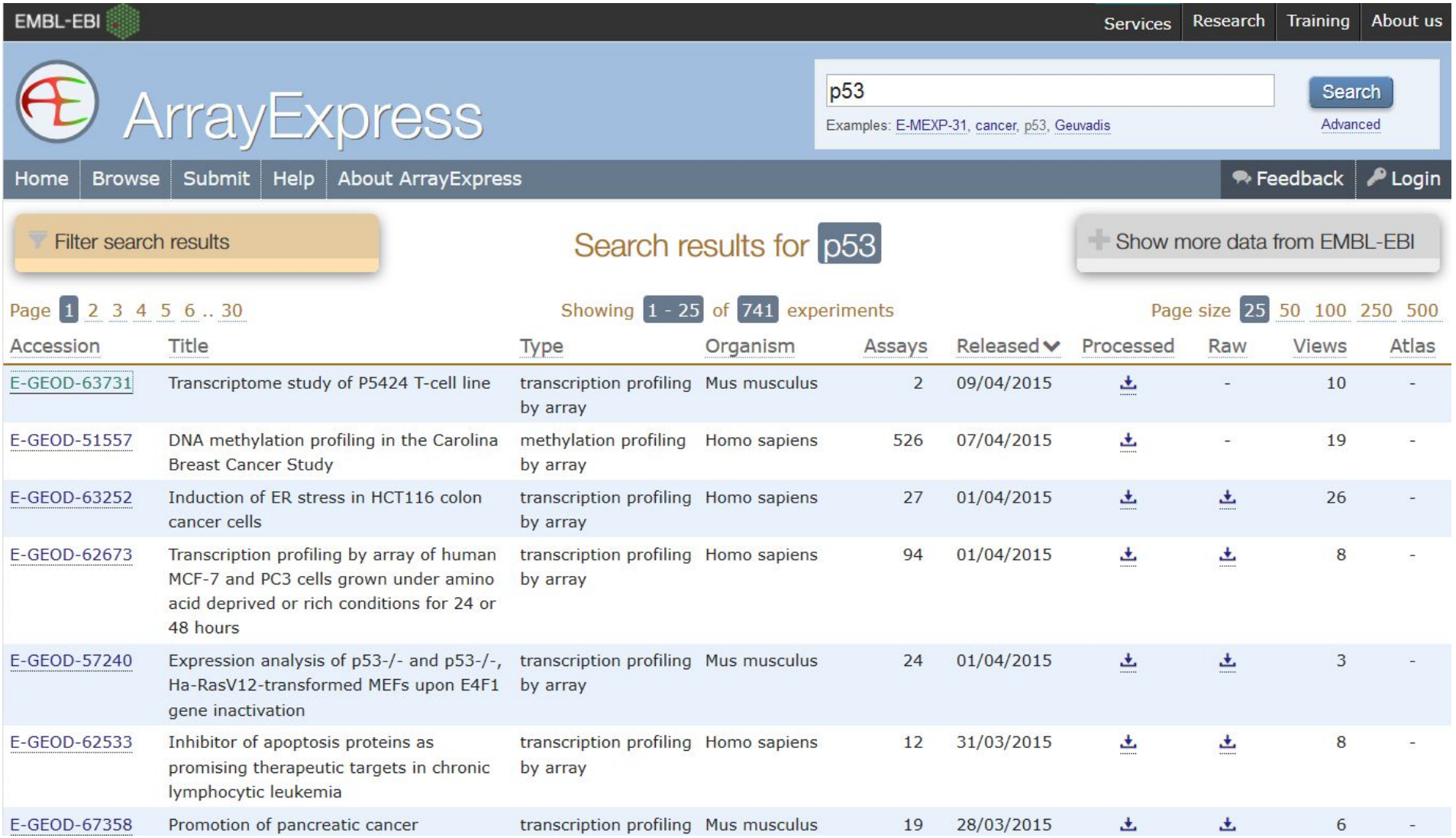


Graph caption help

Sample	Title	Value	Rank
GSM272346	wt_HPC, biological replicate 1	10.5267	93
GSM272347	wt_HPC, biological replicate 2	10.5365	93
GSM272348	wt_HPC, biological replicate 3	10.5658	93
GSM272349	wt_HPC, biological replicate 4	10.4711	93
GSM272353	wt_HPC, biological replicate 5	10.3992	93
GSM272355	wt_HPC, biological replicate 6	10.4112	92

Array Express в European Bioinformatics Institute http://www.ebi.ac.uk/arrayexpress/





Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEOD-63731	Transcriptome study of P5424 T-cell line	transcription profiling by array	Mus musculus	2	09/04/2015		-	10	-
E-GEOD-51557	DNA methylation profiling in the Carolina Breast Cancer Study	methylation profiling by array	Homo sapiens	526	07/04/2015		-	19	-
E-GEOD-63252	Induction of ER stress in HCT116 colon cancer cells	transcription profiling by array	Homo sapiens	27	01/04/2015			26	-
E-GEOD-62673	Transcription profiling by array of human MCF-7 and PC3 cells grown under amino acid deprived or rich conditions for 24 or 48 hours	transcription profiling by array	Homo sapiens	94	01/04/2015			8	-
E-GEOD-57240	Expression analysis of p53-/- and p53-/-, Ha-RasV12-transformed MEFs upon E4F1 gene inactivation	transcription profiling by array	Mus musculus	24	01/04/2015			3	-
E-GEOD-62533	Inhibitor of apoptosis proteins as promising therapeutic targets in chronic lymphocytic leukemia	transcription profiling by array	Homo sapiens	12	31/03/2015			8	-
E-GEOD-67358	Promotion of pancreatic cancer	transcription profiling	Mus musculus	19	28/03/2015			6	-

EMBL-EBI Services Research Training About us

ArrayExpress

p53 Search

Examples: E-MEXP-31, cancer, p53, Geuvadis Advanced

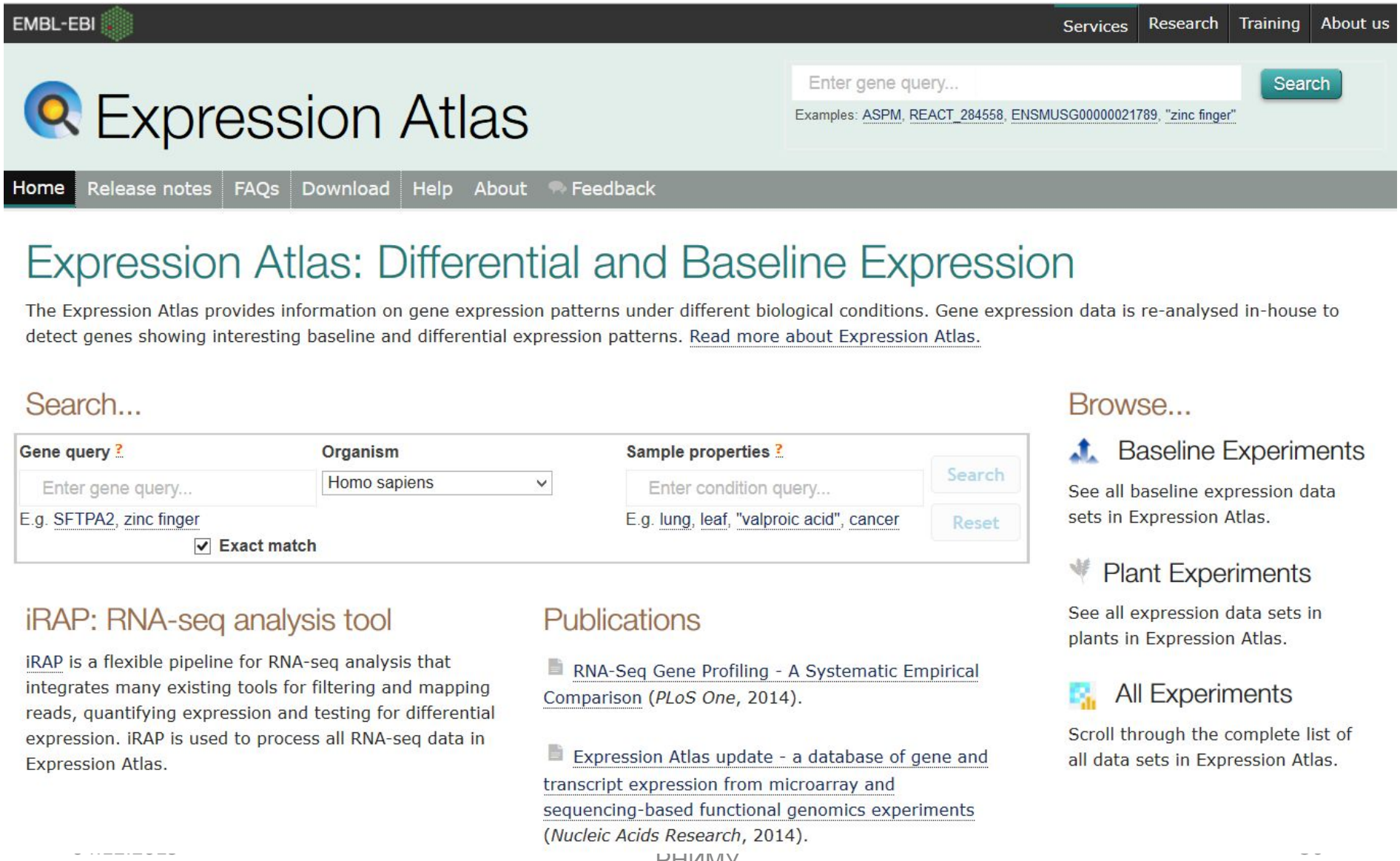
Home Browse Submit Help About ArrayExpress Feedback Login

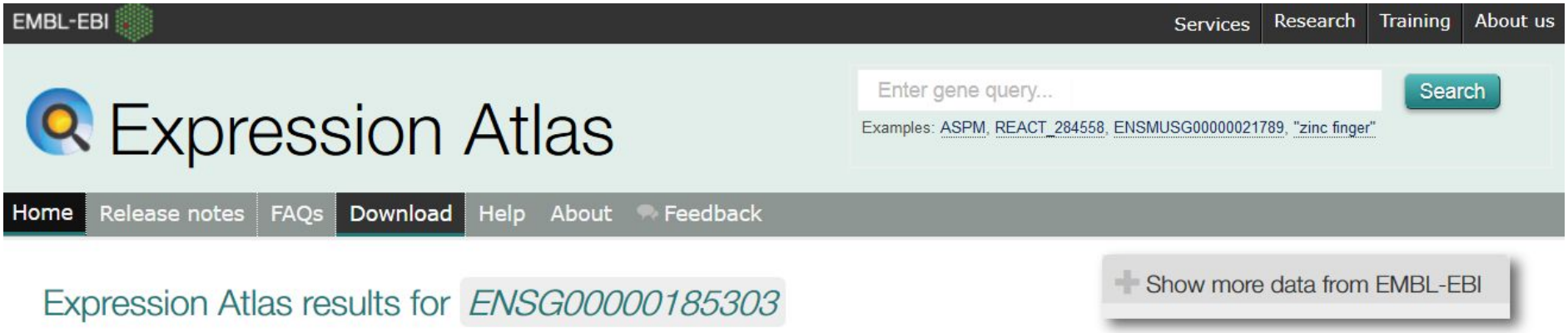
ArrayExpress > Search results for "p53" > E-GEOD-63731

E-GEOD-63731 - Transcriptome study of P5424 T-cell line

Status	*Released on 9 April 2015, last updated on 10 April 2015*
Organism	Mus musculus
Samples (2)	Click for detailed sample information and links to data └ *found inside:* Rag2 and P53 knockout
Array (1)	A-GEOD-13912 - Agilent-028005 SurePrint G3 Mouse GE 8x60K Microarray (Feature Number version)
Protocols (7)	Click for detailed protocol information
Description	P5424 T-cell line was observed transcriptome in two replicates Total RNAs were extracted from 2 samples of P5424 T-cell line and were profiled in replicate after hybridization with Agilent SurePrint G3 Mouse GE 8x60K
Experiment type	transcription profiling by array
Contacts	Aurélien Griffon, Lan T Dao, Laurent Vanhille, Nicolas Fernandez, Salvatore Spicuglia
MIAME	* – – * * Platforms Protocols Variables Processed Raw
Files	Investigation description E-GEOD-63731.idf.txt Sample and data relationship E-GEOD-63731.sdrf.txt

Expression Atlas (EMBL)
https://www.ebi.ac.uk/gxa/home




EMBL-EBI
Services
Research
Training
About us
Expression Atlas
Enter gene query...
Search
Examples: ASPM, REACT_284558, ENSMUSG00000021789, "zinc finger"
Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas results for ENSG00000185303
Show more data from EMBL-EBI
SFTPA2 Homo sapiens surfactant protein A2
Synonyms COLEC5, SP-A2
Orthologs ENSCAFG00000015754 (Canis familiaris), hbl4 (Danio rerio), hbl3 (Danio rerio), hbl1 (Danio rerio), ENSDARG00000070813 (Danio rerio), hbl2 (Danio rerio), SFTPA1 (Equus caballus), LL (Gallus gallus), SFTPA1 (Macaca mulatta), ENSMMUG00000030564 (Macaca mulatta), Sftpa1 (Mus musculus), Sftpa1 (Rattus norvegicus), SFTPA (Sus scrofa), sftpa (Xenopus tropicalis)
Gene Ontology respiratory gaseous exchange, proteinaceous extracellular matrix, extracellular space, carbohydrate binding, protein complex (... and 5 more)
InterPro C-type lectin (domain), Collagen triple helix repeat (repeat), C-type lectin fold (domain)
Ensembl Family PULMONARY SURFACTANT ASSOCIATED A PRECURSOR PSAP PSP A SP A
Ensembl Gene ENSG00000185303
Entrez 729238
UniProt Q8IWL1, R4GMN3, X6REF7
Gene Biotype protein_coding
Design Element 218835_at, 223678_s_at, 3253941, 3254074, 3254080, 3297057, 3297058, 3297060, 3297071, 3297073, 3297075, 3297077, 3297124, 3297130, 3297131, 3297136, 3297137, 3297138, 3297140, A_23_P115652, A_24_P223874, A_33_P3268919, A_33_P3344574, M13686_s_at, M68519_rna1_at, g13346505_3p_at, g190669_3p_s_at

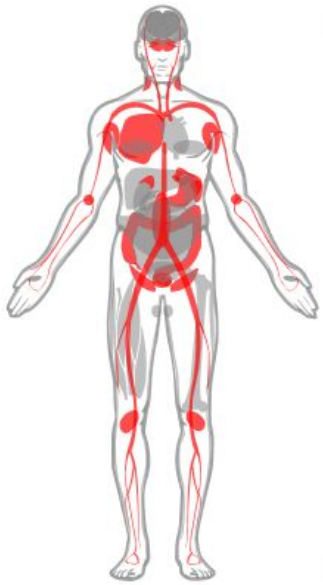
Expression Atlas results for *ENSG00000185303*

Show more data from EMBL-EBI

SFTPA2 *Homo sapiens* surfactant protein A2 +

Baseline Expression Results found −

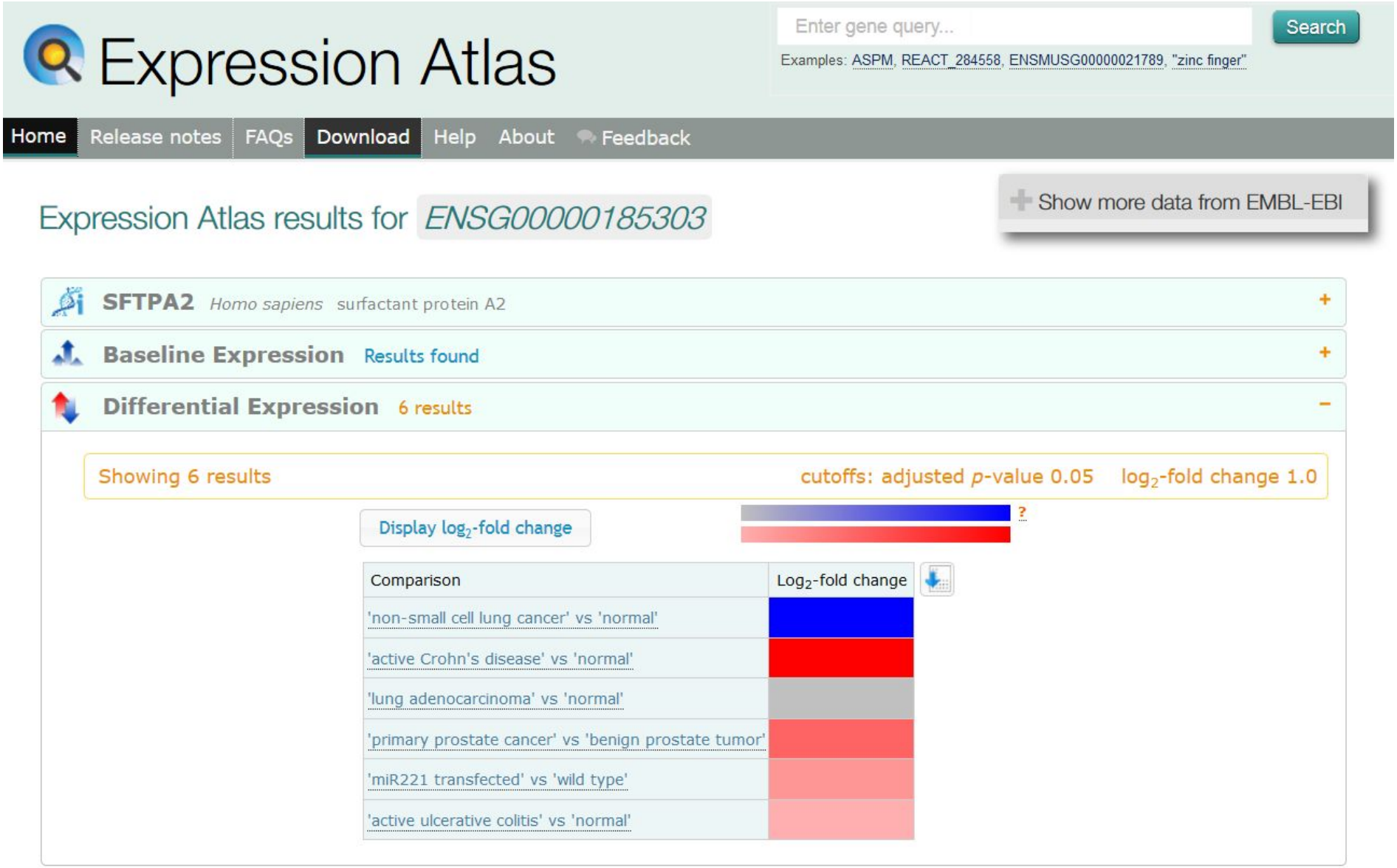
FPKM/TPM (Transcriptomics) > 0.5 Within Sample Abundance (Proteomics) > 0

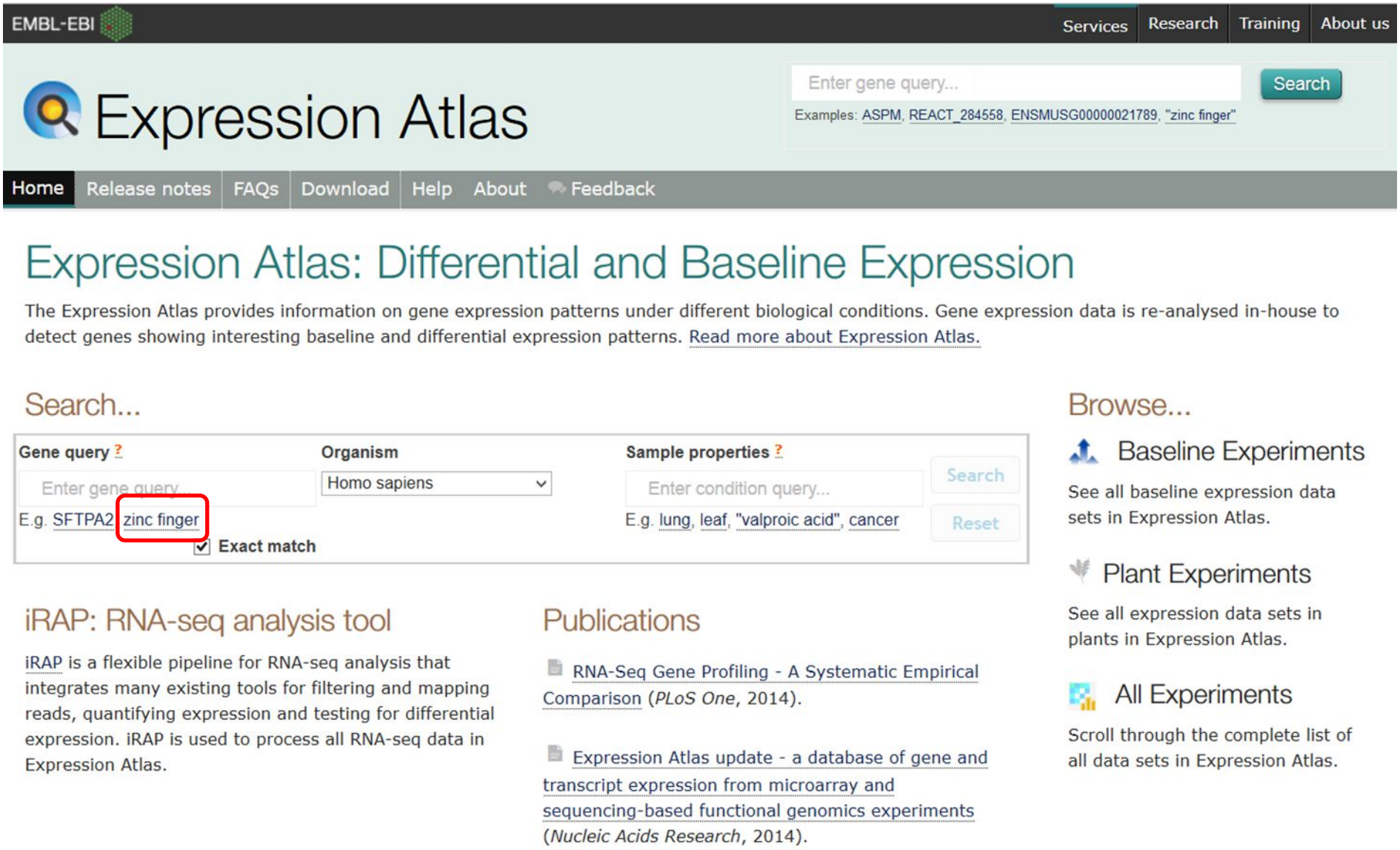


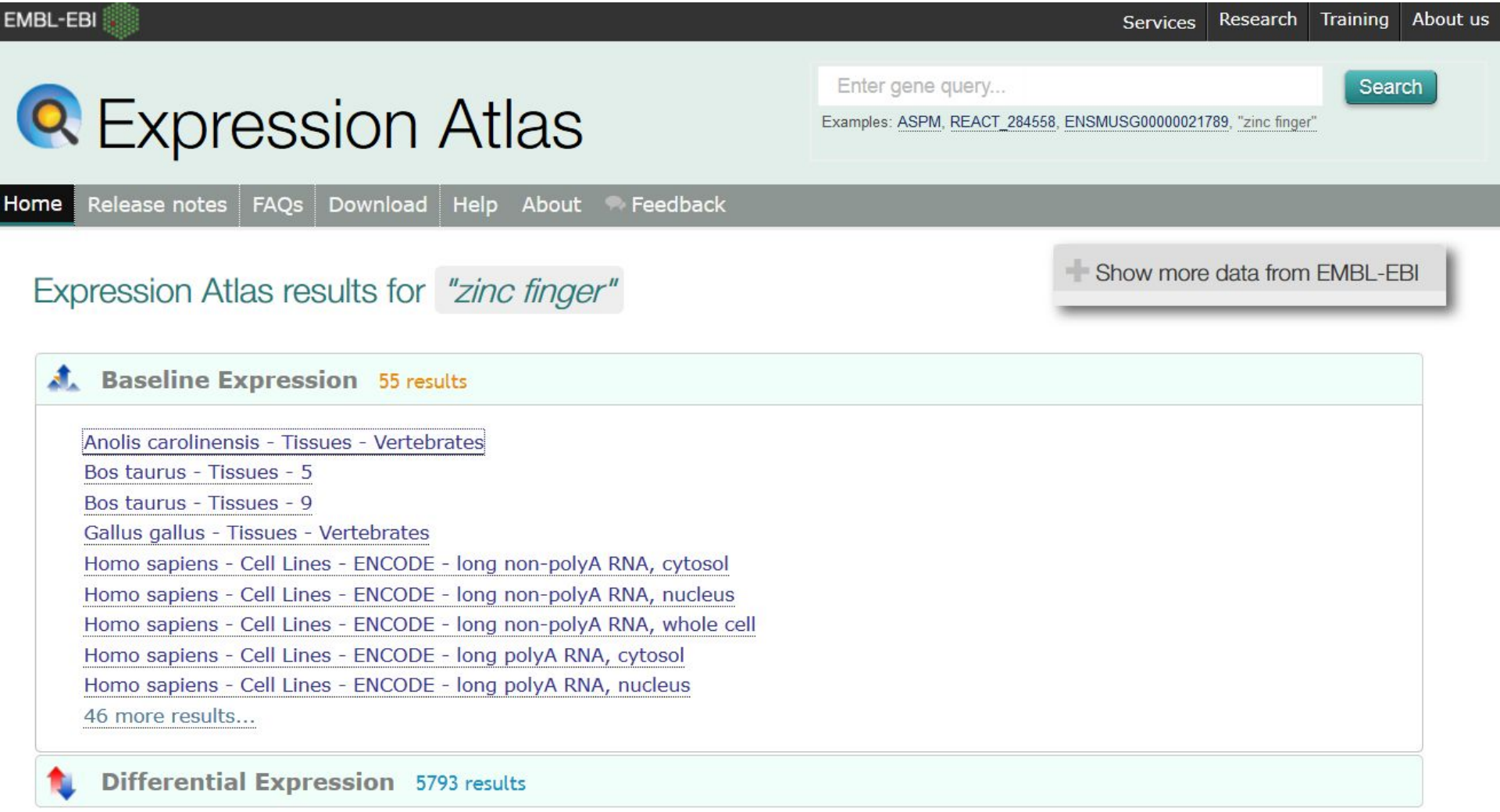
Showing 3 of 3 experiments found:

Display levels

Experiment	adipose tissue	adrenal gland	animal ovary	appendix	bladder	bone marrow	brain	breast	cerebellum	cerebral cortex	colon	duodenum
Tissues - 32 Uhlen's Lab							NA	NA	NA			
Tissues - Illumina Body Map				NA	NA	NA			NA	NA		NA
Tissues - Mammalian Kaessmann	NA	NA	NA	NA	NA	NA	NA	NA		NA	NA	NA


Expression Atlas
Enter gene query...
Search
Examples: ASPM, REACT_284558, ENSMUSG00000021789, "zinc finger"
Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas results for ENSG00000185303
Show more data from EMBL-EBI
SFTPA2 Homo sapiens surfactant protein A2
Baseline Expression Results found
Differential Expression 6 results
Showing 6 results
cutoffs: adjusted p-value 0.05 log2-fold change 1.0
Display log2-fold change
Comparison
Log2-fold change
'non-small cell lung cancer' vs 'normal'
'active Crohn's disease' vs 'normal'
'lung adenocarcinoma' vs 'normal'
'primary prostate cancer' vs 'benign prostate tumor'
'miR221 transfected' vs 'wild type'
'active ulcerative colitis' vs 'normal'


EMBL-EBI
Services
Research
Training
About us
Expression Atlas
Enter gene query...
Search
Examples: ASPM, REACT_284558, ENSMUSG00000021789, "zinc finger"
Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas: Differential and Baseline Expression
The Expression Atlas provides information on gene expression patterns under different biological conditions. Gene expression data is re-analysed in-house to detect genes showing interesting baseline and differential expression patterns. Read more about Expression Atlas.
Search...
Gene query
Enter gene query...
E.g. SFTPA2, zinc finger
Exact match
Organism
Homo sapiens
Sample properties
Enter condition query...
E.g. lung, leaf, "valproic acid", cancer
Search
Reset
Browse...
Baseline Experiments
See all baseline expression data sets in Expression Atlas.
Plant Experiments
See all expression data sets in plants in Expression Atlas.
All Experiments
Scroll through the complete list of all data sets in Expression Atlas.
iRAP: RNA-seq analysis tool
iRAP is a flexible pipeline for RNA-seq analysis that integrates many existing tools for filtering and mapping reads, quantifying expression and testing for differential expression. iRAP is used to process all RNA-seq data in Expression Atlas.
Publications
RNA-Seq Gene Profiling - A Systematic Empirical Comparison (PLoS One, 2014).
Expression Atlas update - a database of gene and transcript expression from microarray and sequencing-based functional genomics experiments (Nucleic Acids Research, 2014).

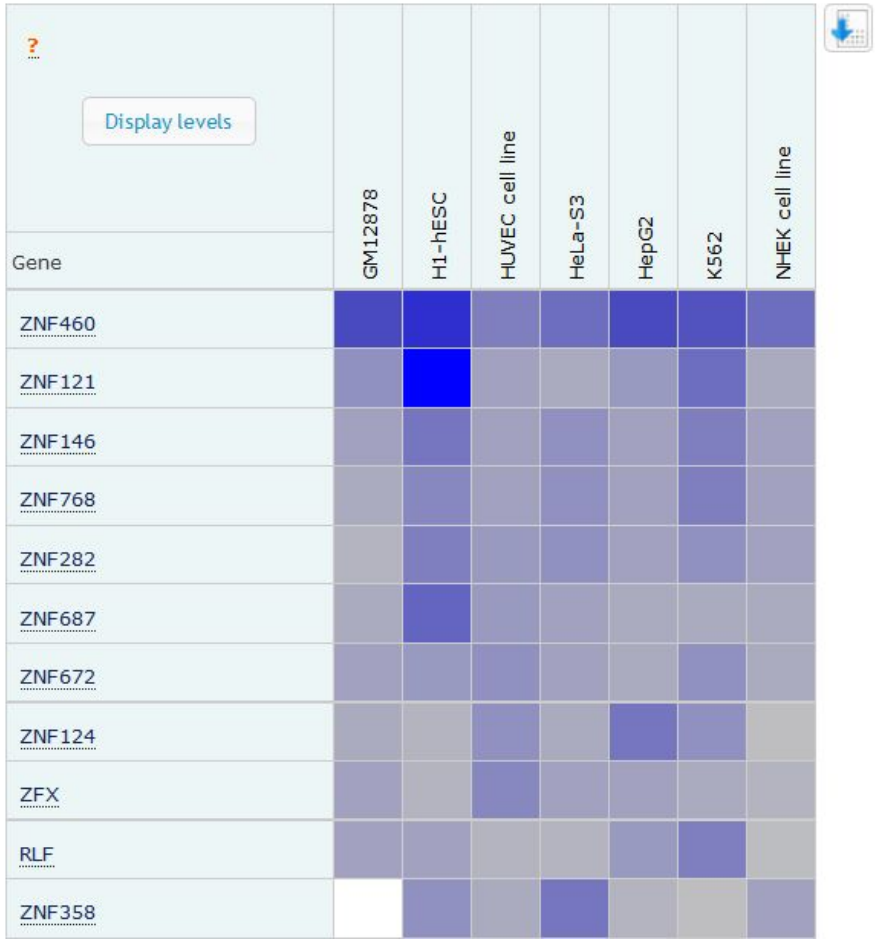

EMBL-EBI
Services
Research
Training
About us
Expression Atlas
Enter gene query...
Search
Examples: ASPM, REACT_284558, ENSMUSG00000021789, "zinc finger"
Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas results for "zinc finger"
Show more data from EMBL-EBI
Baseline Expression 55 results
Anolis carolinensis - Tissues - Vertebrates
Bos taurus - Tissues - 5
Bos taurus - Tissues - 9
Gallus gallus - Tissues - Vertebrates
Homo sapiens - Cell Lines - ENCODE - long non-polyA RNA, cytosol
Homo sapiens - Cell Lines - ENCODE - long non-polyA RNA, nucleus
Homo sapiens - Cell Lines - ENCODE - long non-polyA RNA, whole cell
Homo sapiens - Cell Lines - ENCODE - long polyA RNA, cytosol
Homo sapiens - Cell Lines - ENCODE - long polyA RNA, nucleus
46 more results...
Differential Expression 5793 results

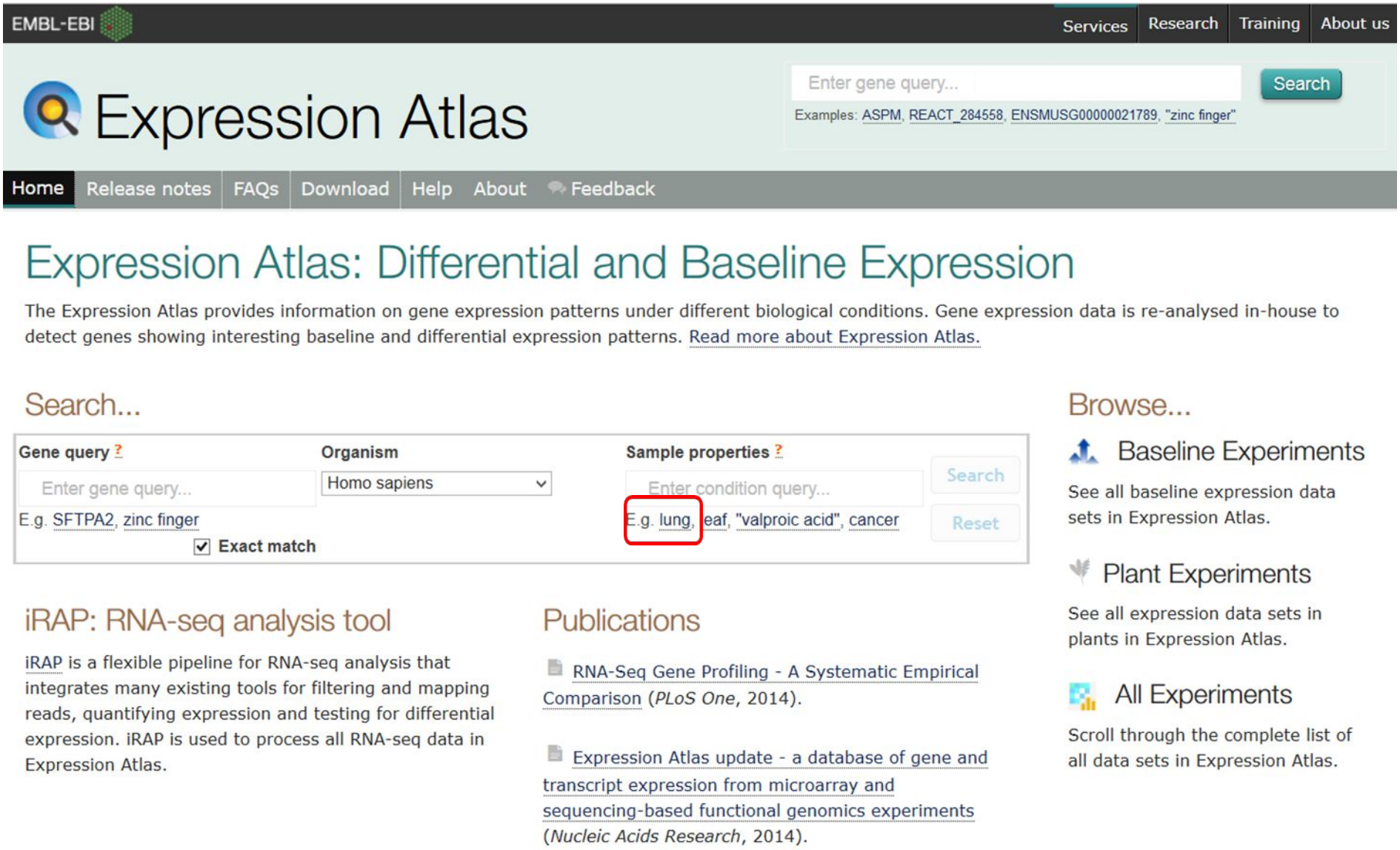
Ensembl Genome Browser

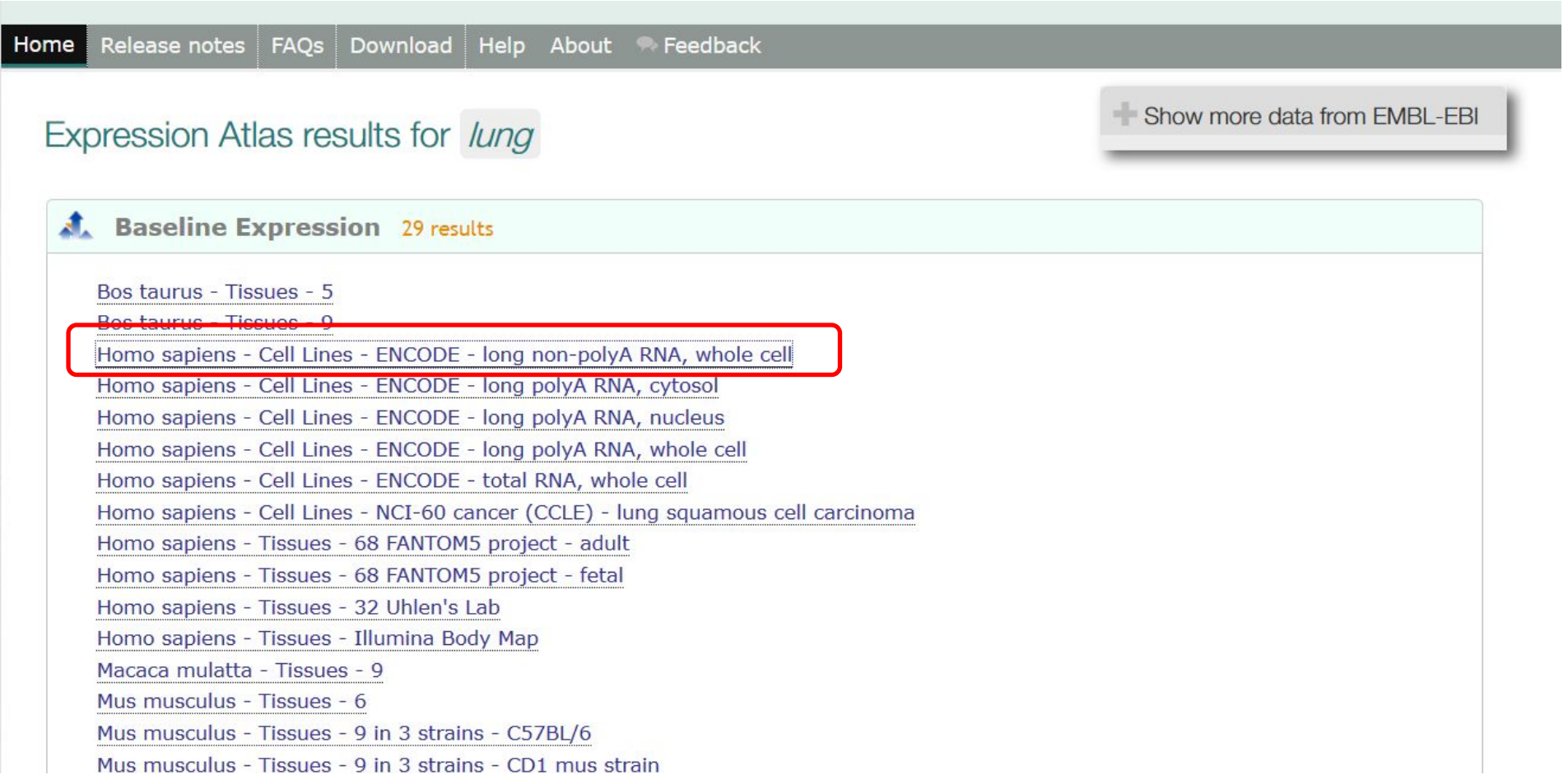
Open

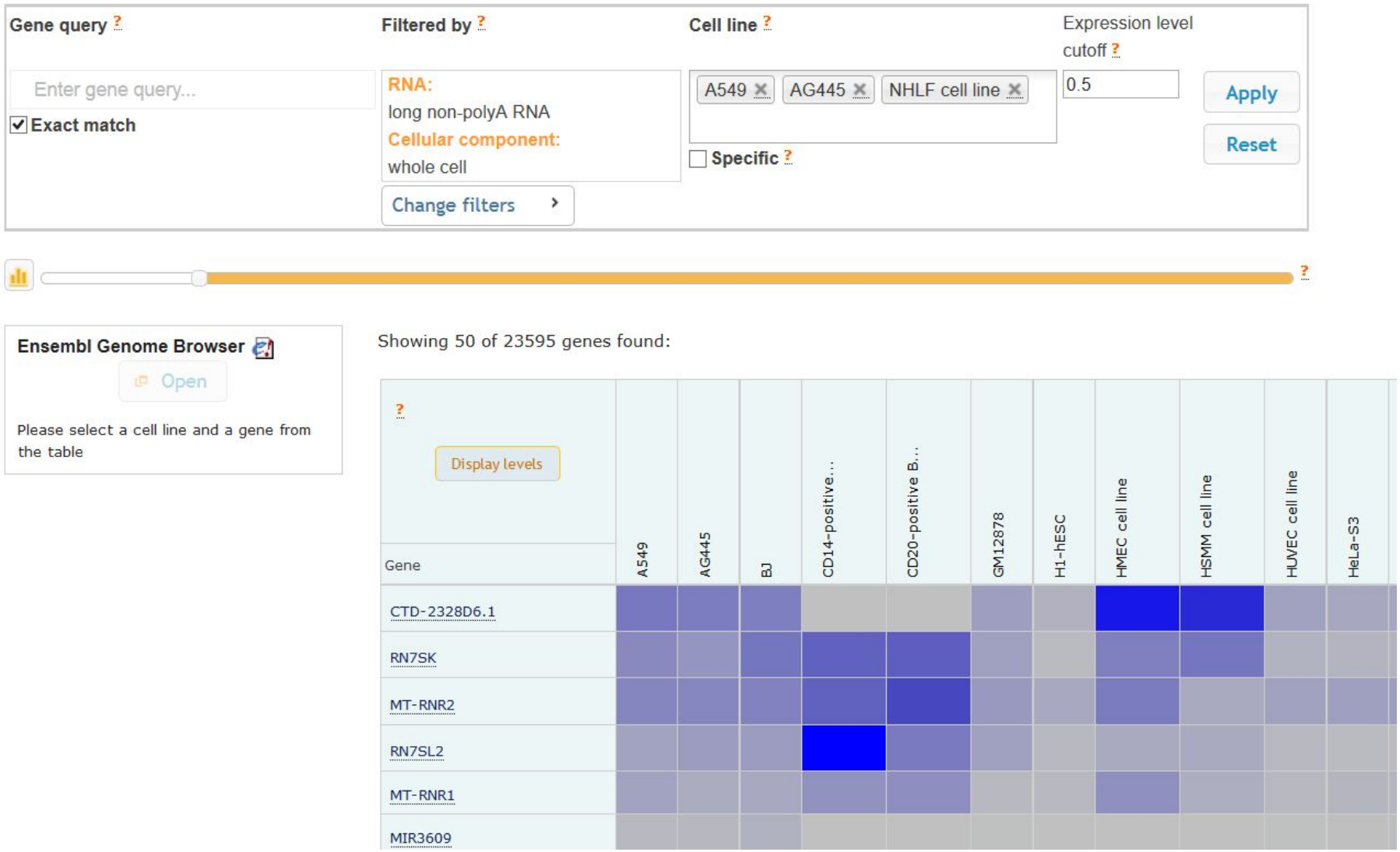
Please select a cell line and a gene from the table

Showing 50 of 324 genes found: (show by gene set)

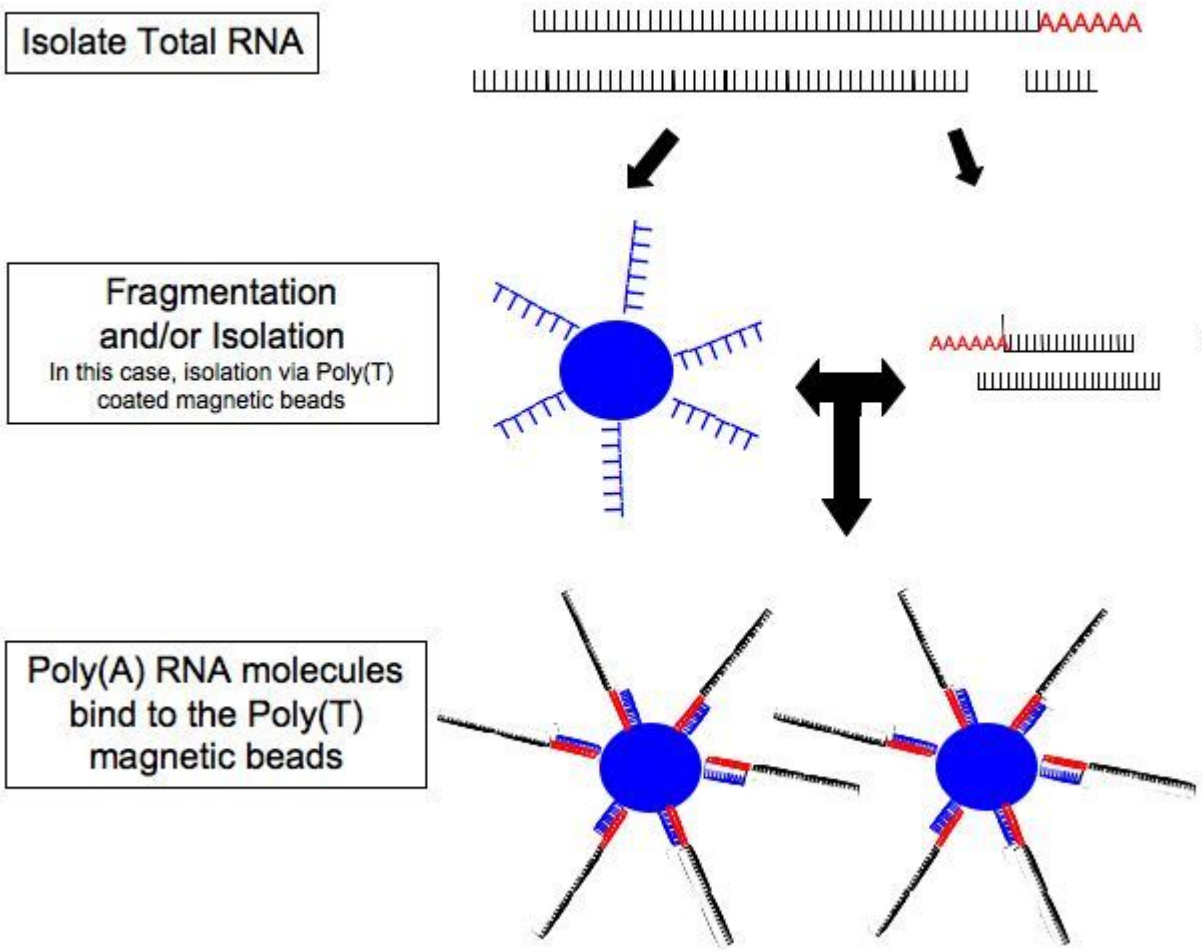



EMBL-EBI
Services
Research
Training
About us
Expression Atlas
Enter gene query...
Search
Examples: ASPM, REACT_284558, ENSMUSG00000021789, "zinc finger"
Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas: Differential and Baseline Expression
The Expression Atlas provides information on gene expression patterns under different biological conditions. Gene expression data is re-analysed in-house to detect genes showing interesting baseline and differential expression patterns. Read more about Expression Atlas.
Search...
Gene query
Enter gene query...
E.g. SFTPA2, zinc finger
Exact match
Organism
Homo sapiens
Sample properties
Enter condition query...
E.g. lung, leaf, "valproic acid", cancer
Search
Reset
Browse...
Baseline Experiments
See all baseline expression data sets in Expression Atlas.
Plant Experiments
See all expression data sets in plants in Expression Atlas.
All Experiments
Scroll through the complete list of all data sets in Expression Atlas.
iRAP: RNA-seq analysis tool
iRAP is a flexible pipeline for RNA-seq analysis that integrates many existing tools for filtering and mapping reads, quantifying expression and testing for differential expression. iRAP is used to process all RNA-seq data in Expression Atlas.
Publications
RNA-Seq Gene Profiling - A Systematic Empirical Comparison (PLoS One, 2014).
Expression Atlas update - a database of gene and transcript expression from microarray and sequencing-based functional genomics experiments (Nucleic Acids Research, 2014).


Home
Release notes
FAQs
Download
Help
About
Feedback
Expression Atlas results for lung
Show more data from EMBL-EBI
Baseline Expression 29 results
Bos taurus - Tissues - 5
Bos taurus - Tissues - 9
Homo sapiens - Cell Lines - ENCODE - long non-polyA RNA, whole cell
Homo sapiens - Cell Lines - ENCODE - long polyA RNA, cytosol
Homo sapiens - Cell Lines - ENCODE - long polyA RNA, nucleus
Homo sapiens - Cell Lines - ENCODE - long polyA RNA, whole cell
Homo sapiens - Cell Lines - ENCODE - total RNA, whole cell
Homo sapiens - Cell Lines - NCI-60 cancer (CCLE) - lung squamous cell carcinoma
Homo sapiens - Tissues - 68 FANTOM5 project - adult
Homo sapiens - Tissues - 68 FANTOM5 project - fetal
Homo sapiens - Tissues - 32 Uhlen's Lab
Homo sapiens - Tissues - Illumina Body Map
Macaca mulatta - Tissues - 9
Mus musculus - Tissues - 6
Mus musculus - Tissues - 9 in 3 strains - C57BL/6
Mus musculus - Tissues - 9 in 3 strains - CD1 mus strain


Gene query ?
Filtered by ?
Cell line ?
Expression level cutoff ?
Enter gene query...
Exact match
RNA:
long non-polyA RNA
Cellular component:
whole cell
Change filters
A549
AG445
NHLF cell line
Specific ?
0.5
Apply
Reset
Ensembl Genome Browser
Open
Please select a cell line and a gene from the table
Showing 50 of 23595 genes found:
Display levels
Gene
A549
AG445
BJ
CD14-positive...
CD20-positive B...
GM12878
H1-hESC
HMEC cell line
HSMM cell line
HUVEC cell line
HeLa-S3
CTD-2328D6.1
RN7SK
MT-RNR2
RN7SL2
MT-RNR1
MIR3609

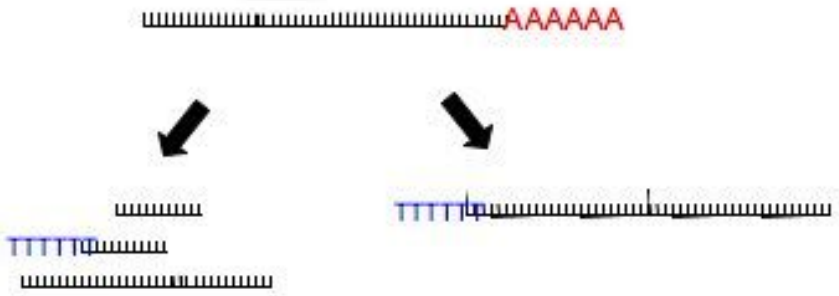
RNASeq



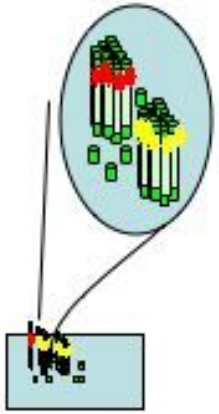
(РНК Секвенирование), также называемый Whole Transcriptome Shotgun Sequencing (WTSS), является технология, которая использует возможности секвенирования следующего поколения выявлять снимок наличия и количества РНК генома в данный момент времени

Magnetically isolate and wash beads

Fragment and/or Reverse Transcribe

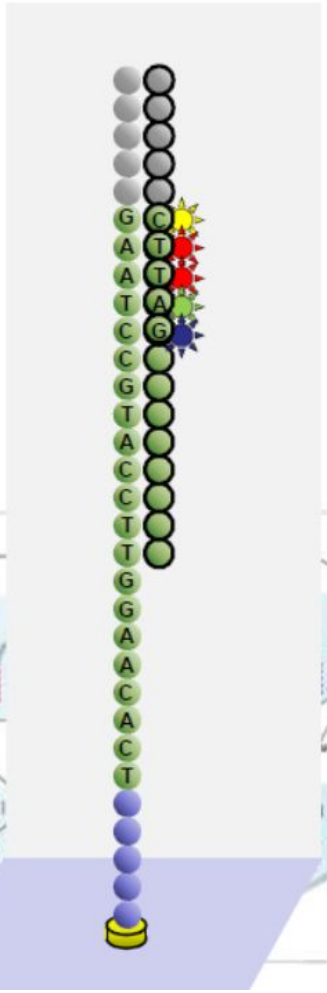


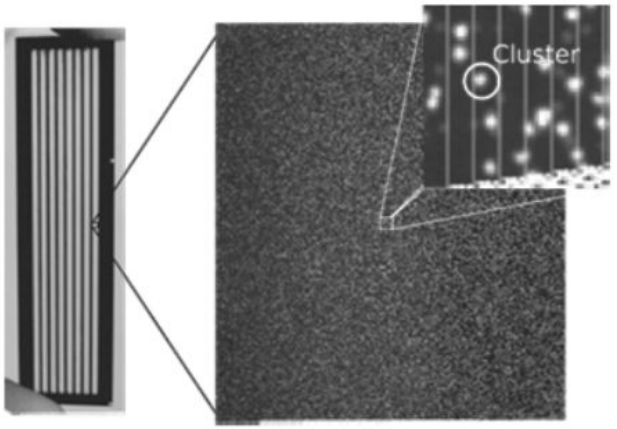
Fragmentation (if not done already), size selection, and sequence

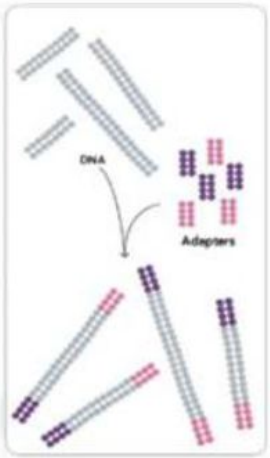


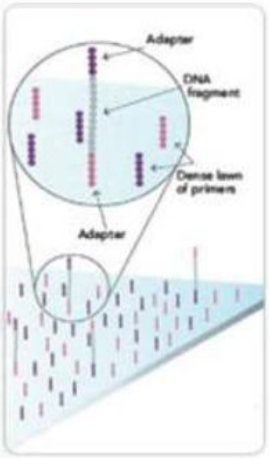
Illumina Solexa, Roche 454, or ABI SOLiD
Graphic shown here is Illumina

illumina®

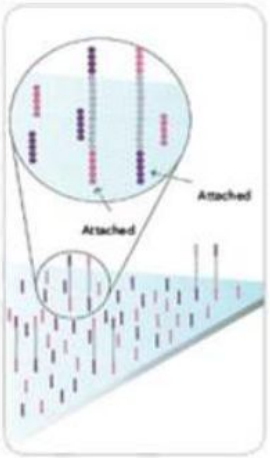

GAATCCGTACCTTGGAACACT
CTTAG

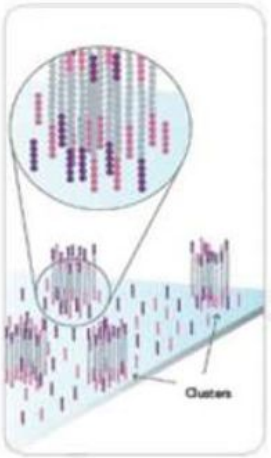

Cluster


DNA
Adapters


Adapter
DNA fragment
Dense lawn of primers
Adapter

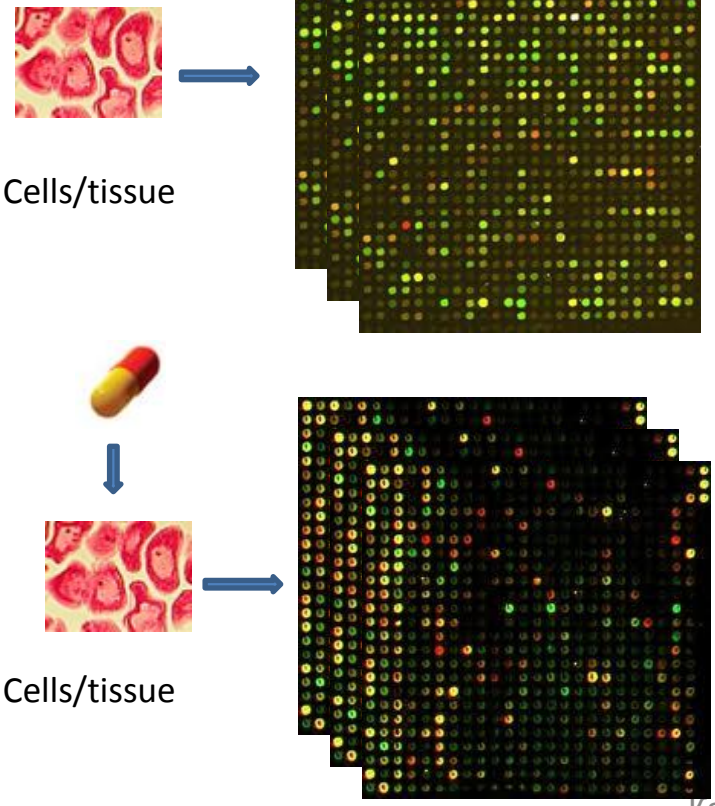
Attached terminus


Attached
Attached


Clusters

Лекарственно-индуцированное изменение генной экспрессии

Исходные данные микроэррейных экспериментов



Normalization
Comparative analysis by statistical methods
(e.g. Student's t-test
or hypergeometric test)

Лекарственно-индуцированное изменение профиля генной экспрессии или генетическая подпись

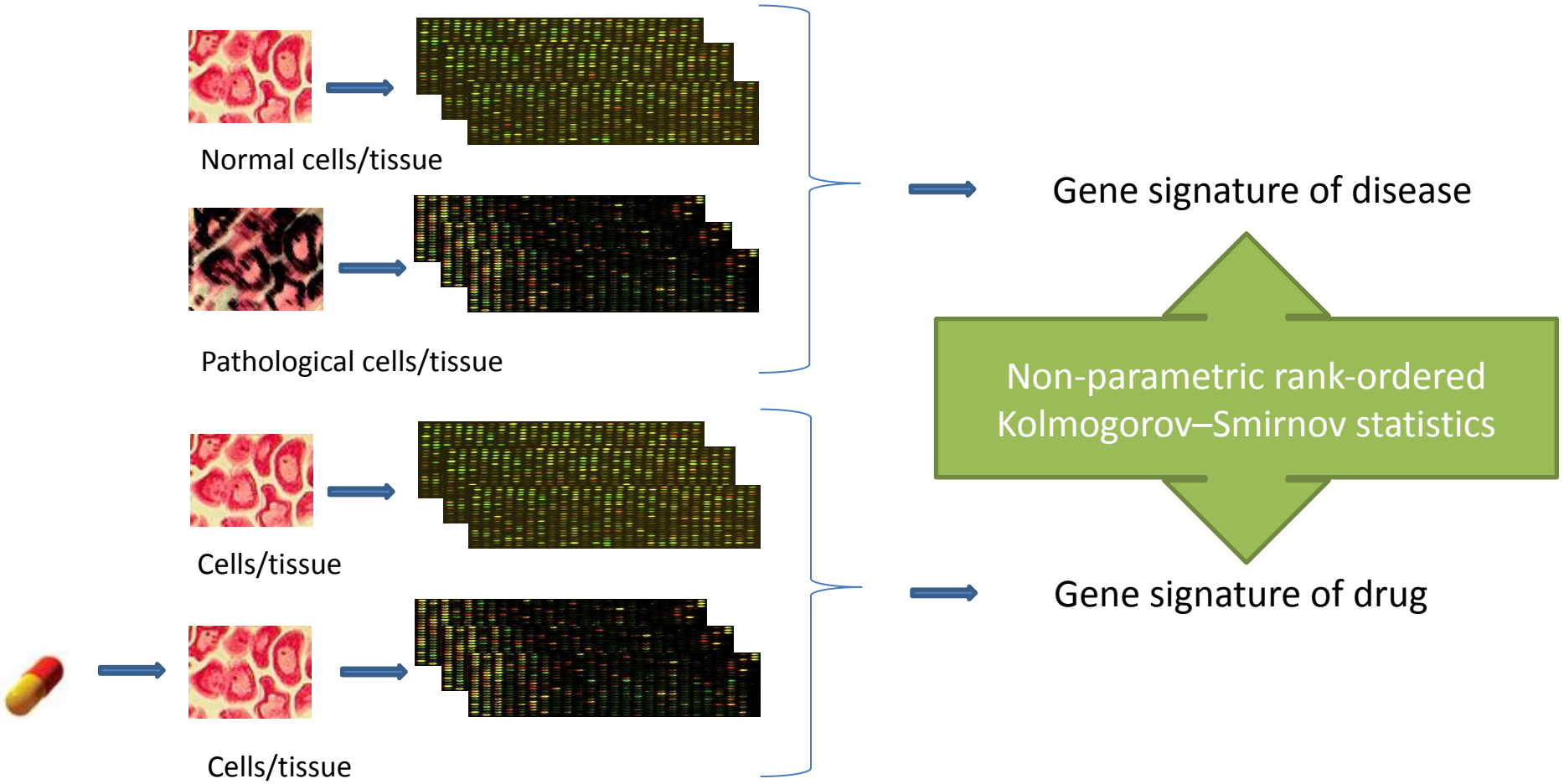
Up regulated genes

CASP3
CDKN1A
CDKN1B
FAS
GPX5
TP53

Down regulated genes

CCND1
CCND2
MYC
NFKB1
PCNA
PTGS2

Connectivity map (CMap) подход



Lamb, J. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935

Применение CMap подхода

Human Disease-Drug Network Based on Genomic Expression Profiles - used Connectivity Map and GEO DataSets for drug repositioning of 395 drugs (*PLoS ONE, 2009, 4(8): e6536*)

Comprehensive gene expression profiles of NK cell neoplasms identify vorinostat as an effective drug candidate – used Connectivity Map data analysis for 6 drugs (*Cancer Lett. 2013 pii: S0304-3835(13)00020-7*)

Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease - GEO DataSets for 143 drugs (*Sci Transl Med. 2011 August 17; 3(96): 96ra76*)

Identification of Identical Transcript Changes in Liver and Whole Blood during Acetaminophen Toxicity – used *in house* data of gene expression profiles for acetaminophen. (*Front Genet. 2012;3:162*)

Prediction of synergistic effects of pairwise drug combinations from gene microarray data - used *in house* data of gene expression profiles from MCF-7 cells for docetaxel and gefitinib in different concentrations. (*Bioinformatics, 2011, 27, i310-i316*).

MIAME

В целях стандартизации представления и анализ данных микроэрреев, Alvis Brazma и его коллеги из 17 учреждений ввели формат Minimum Information About a Microarray Experiment (Минимальные сведения о микроэррейном эксперименте) - MIAME. В рамках MIAMI стандартизируется шесть областей информации:

1. Планирование эксперимента
2. Дизайн микрочипа
3. Проба подготовка
4. Процедура гибридизации
5. Анализ изображений
6. Контроль в отношении нормализации

http://fged.org/projects/miame/

Экспрессия генов: анализ микроэрейных данных

- Экспрессия генов
- Микрочипы (Microarrays)
- Предварительная обработка (препроцессинг)
  - нормализация
  - диаграммы рассеяния (Scatter plots)
- Статистический анализ
  - Т-тест
  - ANOVA
  - расстояния
  - кластеризация
  - анализ главных компонентов (PCA)

Анализ микроэррейных данных

Начинаем с матрицы данных
(значения генной экспрессии в различных образцах)

гены
(уровни транскрипции РНК)

	A	B	C	D
1	Gene Sym	Chromosor	DS_Cereb	DS_Cereb
2	ATP5O	21	10.3957	10.2149
3	CRYBB2	21	5.95712	6.07945
4	C21orf33	21	8.9064	8.74096
5	WRB	21	9.67306	9.3076
6	ALOX5	10	4.35077	4.4185
7	HRMT1L1	21	9.16597	8.91893
8	PTPN1	20	6.32176	6.27589
9	SBF1	22	5.06861	4.94405
10	ATP5J	21	9.27822	8.95333
11	CAMKK2	12	8.13921	7.9926
12	NRTN	19	3.39505	3.33395
13	CTDSPL	3	5.77447	5.69664
14	USP16	21	7.64692	7.42381
15	RUNX1	21	3.52972	3.58988
16	DONSON	21	5.37143	4.72189
17	FLOT1	6	9.50971	9.38725
18	USP25	21	6.94273	7.08138
19	SOD1	21	10.7198	10.3198
20	ATP5O	21	7.4918	7.50782
21	RARA	17	5.1042	4.90355
22	DCTN6	8	8.65151	8.69294
23	TIMM23	10	5.42184	5.42443
24	SP100	2	3.98977	3.97176
25	PDXK	21	7.95603	7.80782
26	PAPD1	10	5.99899	5.87142
27	SUMO3	21	9.36959	9.30012
28	MSC	8	4.59863	4.58913
29	SPRR2C	1	4.50502	4.70578
30	DOCK6	19	4.4069	4.40616

Обычно много генов (>> 20,000) и несколько образцов (~ 10)

Предварительная обработка (препроцессинг)

Наблюдаемые различия в экспрессии генов могут быть связаны с транскрипционными изменения, или же они могут быть вызваны артефактами, такими, как:

- различная эффективность окрашивания Cy3 (зеленый), Cy5 (красный)
- неравномерное распределение ДНК на поверхности массива
- изменения связанные с чистотой или количеством РНК
- изменения связанные с эффективностью отмывки
- изменения связанные с эффективностью сканирования

Предварительная обработка (препроцессинг)

- Основная цель предварительной обработки данных заключается в как можно более полном удалении систематической погрешности в данных, сохраняя при этом различия в экспрессии генов, которое происходит из-за биологически соответствующих изменений в транскрипции.

- Основное предположение большинства процедур нормализации является то, что средний уровень экспрессии генов не меняется в эксперименте.

Глобальная нормализация данных

Глобальная нормализация используется для коррекции двух или более наборов данных. В одном общем случае образцы помеченные Cy3 (зеленый краситель) или Cy5 (красный краситель) гибридизовали с ДНК-элементов на микрочипе. После промывки зонды возбуждаются с помощью лазера и исследуются с помощью сканирующего конфокального микроскопа.

Глобальная нормализация данных

Глобальная нормализация используется для коррекции двух или более наборов данных.

Пример: общая флуоресценция в
Cy3 канал = 4 млн. единиц
Cy 5 канал = 2 млн. единиц

Тогда нескорректированное отношение для гена может показать 2000 единиц в сравнении 1000 единиц. Появление такого артефакта, приводит к тому, что показывает 2-кратное увеличение экспрессии.

Глобальная процедура нормализации

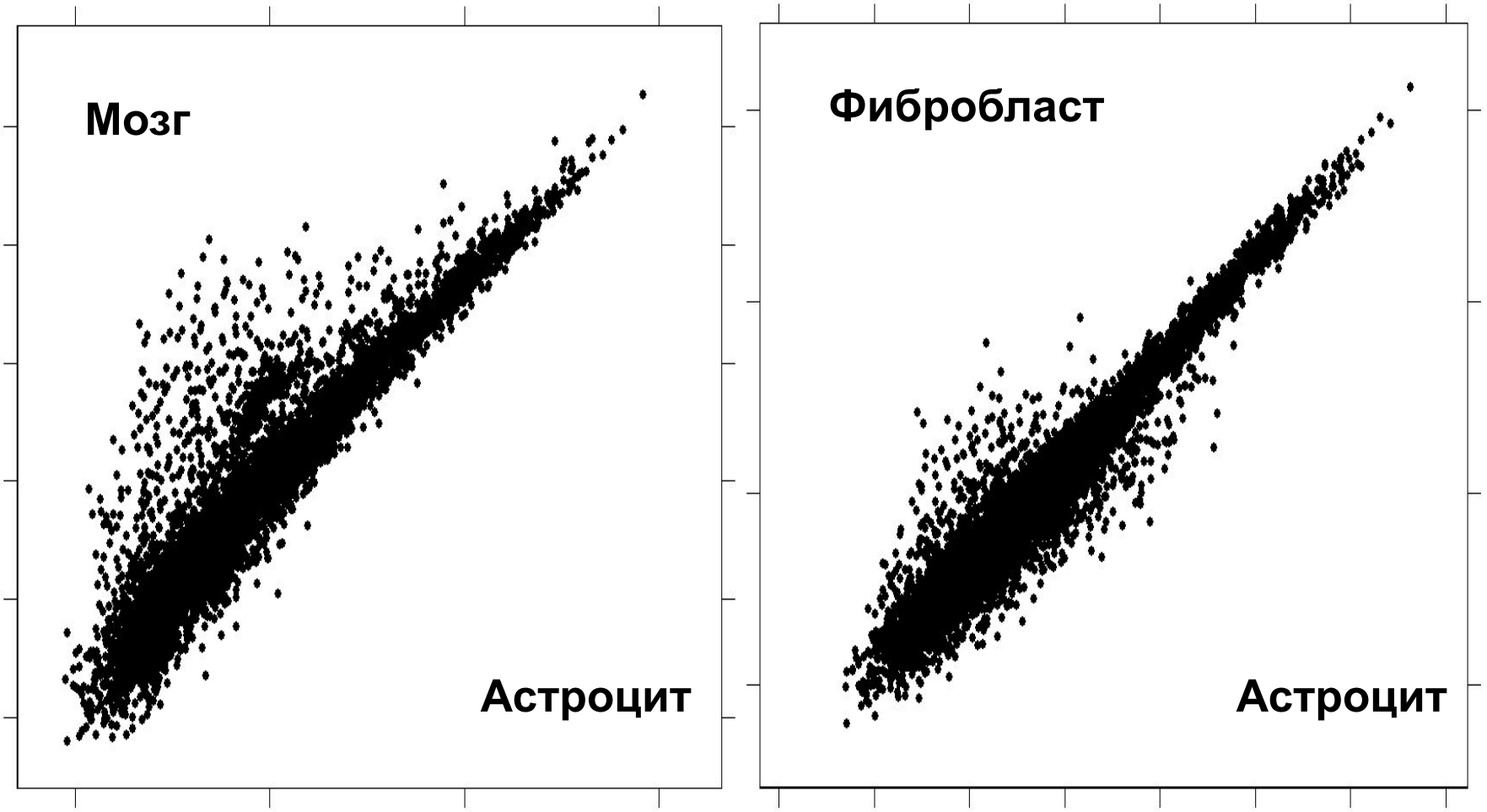
- Шаг 1: вычесть значения интенсивности фона (используется пустая область массива)
- Шаг 2: глобальная нормализация, так чтобы среднее соотношение было равно 1 (применяется к 1-канальным или 2-канальным наборам данных)

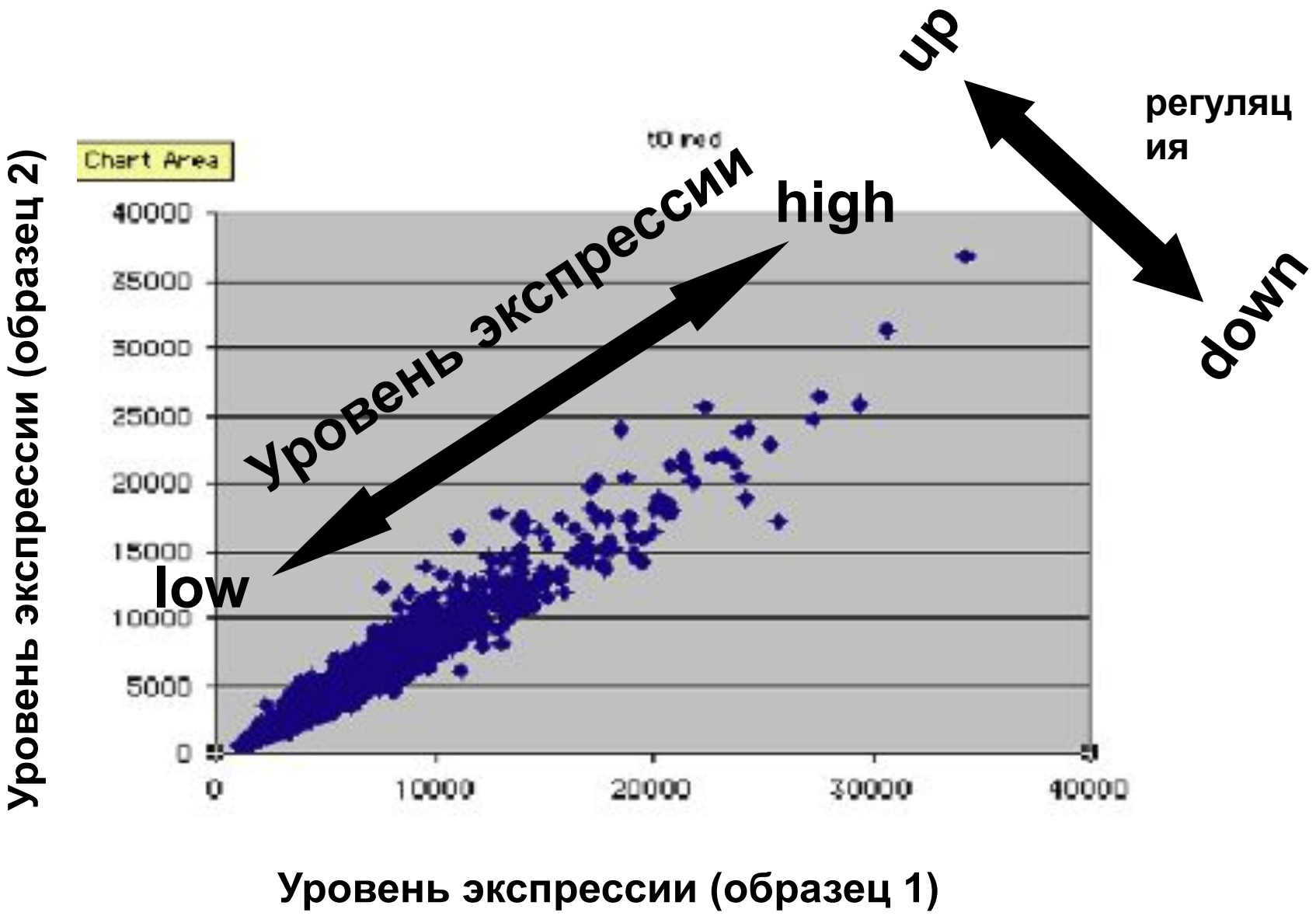
Диаграммы рассеяния (Scatter plots)

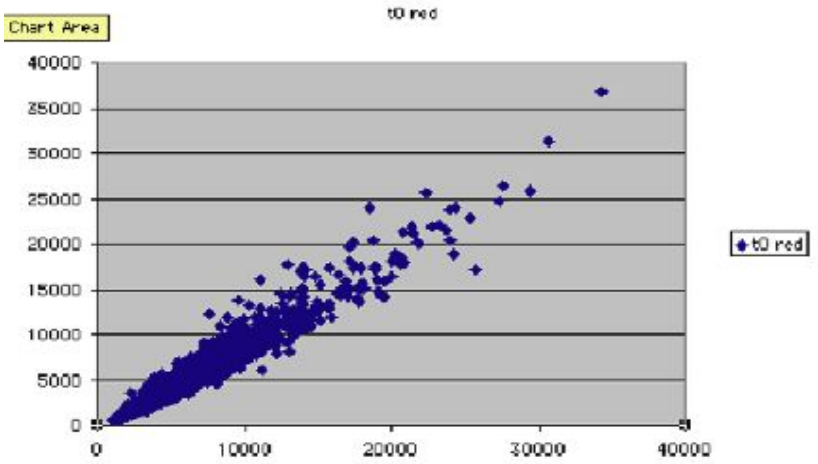
Полезны для представления значений экспрессии генов из двух экспериментов микрочипов (например, контроль, эксперимент)

- Каждая точка соответствует значению экспрессии генов
- Большинство точек находятся вдоль линии
- Выбросы составляют гипо- и гиперэкспрессируемые гены

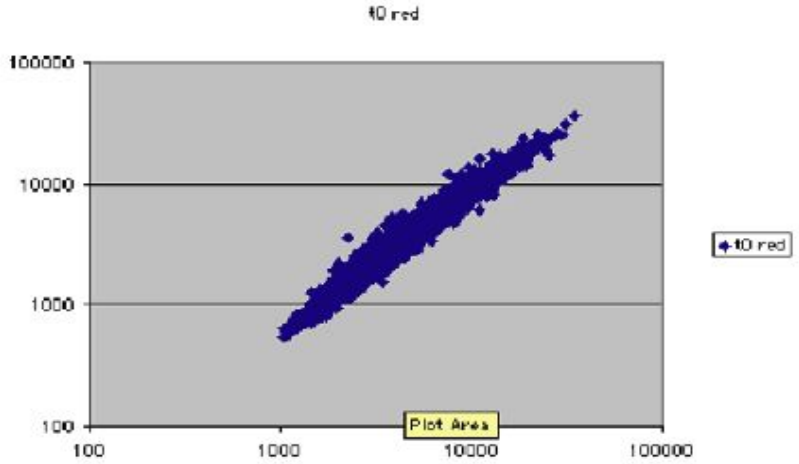
Дифференциальная генная экспрессия в различных тканях и клетках




up
регуляц
ия
down
Chart Area
Уровень экспрессии
high
low
Уровень экспрессии (образец 2)
40000
35000
30000
25000
20000
15000
10000
5000
0
0
10000
20000
30000
40000
Уровень экспрессии (образец 1)



Логарифмическая трансформация данных

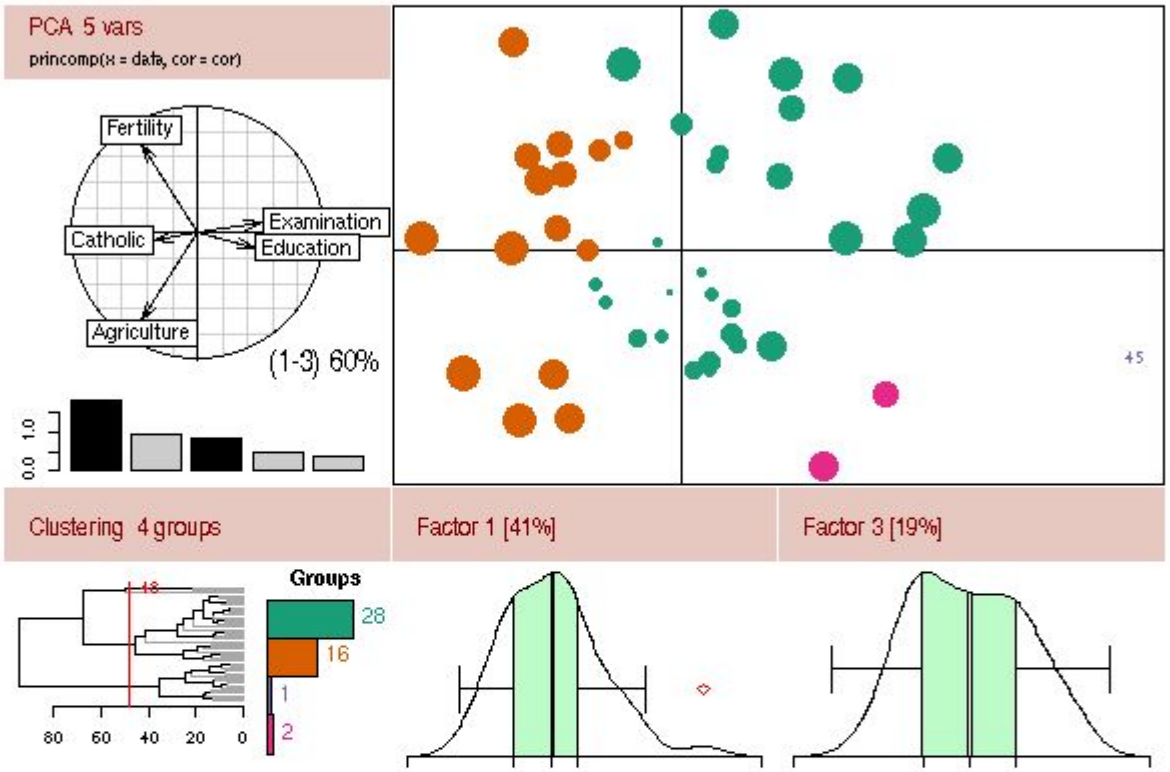


Диаграммы рассеяния (Scatter plots)

Обычно данные изображаются логарифмических координатах

Время	Изменение	исходное значение	$\log_2$ значение
t=0	начальное	1.0	0.0
t=1h	нет изменений	1.0	0.0
t=2h	2-fold up	2.0	1.0
t=3h	2-fold down	0.5	-1.0

The R Project for Statistical Computing



About R
- What is R?
- Contributors
- Screenshots
- What's new?

Download
- CRAN

R Project
- Foundation
- Members & Donors
- Mailing Lists
- Bug Tracking
- Developer Page
- Conferences
- Search

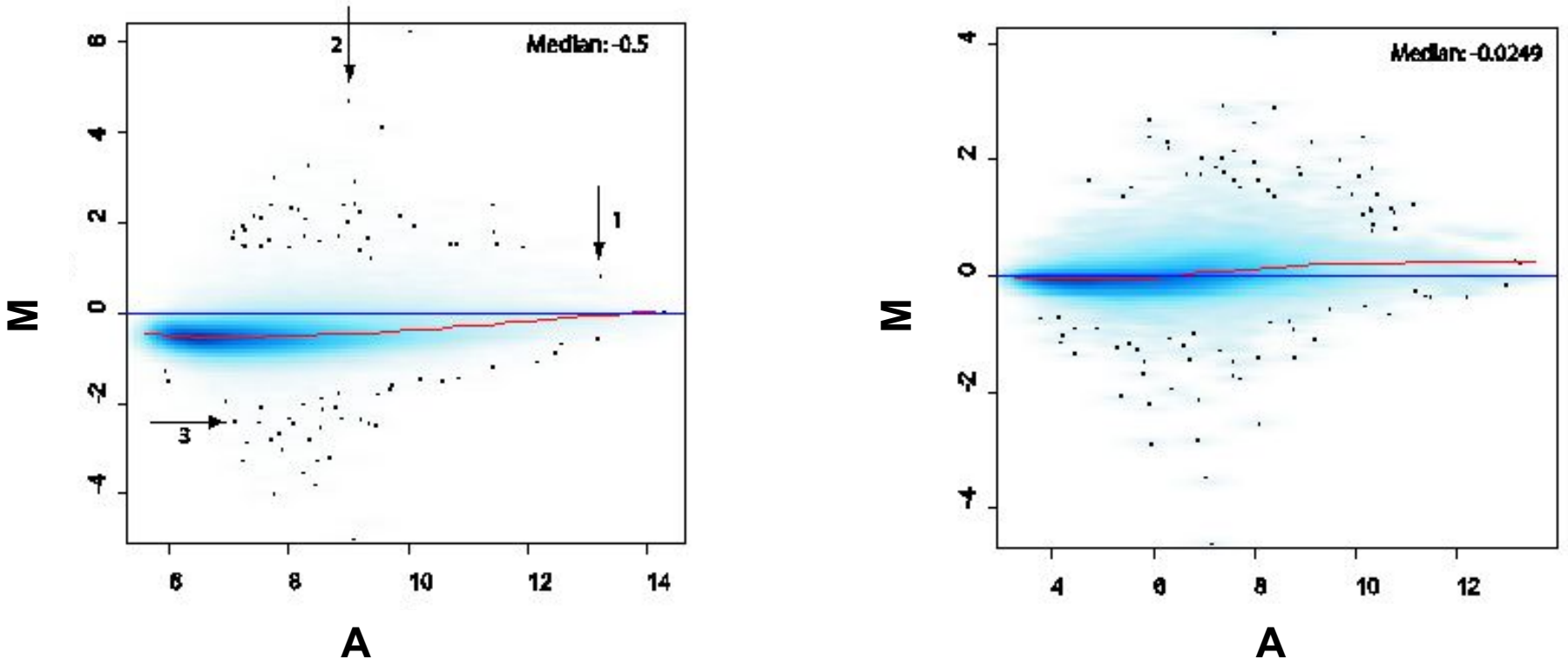
Documentation
- Manuals
- FAQs
- Newsletter
- Wiki
- Books
- Certification
- Other

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

http://www.r-project.org

Эффект нормализации



После RMA (Robust multi-array analysis) процедуры нормализации, the медиана значений близка к нулю и исправлены перекосы.

Экспрессия генов: анализ микроэрейных данных

- Экспрессия генов
- Микрочипы (Microarrays)
- Предварительная обработка (препроцессинг)
  - нормализация
  - диаграммы рассеяния (Scatter plots)
- Статистический анализ
  - Т-тест
  - ANOVA
  - расстояния
  - кластеризация
  - анализ главных компонентов (PCA)

T-тест

T-тест широко используется для оценки различия в средних значениях между двумя группами.

$$t = \frac{x_1 - x_2}{SE} = \frac{\textbf{Разница между средними}}{\textbf{Изменчивость (стандартная ошибка отклонения)}}$$

Вопросы

Адекватный ли размер выборки (N)?
Являются ли данные нормально распределенными?
Есть ли разница между двумя группами?
Уместно ли задать уровень значимости для $p < 0{,}05$?