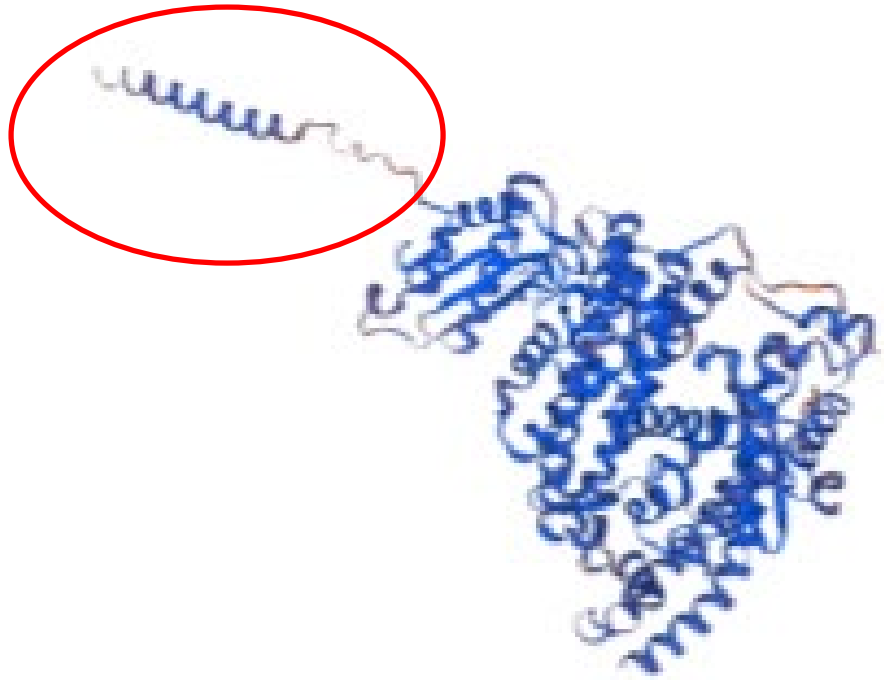
Microarray Data analyses

Kailin Tang

2021.6

REVIEW

- 模板：覆盖率， 相似度
- 多次修正



REVIEW

模型评估

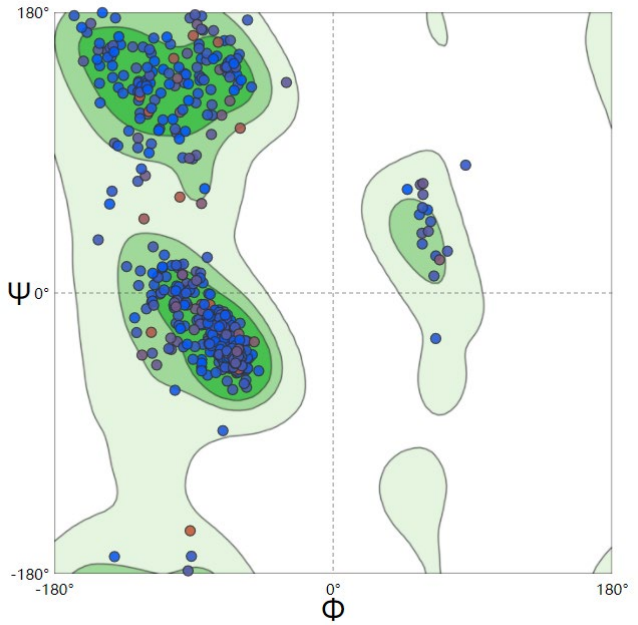
Structure Assessment

G3QWX4_GORGO G3QWX4 Angiotensi converting enzyme; Model 01;

Project Data

Created: Thu 3rd Jun, 16:46;

Ramachandran Plots



General | Glycine | Proline | Pre-Proline | Chain B

MolProbity Results

	MolProbity Score	0.88	
☐	Clash Score	0.25	
	Ramachandran Favoured	96.11%	
☐	Ramachandran Outliers	0.40%	B734 PRO, B170 SER, B679 ILE
☐	Rotamer Outliers	1.06%	B325 GLN, B761 ILE, B39 LEU, B719 ASP, B133 CYS, B341 LYS, B329 GLU
☐	C-Beta Deviations	5	B679 ILE, B761 ILE, B254 SER, B150 GLU, B752 VAL
	Bad Bonds	0 / 6254	
☐	Bad Angles	36 / 8492	(B736 GLN-B737 PRO), (B177 ARG-B178 PRO), B509 ASP, (B499 ASP-B500 PRO), (B450 LEU-B451 PRO), (B589 GLU-B590 PRO), B38 ASP, (B582 ARG-B583 PRO), B295 ASP, B315 PHE, B586 ASN, (B678 ARG-B679 ILE), B523 PHE, B378 HIS, (B234 LYS-B235 PRO), B715 PHE, (B345 HIS-B346 PRO), B345

Results obtained using MolProbity version 4.4

Data analyses-Differential expression

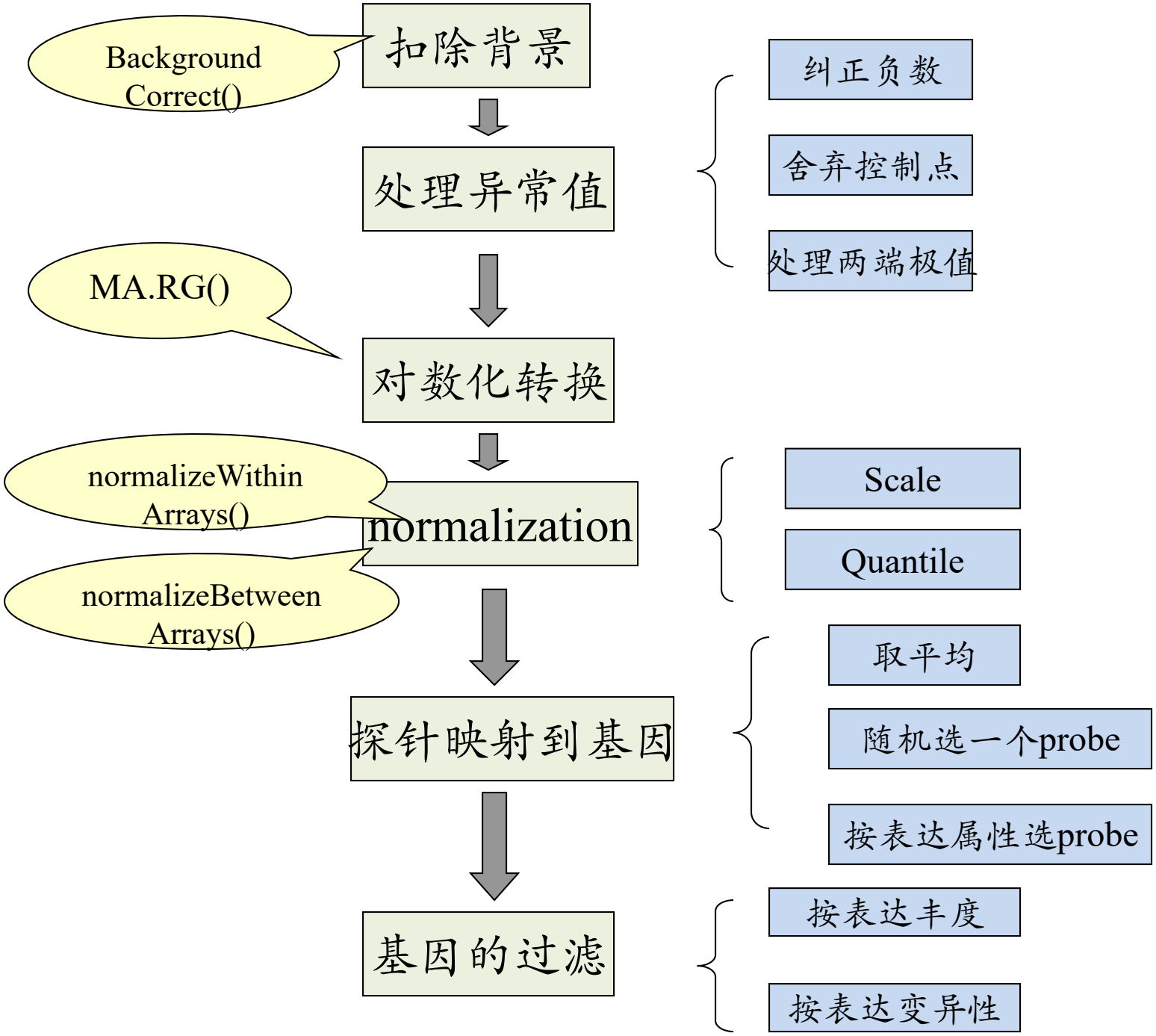
Samples

Probes

GDS4387									
ID_REF	GSM9525	GSM9525	GSM9525	GSM9525	GSM9525	GSM9525	GSM9525	GSM9525	GSM9525
disease state	HBV-asso	HBV-asso	HBV-asso	HBV-asso	HBV-asso	HBV-asso	HBV-asso	HBV-asso	HBV-asso
individual	241	241	241	241	219	219	219	219	219
235915_at	6.84	6.4	4.46	5.04	7.56	8.06	8	8.02	8.08
236652_at	5.02	4.85	4.56	4.59	7.52	8.08	7.88	8.02	7.87
237530_at	4.02	4.08	3.18	3.33	5.12	5.97	6.13	5.99	5.93
238160_at	3.89	3.97	3.74	3.51	8.01	8.76	8.53	8.83	8.53
238835_at	3.42	3.19	3.2	3.04	3.66	3.81	3.69	4.05	3.69
239591_at	2.74	2.75	2.88	2.88	5.75	5.19	5.83	5.16	5.77
240110_at	3.23	3.63	4.09	3.65	5.92	6.49	6.46	6.5	6.13
241914_s_at	3.73	4.66	4.27	4.52	6.88	8.05	7.73	8.29	7.86
242680_at	5.19	4.28	4.39	4.53	5.61	5.2	6.09	5.9	6.21
242817_at	4.03	4.15	4.34	4.09	5.6	6.35	6.13	6.58	6.09
243146_at	4.84	3.87	4.05	4.42	3.92	4.43	4.73	4.38	4.37
243799_x_at	3.77	3.82	4.17	3.9	7.22	7.96	7.55	7.85	7.56
244276_at	4.73	4.13	4.17	3.63	7.18	8.11	7.91	7.95	7.87
244434_at	3.85	4.01	5.95	4.44	4.34	4.11	4.72	3.65	4.79
244562_s_at	5.24	4.87	5.27	4.82	8.36	8.38	8.48	8.65	8.39
244723_at	3.48	3.41	3.65	3.56	4.73	4.63	4.17	4.67	4.16
31835_at	6.24	6.68	6.29	7.08	9.85	10.29	10.31	10.51	10.35
37020_at	8.74	8.88	7.51	7.63	12.34	12.61	12.62	12.86	12.73
37512_at	4.2	4.2	3.98	3.73	8.34	9.34	9.42	9.68	9.39
39763_at	6.2	7.16	6.68	6.63	11.32	11.74	11.72	11.97	11.71

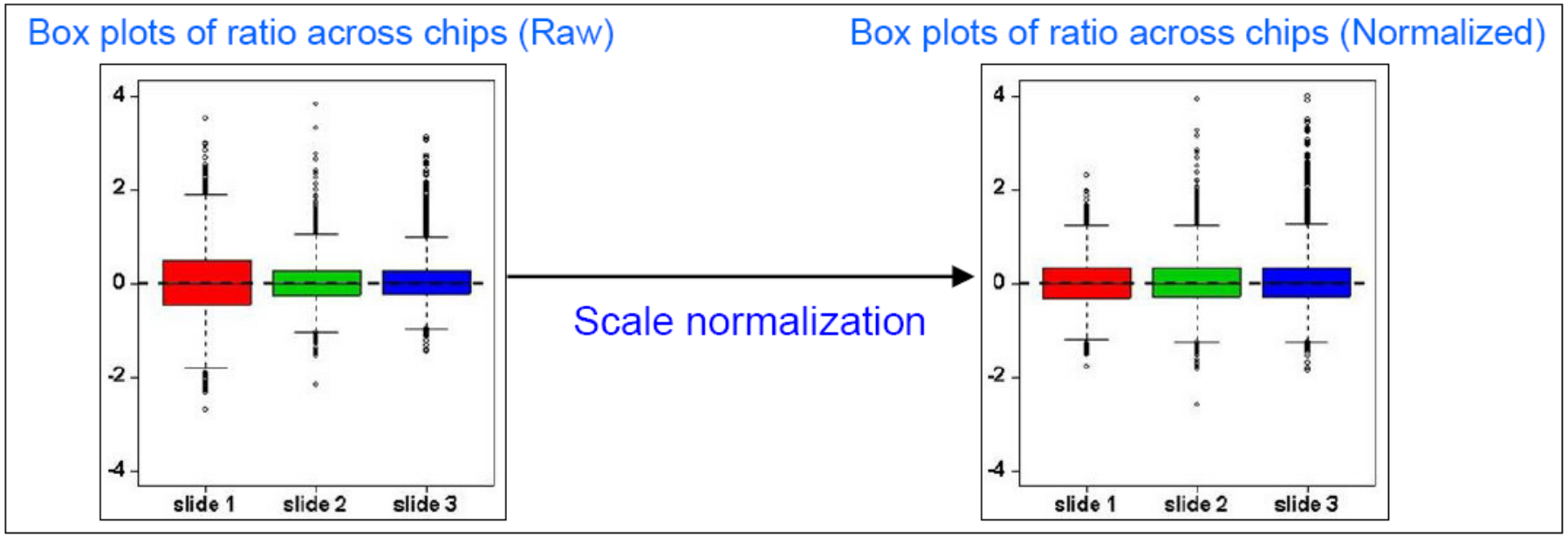
Data table

ID	GB_ACC	SPOT_ID	Species Scientific Name	Annotation Date	Sequence Type	Sequence Source	Target Description
1007_s_at	U48705		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	U48705 /FEATURE=mRNA /DE
1053_at	M87338		Homo sapiens	Oct 6, 2014	Exemplar sequence	GenBank	M87338 /FEATURE= /DEFINITI
117_at	X51757		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	X51757 /FEATURE=cds /DEFIN
121_at	X69699		Homo sapiens	Oct 6, 2014	Exemplar sequence	GenBank	X69699 /FEATURE= /DEFINITI
1255_g_at	L36861		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	L36861 /FEATURE=expanded_
1294_at	L13852		Homo sapiens	Oct 6, 2014	Exemplar sequence	GenBank	L13852 /FEATURE= /DEFINITI
1316_at	X55005		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	X55005 /FEATURE=mRNA /DE
1320_at	X79510		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	X79510 /FEATURE=cds /DEFIN
1405_i_at	M21121		Homo sapiens	Oct 6, 2014	Exemplar sequence	GenBank	M21121 /FEATURE= /DEFINITI
1431_at	J02843		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	J02843 /FEATURE=cds /DEFIN
1438_at	X75208		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	X75208 /FEATURE=cds /DEFIN
1487_at	L38487		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	L38487 /FEATURE=mRNA /DE
1494_f_at	M33318		Homo sapiens	Oct 6, 2014	Exemplar sequence	Affymetrix Proprietary Database	M33318 /FEATURE=mRNA /DE
1552256_a_at	NM_005505		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_005505.2 /DB_XREF=g
1552257_a_at	NM_015140		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_015140.1 /DB_XREF=g
1552258_at	NM_052871		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_052871.1 /DB_XREF=g
1552261_at	NM_080735		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_080735.1 /DB_XREF=g
1552263_at	NM_138957		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_138957.1 /DB_XREF=g
1552264_a_at	NM_138957		Homo sapiens	Oct 6, 2014	Consensus sequence	GenBank	gb:NM_138957.1 /DB_XREF=g


Background Correct()
扣除背景
处理异常值
纠正负数
舍弃控制点
处理两端极值
MA.RG()
对数化转换
normalizeWithin Arrays()
normalization
Scale
Quantile
normalizeBetween Arrays()
探针映射到基因
取平均
随机选一个probe
按表达属性选probe
基因的过滤
按表达丰度
按表达变异性

Between-chip normalization: Scale normalization

- A popular method for between-chip normalization of cDNA chips.
- What it does: It makes the range of values in each chip similar to each other.
- Rationale: It can be assumed that the range of most gene's expression values do not change across replicates and samples.
- How it works: Calculate a scaling factor for each chip and divide values in each chip.



An extra: Detailed procedure as implemented in limma package

1. For each column in the expression data table, get the median of absolute values. Then apply log2 transform.
2. Get the mean of *n* column medians.
3. Calculate a scaling factor for each column (=2^col_median / 2^mean_of_col_medians)
4. Divide values in each column by its scaling factor.
5. Steps 1-4 are applied to "M" and "A" separately

Between-chip normalization: Quantile normalization

- Use this for both one-channel array (and two-channel array also).
- What it does: Make the replicate arrays to have equal distribution.
- Rationale: It can be assumed that distribution of expression values does not change much across all replicates and samples under study.
- How it works: Values for each column are ranked, then the average per rank is taken and is reattributed to each column according to the original rank.

	Chip1	Chip2
Probe A	10	14
Probe B	12	12
Probe C	4	8
Probe D	8	10

Sort the data from each chip

	Chip1	Chip2
Rank1	4 (probe C)	8 (probe C)
Rank2	8 (probe D)	10 (probe D)
Rank3	10 (probe A)	12 (probe B)
Rank4	12 (probe B)	14 (probe A)

For each rank, substitute the data with their average

	Chip1	Chip2
Rank1	6 (probe C)	6 (probe C)
Rank2	9 (probe D)	9 (probe D)
Rank3	11 (probe A)	11 (probe B)
Rank4	13 (probe B)	13 (probe A)

Re-sort to the original probe order

	Chip1	Chip2
Probe A	11	13
Probe B	13	11
Probe C	6	6
Probe D	9	9

Through quantile normalization, all replicate chips get the same distribution. That is,

- The most expressed genes in each chip get the same value.
- The second most expressed genes in each chip get the same value.
- …
- The least expressed genes in each chip get the same value.

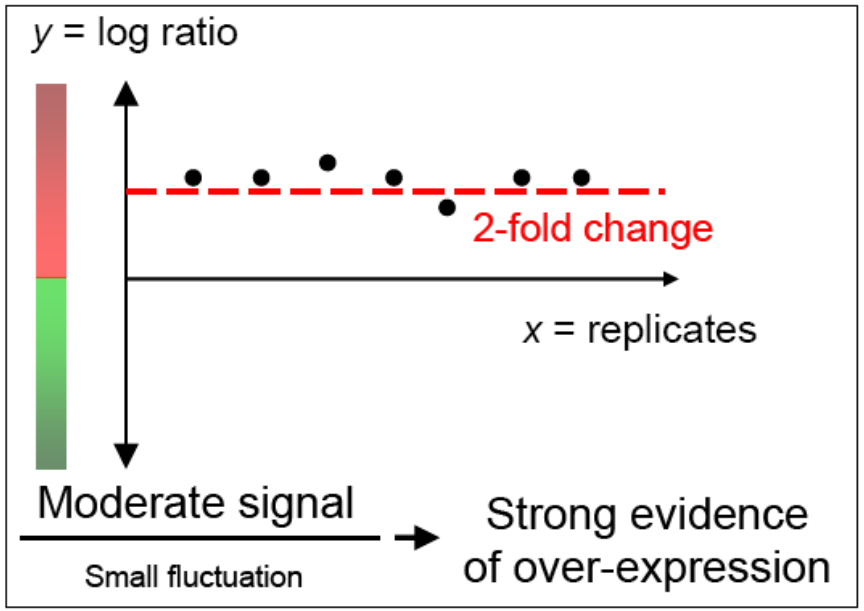
RMA = Quantile normalization + Probe set summarization by Tukey's median polish

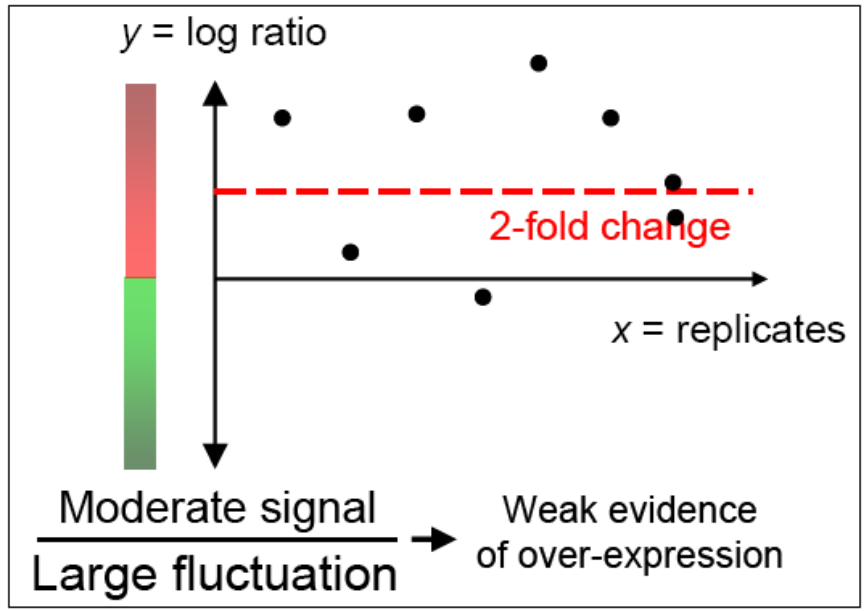
An easy reading: http://www.plexdb.org/modules/documentation/RMAexplained.pdf

数据的初步分析 — 差异基因的选择

- 一般来说，ratio>2或ratio<0.5认为是在两种样本表达有差异。

• Signal: Average of log ratios (Average fold change)
• Noise: Fluctuation (or inconsistency) across replicates
} Signal-to-noise ratio: *t*-statistic





Parametric & non-parametric tests

Three types of tests for differential expression

Parametric test

1. *t*-test
   - Unpaired *t*-test
   - Paired *t*-test

Non-parametric test
(≈ Distribution-free test)

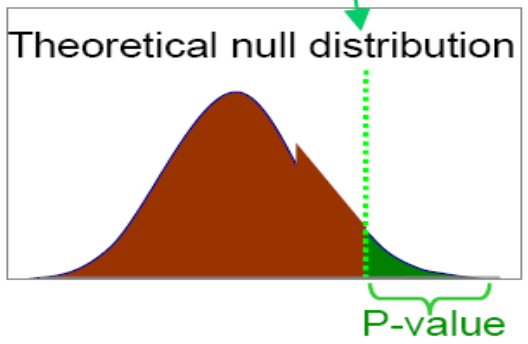
2. Classical non-parametric test
   - Wilcoxon rank-sum test (aka Mann-Whitney test)
   - Wilcoxon signed rank test

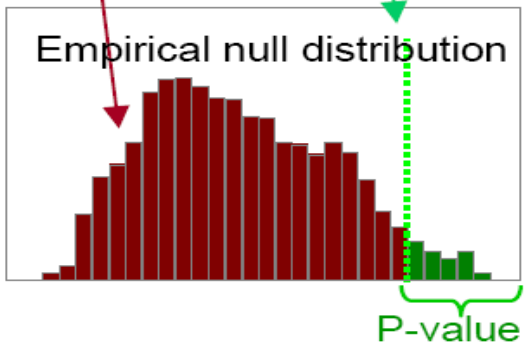
3. Permutation test
   - SAM (Significance Analysis of Microarrays)

Observed *t*-statistic

Generate empirical null distribution through permutation

Observed *t*-statistic





Terminology

- Parameter: Any characteristic of a population.
- Parametric test: We assume that the population follows normal distribution.
- P-value: Probability of getting an event as extreme as, or more extreme than the observed event, given that null hypothesis is true. Small P-value indicates more evidence of differential expression

Do statistical test for each probe

Do statistical test with normalized and log transformed data table, for each probe

Replicate control chips | Replicate case chips

	Rep 1	Rep 2	…	Rep 1	Rep 2	…
Probe 1	5	5.2	…	6.1	6.3	…
Probe 2	4.1	3.9	…	4.1	4.0	…
Probe 3	6.5	6.3	…	4.5	4.3	…
…	…	…	…	…	…	…
Probe N	2.2	2.4	…	2.5	2.2	…

(1) Do test for probe 1 to get its P-value, statistic, and average fold change

	Rep 1	Rep 2	…	Rep 1	Rep 2	…
Probe 1	5	5.2	…	6.1	6.3	…
Probe 2	4.1	3.9	…	4.1	4.0	…
Probe 3	6.5	6.3	…	4.5	4.3	…
…	…	…	…	…	…	…
Probe N	2.2	2.4	…	2.5	2.2	…

(2) Do test for probe 2 to get its P-value, statistic, and average fold change

… (N) Repeat until the last probe N

多重假设检验

- 在进行差异基因挑选时，整个差异基因筛选过程需要做成千上万次假设检验，导致假阳性率的累积增大。对于这种多重假设检验带来的放大的假阳性率，需要进行纠正。

- Bonferroni
  - p/n

- False Discovery Rate (FDR): Benjamini
  - q=p*n/rank

FDR、Q VALUE、ADJUSTED P VALUE

- P value：衡量假阳性率的指标（False positive rate）
-
- q value：衡量错误发现率的指标（False discovery rate，FDR）使用Q value的这个参数预估FDR，通常又被称为adjusted p value。

- Q value = FDR = adjusted p value

MicroArray/Sequencing Quality Control (MAQC/SEQC)

About MAQC/SEQC, links to journal articles featuring MAQC, information about RNA samples, and contact information for questions or suggestions

Subscribe to Email Updates

Share Tweet Linkedin Email Print

MicroArray/Sequencing Quality Control (MAQC/SEQC)

MAQC Publications

MAQC-I
MAQC-II
MAQC-III (also known as SEQC)
MAQC-IV (also known as SEQC2)
Contact Information

Journal Issues Featuring MAQC

Nature Biotechnology – 2014 Special Issue

Nature Biotechnology – August 2010 Issue

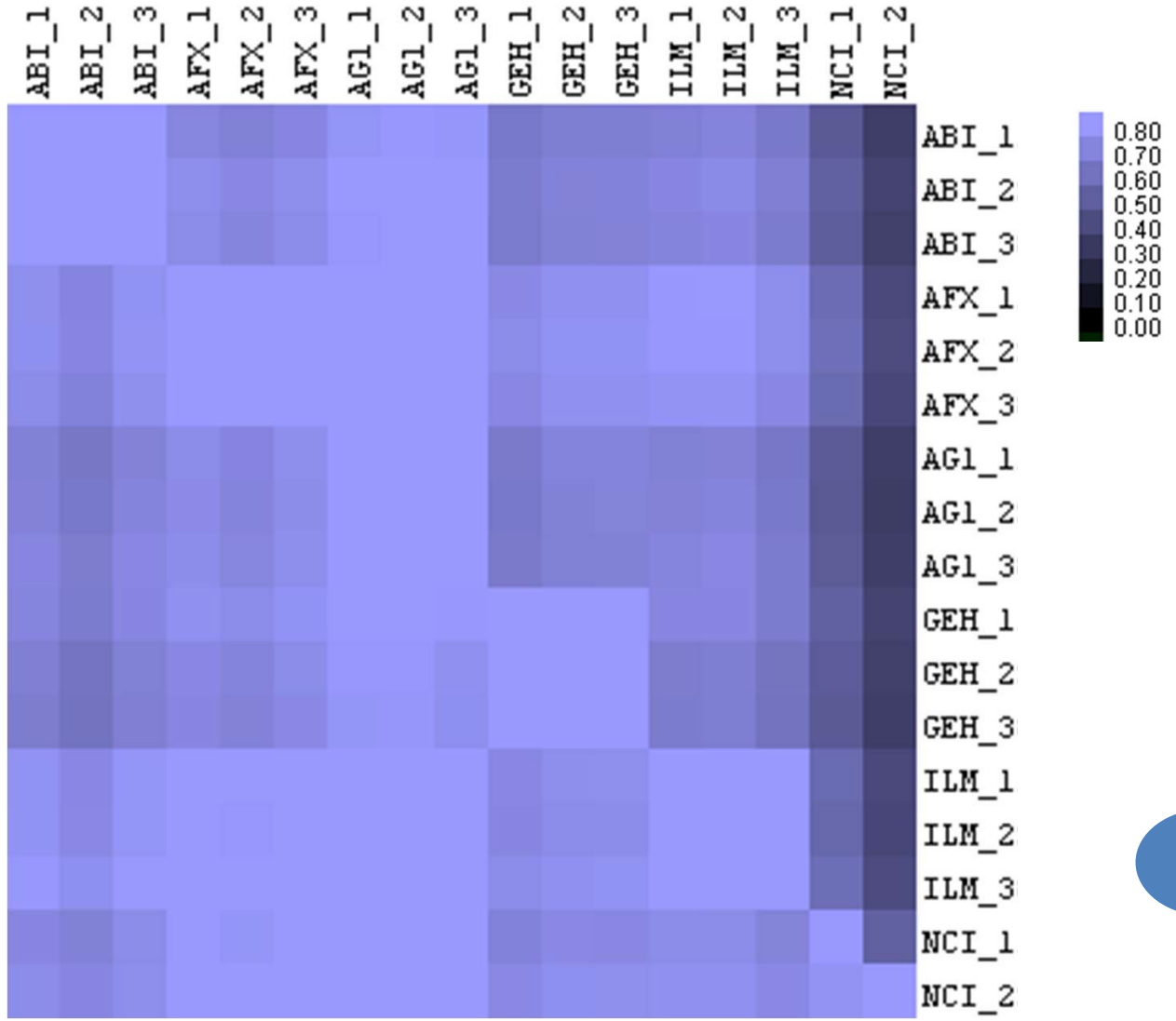
Pharmacogenomics Journal – August 2010 Issue

MAQC I、II

- provide quality control (QC) tools to the microarray community to avoid procedural failures
- develop guidelines for microarray data analysis by providing the public with large reference datasets along with readily accessible reference RNA samples
- establish QC metrics and thresholds for objectively assessing the performance achievable by various microarray platforms
- evaluate the advantages and disadvantages of various data analysis methods
- reach consensus on the “best practices” for development and validation of predictive models based on microarray gene expression and genotyping data for personalized medicine.

MAQC INTER AND INTRA PLATFORM REPRODUCIBILITY

- For the 5 commercial microarray platforms, the minimum overlap is at least 60%
- many achieving >80% between platforms
- >90% within a platform



The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements

MAQC Consortium*

Over the last decade, the introduction of microarray technology has had a profound impact on gene expression research. The publication of studies with dissimilar or altogether contradictory results, obtained using different microarray platforms to analyze identical RNA samples, has raised concerns about the reliability of this technology. The MicroArray Quality Control (MAQC) project was initiated to address these concerns, as well as other performance and data analysis issues. Expression data on four titration pools from two distinct reference RNA samples were generated at multiple test sites using a variety of microarray-based and alternative technology platforms. Here we describe the experimental design and probe mapping efforts behind the MAQC project. We show intraplatform consistency across test sites as well as a high level of interplatform concordance in terms of genes identified as differentially expressed. This study provides a resource that represents an important first step toward establishing a framework for the use of microarrays in clinical and regulatory settings.

Popular Differential Expression Algorithms/Tools

- Marray (http://matrise.uio.no/marray/marray.html)
  - **Wang J. et al., MArray: analysing single, replicated or reversed microarray experiments.** *Bioinformatics*, 2002 Aug;18(8):1139-40.
- SAM (http://www-stat.stanford.edu/~tibs/SAM/)
  - Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001**, 98:**5116-5121.
- LIMMA (http://bioinf.wehi.edu.au/limma/)
  - Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, No. 1, Article 3.
- EDGE (http://www.genomine.org/edge/)
- MeV (http://www.tm4.org/mev/)

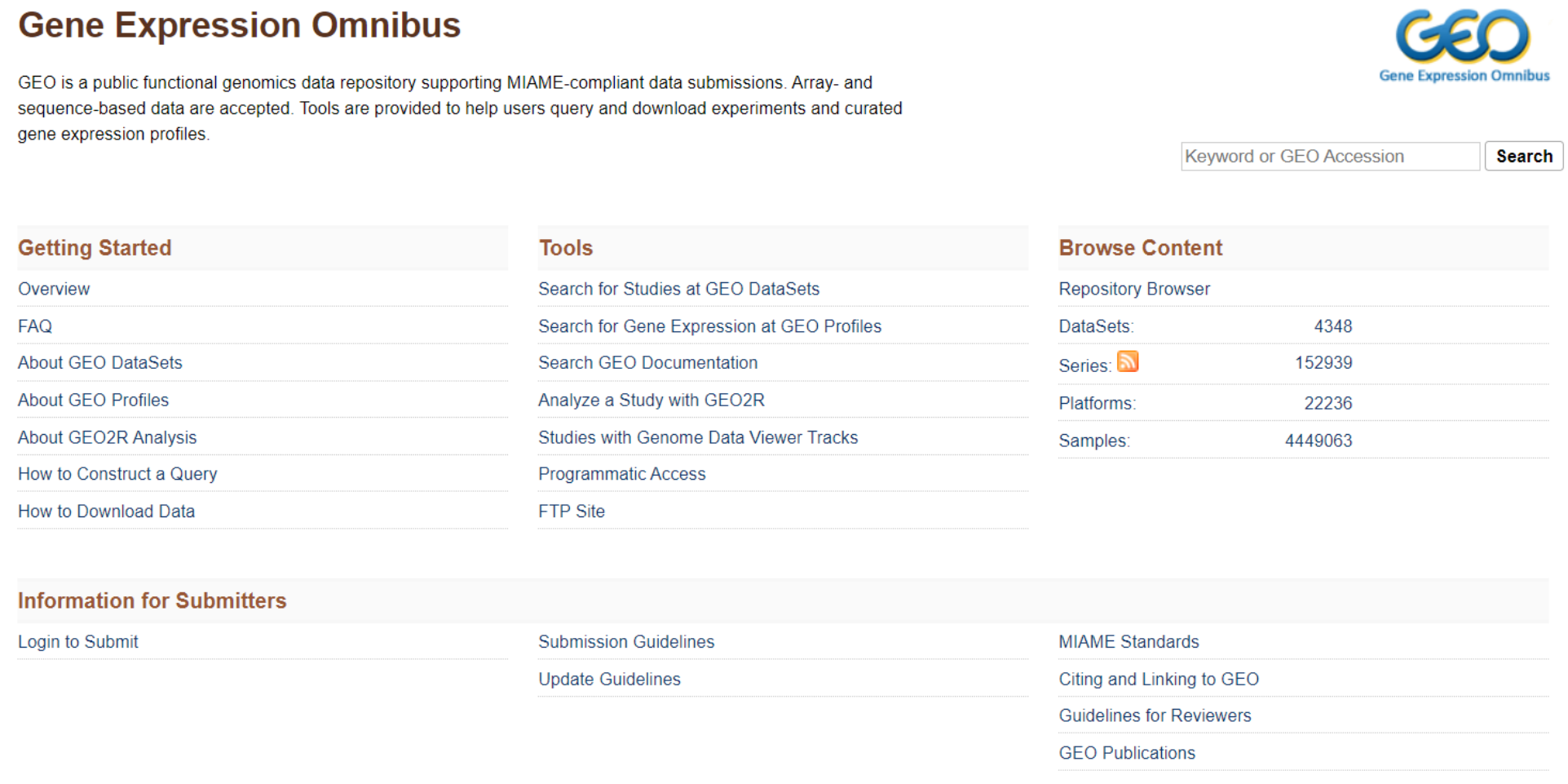
Related databases and softwares

online database

- GEO
- ArrayExpress
- TCGA
- Oncomine

Gene Expression Omnibus

- http://www.ncbi.nlm.nih.gov/projects/geo/



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.

Gene Expression Omnibus

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
Studies with Genome Data Viewer Tracks
Programmatic Access
FTP Site

Browse Content

Repository Browser

DataSets:	4348
Series:	152939
Platforms:	22236
Samples:	4449063

Information for Submitters

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

Series | Platforms | Samples | Organisms | History

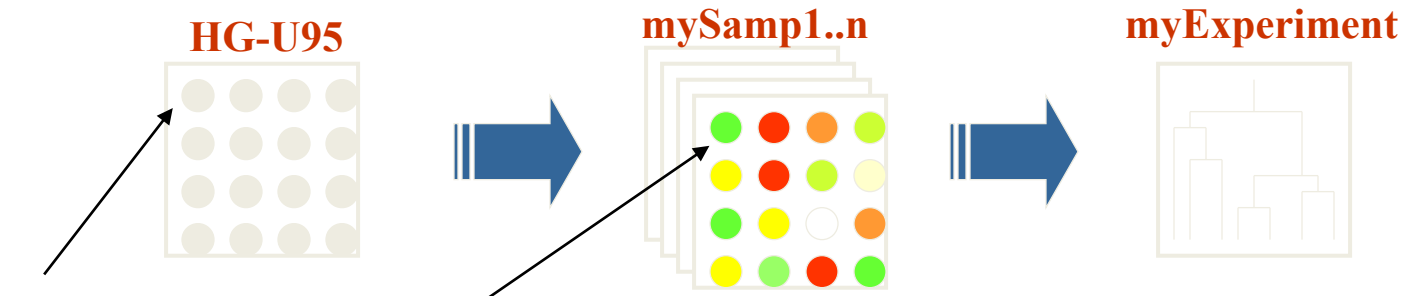
Series type	Count
Expression profiling by array	63,185
Expression profiling by genome tiling array	751
Expression profiling by high throughput sequencing	49,492
Expression profiling by SAGE	239
Expression profiling by MPSS	20
Expression profiling by RT-PCR	809
Expression profiling by SNP array	14
Genome variation profiling by array	835
Genome variation profiling by genome tiling array	1,512
Genome variation profiling by high throughput sequencing	219
Genome variation profiling by SNP array	1,410
Genome binding/occupancy profiling by array	230
Genome binding/occupancy profiling by genome tiling array	2,365
Genome binding/occupancy profiling by high throughput sequencing	22,762
Genome binding/occupancy profiling by SNP array	18
Methylation profiling by array	1,191
Methylation profiling by genome tiling array	1,889

Organism	Series	Platforms	Samples
Homo sapiens	66,540	5,842	2,290,503
Mus musculus	47,364	2,555	1,268,735
Rattus norvegicus	4,539	632	109,600
Drosophila melanogaster	4,124	386	98,433
Arabidopsis thaliana	4,737	403	72,501
Saccharomyces cerevisiae	3,170	621	70,517
Macaca mulatta	575	73	24,644
Sus scrofa	977	153	23,109
Caenorhabditis elegans	1,945	213	20,676
Bos taurus	1,045	198	19,036

Total holdings

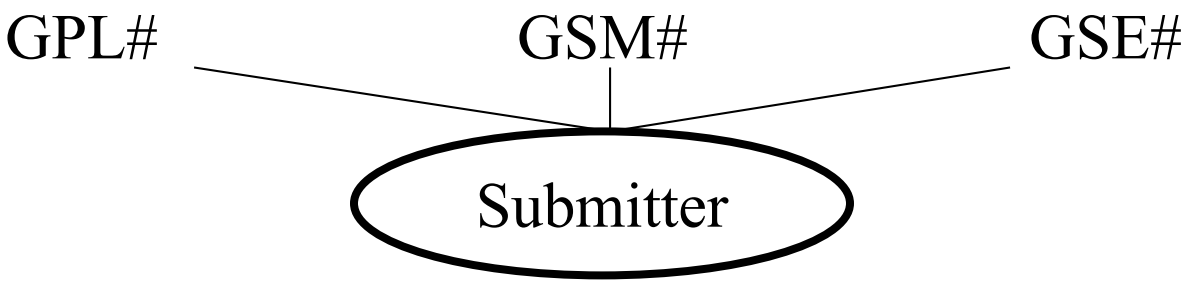
	Public	Unreleased	Total
Series	152,939	17,086	170,025
Platforms	22,236	167	22,403
Samples	4,449,067	535,815	4,984,882

GEO Architecture



ID	GB_ACC
1	AI944549
2	AI945145

ID_REF	VALUE
1	345.45
2	28994.50



GEO profile query

NCBI Resources How To

GEO Profiles GEO Profiles

Limits Advanced

Display Settings: Summary, 20 per page, Sorted by Default order

Results: 1 to 20 of 24833 << First < Prev

Gene name and annotations

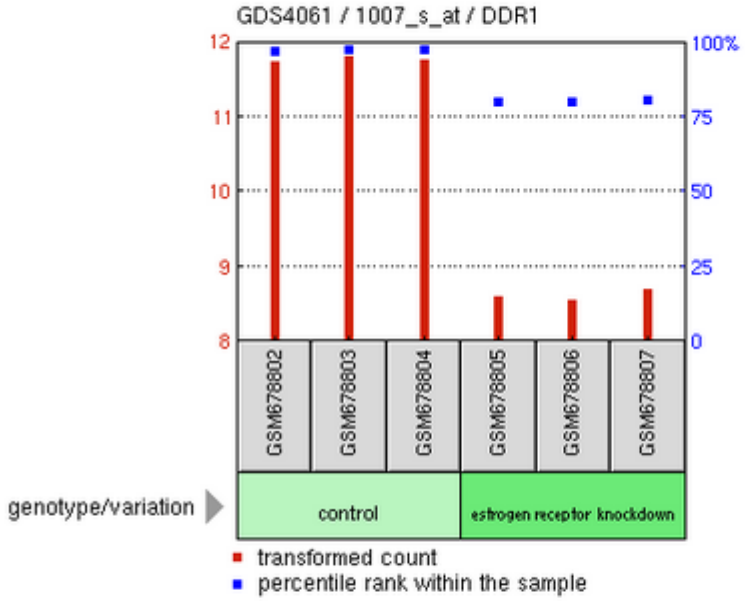
Cross-link information

1. DDR1 - Estrogen receptor alpha-silenced MCF7 breast cancer cells
   Annotation: DDR1, discoidin domain receptor tyrosine kinase 1
   Organism: Homo sapiens
   Reporter: GPL570, 1007_s_at (ID_REF), GDS4061, 780 (Gene ID), U48705
   DataSet type: Expression profiling by array, transformed count, 6 samples
   ID: 77617601
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologe

2. RFC2 - Estrogen receptor alpha-silenced MCF7 breast cancer cells
   Annotation: RFC2, replication factor C (activator 1) 2, 40kDa
   Organism: Homo sapiens
   Reporter: GPL570, 1053_at (ID_REF), GDS4061, 5982 (Gene ID), M87338
   DataSet type: Expression profiling by array, transformed count, 6 samples
   ID: 77617602
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologe

3. UBA7 - Estrogen receptor alpha-silenced MCF7 breast cancer cells
   Annotation: UBA7, ubiquitin-like modifier activating enzyme 7
   Organism: Homo sapiens
   Reporter: GPL570, 1294_at (ID_REF), GDS4061, 7318 (Gene ID), L13852
   DataSet type: Expression profiling by array, transformed count, 6 samples

Profile GDS4061 / 1007_s_at / DDR1
Title Estrogen receptor alpha-silenced MCF7 breast cancer cells
Organism Homo sapiens

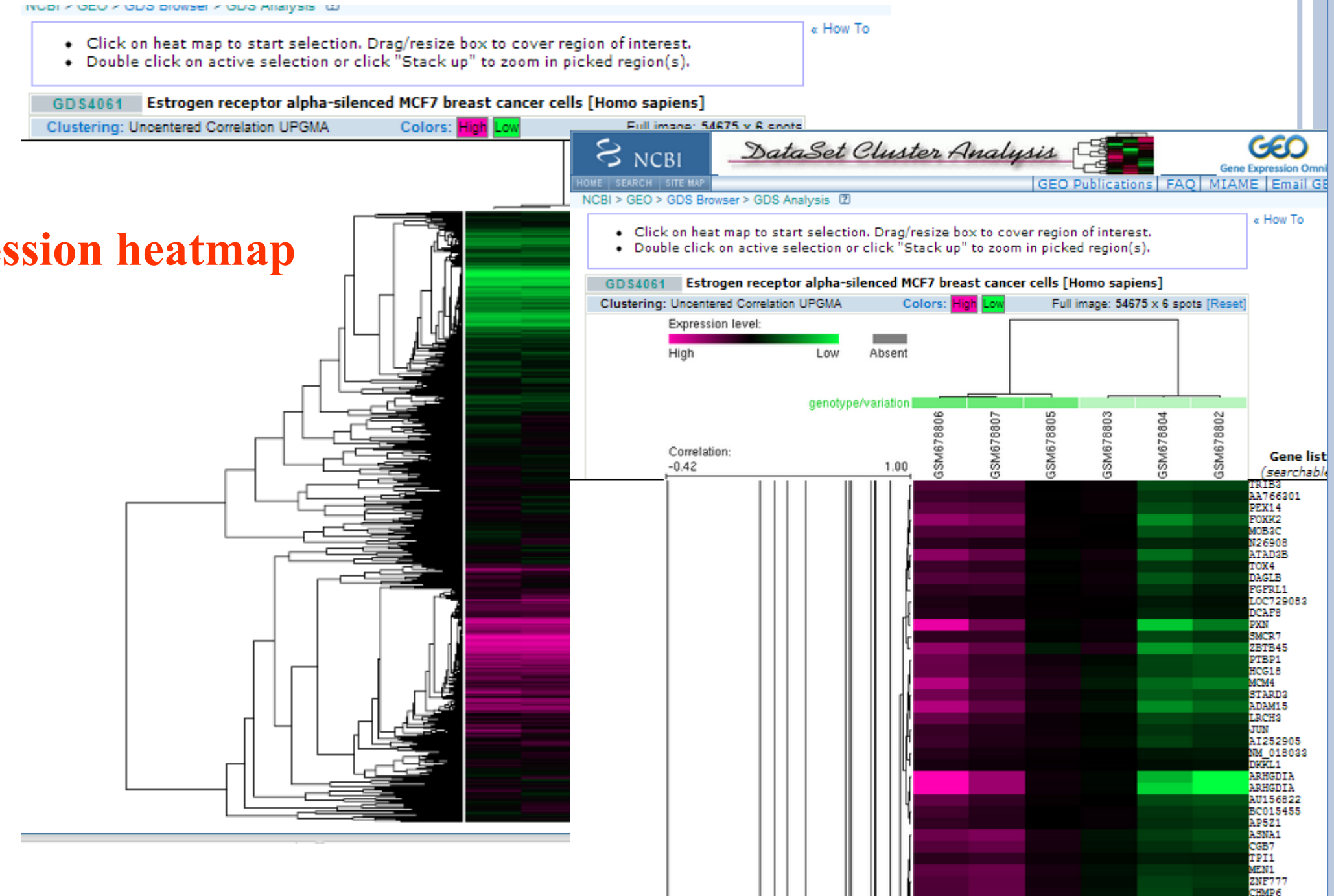


Graph caption help

Sample	Title	Value	Rank
GSM678802	MCF7, biological rep1	11.7581	97
GSM678803	MCF7, biological rep2	11.8284	98
GSM678804	MCF7, biological rep3	11.7697	98
GSM678805	MCF7 silenced Estrogen receptor, biological rep1	8.61565	80
GSM678806	MCF7 silenced Estrogen receptor, biological rep2	8.57259	80
GSM678807	MCF7 silenced Estrogen receptor, biological rep3	8.71401	81

(GDS4095[ACCN]) AND GDS[filter] (1)

1) Expression heatmap



2)Downloadable data GDS4095_full.soft.gz

Soft format

```
#GSM678803 = Value for GSM678803: MCF7, biological rep2; src: MCF7 expressing Estrogen receptor alpha
#GSM678804 = Value for GSM678804: MCF7, biological rep3; src: MCF7 expressing Estrogen receptor alpha
#GSM678805 = Value for GSM678805: MCF7 silenced Estrogen receptor, biological rep1; src: MCF7 silenced Estrogen recepto
#GSM678806 = Value for GSM678806: MCF7 silenced Estrogen receptor, biological rep2; src: MCF7 silenced Estrogen recepto
#GSM678807 = Value for GSM678807: MCF7 silenced Estrogen receptor, biological rep3; src: MCF7 silenced Estrogen recepto
#Gene title = Entrez Gene name
#Gene symbol = Entrez Gene symbol
#Gene ID = Entrez Gene identifier
#UniGene title = Entrez UniGene name
#UniGene symbol = Entrez UniGene symbol
#UniGene ID = Entrez UniGene identifier
#Nucleotide Title = Entrez Nucleotide title
#GI = GenBank identifier
#GenBank Accession = GenBank accession
#Platform_CLONEID = CLONE_ID from Platform data table
#Platform_ORF = ORF from Platform data table
#Platform_SPOTID = SPOT_ID from Platform data table
#Chromosome location = Entrez gene chromosome and location
#Chromosome annotation = Entrez gene chromosome annotation
#GO:Function = Gene Ontology Function term
#GO:Process = Gene Ontology Process term
#GO:Component = Gene Ontology Component term
#GO:Function ID = Gene Ontology Function identifier
#GO:Process ID = Gene Ontology Process identifier
#GO:Component ID = Gene Ontology Component identifier
!dataset_table_begin
ID_REF  IDENTIFIER  GSM678802 GSM678803 GSM678804 GSM678805 GSM678806 GSM678807 Gene title  Gene symbol Gene ID UniGene
1007_s_at DDR1  11.7581 11.8284 11.7[illegible]61565 8.57259 8.71401 discoidin domain receptor tyrosine kinase 1 DDR1  780
1053_at RFC2  10.4523 10.5417 10.4462 [illegible]3092 8.95197 9.02724 replication factor C (activator 1) 2, 40kDa RFC2  5982
117_at  HSPA6 5.6437  5.43553 5.61835 5.59269 5.61428 5.69114 heat shock 70kDa protein 6 (HSP70B')  HSPA6 3310
121_at  PAX8  8.42958 8.36975 8.42138 8.3507  8.54483 8.42141 paired box 8  PAX8  7849  H.sapiens Pax8 mRNA 3842
1255_g_at GUCA1A  4.71405 4.67491 4.53634 4.63943 4.76701 4.89855 guanylate cyclase activator 1A (retina) GUCA1A  2978
1294_at UBA7  5.77325 5.66301 5.91879 6.16913 6.28257 6.27076 ubiquitin-like modifier activating enzyme 7 UBA7  7318
```

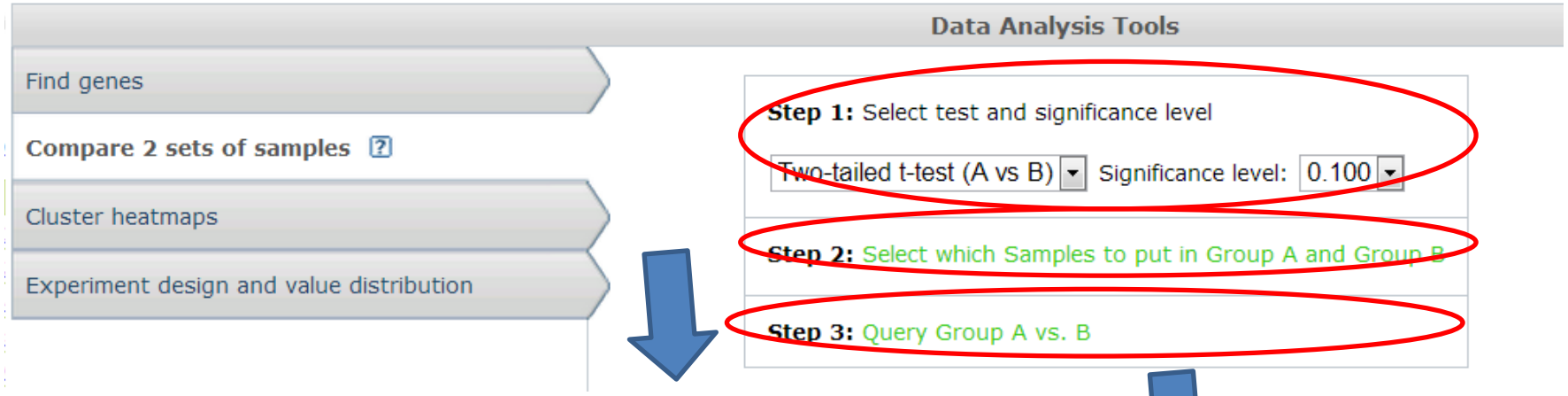
Head informaiont

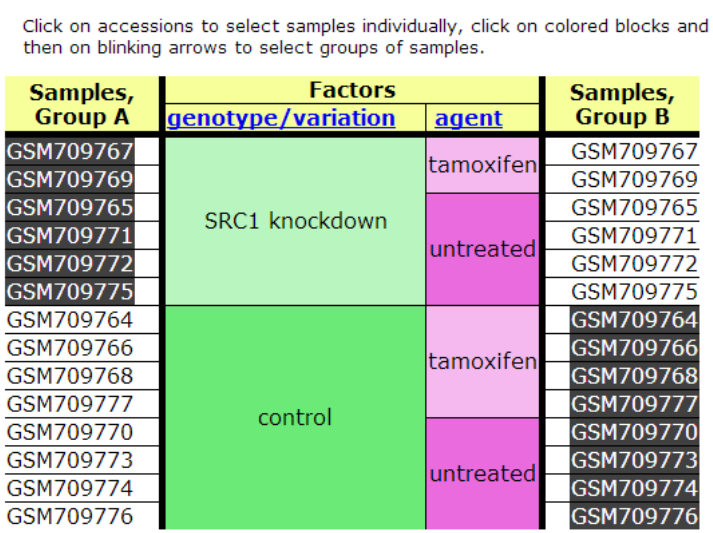
Probe ID

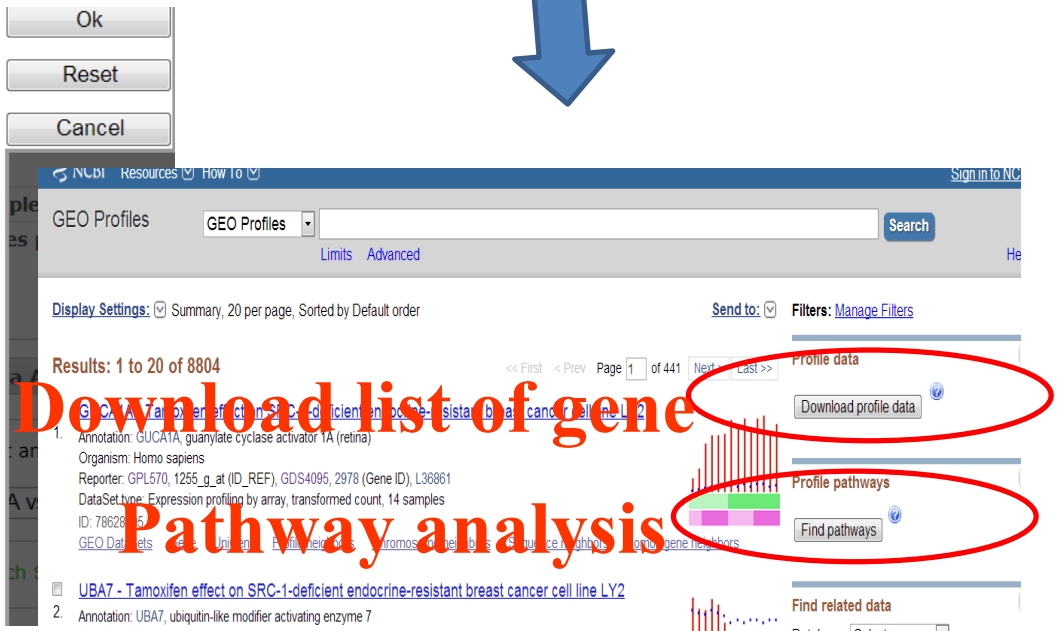
annotation

expression

3)Online analysis tool



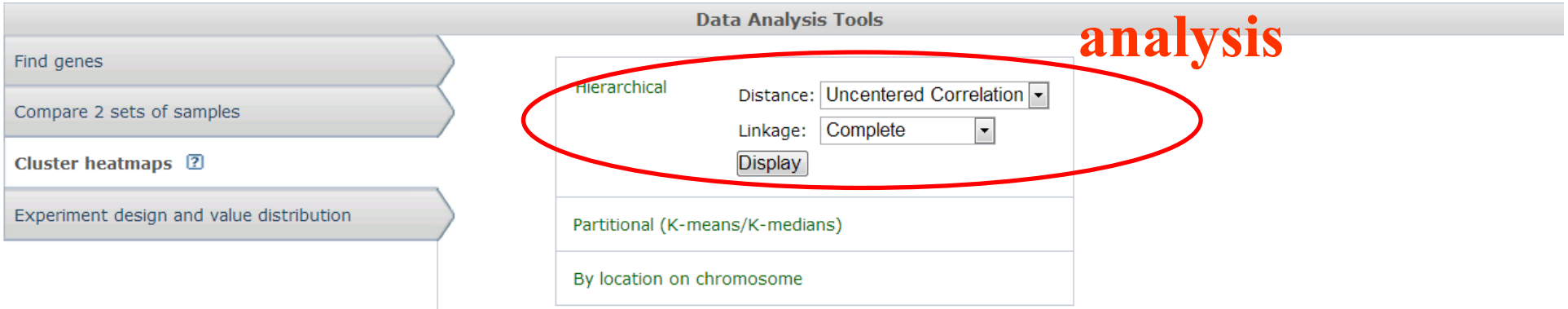


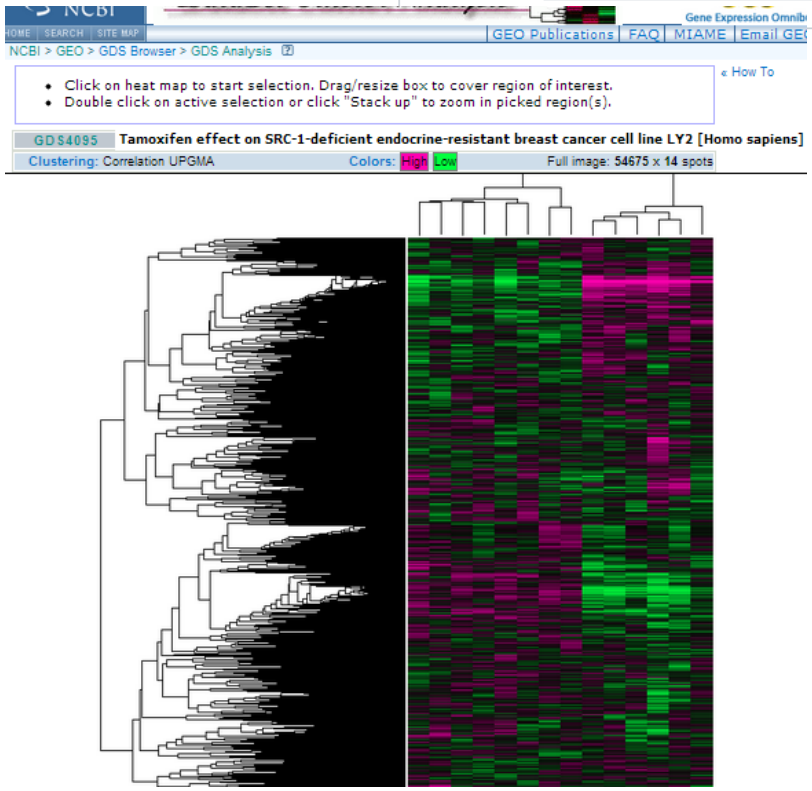


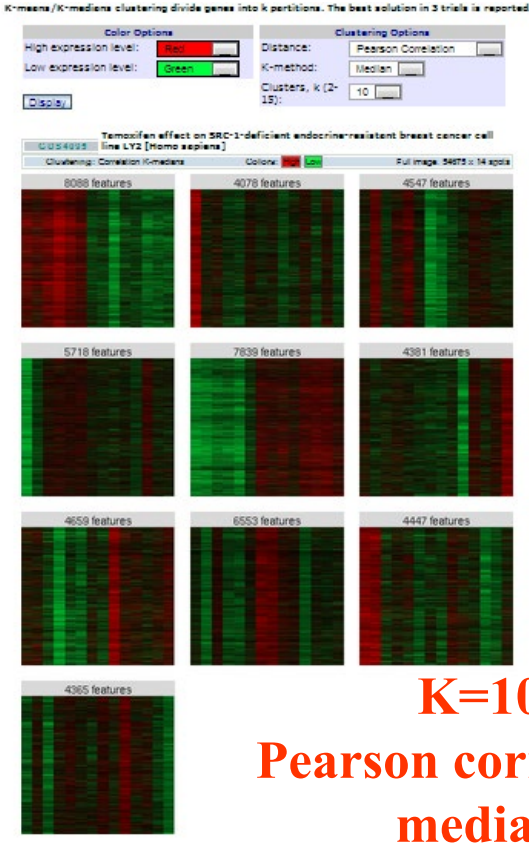
Download list of gene

Pathway analysis

Clustering analysis



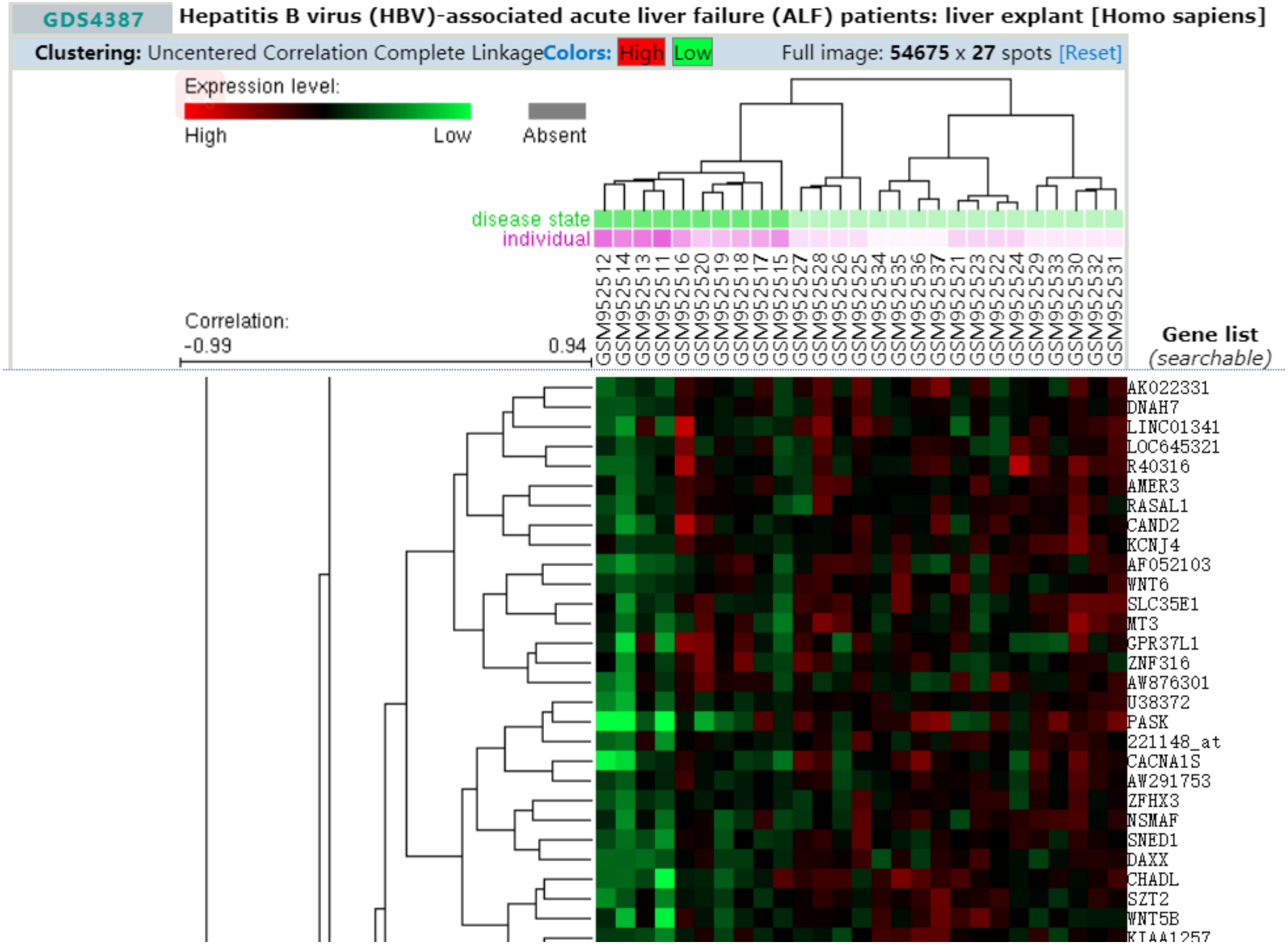




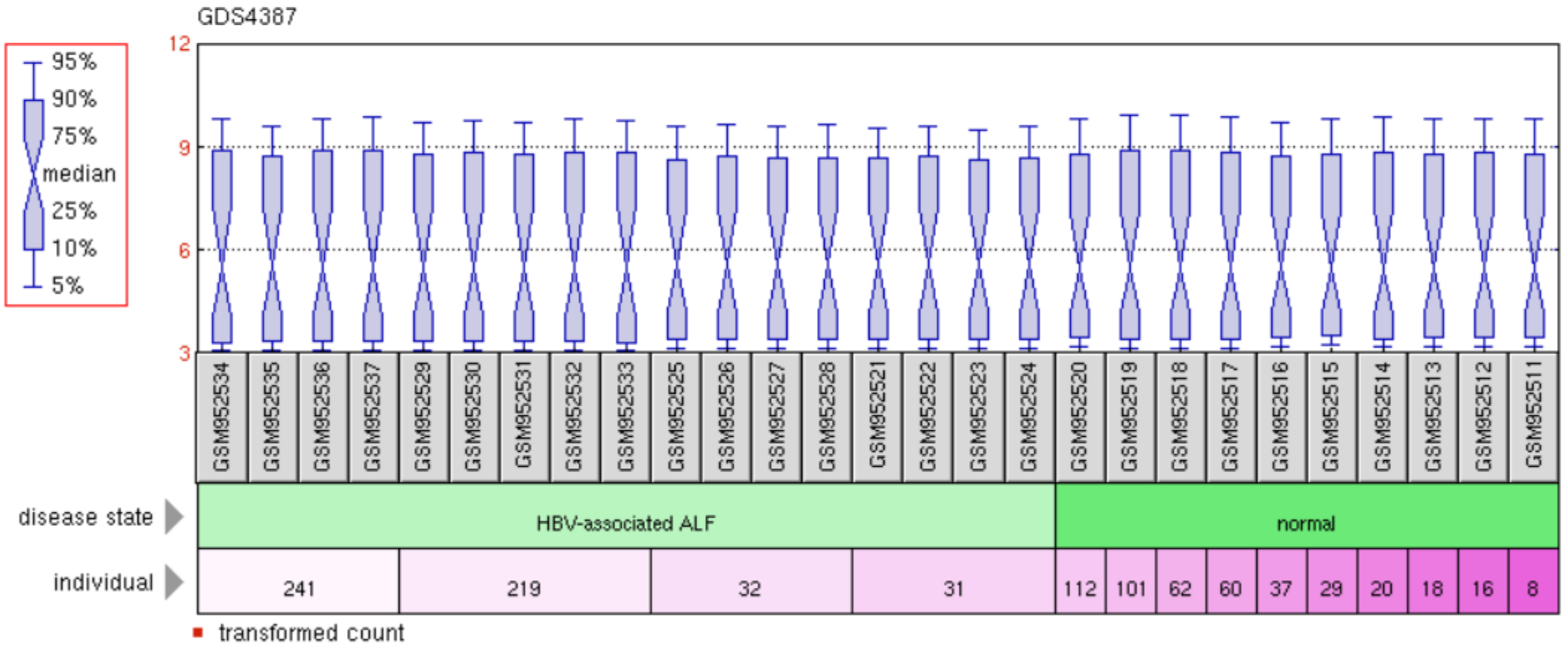
K=10

Pearson correlation

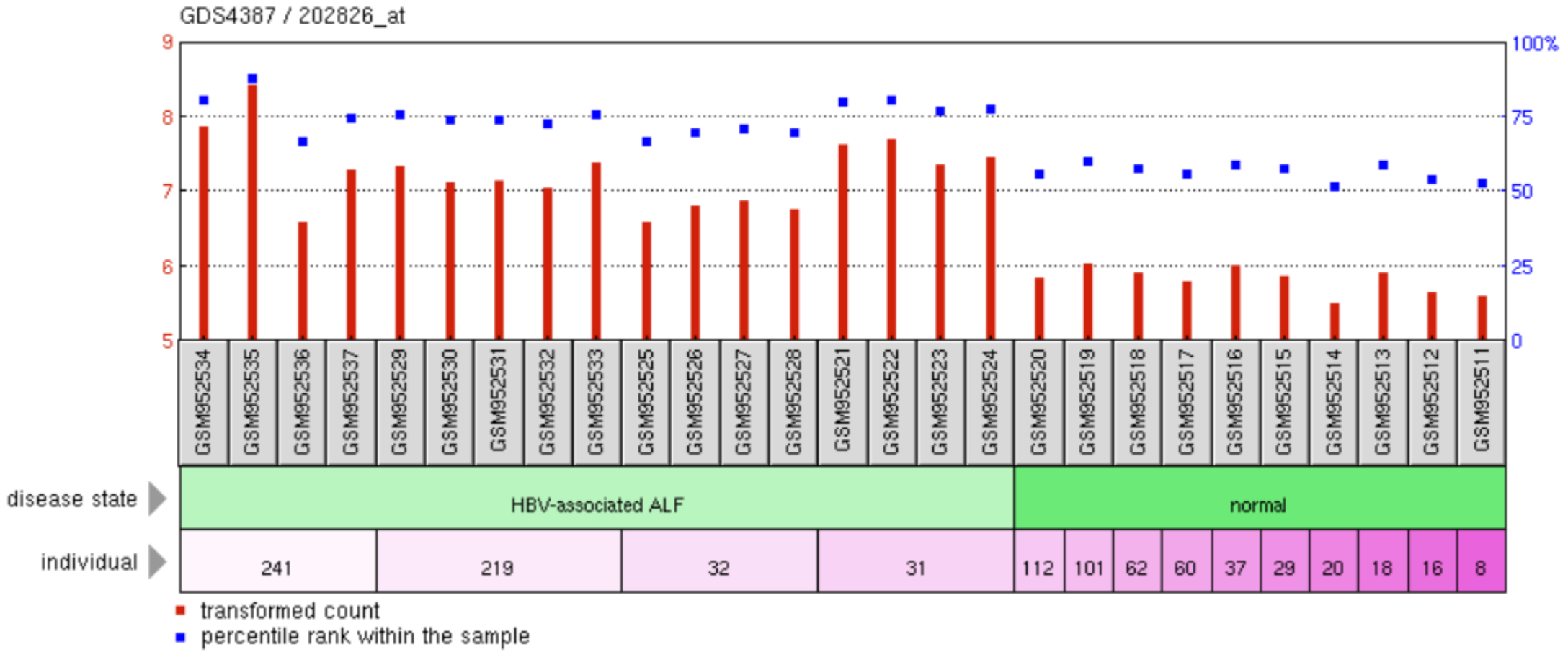
median


GDS4387
Hepatitis B virus (HBV)-associated acute liver failure (ALF) patients: liver explant [Homo sapiens]
Clustering: Uncentered Correlation Complete Linkage
Colors: High Low
Full image: 54675 x 27 spots [Reset]
Expression level:
High
Low
Absent
disease state
individual
GSM952512
GSM952514
GSM952513
GSM952511
GSM952516
GSM952520
GSM952519
GSM952518
GSM952517
GSM952515
GSM952527
GSM952528
GSM952526
GSM952525
GSM952534
GSM952535
GSM952536
GSM952537
GSM952521
GSM952523
GSM952522
GSM952524
GSM952529
GSM952533
GSM952530
GSM952532
GSM952531
Correlation:
-0.99
0.94
Gene list
(searchable)
AK022331
DNAH7
LINC01341
LOC645321
R40316
AMER3
RASAL1
CAND2
KCNJ4
AF052103
WNT6
SLC35E1
MT3
GPR37L1
ZNF316
AW876301
U38372
PASK
221148_at
CACNA1S
AW291753
ZFHX3
NSMAF
SNED1
DAXX
CHADL
SZT2
WNT5B
KIAA1257

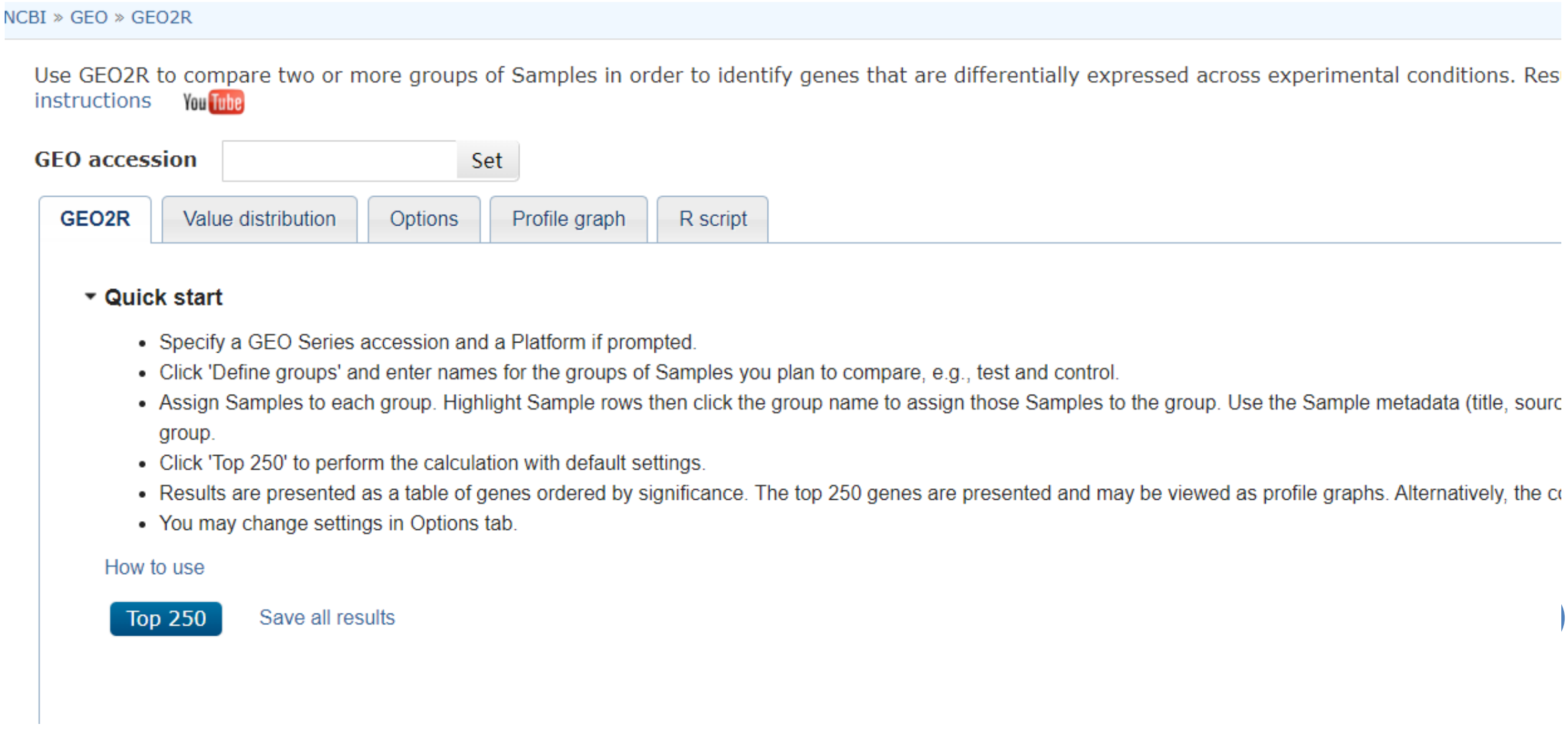
Profile GDS4387
Title Hepatitis B virus (HBV)-associated acute liver failure (ALF) patients: liver explant
Organism Homo sapiens



Profile GDS4387 / 202826_at
Title Hepatitis B virus (HBV)-associated acute liver failure (ALF) patients: liver explant
Organism Homo sapiens

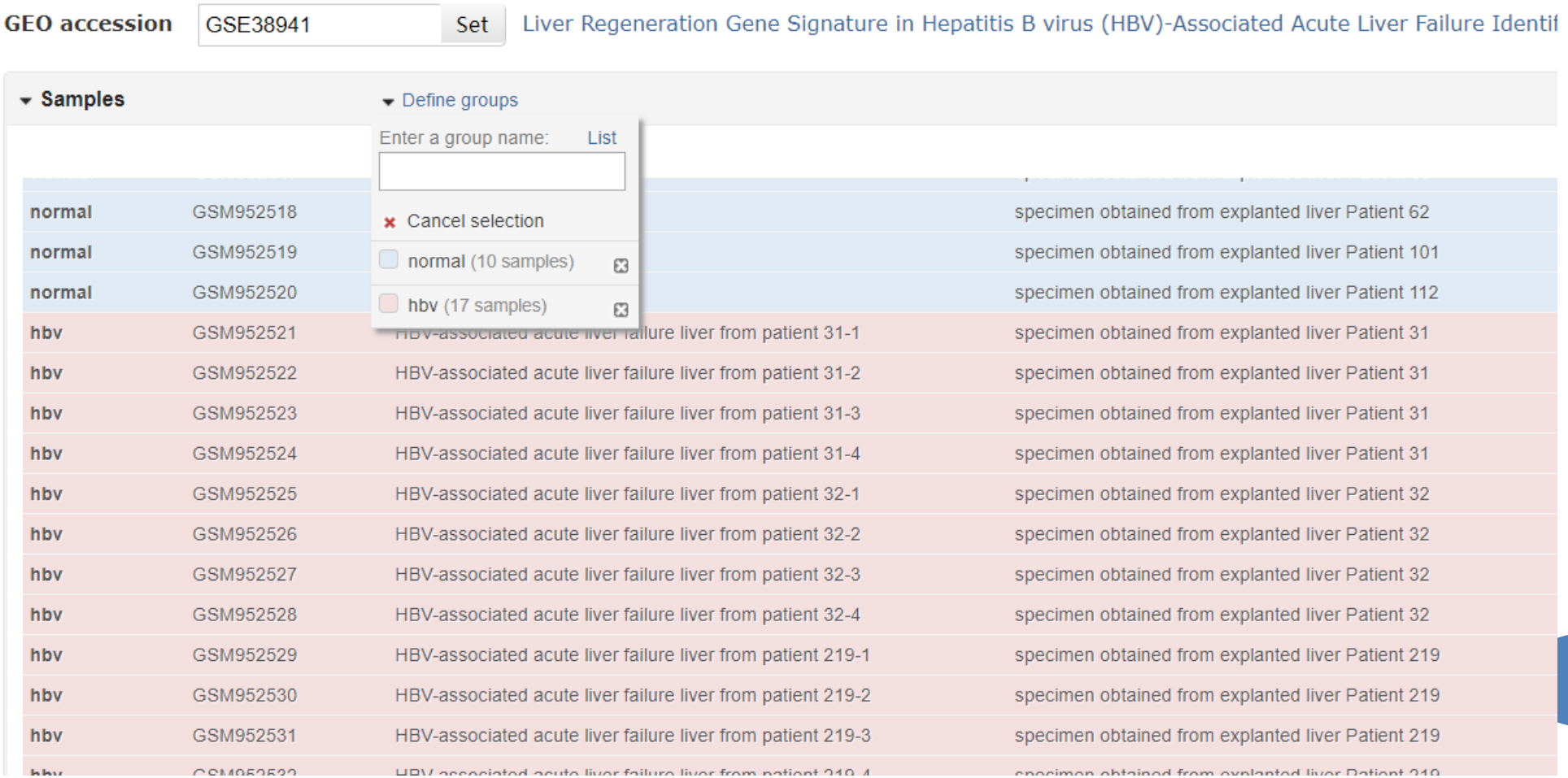


GEO2R


NCBI » GEO » GEO2R
Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Res
instructions
GEO accession
Set
GEO2R
Value distribution
Options
Profile graph
R script
Quick start
Specify a GEO Series accession and a Platform if prompted.
Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, sourc
group.
Click 'Top 250' to perform the calculation with default settings.
Results are presented as a table of genes ordered by significance. The top 250 genes are presented and may be viewed as profile graphs. Alternatively, the c
You may change settings in Options tab.
How to use
Top 250
Save all results

GEO2R

- Define group



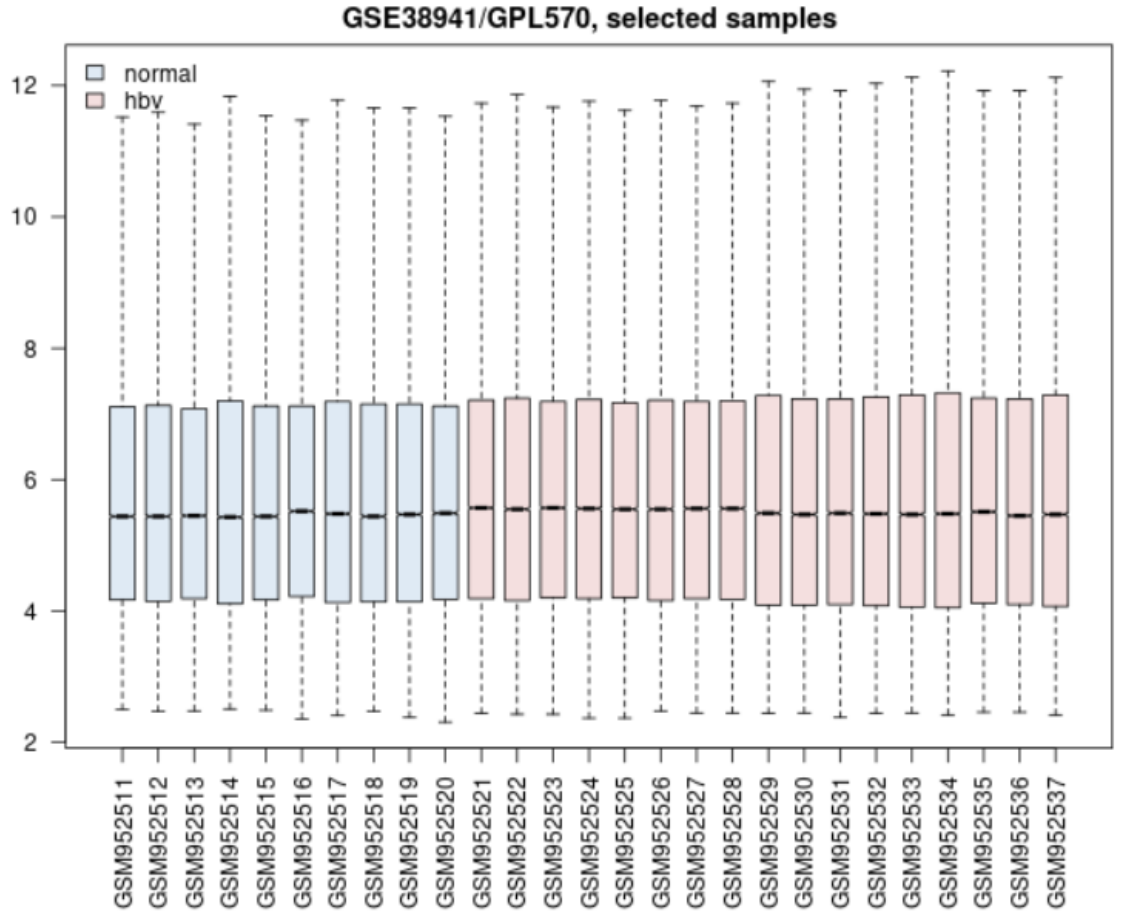
GEO2R

- distribution

GEO2R | Value distribution | Options | Profile graph | R script

Calculate the distribution of value data for the Samples you have selected. Distributions may be viewed graphically as a box plot or exported as a number summary table. The plot is useful for determining if value data are median-centered across Samples, and thus suitable for cross-comparison. More...

View Export



GEO2R | Value distribution | Options | Profile graph | R script

Apply adjustment to the P-values. More...

- (•) Benjamini & Hochberg (False discovery rate)
- ( ) Benjamini & Yekutieli
- ( ) Bonferroni
- ( ) Hochberg
- ( ) Holm
- ( ) Hommel
- ( ) None

Apply log transformation to the data. More...

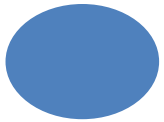
- (•) Auto-detect
- ( ) Yes
- ( ) No

Category of Platform annotation to display on results.

- ( ) Submitter supplied
- (•) NCBI generated

RESULTS

"ID"	"adj.P.Val"	"P.Value"	"t"	"B"	"logFC"	"Gene.symbol"	"Gene.title"
"235129_at"	"3.16e-19"	"5.77e-24"	"-3.21e+01"	"43.77512"	"-3.87"	"PPP1R1A"	"protein phosphatase 1 regulatory inhibitor subunit 1A"
"219093_at"	"1.43e-18"	"5.21e-23"	"-2.97e+01"	"41.825934"	"-3.16"	"PID1"	"phosphotyrosine interaction domain containing 1"
"238835_at"	"2.62e-18"	"1.44e-22"	"-2.86e+01"	"40.913046"	"-6.03"	"AVPR1A"	"arginine vasopressin receptor 1A"
"201425_at"	"2.94e-17"	"2.15e-21"	"-2.59e+01"	"38.441733"	"-1.85"	"ALDH2"	"aldehyde dehydrogenase 2 family (mitochondrial)"
"205152_at"	"6.23e-17"	"5.69e-21"	"-2.50e+01"	"37.540006"	"-4.55"	"SLC6A1"	"solute carrier family 6 member 1"
"210328_at"	"1.54e-16"	"1.69e-20"	"-2.41e+01"	"36.526851"	"-5.19"	"GNMT"	"glycine N-methyltransferase"
"1405_i_at"	"1.54e-16"	"1.97e-20"	"2.39e+01"	"36.381221"	"4.17"	"CCL5"	"C-C motif chemokine ligand 5"
"1556666_a_at"	"3.33e-16"	"5.35e-20"	"-2.31e+01"	"35.442491"	"-2.83"	"TTC6"	"tetratricopeptide repeat domain 6"
"213393_at"	"3.33e-16"	"5.48e-20"	"-2.30e+01"	"35.419137"	"-1.61"	"MFSD9"	"major facilitator superfamily domain containing 9"
"225747_at"	"4.67e-16"	"8.54e-20"	"-2.27e+01"	"34.999421"	"-1.89"	"COQ10A"	"coenzyme Q10A"
"242680_at"	"7.72e-16"	"1.55e-19"	"-2.22e+01"	"34.43374"	"-6.23"	"AVPR1A"	"arginine vasopressin receptor 1A"
"226878_at"	"8.73e-16"	"1.92e-19"	"2.20e+01"	"34.234014"	"3.15"	"HLA-DOA"	"major histocompatibility complex, class II, DO alpha"
"210915_x_at"	"9.40e-16"	"2.24e-19"	"2.19e+01"	"34.087287"	"3.10"	"TRBC1"	"T cell receptor beta constant 1"
"209541_at"	"9.98e-16"	"2.58e-19"	"-2.18e+01"	"33.951832"	"-3.87"	"IGF1"	"insulin like growth factor 1"
"213502_x_at"	"9.98e-16"	"2.74e-19"	"2.17e+01"	"33.894709"	"2.67"	"GUSBP11"	"glucuronidase, beta pseudogene 11"
"206214_at"	"1.04e-15"	"3.03e-19"	"2.16e+01"	"33.796334"	"3.90"	"PLA2G7"	"phospholipase A2 group VII"
"230163_at"	"1.04e-15"	"3.24e-19"	"-2.16e+01"	"33.734478"	"-1.90"	"GFRA1"	"GDNF family receptor alpha 1"
"234764_x_at"	"1.08e-15"	"3.55e-19"	"2.15e+01"	"33.648255"	"4.55"	"IGLJ3///IGLV1-44///CKAP2///IGLV@///IGLC1"	"immunoglobulin lambda j



Popular Differential Expression Algorithms/Tools

- Marray (http://matrise.uio.no/marray/marray.html)
  - **Wang J. et al., MArray: analysing single, replicated or reversed microarray experiments.** *Bioinformatics*, 2002 Aug;18(8):1139-40.
- SAM (http://www-stat.stanford.edu/~tibs/SAM/)
  - Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001**, 98:**5116-5121.
- LIMMA (http://bioinf.wehi.edu.au/limma/)
  - Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, No. 1, Article 3.
- EDGE (http://www.genomine.org/edge/)
- MeV (http://www.tm4.org/mev/)