

# Microarray databases

Mikhail Dozmorov  
Fall 2017

## GEO, Gene Expression Omnibus

<https://www.ncbi.nlm.nih.gov/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.

#### Getting Started

[Overview](#)

[FAQ](#)

[About GEO DataSets](#)

[About GEO Profiles](#)

[About GEO2R Analysis](#)

#### Tools

[Search for Studies at GEO DataSets](#)

[Search for Gene Expression at GEO Profiles](#)

[Search GEO Documentation](#)

[Analyze a Study with GEO2R](#)

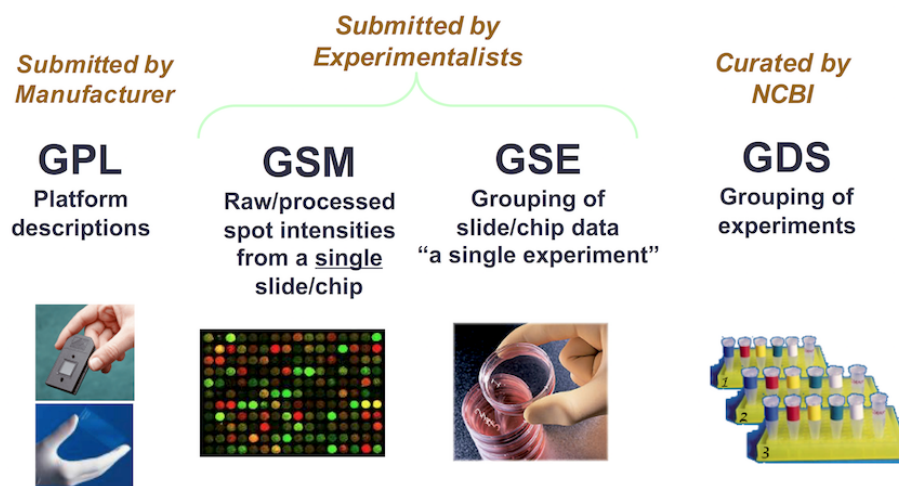
[GEO BLAST](#)

# GEO Structure

- **Platform** (GPL) - Annotation the technology and the features
- **Sample** (GSM) - Submitter-supplied sample data
- **Series** (GSE) - A collection of biologically- and statistically-comparable samples processed using the same platform
- **DataSets** (GDS) - Data collections assembled by GEO staff.

3/11

# GEO Structure



4/11

## Platform page

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570>

Platform details:

- Description
- Type
- Organism
- Manufacturer
- Related samples and series

5/11

## Sample page

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM543663>

Sample details:

- Description
- Type
- Organism
- Source
- Related platforms and series

6/11

# Sample data preview

Field description, truncated view, total rows

Data table header descriptions				
ID_REF	Affy ID			
VALUE	Raw data were normalized by using the global mean method. Probe-set signal values were natural log transformed and scaled by adjusting the mean intensity to a target signal value of log 500.			
ABS_CALL	the call in an absolute analysis that indicates if the transcript was present (P), absent (A), marginal (M) or no call (NC)			
DETECTION P-VALUE	the p-value that indicates the significance level of the detection call			
Data table				
ID_REF	VALUE	ABS_CALL	DETECTION	P-VALUE
212223_at	7.5283865762864	P	0.0048898452144488	
204339_s_at	5.62113156934585	A	0.70350996100991	
212356_at	7.46674066626714	A	0.0907246701401557	
212489_at	8.71669539548175	P	0.00164179881927007	
208850_s_at	7.90635473750185	P	0.00164179881927007	
208973_at	6.31816783076242	P	0.0373956404974136	
210122_at	2.54847298379146	A	0.922978176365498	
208174_x_at	7.8107483698245	P	0.00287512549712787	
210255_at	5.36794275612832	P	0.0202308511531004	
210388_at	4.09414454610406	A	0.95473940454311	
207430_s_at	4.63649721436744	A	0.142524741492000	
206872_at	6.16284326677117	A	0.394634090179955	
207624_s_at	5.94595979188476	P	0.01296767972666	
213315_x_at	7.87805534234773	P	0.0048898452144488	
219850_at	2.84489819756488	A	0.70350996100991	
213509_x_at	7.83250650883439	A	0.0770218236345017	
207818_s_at	1.54654132940278	A	0.857475258508	
200656_s_at	8.08463231173068	P	0.00164179881927007	
217622_at	5.70769317685453	A	0.2664419661303	
204916_at	6.51127285453828	A	0.123405137319250	

Total number of rows: 22283  
Table truncated, full table size 1055 Kbytes  
[View full table...](#)

7/11

## Series page

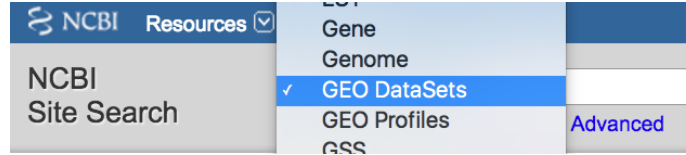
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21862>

Series details:

- Biological sample summary
- Design summary
- Publication information
- Related platforms and samples

8/11

# Search DataSet or Gene profile



- Datasets: **experiment-centric** view and analysis
- Gene profiles: **gene-centric** view

<https://www.ncbi.nlm.nih.gov/>

9/11

# GEO DataSet analysis

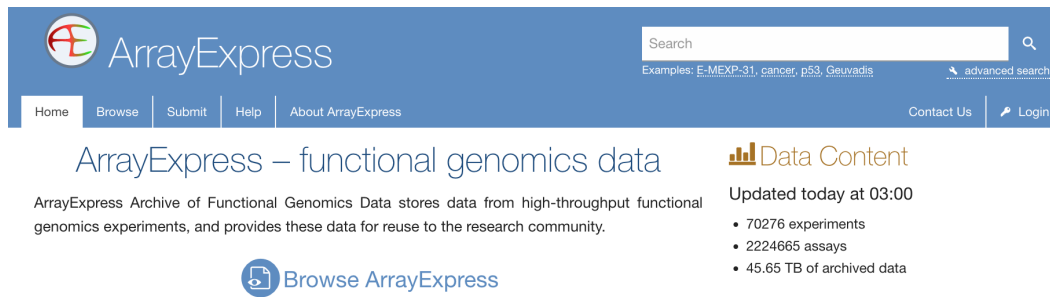
<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS987>

A screenshot of the GEO DataSet Record GDS987 page. The page has a header with tabs for 'Expression Profiles', 'Data Analysis Tools', and 'Sample Subsets'. The main content area is divided into two columns. The left column contains metadata: Title (Kidney transplant response to calcineurin inhibitor-free immunosuppression using sirolimus), Summary (Analysis of kidneys from adult renal transplant recipients...), Organism (Homo sapiens), Platform (GPL96: [HG-U133A] Affymetrix Human Genome U133A Array), Citation (Flechner SM, Kurian SM, Solez K, Cook DJ et al. De novo kidney transplantation without use of calcineurin inhibitors preserves renal structure and function at two years. Am J Transplant 2004 Nov;4(11):1776-85. PMID: 15476476), Reference Series (GSE1743), Sample count (41), Value type (count), and Series published (2004/09/23). The right column features a 'Cluster Analysis' section with a heatmap image and a 'Download' section with links for 'DataSet full SOFT file', 'DataSet SOFT file', 'Series family SOFT file', 'Series family MINIML file', and 'Annotation SOFT file'. Below the main content is a 'Data Analysis Tools' section with a 'Find genes' search box, a 'Compare 2 sets of samples' button, 'Cluster heatmaps' and 'Experiment design and value distribution' buttons, and a search form for finding genes up/down for a condition with an 'agent' checkbox.

10/11

# ArrayExpress

<https://www.ebi.ac.uk/arrayexpress/>



The screenshot shows the ArrayExpress website homepage. At the top left is the ArrayExpress logo, a stylized 'A' with a red and green circle. To its right is the text 'ArrayExpress'. Further right is a search bar with the placeholder text 'Search' and a magnifying glass icon. Below the search bar are examples: 'Examples: E-MEXP-31, cancer, p53, Geuvadis' and a link to 'advanced search'. Below the search bar is a navigation menu with links: 'Home', 'Browse', 'Submit', 'Help', 'About ArrayExpress', 'Contact Us', and 'Login'. The main content area features the heading 'ArrayExpress – functional genomics data' and a sub-heading 'Data Content'. Below the heading is a paragraph: 'ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.' Below this paragraph is a button with a magnifying glass icon and the text 'Browse ArrayExpress'. To the right of the main content is a 'Data Content' section with a bar chart icon, the text 'Updated today at 03:00', and a list of statistics: '70276 experiments', '2224665 assays', and '45.65 TB of archived data'.

... and more [https://en.wikipedia.org/wiki/Microarray\\_databases](https://en.wikipedia.org/wiki/Microarray_databases), <https://datamed.org/>