

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.



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:	4348
-----------	------

Series:	87644
---------	-------

Platforms:	17565
------------	-------

Samples:	2163652
----------	---------

Information for Submitters

[Login to Submit](#)[Submission Guidelines](#)[Update Guidelines](#)[MIAME Standards](#)[Citing and Linking to GEO](#)[Guidelines for Reviewers](#)[GEO Publications](#)

GEO data repository [ncbi.nlm.nih.gov/geo/]

Click on “Platforms”, search for "humanmethylation27"

No tool to extract meta data → use GEOmetadb instead

GEOmetadb: GEO Microarray Search Tool

What is GEOmetadb?

GEOmetadb is an attempt to make access to the metadata associated with the NCBI Gene Expression Omnibus (GEO) samples, platforms, and datasets much more feasible for common biologists and bioinformatics/statisticians. [read more](#)

GEOmetadb Paper ([Paper Link](#))

- Bioinformatics 2008 24(23):2798-2800; doi:10.1093/bioinformatics/btn520.

What's new about GEOmetadb?

- GEOmetadb has been upgraded to version 2.0. New feature includes:
- Database tables and search interfaces have been modified significantly
- Search performance has been improved
- Several user-friendly functions have been added, e.g. drill-down search, download search results, ... [read more](#)

GEOmetadb Web Interface:

- GEO Microarray Online Search Tool

Get Started: [Joint Search](#) | [GSE Search](#)

- Main features:
 - Search by individual data types
 - Search by GSE-GPL-GSM cross data types
 - GEO entities are linked by relationships between them
 - Multiple field query
 - Query within results
 - List creation
 - Flexible display options
 - Export or view details
- [Read More](#)

GEOmetadb Distributions:

- BioConductor Package/SQLite Database:

If you want to find GEO microarray data of interest directly within R by using power of SQL? Please try combination use of GEOmetadb and GEOquery.

- BioConductor package: GEOmetadb (in BioC 2.2 with R2.7)
- SQLite3 database: [GEOmetadb.sqlite.gz](#) (375.8 MB, August 06 2017 07:49:47.)

- Matlab GEOTools:

- Download: [MATLAB_GEOTools.zip](#) (Mac OS X, Intel)
- Document: [GEOmetadb_matlab.pdf](#) (pdf)

- FileMaker distribution:

- Download: [GEOmetaDB.fp7.zip](#) (32.5 MB, 08/01/2008)
- Readme: [SQLite2FileMakerPro.Readme.txt](#)

GEOmetadb [gbnci-abcc.ncifcrf.gov/geo/]

GPL accession for the 27K platform is GPL8490

GEO Sample (GSM) Search

Select Field: Search

contains

Search conditions: 1

Keep searchbox open? ☒ Yes ☐ No Help

[Show/Hide Search](#)
[Display Options](#)
[Download Results](#)

Selected Records: >> With Selection << 1 2 3 4 5 6 7 8 9 10 ... Results 1 - 50 of 2142262 (page 1 of 42846)

No.	<input type="checkbox"/>	ID	GSM Acc	Title	GSM Type	Channels	Organism Ch1	GSE Acc	GPL Acc
1	<input type="checkbox"/>	1	GSM1	Foreskin Fibroblasts	SAGE	1	Homo sapiens	GSE506	GPL4
2	<input type="checkbox"/>	2	GSM2	HCMV-infected foreskin fibroblasts	SAGE	1	Homo sapiens	GSE506	GPL4
3	<input type="checkbox"/>	3	GSM3	testis a	RNA	1	Drosophila melanogaster	GSE462	GPL5
4	<input type="checkbox"/>	4	GSM4	testis b	RNA	1	Drosophila melanogaster	GSE462	GPL5
5	<input type="checkbox"/>	5	GSM5	male a	RNA	1	Drosophila melanogaster	GSE462	GPL5
6	<input type="checkbox"/>	6	GSM6	male b	RNA	1	Drosophila melanogaster	GSE462	GPL5
7	<input type="checkbox"/>	7	GSM7	ovary a	RNA	1	Drosophila melanogaster	GSE462	GPL5
8	<input type="checkbox"/>	8	GSM8	ovary b	RNA	1	Drosophila melanogaster	GSE462	GPL5
9	<input type="checkbox"/>	9	GSM9	female a	RNA	1	Drosophila melanogaster	GSE462	GPL5
10	<input type="checkbox"/>	10	GSM10	female b	RNA	1	Drosophila melanogaster	GSE462	GPL5
11	<input type="checkbox"/>	11	GSM11	NHGRI_A-357	RNA	2	Homo sapiens	GSE1	GPL7
12	<input type="checkbox"/>	12	GSM12	NHGRI_CRL-1634	RNA	2	Homo sapiens	GSE1	GPL7
13	<input type="checkbox"/>	13	GSM13	NHGRI_HA-A	RNA	2	Homo sapiens	GSE1	GPL7
14	<input type="checkbox"/>	14	GSM14	NHGRI_M91-054	RNA	2	Homo sapiens	GSE1	GPL7
15	<input type="checkbox"/>	15	GSM15	NHGRI_M92-001	RNA	2	Homo sapiens	GSE1	GPL7
16	<input type="checkbox"/>	16	GSM16	NHGRI_M93-007	RNA	2	Homo sapiens	GSE1	GPL7

Exercise (10min)

- What is the number of samples and datasets available on GEO using the 450K chip ? (hint: official name is “Illumina HumanMethylation450 BeadChip”)
- Search for datasets that are using the 450K chip, have .idat files and information on age, gender, and BMI of the sample donors (how many samples are available?)

- GPL13534: 63404 samples, 931 datasets
- Look for dataset GSE72556
- Download meta data, make sure to include the columns “characteristics_ch1”
- Download the .tar containing all .idat files from GEO