Gene Expression (humans) hands-on (Prepared for BIFX550 in 2019; may be some menu options in NCBI might be outdated; if you happen to find any issues, please let me know; ravichandran@hood.edu)

Gene Expression (GE) varies from cell to cell.

Every cell has a complete set of genes

In a cell not all genes are expressed at a time.

GE varies during developmental stages, disease (or GE can cause disease), environmental state and final cells location

Gene expression (gene two copies; only one copy if active (methyl transferase is used for silencing the other)

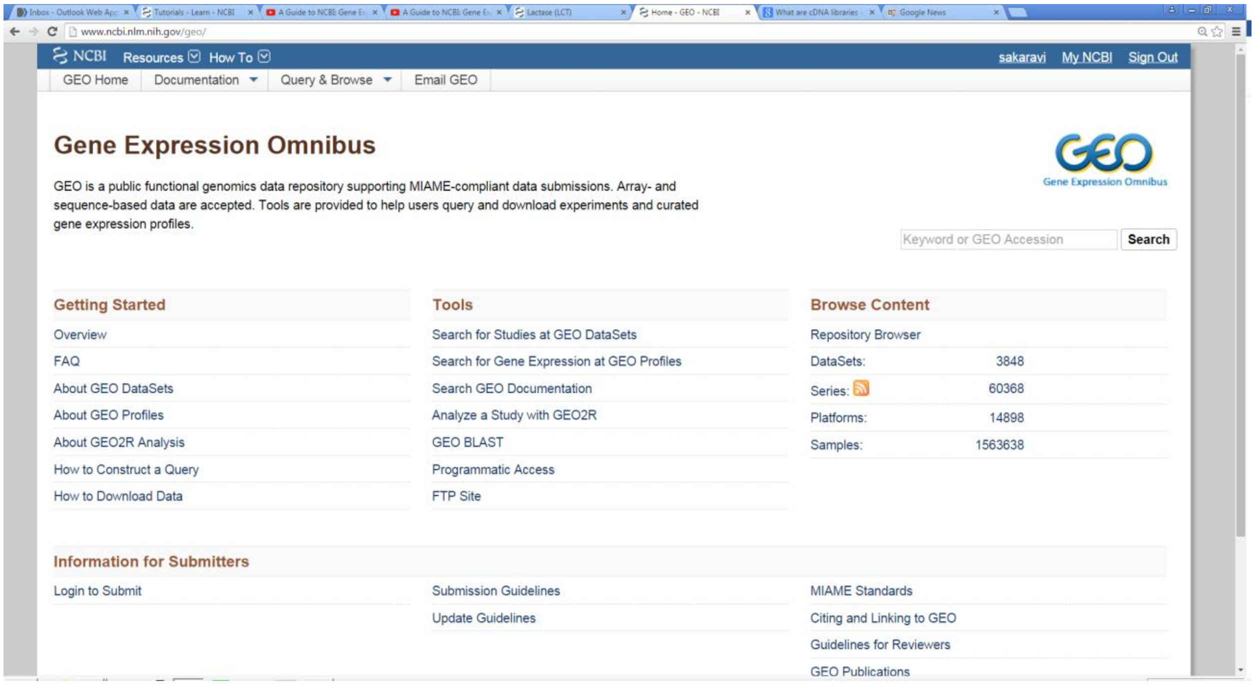
Experiments to identify the Gene expression:
Older methods (not part of NCBI DB but commonly seen in many publications as images)

a. Northern_blot Detect RNA in a sample (http://en.wikipedia.org/wiki/Northern_blot)
b. Western blot: Detect Protein in a sample (http://en.wikipedia.org/wiki/Western_blot)

Expressed Sequences :

1. Primary
    - GenBank and EST (bulk)
        - 200-500 bases;
        - mRNA (tissues/cells) → cDNA
        - Why this is done? To identify genes; To calculate the abundance of proteins
    - Newer: Sequence Read Archive (SRA)

High-Throughput Studies:
Are routinely used to estimate gene Expression, Genome Variation, Epigenomics (DNA methylation, Histone modification etc). Data from these studies are submitted into Gene Expression Omnibus (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
- 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

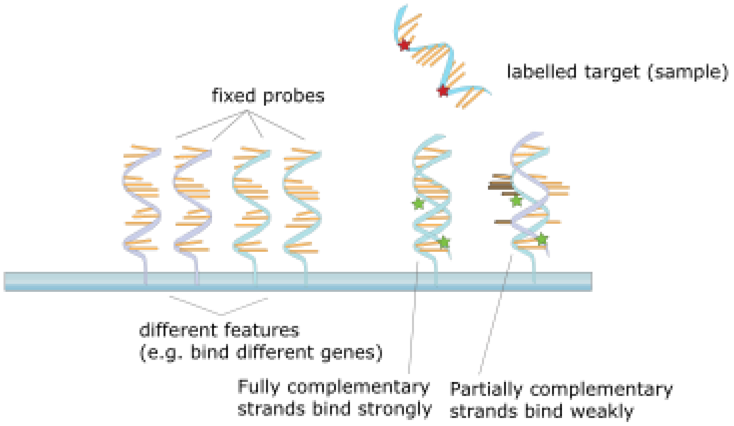
DataSets:	3848
Series:	60368
Platforms:	14898
Samples:	1563638

Information for Submitters

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

- GEO (Microarray)
  - GEO Datasets (RNA-Seq, ChiP Seq, ChIP-ChIP)
  - GEO Profiles (Global expression of protein coding genes-probe-based microarray expt)
  - What is a microarray?
    - Plate with spots. Each spot includes a gene PROBE (in the form of short piece of DNA attached to the plate) that will bind with a short piece of cDNA from a specific gene
    - mRNA that are harvested from a sample (cell/tissue) are transcribed into cDNA. These cDNAs are also labelled into red or green fluorescent dye (treated/control)
    - Light intensity for each color is measured for each spot. Statistics methods are used for processing.

Public domain figure: https://upload.wikimedia.org/wikipedia/commons/thumb/a/a8/NA_hybrid.svg/800px-NA_hybrid.svg.png

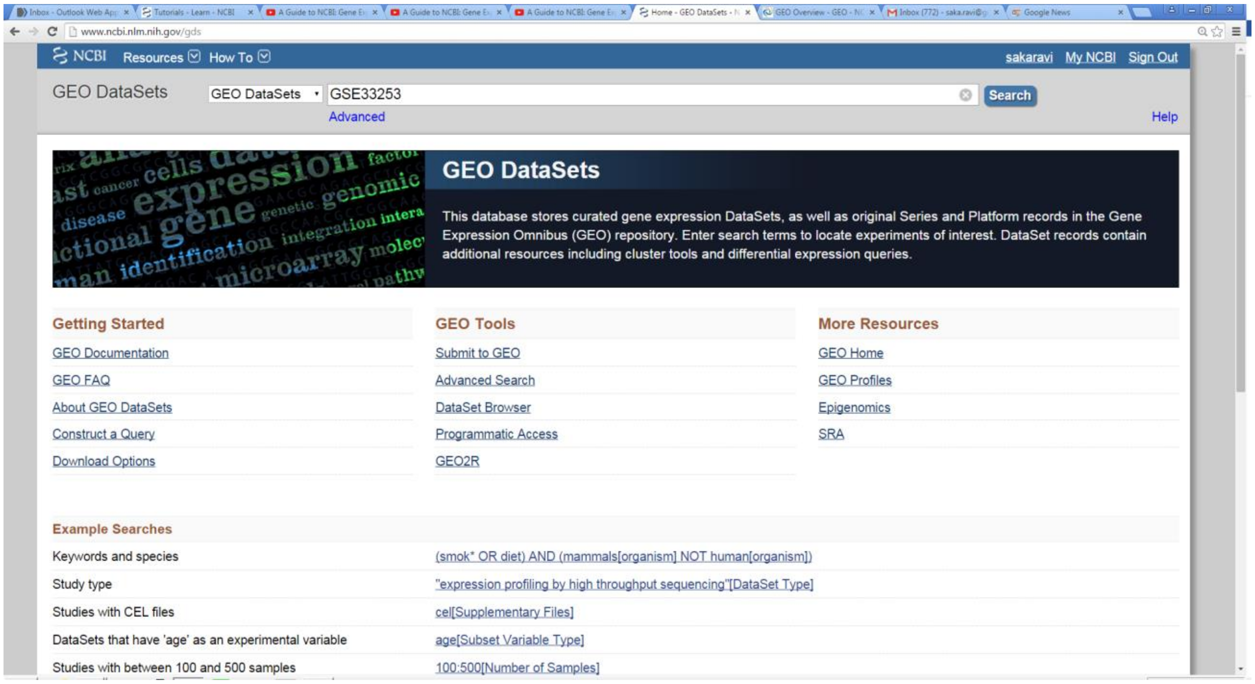


We need to understand how GEO handles data submission and data curation.

a. Investigator carries out an experiment.
b. He submits the details of his experiment by submitting a single series records.
    a. General information about the experiment (how and other details)
c. Samples are separately entered (sample records; ex 2 controls (replicates) and 2 treated (replicates); in this case 4 sample records)
d. Finally, the investigator submits a platform record which provides the details of the Chips used in the experiments.
    a. Chips are also called arrays and in the GEO context they are called Platforms

All these ends up in the GEO DataSets database and gets assigned a number or ID (ex. GSE33253)

Link for GEO DataSets is www.ncbi.nlm.nih.gov/gds



GSE33253

GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

Getting Started	GEO Tools	More Resources
GEO Documentation	Submit to GEO	GEO Home
GEO FAQ	Advanced Search	GEO Profiles
About GEO DataSets	DataSet Browser	Epigenomics
Construct a Query	Programmatic Access	SRA
Download Options	GEO2R	

Example Searches

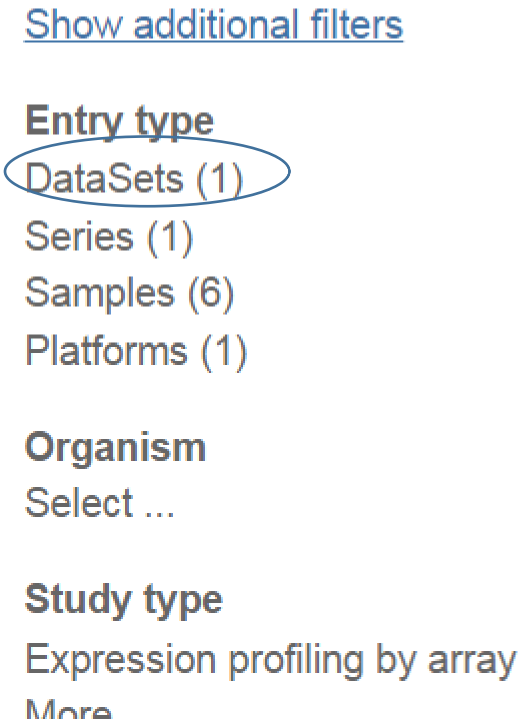
Keywords and species	(smok* OR diet) AND (mammals[organism] NOT human[organism])
Study type	"expression profiling by high throughput sequencing"[DataSet Type]
Studies with CEL files	cel[Supplementary Files]
DataSets that have 'age' as an experimental variable	age[Subset Variable Type]
Studies with between 100 and 500 samples	100:500[Number of Samples]

Note the left side menu shows that there are one **Series** record, four **Samples** record and one **Platforms** record. Look back to see the definitions of these records. Click on each one to explore the details.

Series Accession is GSE33253
Sample Accession IDs are: GSM822870-GSM822873 (4 samples)
Platform Accession: GPL1261 ID: 100001261

Note that there are no reference sequences for GEO. For example, in NCBI, the redundant GenBank sequences are reduced to one RefSeq sequence. The same with other DBs (PubChem etc.). But, NCBI groups the Series entry, their samples, the platform used for the series and groups them into one curated DataSet Entry (called **GDS**). For example, let us look at a curated and non-curated dataset display within GEO DataSets. The left-side entry is for curated and right-side dataset is yet to be curated. GDS entries can be analyzed using Curated DataSet Browser (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser/)

Please note that when you look at this page later, you will see different number of hits

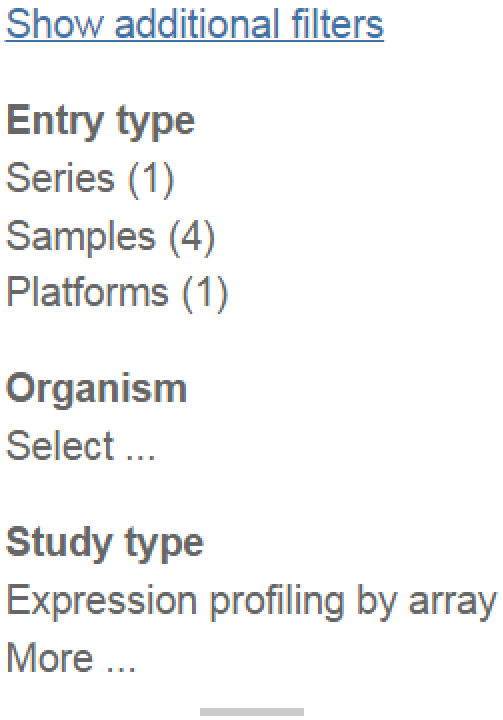


Show additional filters

Entry type
DataSets (1)
Series (1)
Samples (6)
Platforms (1)

Organism
Select ...

Study type
Expression profiling by array
More



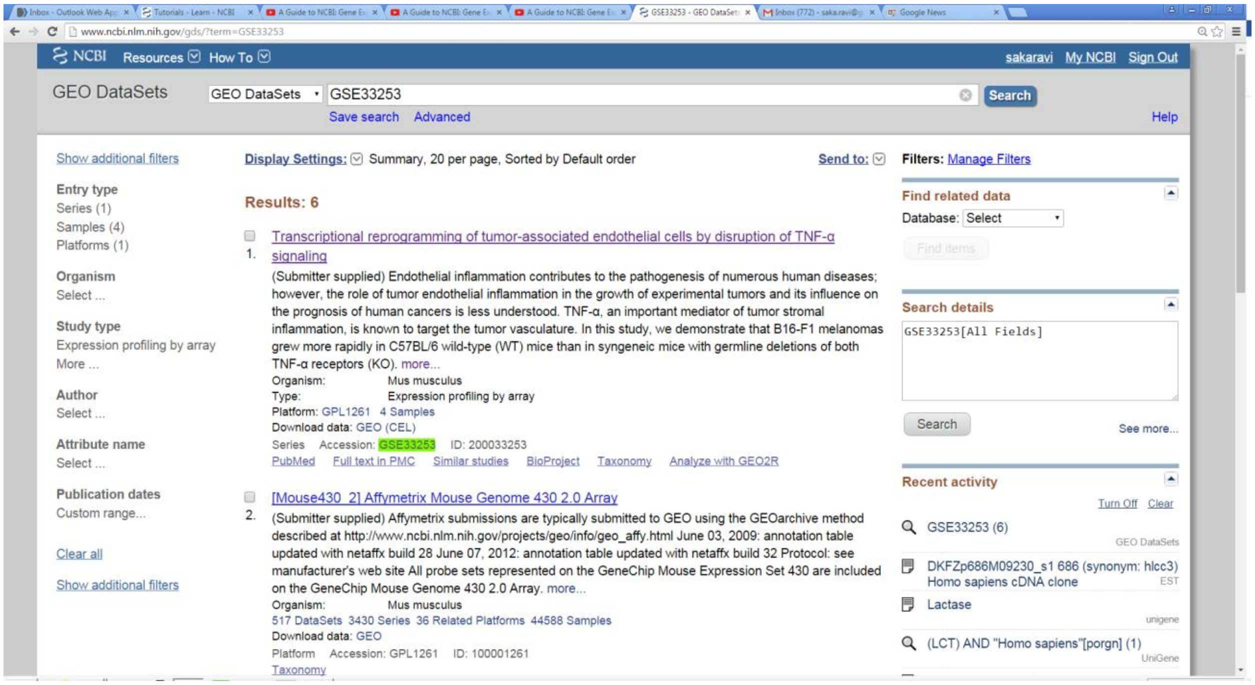
Show additional filters

Entry type
Series (1)
Samples (4)
Platforms (1)

Organism
Select ...

Study type
Expression profiling by array
More ...

Uncurated dataset



GEO DataSets | GEO DataSets | GSE33253

Show additional filters

Entry type
Series (1)
Samples (4)
Platforms (1)

Organism
Select ...

Study type
Expression profiling by array
More ...

Author
Select ...

Attribute name
Select ...

Publication dates
Custom range...

Clear all

Show additional filters

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

Send to:

Results: 6

□ 1. Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF-α signaling

(Submitter supplied) Endothelial inflammation contributes to the pathogenesis of numerous human diseases; however, the role of tumor endothelial inflammation in the growth of experimental tumors and its influence on the prognosis of human cancers is less understood. TNF-α, an important mediator of tumor stromal inflammation, is known to target the tumor vasculature. In this study, we demonstrate that B16-F1 melanomas grew more rapidly in C57BL/6 wild-type (WT) mice than in syngeneic mice with germline deletions of both TNF-α receptors (KO). more...

Organism: Mus musculus
Type: Expression profiling by array
Platform: GPL1261 4 Samples
Download data: GEO (CEL)
Series Accession: GSE33253 ID: 200033253
PubMed Full text in PMC Similar studies BioProject Taxonomy Analyze with GEO2R

□ 2. [Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array

(Submitter supplied) 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 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. more...

Organism: Mus musculus
517 DataSets 3430 Series 36 Related Platforms 44588 Samples
Download data: GEO
Platform Accession: GPL1261 ID: 100001261
Taxonomy

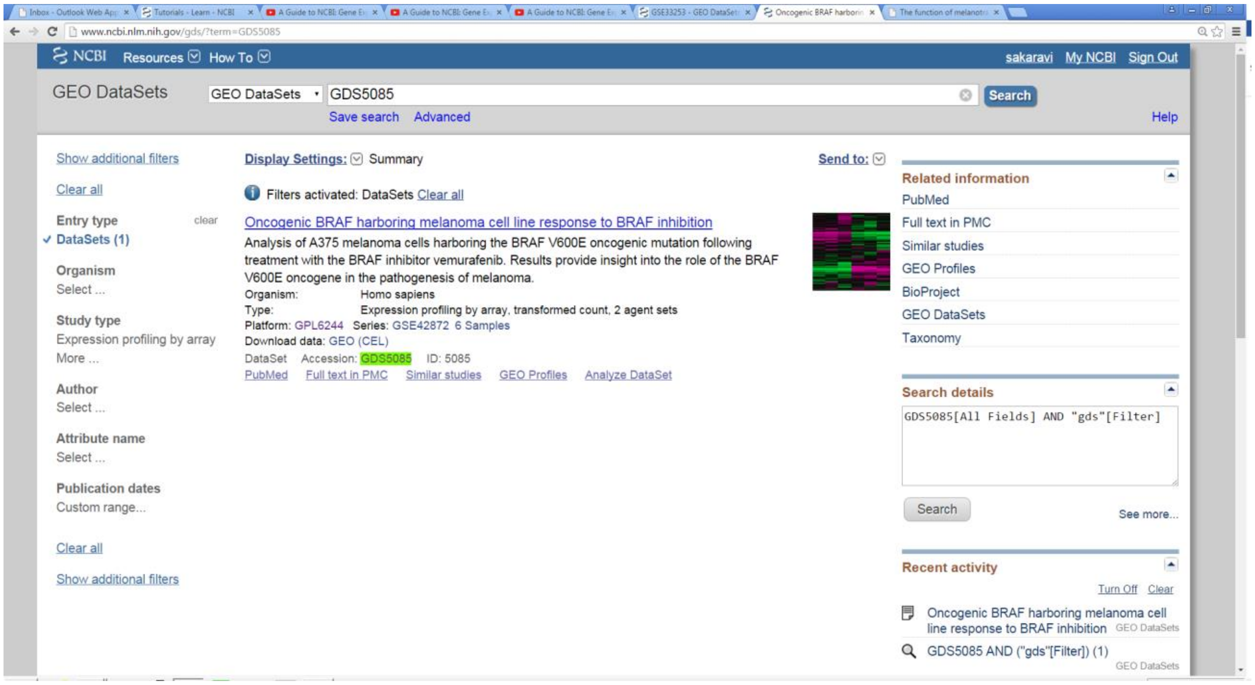
Filters: Manage Filters

Find related data
Database: Select

Search details
```
GSE33253[All Fields]
```

Recent activity
GSE33253 (6) GEO DataSets
DKFZp686M09230_s1 686 (synonym: hlcc3) Homo sapiens cDNA clone EST
Lactase unigene
(LCT) AND "Homo sapiens"[porgn] (1) UniGene

Curated dataset



GEO DataSets | GEO DataSets | GDS5085

Show additional filters

Clear all

Entry type clear
✓ DataSets (1)

Organism
Select ...

Study type
Expression profiling by array
More ...

Author
Select ...

Attribute name
Select ...

Publication dates
Custom range...

Clear all

Show additional filters

Display Settings: Summary

Send to:

Filters activated: DataSets Clear all

Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition

Analysis of A375 melanoma cells harboring the BRAF V600E oncogenic mutation following treatment with the BRAF inhibitor vemurafenib. Results provide insight into the role of the BRAF V600E oncogene in the pathogenesis of melanoma.

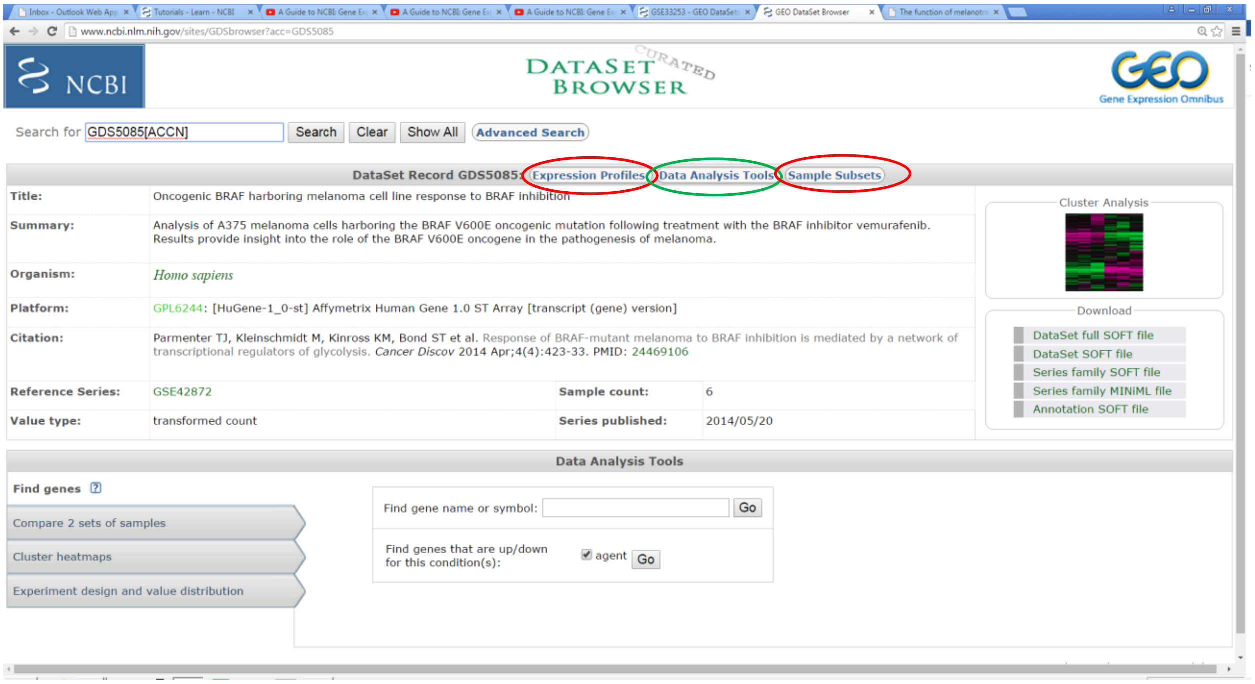
Organism: Homo sapiens
Type: Expression profiling by array, transformed count, 2 agent sets
Platform: GPL6244 Series: GSE42872 6 Samples
Download data: GEO (CEL)
DataSet Accession: GDS5085 ID: 5085
PubMed Full text in PMC Similar studies GEO Profiles Analyze DataSet

Related information
PubMed
Full text in PMC
Similar studies
GEO Profiles
BioProject
GEO DataSets
Taxonomy

Search details
```
GDS5085[All Fields] AND "gds"[Filter]
```

Recent activity
Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition GEO DataSets
GDS5085 AND ("gds"[Filter]) (1) GEO DataSets

Expression Profiles will give the individual Gene profiles associated with the study. The other options are self-explanatory.

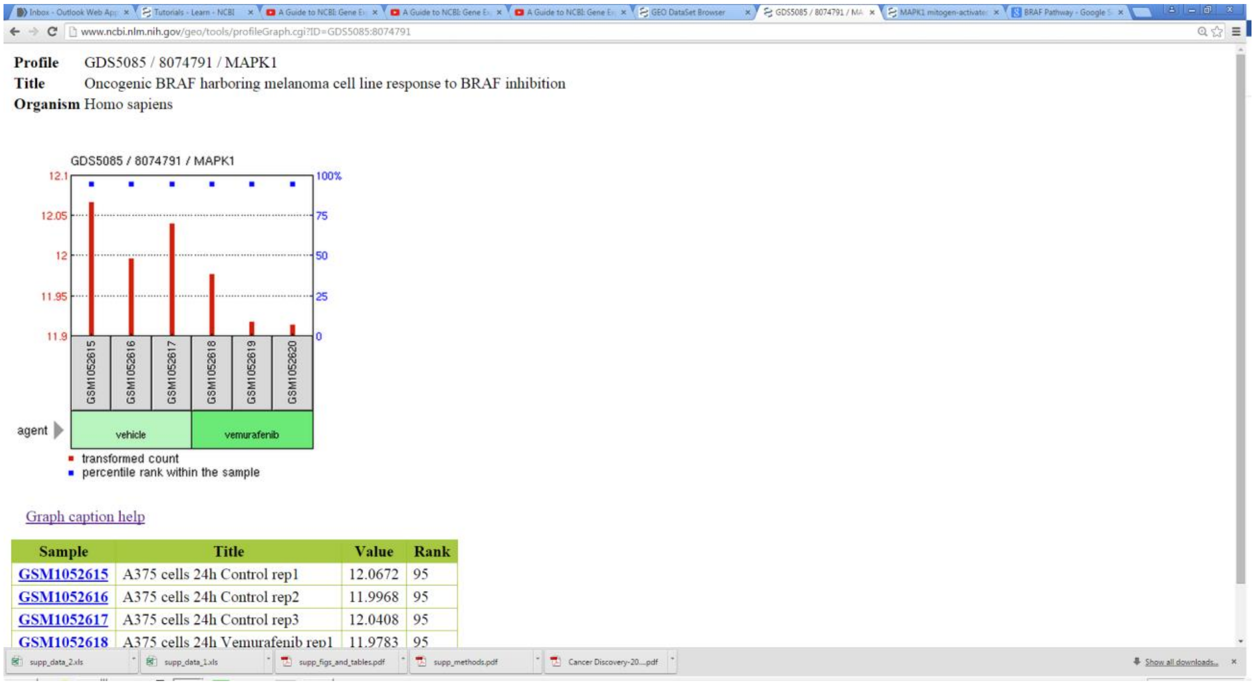


DataSet Record GDS5085: Expression Profiles Data Analysis Tools Sample Subsets

Title:	Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition		
Summary:	Analysis of A375 melanoma cells harboring the BRAF V600E oncogenic mutation following treatment with the BRAF inhibitor vemurafenib. Results provide insight into the role of the BRAF V600E oncogene in the pathogenesis of melanoma.		
Organism:	*Homo sapiens*		
Platform:	GPL6244: [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]		
Citation:	Parmenter TJ, Kleinschmidt M, Kinross KM, Bond ST et al. Response of BRAF-mutant melanoma to BRAF inhibition is mediated by a network of transcriptional regulators of glycolysis. *Cancer Discov* 2014 Apr;4(4):423-33. PMID: 24469106		
Reference Series:	GSE42872	**Sample count:**	6
Value type:	transformed count	**Series published:**	2014/05/20

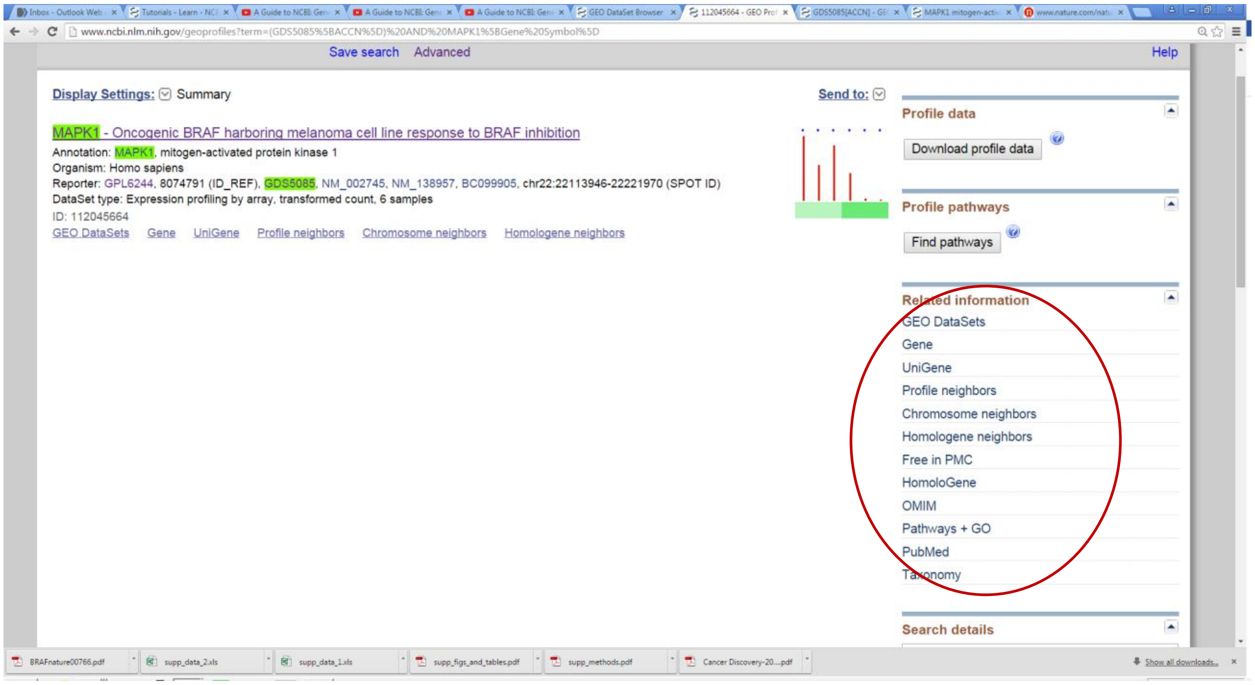
Cluster Analysis
Download
DataSet full SOFT file
DataSet SOFT file
Series family SOFT file
Series family MINiML file
Annotation SOFT file
Data Analysis Tools
Find genes
Compare 2 sets of samples
Cluster heatmaps
Experiment design and value distribution
Find gene name or symbol:
Go
Find genes that are up/down for this condition(s):
☑ agent
Go

Follow the gene profiles to look at the profile of MAP Kinase (ERK)



Profile GDS5085 / 8074791 / MAPK1
Title Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition
Organism Homo sapiens

Sample	Title	Value	Rank
GSM1052615	A375 cells 24h Control rep1	12.0672	95
GSM1052616	A375 cells 24h Control rep2	11.9968	95
GSM1052617	A375 cells 24h Control rep3	12.0408	95
GSM1052618	A375 cells 24h Vemurafenib rep1	11.9783	95



Display Settings: Summary

MAPK1 - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition
Annotation: MAPK1, mitogen-activated protein kinase 1
Organism: Homo sapiens
Reporter: GPL6244, 8074791 (ID_REF), GDS5085, NM_002745, NM_138957, BC099905, chr22:22113946-22221970 (SPOT ID)
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 112045664
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors

We can use GEO2R to carry out some analysis on the non-curated datasets. Let us do that now. Curated datasets can also be used in GEO2R to carry out your own analysis. The big help of having a curated dataset is the links to other parts of the Database.

Other expression data

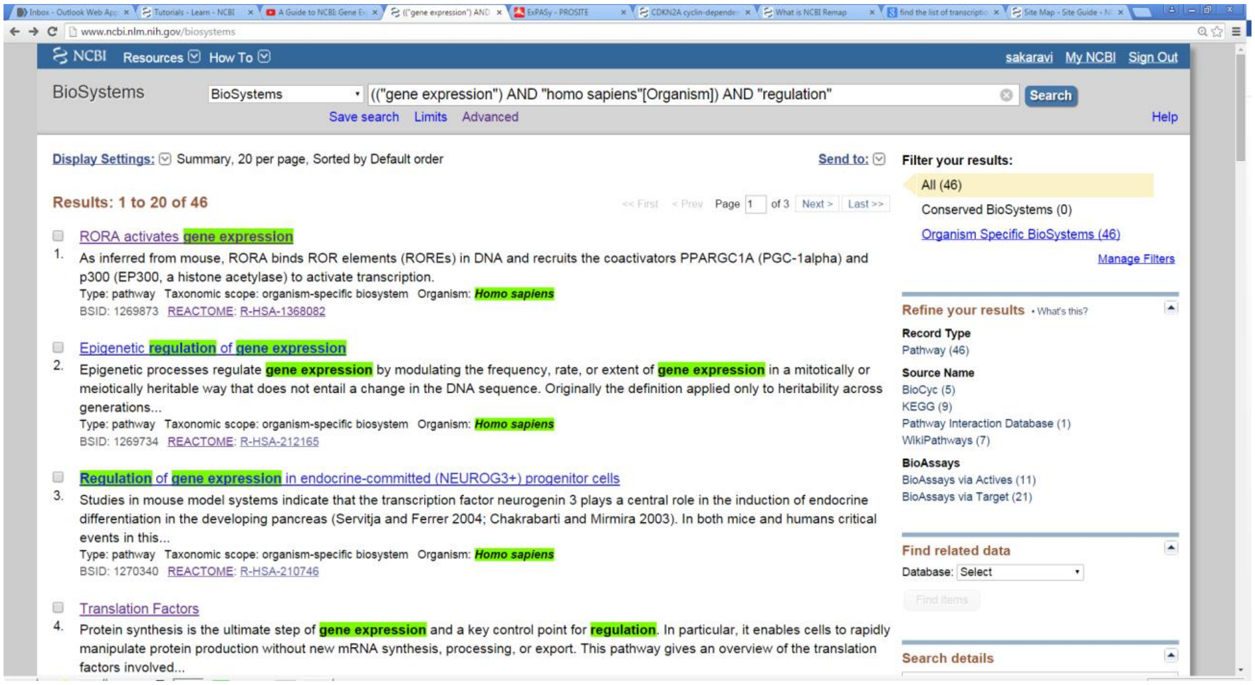
RNA-Seq (Sequence Read Archive)

RNA-Seq can be used just like ESTs. The problem is these are raw reads.

The Sequence Read Archive (SRA) stores raw sequence data from "next-generation" sequencing technologies including 454, IonTorrent, Illumina, SOLiD, Helicos and Complete Genomics.

RNA-seq data is often used to identify exon-intron boundary.

Biosystems can also be used to learn about Gene Expression or regulation or epigenetics etc.



BioSystems: (("gene expression") AND "homo sapiens"[Organism]) AND "regulation"

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

Send to:

Results: 1 to 20 of 46

<< First < Prev Page 1 of 3 Next > Last >>

1. RORA activates gene expression
   As inferred from mouse, RORA binds ROR elements (ROREs) in DNA and recruits the coactivators PPARGC1A (PGC-1alpha) and p300 (EP300, a histone acetylase) to activate transcription.
   Type: pathway Taxonomic scope: organism-specific biosystem Organism: *Homo sapiens*
   BSID: 1269873 REACTOME: R-HSA-1368082

2. Epigenetic regulation of gene expression
   Epigenetic processes regulate gene expression by modulating the frequency, rate, or extent of gene expression in a mitotically or meiotically heritable way that does not entail a change in the DNA sequence. Originally the definition applied only to heritability across generations...
   Type: pathway Taxonomic scope: organism-specific biosystem Organism: *Homo sapiens*
   BSID: 1269734 REACTOME: R-HSA-212165

3. Regulation of gene expression in endocrine-committed (NEUROG3+) progenitor cells
   Studies in mouse model systems indicate that the transcription factor neurogenin 3 plays a central role in the induction of endocrine differentiation in the developing pancreas (Servitja and Ferrer 2004; Chakrabarti and Mirmira 2003). In both mice and humans critical events in this...
   Type: pathway Taxonomic scope: organism-specific biosystem Organism: *Homo sapiens*
   BSID: 1270340 REACTOME: R-HSA-210746

4. Translation Factors
   Protein synthesis is the ultimate step of gene expression and a key control point for regulation. In particular, it enables cells to rapidly manipulate protein production without new mRNA synthesis, processing, or export. This pathway gives an overview of the translation factors involved...

Filter your results:

- All (46)
- Conserved BioSystems (0)
- Organism Specific BioSystems (46)

Manage Filters

Refine your results • What's this?

Record Type
Pathway (46)

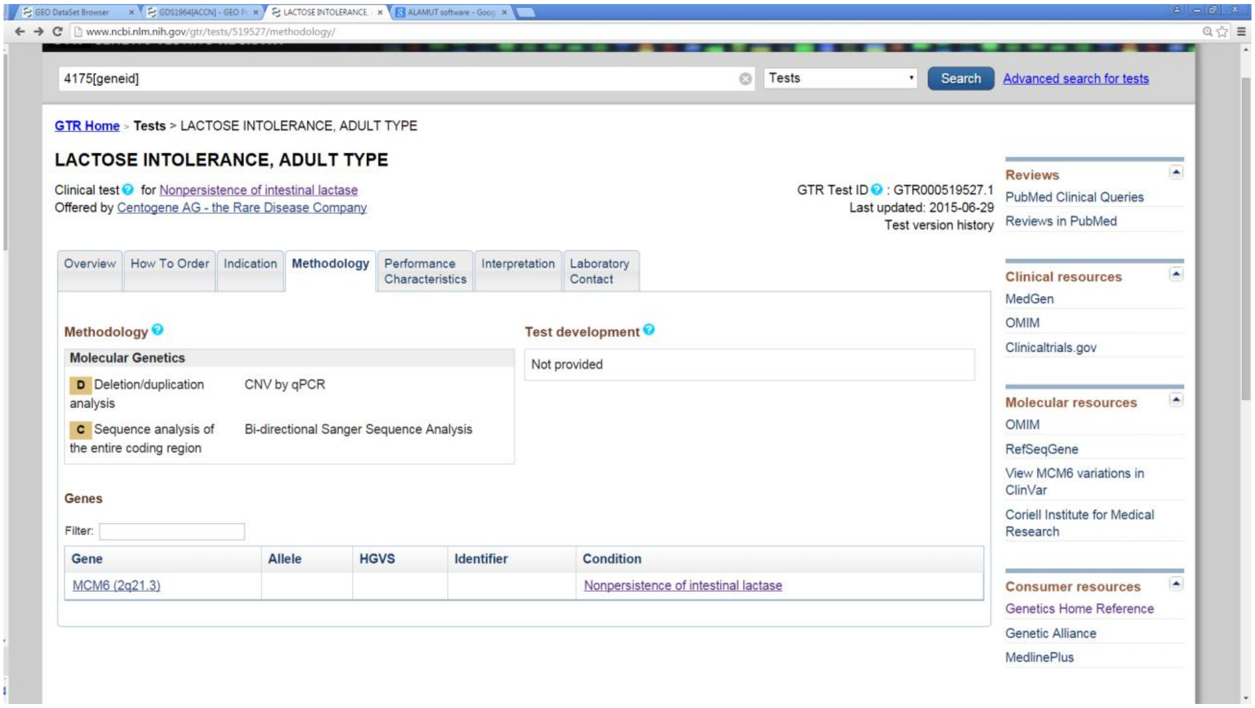
Source Name
BioCyc (5)
KEGG (9)
Pathway Interaction Database (1)
WikiPathways (7)

BioAssays
BioAssays via Actives (11)
BioAssays via Target (21)

Find related data
Database: Select
Find items

Search details

Genomic Testing Registry (GTR)


www.ncbi.nlm.nih.gov/gtr/tests/519527/methodology/
Search
Advanced search for tests

4175[geneid] | Tests

GTR Home > Tests > LACTOSE INTOLERANCE, ADULT TYPE

LACTOSE INTOLERANCE, ADULT TYPE

Clinical test for Nonpersistence of intestinal lactase
Offered by Centogene AG - the Rare Disease Company

GTR Test ID : GTR000519527.1
Last updated: 2015-06-29
Test version history

Overview | How To Order | Indication | Methodology | Performance Characteristics | Interpretation | Laboratory Contact

Methodology

Molecular Genetics

D Deletion/duplication analysis	CNV by qPCR
C Sequence analysis of the entire coding region	Bi-directional Sanger Sequence Analysis

Test development

Not provided

Genes

Filter:

Gene	Allele	HGVS	Identifier	Condition
MCM6 (2q21.3)				Nonpersistence of intestinal lactase

Reviews
PubMed Clinical Queries
Reviews in PubMed

Clinical resources
MedGen
OMIM
Clinicaltrials.gov

Molecular resources
OMIM
RefSeqGene
View MCM6 variations in ClinVar
Coriell Institute for Medical Research

Consumer resources
Genetics Home Reference
Genetic Alliance
MedlinePlus