Web-based tools for Bioinformatics; A (free) introduction to (freely available) NCBI, MUSC and World-wide.

When and Where---Wednesdays at 1pm-2pm Room 438 Library Admin Building Beginning September 10, 2003.

Overview Expression Resources December 3, 2003

Expression experiments can be directed to measure expressed proteins of an organism, organ, or cell type (Proteome) OR they may be directed toward the expression of genes. This talk is aimed at gene expression experiments and the resources which serve such experiments. Because the expense of such experiments is still quite high it becomes critically important to understand the capability and limits of the intended experimental design PRIOR to heading for the bench. It’s not as if experimental design is anything new but because of the inherent limits of the chip methods, it is absolutely essential to have the statistical methods in hand from the earliest planning stages.

Closed Systems: limited to analyzing expression of known sequences. Not ideal for discovery, do not measure absolute numbers, are hard to compare between experiments, but are “cheap” and excellent for high-throughput one and two channel microarrays

Open Systems: do not require prior knowledge of sequences to measure expression, can measure absolute numbers, easy to compare between experiments, but are currently costly and therefore poor for high-throughput Serial Analysis of Gene Expression (SAGE)
TOP LAYER: Variation due to attributes or conditions are a given

SECOND LAYER: biological variation is intrinsic; influenced by genetic & environmental factors, as well as whether samples are from populations or individuals

THIRD LAYER: technical variation results during sample extraction, labeling and hybridization

FOURTH LAYER: measurement variation can arise during laser scanning and fluorescence detection

\[ df = \#\text{units} - \#\text{attributes or conditions} \]
So, in the liver experiment illustrated...

\[ df = 4 \text{ livers} - 2 \text{ groupings (normal vs infected liver)} \]

\[ df = 4 - 2 \]
df = 2

As a rule of thumb, df must be $\geq 5$ to be acceptable.

\[
\text{cost} = n + kC_1 + n \times mC_M
\]

where:

- $n =$ pools composed of individual experimental units
- $k =$ number of individual units
- $m =$ replicates of spots on array
- $C_1 =$ unit cost per experiment
- $C_M =$ cost to measure each array

In the next image we see an example of the power of the microarray method. This is a two channel experiment. Red dots are judged "up" regulated; green "down" and gray unchanged. Thousands of genes are present. The spectrin gene is isolated for further assessment. These images were created with GeneSifter software
Validate array results by an independent method...

- Reverse Transcription-PCR (RT-PCR)
- Northern Blot

Rules of thumb for the uninitiated

- Array experiments are multilayered
- These layers are sources of noise and signal
- Experiments must be designed to distinguish noise from signal
- In general, statistical validity of results improves as the number of samples N increases
But costs also increase with N...

A balance can be struck between costs and validity by a variety of strategies

- Work with a statistician

- Validate array results by other experimental methods

Some References


An anatomy of normal and malignant gene expression Boon K et al PNAS 99(17) 11287-11292


Introduction/Scope

Expression Data Storage and Retrieval Engines

Stanford Microarray Database (SMD)

EBI's ArrayExpress (AE)

Expression Array Manager (EAM)

NCBI Gene Expression Omnibus (GEO).
GEO is a public repository for a wide range of high-throughput experimental data, single and dual channel microarray-based experiments measuring RNA, genomic DNA and protein abundance, and non-array techniques such as serial analysis of gene expression (SAGE), and mass spectrometry proteomic data.

GEO Terms

Submitter
A submitter entity contains contact and authentication information about the submitter. This information is kept only so that the source of data in GEO can be properly referenced. A submitter entity may have relationships to many platforms, many samples and many series.

Platform
A platform record describes the list of elements on the array (e.g., cDNAs, oligonucleotide probesets, ORFs, antibodies) or the list of elements that may be detected and quantified in that experiment (e.g., SAGE tags, peptides). Each platform record is assigned a unique and stable GEO accession number (GPLxxx). A platform may reference many samples that have been submitted by multiple submitters.

Sample
A sample record describes the conditions under which an individual sample was handled, the manipulations it underwent and the abundance measurement of each element derived from it. Each sample record is assigned a unique and stable GEO accession number (GSMxxx). A sample entity must reference only one platform, and may be included in multiple series.

Series
A series record defines a set of related samples considered to be part of a group, how the samples are related and if and how they are ordered. A series provides a focal point and description of the experiment as a whole. Series records may also contain tables describing extracted data, summary conclusions or analyses. Each series record is assigned a unique and stable GEO accession number (GSExxx).

GEODatasets

GEO DataSets (GDSxxx) are curated sets of GEO sample data. A GDS record represents a collection of biologically and statistically comparable GEO samples and forms the basis of GEO’s suite of data display and analysis tools. Samples within a GDS refer to the same platform, that is, they share a common set of probe elements. Value measurements for each sample within a GDS are assumed to be calculated in an equivalent manner, that is, considerations such as background processing and normalization are consistent across the dataset. Information reflecting experimental design is provided through GDS subsets.

Downloading GEO datasets

GEO records Several options are provided on the Accession Display bar (found at the foot of the GEO home page and the top of each GEO record) for the retrieval and display of original GEO records. The Scope feature allows display of a single accession number (Self), or any (Platform, Sample or Series) or all (Family) records related to that accession. Amount dictates the quantity of data displayed, with choices including metadata only, metadata and the first 20 rows of the data table, data table only, or full metadata/data table records. Format controls
whether records are displayed in HTML or in SOFT format. SOFT (Simple Omnibus Format in Text) is a ASCII text format which was designed to be a machine readable representation of data retrieved from, or submitted to, GEO. SOFT is also a line-based format, making it easy to parse using commonly available text processing and formatting languages. For a complete description of SOFT format, see the SOFT guide.

GDS records Each GDS record has three options for the download of that dataset. The complete SOFT document contains all information for that dataset, including dataset description, type, organism, subset allocation, etc, as well as a data table containing identifiers and values. The data only option allows download of the data table only, while the quick view provides dataset descriptive information and the first 20 rows of the data table. The full text tab-delimited data tables provided with these downloads may prove suitable for upload into your favorite microarray analysis software package or database/spreadsheet application.

Both GDS and GEO data are available for bulk download via FTP. GEO DataSets may be downloaded in complete GDS SOFT format, while complete original GEO records, partitioned by GEO platform, may be downloaded in SOFT format.

**Non-Commercial Analysis Software**

EBI Expression Profiler ([EP](http://www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/index.smf))

Bio Array Software Environment ([BASE](http://www.rosettabio.com/products/resolver/default.htm))

**Commercial Expression Analysis Software**

GeneSpring from SiliconGenetics:

[http://www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/index.smf](http://www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/index.smf)

Resolver from Rosetta Biosoftware:


GeneSifter from VizX Labs:

[http://www.genesifter.net/](http://www.genesifter.net/)

The Cancer Genome Anatomy Project


---

**A worked Example**

http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm

12/3/2003
Looking for CASP7 Results at GEO
The **Gene Expression Omnibus** is a gene expression and hybridization array data repository, as well as a curated, online resource for gene expression data browsing, query and retrieval. GEO was the first fully public high-throughput gene expression data repository, and became operational in July 2000.

### Gene profiles

This tool queries Entrez GEO molecular abundance profiles by annotation or pre-computed profile characteristics. Enter as much information as desired in the boxes below and click Submit.

- **Gene**: CASP7
- **Accession**: 
- **Any text**: 
- **Organism**: HUMAN
- **Effect**: any or no effect, age, agent
- **Abundance**: any max value rank, max value 99-100%, any min value rank, min value 99-100%
- **Variability**: any std dev rank, standard deviation 99-100%
- **Outliers**: keep, discard

### Retrieve GEO accession

- **Scope**: Self, In: HTML
- **Depositors only**
- **User**: 
- **Password**: 

[http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm](http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm)
This opens the ENTREZ GEO page with 26 entries

The small graph opens to a larger image with more details. The red line shows user-
submitted normalized or transformed values of molecular abundance. The blue line is percentile ranked values for the expression. Note from the previous page that GSM1938 is flagged as an outlier.

The details for the GDS90 dataset are in the linked table:
**Record for GDS90**

**Title:** Mammary epithelial cells and breast cancer

**Summary:** Comparison of gene expression in mammary epithelial cells growing in culture and primary breast tumors. Expression patterns associated with complex physiological properties and cellular composition of the tumors identified.

**Base platform:** GPL170

**Organism:** Homo sapiens

**Type:** dual channel nucleotide

**Value type:** log ratio

**Number of probes:** 5760

**Reference Series:** GSE53

**Last update:** April 06 2003

### 26 dataset samples, order: none

<table>
<thead>
<tr>
<th>Sample selection</th>
<th>Data</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>check all</td>
<td>download</td>
<td>close</td>
</tr>
<tr>
<td>uncheck all</td>
<td>analysis</td>
<td></td>
</tr>
<tr>
<td>toggle check</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2 Assigned subsets

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>? toggle</th>
</tr>
</thead>
<tbody>
<tr>
<td>disease</td>
<td>in vitro mammary epithelial cells</td>
<td>(3)</td>
</tr>
<tr>
<td>disease</td>
<td>primary breast tumor</td>
<td>(18)</td>
</tr>
</tbody>
</table>

- GSM1932 : HMEC X HMEC -EGF for 2 days
- sro1: control HMEC
- sro2: HMEC -EGF for 2 days
- GSM1935 : HMEC X HMEC + TGFbeta 24hrs
- sro1: control HMEC
- sro2: HMEC + TGFbeta 24hrs
- GSM1950 : HMEC control X HMEC+Interferon gamma 24hrs
- sro1: HMEC control
- sro2: HMEC + INF gamma 24hrs
- GSM1952 : HMEC X Matrigel
- sro1: HMEC
- sro2: HMEC on Matrigel 1 day
- GSM1929 : HMEC X NE2
- sro1: HMEC
- sro2: ND (pooled2)
- GSM1930 : HMEC X BC24
- sro1: HMEC
- sro2: BC24
- GSM1934 : HMEC X LN5
- sro1: HMEC
- sro2: LN5
- GSM1937 : HMEC X BC1498
- sro1: HMEC

You can review details of the experiment or the platform

### Platform GPL170

<table>
<thead>
<tr>
<th><strong>Status</strong></th>
<th>Public on Jun 27 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>5K</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>non-commercial nucleotide</td>
</tr>
<tr>
<td><strong>Organism</strong></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>A microarray with 5760 spot features. Tip Configuration: Standard 4-tip Columns per Sector: 38 Rows per Sector: 38 Column Spacing: 175 Row Spacing: 175</td>
</tr>
<tr>
<td><strong>Submission date</strong></td>
<td>May 22 2002</td>
</tr>
<tr>
<td><strong>Submitter name</strong></td>
<td>Sherlock, Gavin</td>
</tr>
<tr>
<td><strong>Submitter email</strong></td>
<td><a href="mailto:sherlock@genome.stanford.edu">sherlock@genome.stanford.edu</a></td>
</tr>
<tr>
<td><strong>Submitter institute</strong></td>
<td>Stanford University, School of Medicine</td>
</tr>
<tr>
<td><strong>Submitter department</strong></td>
<td>Department of Genetics</td>
</tr>
<tr>
<td><strong>Submitter address</strong></td>
<td>300 Pasteur Drive</td>
</tr>
<tr>
<td><strong>Submitter city</strong></td>
<td>Stanford, CA 94305 USA</td>
</tr>
<tr>
<td><strong>Submitter phone</strong></td>
<td>650-498-6012</td>
</tr>
<tr>
<td><strong>Submitter web link</strong></td>
<td>genome-www.stanford.edu/~sherlock/</td>
</tr>
</tbody>
</table>

### Data table header descriptions

<table>
<thead>
<tr>
<th><strong>ID</strong></th>
<th>unique spot identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NAME</strong></td>
<td>name of spot</td>
</tr>
<tr>
<td><strong>CLONE_ID</strong></td>
<td>cDNA clone identifier</td>
</tr>
<tr>
<td><strong>GENE_SYM</strong></td>
<td>gene symbol</td>
</tr>
<tr>
<td><strong>GENE_NAME</strong></td>
<td>gene name, description</td>
</tr>
<tr>
<td><strong>UNIGENE</strong></td>
<td>NCBI UniGene cluster identifier</td>
</tr>
<tr>
<td><strong>GB_ACC</strong></td>
<td>GenBank accession number</td>
</tr>
</tbody>
</table>

[Go to Query DataSets for GPL170](http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm)
**Series GSE53**

<table>
<thead>
<tr>
<th><strong>Status</strong></th>
<th>Public on Jun 27 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>Human mammary epithelium and breast cancer</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>other</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. cDNA microarrays and a clustering algorithm were used to identify patterns of gene expression in human mammary epithelial cells growing in culture and in primary human breast tumors. Clusters of coexpressed genes identified through manipulations of mammary epithelial cells in vitro also showed consistent patterns of variation in expression among breast tumor samples. By using immunohistochemistry with antibodies against proteins encoded by a particular gene in a cluster, the identity of the cell type within the tumor specimen that contributed the observed gene expression pattern could be determined. Clusters of genes with coherent expression patterns in cultured cells and in the breast tumors samples could be related to specific features of biological variation among the samples. Two such clusters were found to have patterns that correlated with variation in cell proliferation rates and with activation of the IFN-regulated signal transduction pathway, respectively. Clusters of genes expressed by stromal cells and lymphocytes in the breast tumors also were identified in this analysis. These results support the feasibility and usefulness of this systematic approach to studying variation in gene expression patterns in human cancers as a means to dissect and classify solid tumors. This study is described more fully in Perou CM, et al. 1999. Proc Natl Acad Sci USA 96:9212-7</td>
</tr>
<tr>
<td><strong>Author</strong></td>
<td>Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D</td>
</tr>
<tr>
<td><strong>PubMed id</strong></td>
<td>10430922</td>
</tr>
<tr>
<td><strong>Submission date</strong></td>
<td>Jun 18 2002</td>
</tr>
<tr>
<td><strong>Submitter name</strong></td>
<td>Sherlock, Gavin</td>
</tr>
<tr>
<td><strong>Submitter email</strong></td>
<td><a href="mailto:sherlock@genome.stanford.edu">sherlock@genome.stanford.edu</a></td>
</tr>
<tr>
<td><strong>Submitter institute</strong></td>
<td>Stanford University, School of Medicine</td>
</tr>
<tr>
<td><strong>Submitter department</strong></td>
<td>Department of Genetics</td>
</tr>
<tr>
<td><strong>Submitter address</strong></td>
<td>300 Pasteur Drive</td>
</tr>
</tbody>
</table>

You may download all the dataset or only some subset of the data
**Organism:** Homo sapiens  
**Type:** dual channel nucleotide  
**Value type:** log ratio  
**Number of probes:** 5760  
**Reference Series:** GSE53  
**Last update:** April 06 2003

### 26 dataset samples, order: none

<table>
<thead>
<tr>
<th>Sample selection</th>
<th>Data</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>check all</td>
<td></td>
<td>complete DataSet</td>
</tr>
<tr>
<td>uncheck all</td>
<td></td>
<td>only checked samples</td>
</tr>
<tr>
<td>toggle check</td>
<td></td>
<td>checked data only</td>
</tr>
</tbody>
</table>

#### 2 Assigned subsets

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>disease</td>
<td>in vitro mammary epithelial cells</td>
</tr>
<tr>
<td>disease</td>
<td>primary breast tumor</td>
</tr>
</tbody>
</table>

☑ GSM1932 : HMEC + HMEC - EGF for 2 days  
src1: control HMEC  
src2: HMEC - EGF for 2 days  
☑ GSM1935 : HMEC X HMEC + TGFbeta 24hrs  
src1: control HMEC  
src2: HMEC + TGFbeta 24hrs  
☑ GSM1950 : HMEC control X HMEC+Interferon gamma 24hrs  
src1: HMEC control  
src2: HMEC + INF gamma 24hrs  
☑ GSM1952 : HMEC X Matrigel  
src1: HMEC  
src2: HMEC on Matrigel 1day  
☑ GSM1928 : HMEC X NB2  
☑ GSM1933 : HMEC - EGF then refed EGF for 1.5hrs  
src1: HMEC - EGF for 2days  
src2: HMEC refed EGF 1.5hrs  
☑ GSM1949 : HMEC X Confluent HMEC  
src1: control HMEC  
src2: 100% Confluent HMEC  
☑ GSM1951 : HMEC X HMEC+Interferon alpha 24hrs  
src1: HMEC  
src2: HMEC+INF Alpha 24hrs  
☑ GSM1953 : HMEC X Senescent HMEC-2  
src1: control HMEC  
src2: Senescent HMEC p19  
☑ GSM1929 : HMEC X MCF7-NCI

---

http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm  
12/3/2003
The Cancer Genome Anatomy Project is another place to launch a search CGAP

CGAP's Goals

The goal of the NCI's Cancer Genome Anatomy Project is to determine and cancer cells, leading eventually to improved detection, diagnosis, and treatment of cancer. Scientists worldwide, such as the Ludwig Institute for Cancer Research, are using the project to expand their databases for the benefit of all cancer researchers.

The Web Site

The information is organized in a "biological sense" as follows:

- **Genes**: Information on specific genes and collections of genes.
- **Tissues**: Information on CGAP and other cDNA libraries, gene expression, and SNPs.
- **Pathways**: Diagrams of biological pathways and protein complexes, with links to genetic resources for each known protein/enzyme.

The CGAP web site provides researchers with access to all CGAP data:

- Genomic data for human and mouse, including expressed sequence tag (EST) nucleotide polymorphisms (SNPs), cluster assemblies, and cytogenetic maps.
- Informatics tools to query and analyze the data.
- Information on methods and resources for reagents developed by the project.

If you have any questions, comments, or need information about CGAP, please contact:

http://people.musc.edu/~hazards/WebBioInformatics/ExpressionResources.htm

12/3/2003
Gene Finder

What the Gene Finder Tool Can Do

The Gene Finder tool finds one gene or list of genes, based on selected criteria. It works best if you start from the Gene Info page which provides selected gene data and links to various resources.

Need help! if this is your first time using the Gene Finder tool and you don't know how to use it, use the Gene Finder Tool.

Use the Gene Finder

Specific: By Gene Symbol, Accession Number

1. Select organism: Homo sapiens
2. Enter a unique identifier:* CASP7
3. Submit query: Submit Query

*A unique identifier is either a gene symbol, GenBank accession number, or Entrez Gene ID.

General: By Tissue, Function, Location

1. Select organism: Homo sapiens
2. Search on one or more fields:
   - Tissue Type: (Any)
   - Gene Ontology Term Help:
   - Cytogenetic Location:
   - Keyword in Gene Name:
3. Submit the query: Submit Query
   or
4. Reset the form: Reset Form
Gene Finder Results For: Hs; CASP7
UniGene Build: Hs.163/Mm.131

Displaying 1 thru 1 of 1 items

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Sequence ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP7</td>
<td>caspase 7, apoptosis-related cysteine protease</td>
<td>NM_001227, NM_003340, NM_003339, NM_003338</td>
</tr>
</tbody>
</table>
Gene Info

Gene Information For: Hs. CASP7, caspase 7, apoptosis-related cysteine protease

Sequence ID:  
- NM_001227  
- NM_033340
- NM_033339  
- NM_033338

Database Links

UniGene LocusLink OMIM DTP SNPViewer Assemblies

Gene Expression Data

- This gene is found in these cDNA libraries from the following tissue types: adipose, adrenal cortex, b-cell, bone, bone marrow, brain, cartilage, gastrointestinal tract, genitourinary, germ cell, head and neck, heart, lymphoepithelial, mammary gland, muscle, nervous, ovary, pancreas, peritoneum, placenta, pooled tissue, prostate, retina, salivary gland, skin, spleen, testis, uncharacterized tissue, uterus, vascular, white blood cell.

- SAGE Anatomic Viewer
  - Tissues only
  - Cell lines only
  - Tissues and cell lines

- SAGE Digital Northern

- Monochromatic SAGE/cDNA Virtual Northern

- Two-dimensional array displays (similar expression pattern in NCI60 microarray)

Cytogenetic Location (from Unit)

Cytogenetic Location: 10q25 Mitelman Breakpoint Data

Full-Length MGC Clones for This
**SAGE Anatomic Viewer**

*Search query: ACAAGAACA, Tissues and cell lines*

Colored organ image is hyperlinked to Digital Northern. "Brain" label is hyperlinked.

<table>
<thead>
<tr>
<th>Normal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="brain.png" alt="Brain" /></td>
<td>Brain</td>
</tr>
<tr>
<td><img src="retina.png" alt="Retina" /></td>
<td>Retina</td>
</tr>
<tr>
<td><img src="thyroid.png" alt="Thyroid" /></td>
<td>Thyroid</td>
</tr>
<tr>
<td><img src="lung.png" alt="Lung" /></td>
<td>Lung</td>
</tr>
<tr>
<td><img src="heart.png" alt="Heart" /></td>
<td>Heart</td>
</tr>
<tr>
<td><img src="breast.png" alt="Breast" /></td>
<td>Breast</td>
</tr>
<tr>
<td><img src="stomach.png" alt="Stomach" /></td>
<td>Stomach</td>
</tr>
<tr>
<td><img src="pancreas.png" alt="Pancreas" /></td>
<td>Pancreas</td>
</tr>
<tr>
<td><img src="liver.png" alt="Liver" /></td>
<td>Liver</td>
</tr>
<tr>
<td><img src="kidney.png" alt="Kidney" /></td>
<td>Kidney</td>
</tr>
<tr>
<td><img src="colon.png" alt="Colon" /></td>
<td>Colon</td>
</tr>
<tr>
<td><img src="peritoneum.png" alt="Peritoneum" /></td>
<td>Peritoneum</td>
</tr>
<tr>
<td><img src="spinal_cord.png" alt="Spinal Cord" /></td>
<td>Spinal Cord</td>
</tr>
</tbody>
</table>
Search query: ACAAGAACAA, Tissues only
### Digital Northern Resources

**Search query:** ACAAGAACAA

<table>
<thead>
<tr>
<th>Color Code</th>
<th>&lt;2</th>
<th>&lt;4</th>
<th>&lt;8</th>
<th>&lt;16</th>
<th>&lt;32</th>
<th>&lt;64</th>
<th>&lt;128</th>
<th>&lt;256</th>
<th>&lt;512</th>
<th>&gt;512</th>
</tr>
</thead>
</table>

#### Library

<table>
<thead>
<tr>
<th>Library</th>
<th>Total in Library</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAGE_Colon_normal_B_NC2</td>
<td>4847</td>
</tr>
<tr>
<td>SAGE_Pancreas_adenocarcinoma_B_96-6252</td>
<td>3321</td>
</tr>
<tr>
<td>SAGE_Colon_adenocarcinoma_B_Tu98</td>
<td>4137</td>
</tr>
<tr>
<td>SAGE_Breast_metastatic_carcinoma_B_95-260</td>
<td>4506</td>
</tr>
<tr>
<td>SAGE_Breast_normal_epithelium_AP_1</td>
<td>4872</td>
</tr>
<tr>
<td>SAGE_Kidney_embryonic_CL_293+beta-caterin</td>
<td>2351</td>
</tr>
<tr>
<td>SAGE_Brain_ependymoma_B_1394</td>
<td>5631</td>
</tr>
<tr>
<td>SAGE_Peritoneum_normal_B_13</td>
<td>5352</td>
</tr>
<tr>
<td>SAGE_Colon_adenocarcinoma_CL_Caco2</td>
<td>6066</td>
</tr>
<tr>
<td>SAGE_Peritoneum_mesothelioma_B_12</td>
<td>3252</td>
</tr>
<tr>
<td>SAGE_Brain_glioblastoma_B_H1110</td>
<td>6896</td>
</tr>
<tr>
<td>SAGE_Breast_carcinoma_myoeithelium_AP_DC167</td>
<td>3743</td>
</tr>
<tr>
<td>SAGE_Pancreas_adenocarcinoma_CL_CAPAN1</td>
<td>3767</td>
</tr>
<tr>
<td>SAGE_Brain_ependymoma_B_239</td>
<td>4665</td>
</tr>
<tr>
<td>SAGE_Leukocytes_normal_B_1</td>
<td>4787</td>
</tr>
<tr>
<td>SAGE_Ovary_normal_CL_ICSE29EC-11</td>
<td>4715</td>
</tr>
<tr>
<td>SAGE_Brain_ependymoma_B_1160</td>
<td>6237</td>
</tr>
<tr>
<td>SAGE_Brain_ependymoma_B_R455</td>
<td>5182</td>
</tr>
<tr>
<td>SAGE_Brain_ependymoma_B_R582</td>
<td>5218</td>
</tr>
<tr>
<td>SAGE_Brain_glioblastoma_B_R336</td>
<td>1023</td>
</tr>
<tr>
<td>SAGE_Brain_glioblastoma_B_pooled</td>
<td>5642</td>
</tr>
<tr>
<td>SAGE_Brain_madulloblastoma_B_H484</td>
<td>5746</td>
</tr>
<tr>
<td>SAGE_Brain_normal_cerebellum_B_1</td>
<td>5036</td>
</tr>
</tbody>
</table>
Sample Questions/Data

Try the CGAP search on the gene GFAP.

Created by ESH 8-18-2003; updated 12-3-2003 12:40

e-mail to Starr about this page