

## BTEP Workshop 2016: Using NCBI's Gene Expression Omnibus (GEO) to Explore Gene Expression; June 6, 2016

### Exercise 1: Explore a Study that was submitted to GEO

#### (Independent practice)

#### 1. Access the GSE33253 study and answer any questions along the way

- On the NCBI home page select the **GEO DataSets** database from the **All Databases** pull down menu.

Q: What is the name of another GEO database that you notice on the pull down menu? \_\_\_\_\_

(You will explore this second database later.)

- You can search the database with a text term, or a GEO accession number. Here, enter the following GEO accession number: **GSE33253** and click on the **Search** button.

The screenshot shows the NCBI homepage with the search bar containing 'GSE33253'. The 'All Databases' dropdown menu is open, and 'GEO DataSets' is highlighted. The page also features sections for 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research', along with 'Popular Resources' and 'NCBI Announcements'.

#### 2. Examine your search results

- How many records have you retrieved in total? \_\_\_\_\_
- Note the **Entry type** filter on the left side of the screen.  
Q: What type of entries do you see? \_\_\_\_\_

NCBI Resources How To

GEO DataSets GEO DataSets GSE33253

Create alert Advanced

Summary 20 per page Sort by Default order Send to:

**Entry type**  
 DataSets (0)  
 Series (1)  
 Samples (4)  
 Platforms (1)

Organism  
 Customize ...

Study type  
 Expression profiling by array  
 Methylation profiling by array  
 Customize ...

Author  
 Customize ...

Attribute name  
 tissue (0)  
 strain (0)  
 Customize ...

Publication dates  
 30 days  
 1 year  
 Custom range...

[Clear all](#)  
[Show additional filters](#)

[Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  signaling](#)  
 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....  
 Species: Mus musculus Type: Expression profiling by array  
 Dataset: GSE33253  
[PubMed](#)

**Search results**

Items: 6

[Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  signaling](#)

1. [Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  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- $\alpha$ , 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- $\alpha$  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](#) [Analyze with GEO2R](#)

[\[Mouse430\\_2\] Affymetrix Mouse Genome 430 2.0 Array](#)

2. (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](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

### 3. Examine individual records

#### a) Series

- Use the **Entry type** filter on the left side of the screen. Click on **Series** to select the series record that describes the overall design of the study.  
 Q: Looking at the summary of the series record, can you confirm that the technology used in the study was indeed **Expression profiling by array**? \_\_\_\_\_
- Click on the series record title to open up and examine the record itself.  
 Q: Would you agree that the series record describes the overall design of the study? \_\_\_\_\_

## b) Platform

- Go back in your browser to the search page and deselect **Series** and select **Platforms**.

Q: Looking at the summary of the Platform record, what is the Platform Accession number (Hint: it is listed at the bottom of the summary)?  
\_\_\_\_\_

- Click on the Platform record title to examine the record itself.

Q: Is this a commercial array or a custom-made array? \_\_\_\_\_

Q: Who is the manufacturer of the array? \_\_\_\_\_

Q: Was this array used in other studies and how many?  
(Hint: scroll down the page to find how many series are listed for the platform. For this purpose you can assume 1 series = 1 study but be aware that it is a bit more complex as there are super series that include several series.) \_\_\_\_\_

- Scroll down the Platform record (past **Relations** that indicate that this particular array relates to many other arrays) to access the **Data table**:

<b>Gene Title</b>	best associated with the transcribed region being interrogated by the probe set. Refer to the "Sequence Source" field to determine the database used.						
<b>Gene Symbol</b>	Title of Gene represented by the probe set.						
<b>ENTREZ_GENE_ID</b>	A gene symbol, when one is available (from UniGene).						
<b>RefSeq Transcript ID</b>	Entrez Gene Database UID						
<b>Gene Ontology Biological Process</b>	References to multiple sequences in RefSeq. The field contains the ID and Description for each entry, and there can be multiple entries per ProbeSet.						
<b>Gene Ontology Cellular Component</b>	Gene Ontology Consortium Biological Process derived from LocusLink. Each annotation consists of three parts: "Accession Number // Description // Evidence". The description corresponds directly to the GO ID. The evidence can be "direct", or "extended".						
<b>Gene Ontology Molecular Function</b>	Gene Ontology Consortium Cellular Component derived from LocusLink. Each annotation consists of three parts: "Accession Number // Description // Evidence". The description corresponds directly to the GO ID. The evidence can be "direct", or "extended".						
<b>Data table</b>							
ID	GB_ACC	SPOT_ID	Species Scientific Name	Annotation Date	Sequence Type	Sequence Source	Target Description
<a href="#">1415670_at</a>	<a href="#">BC024686</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:BC024686.1 /DB_XREF=gi:19354080 /FEA=FLmRNA
<a href="#">1415671_at</a>	<a href="#">NM_013477</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_013477.1 /DB_XREF=gi:7304908 /GEN=Atp6v0
<a href="#">1415672_at</a>	<a href="#">NM_020585</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_020585.1 /DB_XREF=gi:10181207 /GEN=AB04:
<a href="#">1415673_at</a>	<a href="#">NM_133900</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_133900.1 /DB_XREF=gi:19527115 /GEN=AI48C
<a href="#">1415674_a_at</a>	<a href="#">NM_021789</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_021789.1 /DB_XREF=gi:11140824 /GEN=Sbdn
<a href="#">1415675_at</a>	<a href="#">BC008256</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:BC008256.1 /DB_XREF=gi:14198389 /FEA=FLmRNA
<a href="#">1415676_a_at</a>	<a href="#">NM_011186</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_011186.1 /DB_XREF=gi:6755203 /GEN=Psmb5
<a href="#">1415677_at</a>	<a href="#">NM_026819</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_026819.1 /DB_XREF=gi:13386167 /GEN=1110C
<a href="#">1415678_at</a>	<a href="#">BC008595</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:BC008595.1 /DB_XREF=gi:14250330 /FEA=FLmRNA
<a href="#">1415679_at</a>	<a href="#">NM_025498</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_025498.1 /DB_XREF=gi:13384921 /GEN=1700C
<a href="#">1415680_at</a>	<a href="#">NM_008569</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_008569.1 /DB_XREF=gi:6678833 /GEN=Mcpr /I
<a href="#">1415681_at</a>	<a href="#">NM_053164</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_053164.1 /DB_XREF=gi:17298675 /GEN=Mrpl4
<a href="#">1415682_at</a>	<a href="#">NM_023045</a>		Mus musculus	Jun 9, 2011	Consensus sequence	GenBank	gb:NM_023045.1 /DB_XREF=gi:12746421 /GEN=Ranp4

- The main purpose of the table is to correspond the microarray identifiers with the gene information (IDs, names, ontology, ...) . Each column in the data table contains information as described in the **Data table header descriptions**.
  - Focus on the first two columns. The first column provides the microarray **ID** and the second column tells the accession number of the sequence that was used to design the oligos attached to the spot.  
Q: Can you recognize which of the **GB\_ACC** in the second column are the NCBI reference accessions and which one are for primary GenBank records?  
NCBI reference sequence accession example \_\_\_\_\_  
Primary GenBank record accession example \_\_\_\_\_
  - Scrolling along the table, find the Gene Symbol and ENTREZ\_GENE\_ID columns.
- A partial table that corresponds microarray IDs with sequences and genes is copied here (you can refer to this table when working with GEO2R):

Data table

ID	GB_ACC	Gene Symbol	ENTREZ_GENE_ID
1415670_at	BC024686	Copg	54161
1415671_at	NM_013477	Atp6v0d1	11972
1415672_at	NM_020585	Golga7	57437
1415673_at	NM_133900	Psph	100678
1415674_a_at	NM_021789	Trappc4	60409
1415675_at	BC008256	Dpm2	13481
1415676_a_at	NM_011186	Psmb5	19173
1415677_at	NM_026819	Dhrs1	52585
1415678_at	BC008595	Ppm1a	19042
1415679_at	NM_025498	Psenen	66340
1415680_at	NM_008569	Anapc1	17222
1415681_at	NM_053164	Mrpl43	94067
1415682_at	NM_023045	Xpo7	65246
1415683_at	BC016526	Nmt1	18107
1415684_at	AV168389	Atg5	11793
1415685_at	NM_133767	Mtif2	76784
1415686_at	AV339290	Rab14	68365
1415687_a_at	BM212050	Psap	19156
1415688_at	NM_025985	Ube2g1	67128
1415689_s_at	BC007473	Zkscan3	72739

- Back on the web, check the number of rows that are listed for the table.  
Q: If you consider that about 80 rows in the table are dedicated to descriptors and control, approximately how many genes are assayed on this array? \_\_\_\_\_
- (To find information for your gene of interest you would need to download and parse the full table. )

### c) Samples

- From the Platform record, go back in the browser to access the search results and uncheck Platforms and check **Samples**.  
Q: Each of the samples have its own accession (at the bottom of each of the summary. What three letters are used in the sample accession format?  
\_\_\_\_\_
- To check one of the samples click, for example, on the **Wild-type-2** title to access the information in the record. Confirm that you are able to find the growth- and sample preparation protocols in the record.
- Scroll to the bottom of the record to the **Data table** and note that the table is available for download.
- Click on the **View full table...** button. In most browsers, the table should open in a new tab. Scroll past the “AFFX-...” entries that are various controls, until you come to the spot IDs that you will recognize because you have just seen them in the Platform table.  
  
Q: What was the expression value for the Dpm2 gene? \_\_\_\_\_  
(Hint: Help yourself with the simplified platform table on page 4 of this document or the platform record on the web)
- Q: Is the Dpm gene expression value for this sample higher or lower than those for the preceding spots/genes in the table? \_\_\_\_\_

## 4. Gene expression analysis on demand

### (Guided practice)

- **GSE33253** is not a curated study (you will learn about GEO curation shortly), but you can still analyze the data on demand by using the **GEO2R** web analysis tool. The **GSE33253** series record provides the **Analyze with GEO2R** link to the tool.

**Items: 6**


**Filters activated:** Samples, Platforms, Series. [Clear all](#) to show 6 items.

[Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  signaling](#)

1. [Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  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- $\alpha$ , 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- $\alpha$  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](#) [Analyze with GEO2R](#)



- **Define groups:** Note that this study only have two treatments, so you do not have options. In other studies with several treatments/samples, GEO2R allows you to define your own groups and select samples for comparison.

CBI » GEO » GEO2R » GSE33253

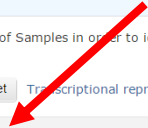
Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession:  Set [Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  signaling](#)

**Samples** Define groups Selected 0 out of 4 samples

Enter a group name:  List

Group	Accession	Title	Source name	Background strain	Genotype/variation	Tumor type	Cell type
-	GSM822870	Knockout-1	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in TNFR 1, 2 -/- mouse	C57BL/6	TNFR 1, 2 -/-	B16F1 melanoma	tumor endothelial cells
-	GSM822871	Knockout-2	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in TNFR 1, 2 -/- mouse	C57BL/6	TNFR 1, 2 -/-	B16F1 melanoma	tumor endothelial cells
-	GSM822872	Wild-type-1	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in C57BL/6 mouse	C57BL/6	wild type	B16F1 melanoma	tumor endothelial cells
-	GSM822873	Wild-type-2	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in C57BL/6 mouse	C57BL/6	wild type	B16F1 melanoma	tumor endothelial cells



CBI » GEO » GEO2R » GSE33253

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

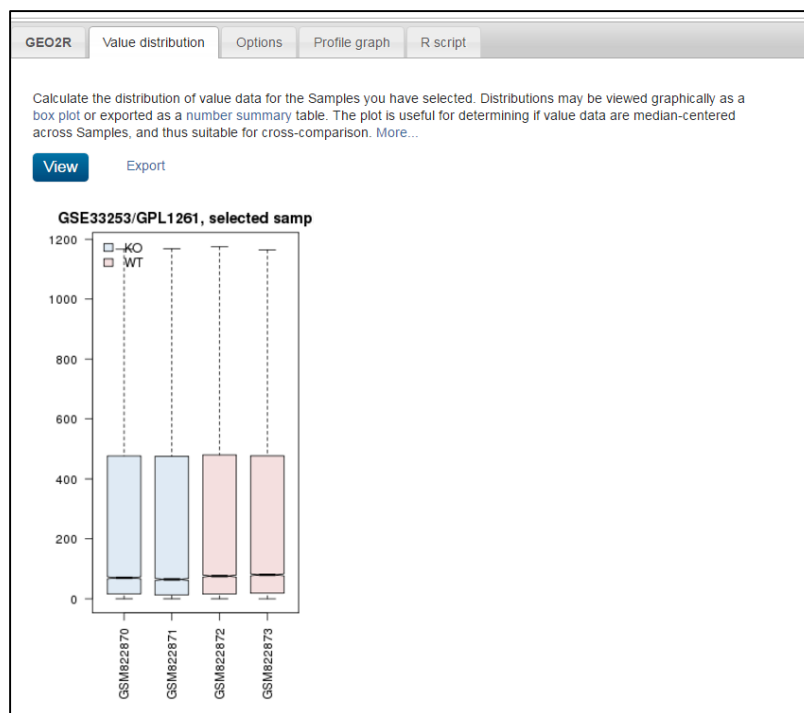
GEO accession:  Set [Transcriptional reprogramming of tumor-associated endothelial cells by disruption of TNF- \$\alpha\$  signaling](#)

**Samples** Define groups Selected 4 out of 4 samples

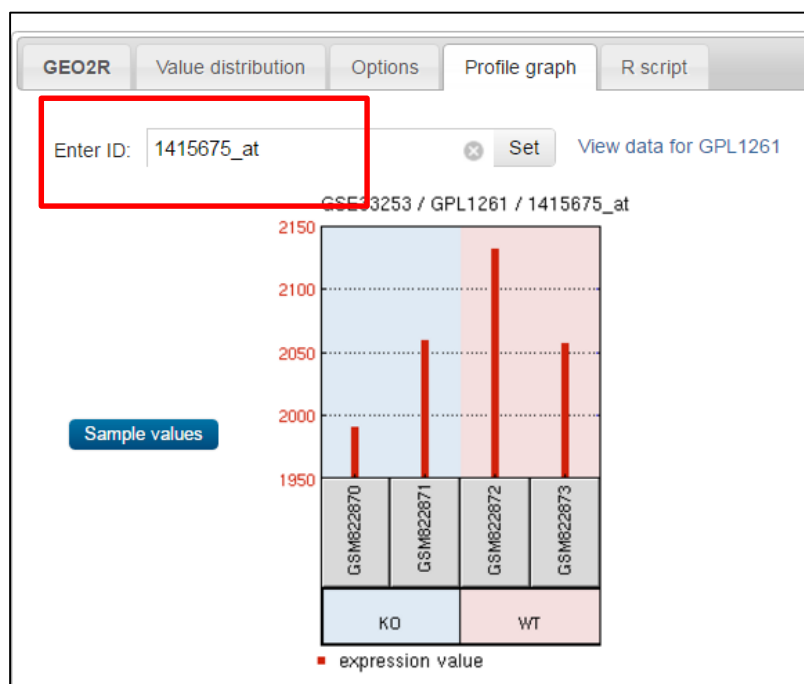
Group	Accession	Title	Source name	Background strain	Genotype/variation	Tumor type	Cell type
KO	GSM822870	Knockout-1	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in TNFR 1, 2 -/- mouse	C57BL/6	TNFR 1, 2 -/-	B16F1 melanoma	tumor endothelial cells
KO	GSM822871	Knockout-2	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in TNFR 1, 2 -/- mouse	C57BL/6	TNFR 1, 2 -/-	B16F1 melanoma	tumor endothelial cells
WT	GSM822872	Wild-type-1	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in C57BL/6 mouse	C57BL/6	wild type	B16F1 melanoma	tumor endothelial cells
WT	GSM822873	Wild-type-2	Tumor-associated endothelial cells from B16F1 melanoma tumor grown in C57BL/6 mouse	C57BL/6	wild type	B16F1 melanoma	tumor endothelial cells

If you are going to use GEO2R in your research, please read Full Instructions!

- **View Value distribution:** This tells you if the samples are centered at the median.



- Check **Options** (refer to Full Instructions)
- Find the Profile graph for the Dpm2 (1415675\_at) gene:



## – Run GEO2R:

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. [Full instructions](#) [YouTube](#)

GEO accession  Set

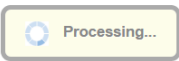
▶ Samples ▶ Define groups Selected 4 out of 4 samples

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, source and characteristics) columns to help determine which Samples belong to which 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 complete results table may be saved.
- You may change settings in Options tab.

How to use



- On the results page note that Log-transformation has been applied to the data.
- **Select columns** so that you display the gene symbol and the GO. Function and GO. Process columns

GEO2R Value distribution Options Profile graph R script

▶ Quick start

Log-transformation has been applied to the data. You can change this in the Options tab.

Recalculate if you changed any options.

ID	Gene.symbol	Gene.ID	GO.Function	GO.Process
▶ 1449025_at	Ifit3	15959	identical protein binding//protein binding	cellular response to interferon-alpha//c...
▶ 1451426_at	Dhx58	80861	ATP binding//DNA binding//RNA bindi...	defense response to virus//immune sy...
▶ 1449009_at	Tgtp2//Tgtp1	100039796//21822	molecular_function//GTP binding//GT...	biological_process//GTP catabolic pro...
▶ 1449963_at	Krtap9-3	75586	molecular_function	biological_process
▶ 1435906_x_at	Gbp2	14469	GTP binding//GTPase activity//nucleo...	GTP catabolic process//adhesion of sy...
▶ 1417961_a_at	Trim30a	20128	DNA binding//metal ion binding//zinc ...	negative regulation of NLRP3 inflammm...
▶ 1431591_s_at	Isg15//Gm9706	100038882//677168	protein binding//protein tag	ISG15-protein conjugation//ISG15-pro...
▶ 1448591_at	Ctss	13040	collagen binding//cysteine-type endop...	basement membrane disassembly//ba...
▶ 1450783_at	Ifit1	15957	RNA binding//protein binding	cellular response to interferon-alpha//c...
▶ 1418191_at	Usp18	24110	ISG15-specific protease activity//cyste...	ISG15-protein conjugation//proteolysis...
▶ 1447927_at	Gbp10//Gbp6	626578//100702	molecular_function//molecular_function	cellular response to interferon-gamma/...



**(Independent practice)**

- Many of the top five differentially expressed genes (skip Krtap9-3 for which GO annotation is missing) have at least a couple of things in common.  
Q: Is the expression of the top five genes higher in KO or WT? What is the common function of these genes (the GO information is copied below because it is hard to read from the web)\_\_\_\_\_

**"1449025\_at"      "Ifit3"      "15959"**

"identical protein binding///protein binding"  
"cellular response to interferon-alpha///cellular response to interferon-beta///defense response to virus///immune system process///innate immune response"

**"1451426\_at"      "Dhx58"      "80861"**

"ATP binding///DNA binding///RNA binding///double-stranded RNA binding///double-stranded RNA binding///helicase activity///hydrolase activity///hydrolase activity, acting on acid anhydrides///metal ion binding///nucleotide binding///single-stranded RNA binding///zinc ion binding"

"defense response to virus///immune system process///innate immune response///negative regulation of MDA-5 signaling pathway///negative regulation of RIG-I signaling pathway///negative regulation of RIG-I signaling pathway///negative regulation of innate immune response///negative regulation of innate immune response///negative regulation of type I interferon production///negative regulation of type I interferon production///positive regulation of MDA-5 signaling pathway///positive regulation of RIG-I signaling pathway///positive regulation of type I interferon production///regulation of innate immune response///response to virus"

**"1449009\_at"      "Tgtp2///Tgtp1"      "100039796///21822"**

"molecular\_function///GTP binding///GTPase activity"  
"biological\_process///GTP catabolic process///cellular response to interferon-beta///response to interferon-alpha///response to interferon-gamma"

**"1449963\_at"      "Krtap9-3"      "75586"**

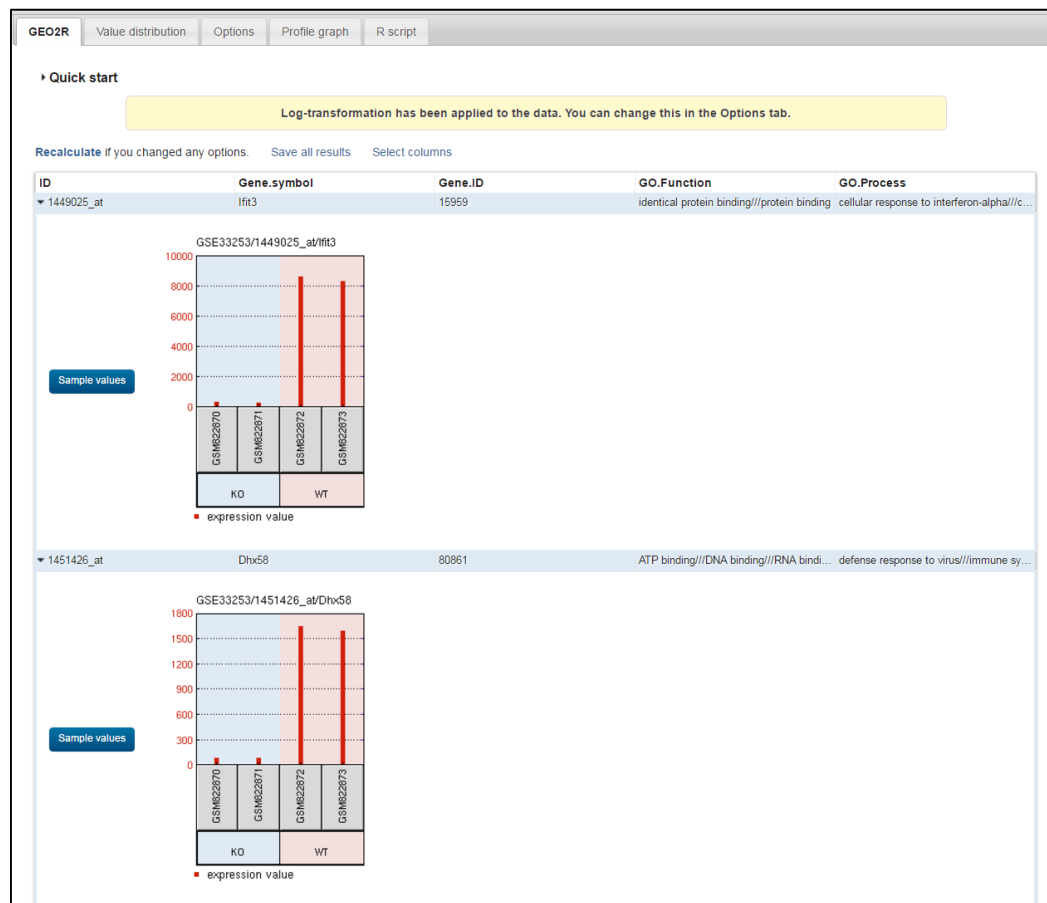
"molecular\_function"  
"biological\_process"

"1435906\_x\_at" "Gbp2" "14469"  
 "GTP binding///GTPase activity///nucleotide binding"

"GTP catabolic process///adhesion of symbiont to host///cellular response to interferon-beta///cellular response to interferon-gamma///cellular response to lipopolysaccharide///defense response to Gram-positive bacterium///defense response to protozoan"

"1417961\_a\_at" "Trim30a" "20128"  
 "DNA binding///metal ion binding///zinc ion binding"

"negative regulation of NLRP3 inflammasome complex assembly///negative regulation of interleukin-6 production///negative regulation of reactive oxygen species metabolic process///negative regulation of toll-like receptor signaling pathway///negative regulation of tumor necrosis factor production///positive regulation of protein catabolic process///positive regulation of viral entry into host cell///protein autoubiquitination///regulation of transcription, DNA-templated///transcription, DNA-templated"



## Exercise 2: Explore a Curated GEO DataSets record: Analysis Tools, Individual Gene Profiles, and the Gene- and Biosystems databases

### (Guided practice)

1. Search for GSE42872 and examine your search results:

The screenshot shows the NCBI GEO DataSets search results for GSE42872. The search filters on the left include:

- Entry type** (highlighted with a red box): DataSets (1), Series (1), Samples (6), Platforms (1)
- Organism**: Customize ...
- Study type**: Expression profiling by array, Expression profiling by high throughput sequencing, Customize ...
- Author**: Customize ...
- Attribute name**: tissue (0), strain (0), Customize ...
- Publication dates**: 30 days, 1 year, Custom range...

The search results section shows 9 items. The first two items are highlighted:

- A375 cells 24h Vemurafenib rep3**  
 1. Organism: Homo sapiens  
 Source name: A375 melanoma cells treated with vemurafenib  
 Platform: GPL6244 Series: **GSE42872** Dataset: GDS5085  
 Download data: GEO (CEL)  
 Sample Accession: GSM1052620 ID: 301052620
- A375 cells 24h Vemurafenib rep2**  
 2. Organism: Homo sapiens  
 Source name: A375 melanoma cells treated with vemurafenib  
 Platform: GPL6244 Series: **GSE42872** Dataset: GDS5085  
 Download data: GEO (CEL)  
 Sample Accession: GSM1052619 ID: 301052619

The summary box contains the following text:

[Expression data from BRAFV600E A375 melanoma cells treated with vehicle or vemurafenib](#)  
 Vemurafenib is a BRAF inhibitor with specificity for the most common BRAF mutant encountered in melanomas (BRAFV600E). Vemurafenib suppresses the proliferation of BRAF mutant human melanoma cells by suppressing downstream activation of the MEK/ERK...  
 Species: Homo sapiens Type: Expression profiling by array  
 Dataset: GSE42872  
[PubMed](#)

- Clear any of the set filters from the previous search.
- In addition to the submitted records (series, sample, and platform) you will also retrieve a **DataSet** record that was generated as a product of GEO curation.
- The study included six samples. Each of the sample was the A375 cell line that endogenously expresses a mutant BRAF protein. The mutation is designated as BRAF(V600E), meaning that the gene is mutated so that valine at position 600 of the protein expressed by the BRAF gene is replaced by glutamic acid. BRAF(V600E) is an oncogenic mutation: <http://www.ncbi.nlm.nih.gov/clinvar/variation/13961/>

- In the study, three of the samples were treated with vemurafenib (<https://pubchem.ncbi.nlm.nih.gov/compound/42611257>) that is an inhibitor of the BRAF(V600E) kinase. (Control in this case means that the mutant kinase stays active in the cell).
- Select the DataSets record:
  - Check the summary display (do not go to the record just yet) and note the accession number of the curated dataset record: **GDS5085**

GEO DataSets GEO DataSets ▼ GSE42872

[Create alert](#) [Advanced](#)

---

Entry type clear Summary ▼ Send to: ▼

✓ DataSets (1)

Series (0)

Samples (0)

Platforms (0)

Organism

Customize ...

Study type

Expression profiling by array

Expression profiling by high throughput sequencing

Customize ...

Author

Customize ...

**Filters activated:** DataSets. [Clear all](#) to show 9 items.

[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

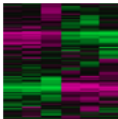
Type: Expression profiling by array, transformed count, 2 agent sets

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](#)



## 2. Analysis of a curated set

- Click on the title of the curated study to access and examine the record:
- The **Sample Subset** button displays the sample information.
- **Data Analysis Tools** allows for visualizing clusters of genes with similar expression, and find up and down regulated genes.
- The **Expression Profiles** button leads in the **Geo Profiles** database for either all of the individual-gene profiles in the study, or those that are up/down in a sample comparison.

NCBI DATASET BROWSER

Search for

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

**Sample Subsets**

Samples	Factors agent	Title
GSM1052615	vehicle	A375 cells 24h Control rep1
GSM1052616	vehicle	A375 cells 24h Control rep2
GSM1052617	vehicle	A375 cells 24h Control rep3
GSM1052618	vemurafenib	A375 cells 24h Vemurafenib rep1
GSM1052619	vemurafenib	A375 cells 24h Vemurafenib rep2
GSM1052620	vemurafenib	A375 cells 24h Vemurafenib rep3

- This study has numerous differentially expressed genes, and we will select only those where there is a four-fold difference in expression between “vehicle” (the BRAF V600E kinase is active) and vemurafenib (the BRAF V600E kinase is inhibited):
  - Use the **Compare 2 sets of samples** widget, and from there **Value means difference** and set it to **4+ fold**. In the first round select **higher**.
  - Select “vehicle” as group A by clicking on all of the samples for vehicle. Select all of the vemurafenib samples as Group B and **OK**.

Click on accessions to select samples individually, click on colored blocks and then on blinking arrows to select groups of samples.

Samples, Group A	Factors agent	Samples, Group B
GSM1052615	vehicle	GSM1052615
GSM1052616		GSM1052616
GSM1052617		GSM1052617
GSM1052618	vemurafenib	GSM1052618
GSM1052619		GSM1052619
GSM1052620		GSM1052620

- Query Group A vs. B.
- The list of genes that have four fold (or more) higher expression in the cells with active BRAF mutant will be displayed in the GEO Profiles database:

NCBI Resources How To

GEO Profiles GEO Profiles Advanced

Gene symbol Summary 20 per page Sort by Default order Send to:

Gene keyword **Selected items**

Organism **Items: 13**

Gene ontology

Differential expression

DataSet keyword

GEO accession


Clear all

Show additional filters

1. [DTL - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition](#)  
 Annotation: DTL, denticleless E3 ubiquitin protein ligase homolog (Drosophila)  
 Organism: Homo sapiens  
 Reporter: GPL6244, 7909568 (ID\_REF), GDS5085, NM\_016448, AK292343, chr1:212208919-212278348 (SPOT ID)  
 DataSet type: Expression profiling by array, transformed count, 6 samples  
 ID: 112029165  
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Homologene neighbors](#)

2. [FAM111B - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition](#)  
 Annotation: FAM111B, family with sequence similarity 111, member B  
 Organism: Homo sapiens  
 Reporter: GPL6244, 7940147 (ID\_REF), GDS5085, NM\_198947, NM\_001142703, NM\_001142704, AY457926, chr11:58874710-58894867 (SPOT ID)  
 DataSet type: Expression profiling by array, transformed count, 6 samples  
 ID: 112032232  
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Homologene neighbors](#)

3. [DUSP6 - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition](#)  
 Annotation: DUSP6, dual-specific phosphatase 6



- Check the profile of the DTL gene and its expression values. From there you can link to the individual sample information that states that the values are given as log base 2
- A web-based log calculator (for example <http://www.1728.org/logrithm.htm>) will help you get the non-log values; where 12.1 is 4,400 and 8.9 is about 500.
- Find the biological pathways in which the 13 genes participate:

How To grafianm My NCBI Sign Out

GEO Profiles   Help

Advanced

Summary 20 per page Sort by Default order Send to: **Filters:** [Manage Filters](#)

**Selected items**

Items: 13

[DTL - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition](#)

1. [inhibition](#)

Annotation: DTL, denticleless E3 ubiquitin protein ligase homolog (Drosophila)  
 Organism: Homo sapiens  
 Reporter: GPL6244, 7909568 (ID\_REF), GDS5085, NM\_016448, AK292343, chr1:212208919-212278348 (SPOT ID)  
 DataSet type: Expression profiling by array, transformed count, 6 samples  
 ID: 112029165

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Homologene neighbors](#)

**Profile data**

**Profile pathways**

**Find related data**

Database:

- The **FLink (Frequency weighted Links)** results indicate that six of the 13 genes are involved in the Signal Transduction pathway, BSID 1269379. The biological pathways in the NCBI **BioSystems** database are imported from external resources, in this case REACTOME. (If you open the BSID 1269379 record, you can get the pathway description. There are over 2500 participating genes, including BRAF and TNF.)

NCBI **FLink**

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains PubChem BioSystems

**FLink - Frequency weighted Links**

[Links from geoprofiles records to biosystems records weighted by frequency \(click to see details\)](#)

Gene  BioSystems

<input type="checkbox"/>	Frequency	BSID	Source	Name
<input checked="" type="checkbox"/>	6	1269379	REACTOME	Signal Transduction
<input type="checkbox"/>	3	1269501	REACTOME	MAPK family signaling cascades
<input type="checkbox"/>	3	1269419	REACTOME	Signaling by FGFR4
<input type="checkbox"/>	3	1269409	REACTOME	Signaling by FGFR3
<input type="checkbox"/>	3	1269399	REACTOME	Signaling by FGFR2
<input type="checkbox"/>	3	1269387	REACTOME	Signaling by FGFR1
<input type="checkbox"/>	3	1269386	REACTOME	Signaling by FGFR
<input type="checkbox"/>	3	1269380	REACTOME	Signaling by EGFR
<input type="checkbox"/>	3	523016	KEGG	Transcriptional misregulation in cancer
<input type="checkbox"/>	3	522987	KEGG	Transcriptional misregulation in cancer

- To see what are the six genes, and to learn more about their function, click on the **Frequency** number to link to the records in the Gene database:

How To ☑

Gene

Advanced

Tabular ▾ 20 per page ▾ Sort by Relevance ▾ Send to: ▾

**Selected items**

Items: 6

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">DUSP6</a> ID: 1848	dual specificity phosphatase 6 [ <i>Homo sapiens</i> (human)]	Chromosome 12, NC_000012.12 (89347825..89352859, complement)	HH19, MKP3, PYST1	602748
<input type="checkbox"/> <a href="#">ETV4</a> ID: 2118	ETS variant 4 [ <i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (43527843..43546432, complement)	E1A-F, E1AF, PEA3, PEAS3	600711
<input type="checkbox"/> <a href="#">ITGB3</a> ID: 3690	integrin subunit beta 3 [ <i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (47253842..47312711)	BDPLT16, BDPLT2, CD61, GP3A, GPIIIa, GT	173470
<input type="checkbox"/> <a href="#">LDLR</a> ID: 3949	low density lipoprotein receptor [ <i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (11089362..11133830)	FH, FHC, LDLCQ2	606945
<input type="checkbox"/> <a href="#">SPRY2</a> ID: 10253	sprouty RTK signaling antagonist 2 [ <i>Homo sapiens</i> (human)]	Chromosome 13, NC_000013.11 (80335976..80340951, complement)	IGAN3, hSPRY2	602466
<input type="checkbox"/> <a href="#">FST</a> ID: 10468	follistatin [ <i>Homo sapiens</i> (human)]	Chromosome 5, NC_000005.10 (53480341..53487134)	FS	136470

Tabular ▾ 20 per page ▾ Sort by Relevance ▾ Send to: ▾

- You can display the summary of the records as text and copy the information to be used in Part II of this module:

## Genes with higher expression when mutant BRAF is active

### 1. DUSP6

Official Symbol: DUSP6 and Name: dual specificity phosphatase 6 [*Homo sapiens* (human)]

Other Aliases: HH19, MKP3, PYST1

Other Designations: MAP kinase phosphatase 3; dual specificity protein phosphatase PYST1; mitogen-activated protein kinase phosphatase 3; serine/threonine specific protein phosphatase

Chromosome: 12; Location: 12q21.33

Annotation: Chromosome 12 NC\_000012.12 (89347825..89352859, complement)

MIM: 602748

ID: 1848

### 2. ETV4

Official Symbol: ETV4 and Name: ETS variant 4 [*Homo sapiens* (human)]

Other Aliases: E1A-F, E1AF, PEA3, PEAS3



Other Designations: EWS protein/E1A enhancer binding protein chimera; adenovirus E1A enhancer-binding protein; ets variant gene 4 (E1A enhancer-binding protein, E1AF); polyomavirus enhancer activator 3 homolog  
Chromosome: 17; Location: 17q21.31  
Annotation: Chromosome 17 NC\_000017.11 (43527843..43546432, complement)  
MIM: 600711  
ID: 2118

### 3. ITGB3

Official Symbol: ITGB3 and Name: integrin subunit beta 3 [Homo sapiens (human)]  
Other Aliases: BDPLT16, BDPLT2, CD61, GP3A, GPIIIa, GT  
Other Designations: antigen CD61; integrin beta 3; integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61); platelet membrane glycoprotein IIIa  
Chromosome: 17; Location: 17q21.32  
Annotation: Chromosome 17 NC\_000017.11 (47253842..47312711)  
MIM: 173470  
ID: 3690

### 4. LDLR

Official Symbol: LDLR and Name: low density lipoprotein receptor [Homo sapiens (human)]  
Other Aliases: FH, FHC, LDLCQ2  
Other Designations: LDL receptor; low-density lipoprotein receptor class A domain-containing protein 3  
Chromosome: 19; Location: 19p13.2  
Annotation: Chromosome 19 NC\_000019.10 (11089362..11133830)  
MIM: 606945  
ID: 3949

### 5. SPRY2

Official Symbol: SPRY2 and Name: sprouty RTK signaling antagonist 2 [Homo sapiens (human)]  
Other Aliases: IGAN3, hSPRY2  
Chromosome: 13; Location: 13q31.1  
Annotation: Chromosome 13 NC\_000013.11 (80335976..80340951, complement)  
MIM: 602466  
ID: 10253

### 6. FST

Official Symbol: FST and Name: follistatin [Homo sapiens (human)]  
Other Aliases: FS  
Other Designations: activin-binding protein; follistatin isoform FST317  
Chromosome: 5; Location: 5q11.2  
Annotation: Chromosome 5 NC\_000005.10 (53480341..53487134)  
MIM: 136470  
ID: 10468

## (Independent practice)

- Return to the DataSet Browser for the GDS5085 record and this time select for those genes with 4+ fold lower expression in vehicle (the BRAF V600E kinase is active):

The screenshot shows the NCBI GEO DataSet Browser interface for GDS5085. The title is 'Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition'. The summary describes the analysis of A375 melanoma cells. The organism is *Homo sapiens*. The platform is GPL6244: [HuGene-1\_0-st] Affymetrix Human Gene 1.0 ST Array. The citation is 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. The reference series is GSE42872, with a sample count of 6. The value type is 'transformed count' and the series published date is 2014/05/20. The 'Data Analysis Tools' section is active, showing a configuration for a query: 'Value means difference' set to '4+ fold' and 'higher', with 'Group A' selected as 'lower'. The selected samples for Group A are GSM1052615, GSM1052616, and GSM1052618. The selected samples for Group B are GSM1052618, GSM1052619, and GSM1052620. The query is 'Query Group A vs. B'.

- One of the genes that you should have retrieved is TCN1.  
Q: How does this gene rank among all of the genes that were expressed on the platform? How much does the rank changes when the mutant BRAF is not active?
- From the FLink results, select the three genes that participate in the Metabolism (BSID 1269956)

## Gene with higher expression when mutant BRAF is inhibited

1. CD36

Official Symbol: CD36 and Name: CD36 molecule [Homo sapiens (human)]

Other Aliases: BDPLT10, CHDS7, FAT, GP3B, GP4, GPIV, PASIV, SCARB3

Other Designations: CD36 antigen (collagen type I receptor, thrombospondin receptor); CD36 molecule (thrombospondin receptor); GPIIIB; PAS IV; PAS-4 protein; cluster determinant 36; fatty acid translocase; glycoprotein IIIb; leukocyte differentiation antigen CD36; platelet glycoprotein IV; scavenger receptor class B, member 3

Chromosome: 7; Location: 7q11.2

Annotation: Chromosome 7 NC\_000007.14 (80602188..80679277)

MIM: 173510

ID: 948

## 2. DCT

Official Symbol: DCT and Name: dopachrome tautomerase [Homo sapiens (human)]  
Other Aliases: TRP-2, TYRP2  
Other Designations: DT; L-dopachrome Delta-isomerase; L-dopachrome isomerase;  
TRP2; dopachrome delta-isomerase; dopachrome tautomerase (dopachrome delta-  
isomerase, tyrosine-related protein 2); tyrosinase related protein-2;  
tyrosinase-related protein 2; tyrosine-related protein 2  
Chromosome: 13; Location: 13q32  
Annotation: Chromosome 13 NC\_000013.11 (94437304..94550265, complement)  
MIM: 191275  
ID: 1638

## 3. TCN1

Official Symbol: TCN1 and Name: transcobalamin 1 [Homo sapiens (human)]  
Other Aliases: HC, TC-1, TC1, TCI  
Other Designations: haptocorin; haptocorrin; protein R; transcobalamin I  
(vitamin B12 binding protein, R binder family)  
Chromosome: 11; Location: 11q11-q12  
Annotation: Chromosome 11 NC\_000011.10 (59852808..59866568, complement)  
MIM: 189905  
ID: 6947

## Exercise 3: The contents of the GEO DataSets database and RNA-Seq studies

### (Guided practice)

#### 1. Study (technology) types in GEO database

- Return to the GEO DataSets search page and check the total number of records in the database:

NCBI Resources How To

GEO DataSets GEO DataSets all[filter] Create alert Advanced

Entry type  
DataSets (3,848)  
Series (69,045)  
Samples (1,811,664)  
Platforms (15,899)

Organism  
Customize ...

Study type  
Expression profiling by array  
Methylation profiling by array  
Customize ...

Author  
Customize ...

Attribute name  
tissue (649,009)  
strain (294,350)  
Customize ...

Publication dates  
30 days  
1 year  
Custom range...

[Clear all](#)  
[Show additional filters](#)

Summary 20 per page Sort by Number of Samples (Low to High) Send to:

**Search results**  
Items: 1 to 20 of 1900726 << First < Prev Page 1 of 95037 Next > Last >>

[Identification of the long, edited dsRNAome in LPS-stimulated primary human peripheral blood monocytes](#)  
1. [monocytes](#)  
(Submitter supplied) Endogenous double-stranded RNA (dsRNA) is intricately regulated in mammals to prevent aberrant activation of host inflammatory pathways by cytosolic dsRNA binding proteins. We define the endogenous dsRNA repertoire in human peripheral blood monocytes using published data (GSE60216) derived from 18 RNA-Seq experiments from neonates, adults, and elderly patients during the inflammatory response to bacterial lipopolysaccharide. [more...](#)  
Organism: Homo sapiens  
Type: Third-party reanalysis; Expression profiling by high throughput sequencing  
Download data: GEO (BW, TXT)  
Series Accession: GSE75149 ID: 200075149  
[PubMed](#) [Similar studies](#)

[Transcriptomic analysis of pigs exhibiting differential susceptibility to swine influenza \(SIV\) pathology in the lung](#)  
2. [pathology in the lung](#)  
(Submitter supplied) Comparison of gene expression in the lungs of pigs classified as Resistant (RES) or Susceptible (SUS) to influenza pathology based on clinical lesion scores. The aim of the experiment is to identify genes whose expression is associated with resistance/susceptibility to influenza pathology. These are a subset of animals selected from a larger experiment that investigated the effect of low (LBW) or high (HBW) litter birth weight phenotype on influenza pathology.  
Organism: Sus scrofa

- Search for studies that relate to melanoma:

The screenshot displays the GEO DataSets search interface. At the top, the search term 'melanoma' is entered in the search bar. The search results show 24339 items. A 'Study type' filter is open, showing 'Expression profiling by array' and 'Expression profiling by high throughput sequencing' selected. The search results list several studies, including 'Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition' and 'melanoma cell line'. The 'Filter your results' section shows 'All (24339)' selected, with options for 'track (34)' and 'gds pubmed (595)'. The 'Recent activity' section shows 'melanoma (24339)' and 'melanoma AND BRAF AND "has track" [properties] (5)'.

- Customize the left-side **Study Type** filter so that it displays filtering for **“Expression profiling by array”** and also for **“Expression profiling by high throughput sequencing”** (RNA-Seq). These are two of the prevalent study types among several that are accepted to be archived in the GEO DataSet database.
- Select the “Expression profiling by high throughput sequencing”. To find out if there are any RNA-Seq studies that deal with melanoma and also specifically with the BRAF mutant, add the term BRAF to the search.
- To be able to do any visualization and examination of the RNA-Seq data on the web, you also need to select a study that “has track”:

*melanoma AND BRAF AND "has track"[properties]*

NCBI Resources How To gralianm My NCBI Sign Out

GEO DataSets GEO DataSets melanoma AND BRAF AND "has track"[properties] Search

Create alert Advanced Help

Entry type Summary Sort by Default order Send to: Filter your results:

DataSets (0) All (3)

Series (3) track (3)

Samples (0) gds pubmed (3)

Platforms (0) Manage Filters

Organism Customize ...

Study type clear

Expression profiling by array

Expression profiling by high throughput sequencing

Customize ...

Author Customize ...

Attribute name tissue (0) strain (0) Customize ...

Publication dates 30 days 1 year Custom range...

Clear all Show additional filters

**Search results**

Items: 3

1 Filters activated: Expression profiling by high throughput sequencing. [Clear all](#) to show 5 items.

1. [Wnt-signaling potentiates nevogenesis.](#)  
 (Submitter supplied) This SuperSeries is composed of the SubSeries listed below.  
 Organism: Homo sapiens  
 Type: Expression profiling by array: **Expression profiling by high throughput sequencing**  
 Platforms: GPL570 GPL10999 18 Samples  
 Download data: GEO (BIGWIG, CEL)  
 Series Accession: GSE46818 ID: 200046818  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [Analyze with GEO2R](#)

2. [Expression data from 7 Human Melanomas](#)  
 (Submitter supplied) Melanocytes within benign human nevi are the paradigm for tumor suppressive senescent cells in a pre-malignant neoplasm. These cells typically contain mutations in either the **BRAF** or N-RAS oncogene and express markers of senescence, including p16. However, a nevus can contain 10s to 100s of thousands of clonal melanocytes and approximately 20-30% of **melanoma** are thought to arise in association with a pre-existing nevus. more...  
 Organism: Homo sapiens  
 Type: **Expression profiling by high throughput sequencing**  
 Platform: GPL10999 7 Samples  
 Download data: GEO (BIGWIG), SRA SRP022260  
 Series Accession: GSE46817 ID: 200046817  
[PubMed](#) [Full text in PMC](#) [Similar studies](#)

3. [Expression data from Uninfected and BRAF600V induced melanocytes](#)  
 (Submitter supplied) Melanocytes within benign human nevi are the paradigm for tumor suppressive senescent cells in a pre-malignant neoplasm. These cells typically contain mutations in either the **BRAF** or N-RAS oncogene and express markers of senescence, including p16. However, a nevus can contain 10s to 100s of thousands of clonal melanocytes and approximately 20-30% of **melanoma** are thought to arise in association with a pre-existing nevus. more...  
 Organism: Homo sapiens  
 Type: **Expression profiling by high throughput sequencing**  
 Platform: GPL10999 2 Samples  
 Download data: GEO (BIGWIG), SRA SRP022259  
 Series Accession: GSE46805 ID: 200046805  
[PubMed](#) [Full text in PMC](#) [Similar studies](#)

Find related data Database: Select Find items

Search details ("melanoma"[MeSH Terms] OR melanoma[All Fields]) AND BRAF[All Fields] AND "has track"[properties] AND "Expression profiling by high throughput sequencing"[Filter] Search See more...

Recent activity Turn Off Clear

melanoma AND BRAF AND "has track"[properties] AND ("Expressio GEO DataSets

melanoma AND BRAF AND "has tracks"[properties] AND ("Expressio GEO DataSets

melanoma AND ("Expression profiling by high throughput sequenci GEO DataSets

melanoma AND (("Expression profiling by array"[Filter] OR "Expres... GEO DataSets

melanoma AND ("Expression profiling by array"[Filter]) (564) GEO DataSets See more...

Important Links

- Even though there are only three studies in your search results, sort the results by **Number of Samples (Low to High)**.
- The “Expression data from Uninfected and BRAF600V induced melanocytes” study (GSE46805) has two samples.

How To

GEO DataSets  [Create alert](#) [Advanced](#)

Summary  Sort by Number of Samples (Low to High)  [Send to:](#)

### Search results

Items: 4

[BRAF](#)

1. Organism: Homo sapiens  
Source name: BRAF  
Platform: GPL10999 Series: **GSE46805** GSE46818  
Download data: GEO (BIGWIG), SRA SRX276897  
Sample Accession: GSM1138581 ID: 301138581

[Uninfected](#)

2. Organism: Homo sapiens  
Source name: Uninfected  
Platform: GPL10999 Series: **GSE46805** GSE46818  
Download data: GEO (BIGWIG), SRA SRX276896  
Sample Accession: GSM1138580 ID: 301138580

[Expression data from Uninfected and BRAF600V induced melanocytes](#)

3. (Submitter supplied) Melanocytes within benign human nevi are the paradigm for tumor suppressive senescent cells in a pre-malignant neoplasm. These cells typically contain mutations in either the BRAF or N-RAS oncogene and express markers of senescence, including p16. However, a nevus can contain 10s to 100s of thousands of clonal melanocytes and approximately 20-30% of melanoma are thought to arise in association with a pre-existing nevus. [more...](#)  
Organism: Homo sapiens  
Type: Expression profiling by high throughput sequencing  
Platform: GPL10999 2 Samples  
Download data: GEO (BIGWIG), SRA SRP022259  
Series Accession: **GSE46805** ID: 200046805  
[PubMed](#) [Full text in PMC](#) [Similar studies](#)

[Illumina Genome Analyzer Ix \(Homo sapiens\)](#)

4. Organism: Homo sapiens  
680 Series 7341 Samples  
Download data: GEO  
Platform Accession: GPL10999 ID: 100010999

- In the “BRAF” sample, human epidermal melanocytes were infected with lentivirus with the BRAF mutant. (This is somewhat comparable with the “vehicle cells” in the A375 cell line as the mutant protein is expressed there as well.) The “uninfected” sample could be considered as an equivalent for BRAF suppression of the A375 cell line.
- The sequencing was on the Illumina Genome Analyzer Ix (Homo sapiens) platform.
- Access the series record and from there use the “**See the data on Genome Data Viewer**” link.

ZIP/Postal code	G61 1BD		
Country	United Kingdom		
Platforms (1)	<a href="#">GPL10999</a> Illumina Genome Analyzer IIX (Homo sapiens)		
Samples (2)	<a href="#">GSM1138580</a> Uninfected <a href="#">GSM1138581</a> BRAF		
This SubSeries is part of SuperSeries: <a href="#">GSE46818</a> Wnt-signaling potentiates neovogenesis.			
<b>Relations</b>			
BioProject	<a href="#">PRJNA202399</a>		
SRA	<a href="#">SRP022259</a>		
<a href="#">See the data on Genome Data Viewer</a>			
<b>Download family</b>	<b>Format</b>		
<a href="#">SOFT formatted family file(s)</a>	SOFT <a href="#">?</a>		
<a href="#">MINiML formatted family file(s)</a>	MINiML <a href="#">?</a>		
<a href="#">Series Matrix File(s)</a>	TXT <a href="#">?</a>		
<b>Supplementary file</b>	<b>Size</b>	<b>Download</b>	<b>File type/resource</b>
<a href="#">GSE46805_RAW.tar</a>	202.3 Mb	<a href="#">(http)(custom)</a>	TAR (of BIGWIG)
<a href="#">SRP/SRP022/SRP022259</a>		<a href="#">(ftp)</a>	SRA Study
<i>Raw data provided as supplementary file</i>			
<i>Processed data provided as supplementary file</i>			

### (Independent practice)

- Use the **Genome Data Viewer** to see if you can corroborate the significance of the differentially expressed genes from Exercise 2 in this study. The genes with the following symbols: DUSP6, ETV4, ITGB3, LDLR, SPRY2, and FST were those with much higher expression in the presence of the active BRAF mutant protein.
- Select one of the genes and search the Genome Data Viewer to locate the gene, so that you can examine its expression tracks.

Q: What is the genomic assembly to which the RNA-Seq BRAF expression data were mapped? \_\_\_\_\_

Q: Is this the current assembly for the human reference genome? \_\_\_\_\_

Q: What can you tell about the expression of the gene that you checked? \_\_\_\_\_

Q: Does the expression corroborate the findings from the array study in Exercise 2? \_\_\_\_\_



The screenshot shows the NCBI Genome Data Viewer interface. The top navigation bar includes 'NCBI Resources', 'How To', and user options like 'gratianm', 'My NCBI', and 'Sign Out'. The main title is 'Genome Data Viewer' and the current view is 'Homo sapiens: NCBI36 (GCF\_000001405.12) Chr 12 (NC\_000012.10): 88.27M - 88.27M'. The left sidebar has sections for 'Pick Assembly', 'Ideogram View' (showing chromosomes 1-22, X, Y), and 'Search'. The search input contains 'DUSP6' and a table of results is displayed below it. Two red arrows point from the search input to the first two rows of the table.

Name	Location
NP_073143.2	Chr12 88.27M - 88.27M
NP_001937.2	Chr12 88.27M - 88.27M
AB189400.1	Chr12 88.27M - 88.27M
AB013382.1	Chr12 88.27M - 88.27M

The main viewing area shows a genomic track for 'NC\_000012.10: 88M..88M (5.4Kbp)'. It includes tracks for 'CpG Islands', 'NCBI Genes' (showing the DUSP6 gene structure), and two tracks for 'GSM1138581 BRAF NA000028230.1' and 'GSM1138580 Uninfected NA000028229.1' showing expression levels.

- The genes with the following symbols: CD36, DCT, and TCN1 were those with much lower expression in the presence of the active BRAF mutant.
- Locate the TCN1 gene in the Genome Data Viewer:

Q: What can you tell about the expression of TCN1 (transcobalamin 1)? \_\_\_\_\_

## Exercise 4: Processed RNA-Seq data in the Gene database (with a quick detour to UniGene)

### (Independent practice)

1. Visualizing gene expression in normal tissues with NCBI Sequence Viewer gene expression tracks
  - From the Genome Data Viewer link to the TCN1 gene record in the Gene database.

The screenshot displays the NCBI Genome Data Viewer interface. The main window shows the Homo sapiens genome with a focus on Chromosome 11 (NC\_000011.8) in the 59,38M - 59,39M region. The search bar on the left contains 'TCN1', and the search results table below it lists several entries, including TCN1 on Chromosome 11. A pop-up window for the TCN1 gene record is open, showing the following information:

- Gene:** TCN1
- Location:** complement(59,376,857..59,390,617)
- Length:** 13,761
- Merged features:** NP\_001053.2 and NM\_001062.3
- Download:** [NP\\_001053.2](#), [NM\\_001062.3](#)
- Links & Tools:**
  - [View CCDS: CCDS7978.1](#)
  - [View GeneID: 6947 \(TCN1\)](#)
  - [View MIM: 189905](#) (indicated by a red arrow)
  - [View HGNC: 11652](#)
- GenBank View:** [NC\\_000011.8 \(59,376,857..59,390,617\)](#)
- FASTA View:** [NC\\_000011.8 \(59,376,857..59,390,617\)](#)
- BLAST Genomic:** [NC\\_000011.8 \(59,376,857..59,390,617\)](#)

Gene  [Advanced](#)

Full Report

### TCN1 transcobalamin 1 [ *Homo sapiens* (human) ]

Gene ID: 6947, updated on 8-May-2016

**Summary**

**Official Symbol** TCN1 provided by [HGNC](#)

**Official Full Name** transcobalamin 1 provided by [HGNC](#)

**Primary source** [HGNC:HGNC:11652](#)

**See related** [Ensembl:ENSG00000134827](#) [HPRD:01795](#); [MIM:189905](#); [Vega:OTTHUMG00000167400](#)

**Gene type** protein coding

**RefSeq status** REVIEWED

**Organism** [Homo sapiens](#)

**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

**Also known as** HC; TC1; TCI; TC-1

**Summary** This gene encodes a member of the vitamin B12-binding protein family. This family of proteins, alternatively referred to as R binders, is expressed in various tissues and secretions. This protein is a major constituent of secondary granules in neutrophils and facilitates the transport of cobalamin into cells. [provided by RefSeq, Jul 2008]

**Orthologs** [all](#)

- Scroll to the **Genomic context** section of the record.

Q: What is the chromosome on which the TCN1 gene is annotated? \_\_\_\_\_

Q: Is the gene annotated on the plus strand or on the reverse complement? \_\_\_\_\_

Q: Are you visualizing annotation on the current human genome reference assembly? \_\_\_\_\_

Q: How many transcripts are known for the Gene? \_\_\_\_\_

Q: What is the accession number for the transcript (mRNA)? \_\_\_\_\_

- Verify your answers just below the **Genomic context** image in this document:

**Genomic context**

Location: 11q11-q12 See TCN1 in [Genome Data Viewer](#) [Epigenomics](#) [Map Viewer](#)

Exon count: 9

Annotation release	Status	Assembly	Chr	Location
<a href="#">107</a>	current	GRCh38.p2 ( <a href="#">GCF_000001405.28</a> )	11	NC_000011.10 (59852808..59866568, complement)
<a href="#">105</a>	previous assembly	GRCh37.p13 ( <a href="#">GCF_000001405.25</a> )	11	NC_000011.9 (59620281..59634041, complement)

**Chromosome 11 - NC\_000011.10**

The *TCN1* (transcobalamin 1) gene is annotated on chromosome 11 (NC\_000011.10). It is located on the reverse complement strand and it is spanning from position 59852808 to 59866568. The annotation that is on display by default is for the current assembly GRCh38.p2. The current annotation release is 107. There is a single transcript (mRNA) that is annotated for this gene, meaning that no splice variants are described. The accession number for the transcript is NM\_001062.3 and it encodes the transcobalamin-1 precursor protein, designated with the NP\_001053.2 accession.

### (Guided practice)

- Scroll to the **Genomic regions, transcripts, and products** section of the record.

**Genomic regions, transcripts, and products**

[Go to reference sequence details](#)

Genomic Sequence: NC\_000011.10 Chromosome 11 Reference GRCh38.p2 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

NC\_000011.10: 60M..60M (18Kbp) C

Tools | Tracks

59,868 K | 59,866 K | 59,864 K | 59,862 K | 59,860 K | 59,858 K | 59,856 K | 59,854 K | 59,852 K

Genes, NCBI Homo sapiens Annotation Release 107, 2015-03-19

Genes, Ensembl release 84

dbSNP Build 147 (Homo sapiens Annotation Release 107) all data

ClinVar Short Variations based on dbSNP Build 147 (Homo sapiens Annotation Release 107)

Cited Variants, dbSNP Build 147 (Homo sapiens Annotation Release 107)

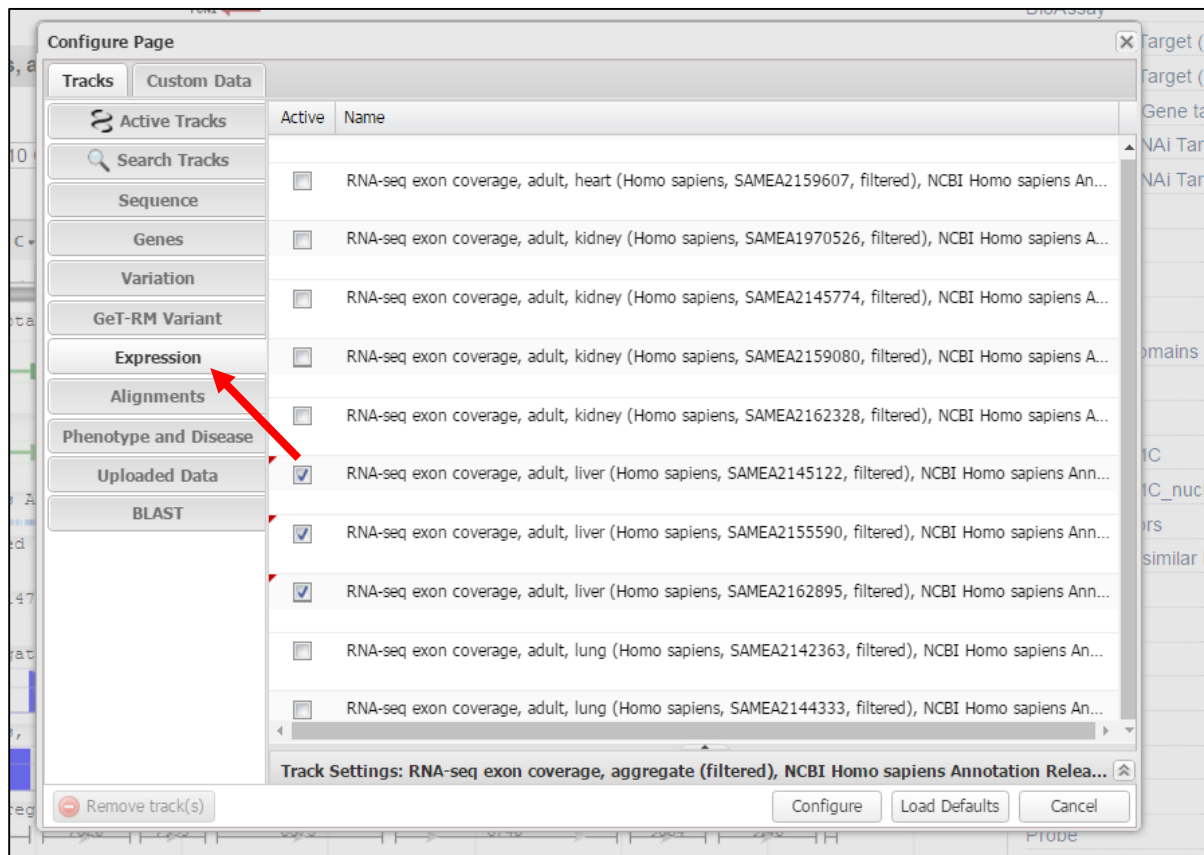
RNA-seq exon coverage, aggregate (filtered), NCBI Homo sapiens Annotation Release 107 - log base 2 scaled

RNA-seq intron-spanning reads, aggregate (filtered), NCBI Homo sapiens Annotation Release 107 - log base 2 scaled

RNA-seq intron features, aggregate (additional filtering)

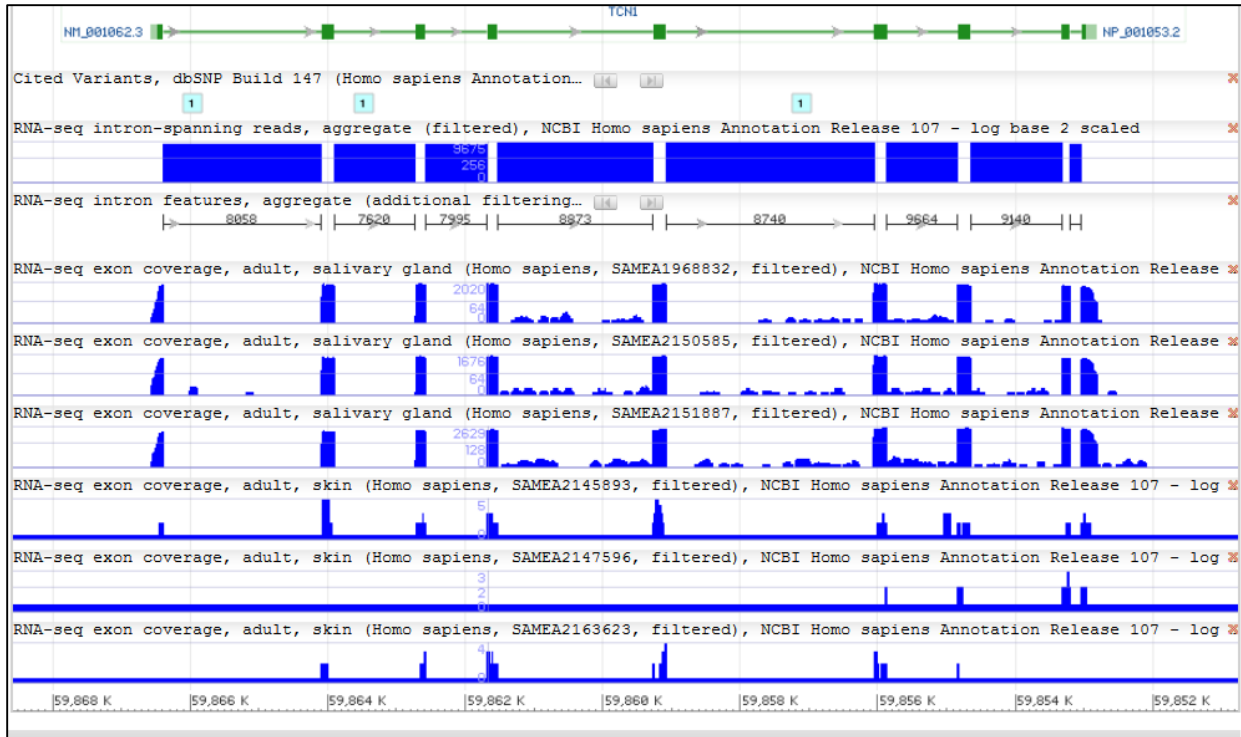
59,868 K | 59,866 K | 59,864 K | 59,862 K | 59,860 K | 59,858 K | 59,856 K | 59,854 K | 59,852 K

- In the Graphics, note the RNA-Seq exon coverage and RNA-Seq intron-spanning reads. This information is from analyzed (“processed”) RNA-Seq studies in the **SRA (Sequence Reads Archive)** database.
- Use the **Tracks** button to **Configure Expression** tracks. By default some of the aggregate tracks are selected.
- Scroll down to the second category which is RNA-Seq from individual samples. Select the RNA-Seq exon coverage tracks for salivary gland and skin:



- Once you upload the tracks, also uncheck all of the variation and other tracks that are not relevant for gene expression:





- Compare the expression levels in skin with that in salivary gland.

Q: Would you consider expression of this gene in the skin as high? \_\_\_\_\_

## 2. Using UniGene as a semi-quantitative assessment of TCN1 expression in various tissues

**Related articles in PubMed**

1. [Transcriptomic profile reveals gender-specific molecular mechanisms driving multiple sclerosis progression.](#)  
Irizar H, et al. PLoS One. 2014. PMID 24587374. [Free PMC Article](#)
2. [Role of serum holotranscobalamin \(holoTC\) in the diagnosis of patients with low serum cobalamin. Comparison with methylmalonic acid and homocysteine.](#)  
Remacha AF, et al. Ann Hematol. 2014 Apr. PMID 24057896
3. [Structural basis for universal corrinoid recognition by the cobalamin transport protein haptocorrin.](#)  
Furger E, et al. J Biol Chem. 2013 Aug 30. PMID 23846701. [Free PMC Article](#)

**Links to other resources**

- Protein
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed (GeneRIF)
- PubMed (OMIM)
- PubMed(nucleotide/PMC)
- RefSeq Proteins
- RefSeq RNAs
- SNP
- SNP: GeneView
- Taxonomy
- UniGene
- Variation Viewer
- Links to other resources
- HGNC

- RNA-Seq data for TCN1 expression in normal tissues indicate a very high expression in salivary gland and a very low expression in skin.

- Check the UniGene database if you could corroborate the RNA-Seq finding with UniGene expression profiles which are based on EST sequences.
- Use the UniGene link from the Gene record that takes you to the Hs.2012 UniGene Cluster. Each UniGene record is a cluster of RNA sequences that are products of the same gene.

NCBI Resources How To

UniGene UniGene Limits Advanced

Summary

**Links from Gene**

[Transcobalamin I \(vitamin B12 binding protein, R binder family\)](#)

TCN1, Homo sapiens  
Hs.2012: 81 sequences.

- Open up the record and access the **EST Profile**:

UGID:131135 UniGene Hs.2012 Homo sapiens (human) TCN1 [Links](#)

### Transcobalamin I (vitamin B12 binding protein, R binder family) (TCN1)

Human protein-coding gene TCN1. Represented by 77 ESTs from 42 cDNA libraries. EST representation biased toward adult. Corresponds to reference sequence NM\_001062.3. [UniGene 131135 - Hs.2012]

#### SELECTED PROTEIN SIMILARITIES

Comparison of cluster transcripts with RefSeq proteins. The alignments can suggest function of the cluster.

Best Hits and Hits from model organisms		Species	Id(%)	Len(aa)
<a href="#">NP_001053.2</a>	TCN1 gene product	<i>H. sapiens</i>	100.0	432
Other hits (2 of 11) <a href="#">[Show all]</a>		Species	Id(%)	Len(aa)
<a href="#">XP_001136280.2</a>	PREDICTED: transcobalamin-1	<i>P. troglodytes</i>	98.9	372
<a href="#">XP_003909825.1</a>	PREDICTED: transcobalamin-1, partial	<i>P. anubis</i>	90.7	312

#### GENE EXPRESSION

Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

Restricted Expression: adult [\[show more like this\]](#)

**EST Profile:** Approximate expression patterns inferred from EST sources. [\[Show more entries with profiles like this\]](#)

**GEO Profiles:** Experimental gene expression data (Gene Expression Omnibus).

**cDNA Sources:** stomach; mammary gland; intestine; mixed; pancreas; uncharacterized tissue; pharynx; salivary gland; testis; larynx; prostate; uterus; eye; bone marrow; trachea; blood; lung; muscle; liver; heart; parathyroid; brain

prostate	26		5 / 189536
salivary gland	148		3 / 20265
skin	0		0 / 210759
spleen	0		0 / 53397
stomach	114		11 / 95679
testis	2		1 / 435204

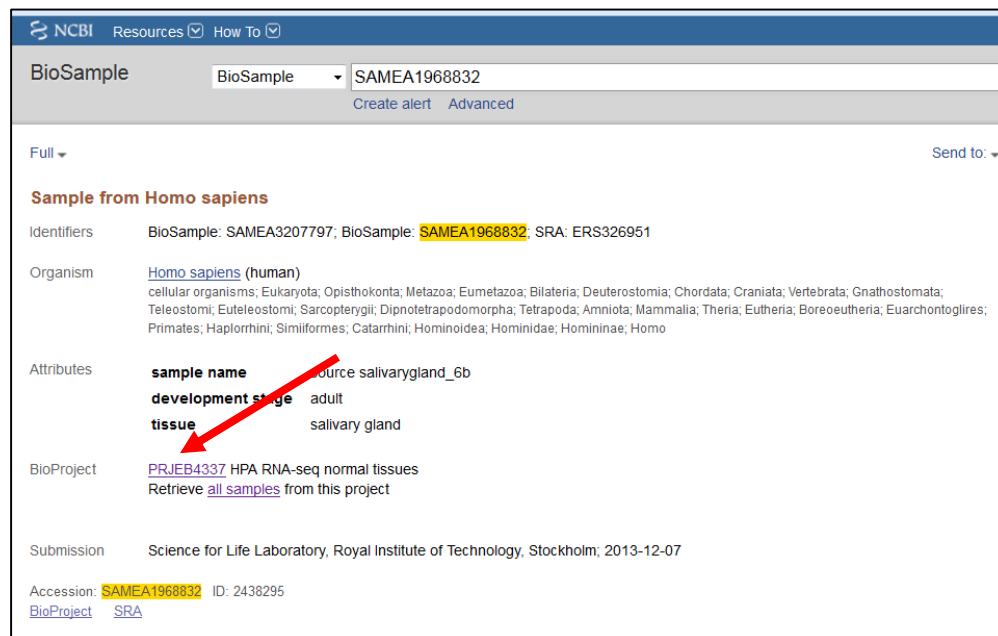
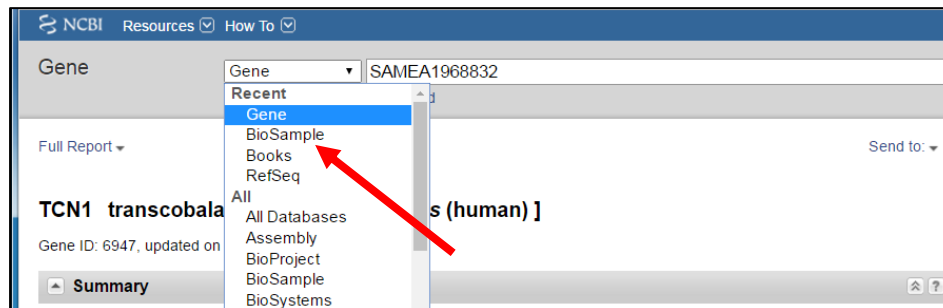
## Exercise 5 (Optional): A tour of BioSample-, BioProject-, and SRA databases.

### 1. RNA-Seq information in Gene: Where is it coming from?

- Return back to Gene. Each of the tracks in the Sequence Viewer (Graphic Display) is labeled. One of the labels states:

“RNA-seq exon coverage, adult, salivary gland (Homo sapiens, SAMEA1968832, filtered), NCBI Homo sapiens Annotation Release 107 - log 2 scaled”

- The SAMEA1968832 designation is sample accession of the sample that was used in the SRA experiment.
- Search the **BioSample** database with the accession. The BioSample database is one of the two meta-databases used for registering large research projects. The second one is the **BioProject** database.





- The SAMEA1968832 sample is one of the several samples that was used in a project designated with the PRJEB4337 accession.
- Link to the project in the BioProject database, to see what the study was all about:

BioProject

Display Settings: ▾ Send to: ▾

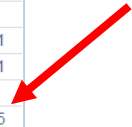
**HPA RNA-seq normal tissues** Accession: PRJEB4337 ID: 231263

RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes.

Accession	PRJEB4337
Data Type	Transcriptome or Gene expression
Scope	Monoisolate
Publications	Fagerberg L <i>et al.</i> , "Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.", <i>Mol Cell Proteomics</i> , 2013 Dec 5;13(2):397-406
Submission	Registration date: 12-Dec-2013 <b>Science for Life Laboratory, Stockholm, Sweden</b>

**Project Data:**

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	171
PUBLICATIONS	
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	95



- The BioProject record also links you back to all 95 sample records in the BioSample database:

NCBI Resources ▾ How To ▾ gratianm My NCBI Sign Out

BioSample   Help

Advanced

Organism  20 per page ▾ Send to: ▾ [Filters: Manage Filters](#)

Attribute name **Links from BioProject**

tissue **Items: 1 to 20 of 95** << First < Prev Page 1 of 5 Next > Last >>

Access Public (95)

Other Used by SRA (95) [Clear all](#) [Show additional filters](#)

[Sample from Homo sapiens](#)

- Identifiers: BioSample: SAMEA3207834; BioSample: SAMEA2153031; SRA: ERS327025  
Organism: Homo sapiens  
Accession: SAMEA2153031 ID: 2438369  
[BioProject](#) [SRA](#)
- Identifiers: BioSample: SAMEA3207842; BioSample: SAMEA2154665; SRA: ERS327024  
Organism: Homo sapiens  
Accession: SAMEA2154665 ID: 2438368  
[BioProject](#) [SRA](#)

**Find related data**

Database:

**Recent activity** [Turn Off](#) [Clear](#)

- BioSample for BioProject (Select 231263) (95) BioSample
- HPA RNA-seq normal tissues BioProject
- SAMEA1968832 (1) BioSample
- Sample from Homo sapiens

- Configure the display from summary to Full and 100 items per page:

The screenshot shows the NCBI BioSample search results page. The search term is 'BioSample'. The display is set to 'Full' and '100 per page'. A dropdown menu is open, showing options for 'Items per page' (5, 10, 20, 50, 100, 200). The search results show a list of samples, with the first one selected. The organism is 'Homo sapiens (human)'. The attributes are 'sample name' and 'source spleen\_3a'.

- Use the browser's search function to locate the tissue that you want, for example skin.

The screenshot shows the search results for 'skin'. The search term 'skin' is entered in the search box. The results show two samples selected (checked). The first sample is 'Sample from Homo sapiens' with ID 83. The attributes are 'sample name', 'source skin\_5f', 'development stage adult', and 'tissue skin'. The second sample is 'Sample from Homo sapiens' with ID 84. The search results are displayed in a table format.

- Select (checkmark) the sample of your interest. Here two samples are selected:

NCBI Resources How To gratianm

BioSample BioSample Search

Advanced

Organism Full 100 per page Send to: Filters: Manage Filters

Attribute name Links from BioProject  
tissue Items: 95 Selected: 2

Access Public (95)  Sample from Homo sapiens

Other Used by SRA (95) 1. Identifiers BioSample: SAMEA3207834; BioSample: SAMEA2153031; SRA: ERS327025

Organism [Homo sapiens \(human\)](#)  
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorhini; Simiiformes; Catarrhini; Hominoidea; Hominiidae; Homininae; Homo

Attributes **sample name** source spleen\_3a  
**development stage** adult

Find related data  
Database: Select  
Select  
Assembly  
BioProject  
dbGaP  
dbVar  
EST  
GEO DataSets  
GSS  
Nucleotide  
OMIM  
PubMed  
SNP  
SRA  
Taxonomy

Recent activity  
BioSample 23126  
Sample  
SAMEA  
TCN1 transcobalamin 1 [H

- From the samples, link to the SRA experiments in the SRA database: these are the records for RNA-Seq reads that were processed by NCBI and made available:

NCBI Resources How To gratianm My NCBI Sign Out

SRA SRA Search

Advanced Help

Access Public (4) Summary Send to: Filters: Manage Filters

Source RNA (4) Send results to Blast

Clear all Show additional filters

Links from BioSample Items: 4

[HPA RNA-seq normal tissues](#)

1. 1 ILLUMINA (Illumina HiSeq 2000) run: 10.1M spots, 2G bases, 937.9Mb downloads  
Accession: ERX288599

[HPA RNA-seq normal tissues](#)

2. 1 ILLUMINA (Illumina HiSeq 2000) run: 14.4M spots, 2.9G bases, 1.3Gb downloads  
Accession: ERX288593

[HPA RNA-seq normal tissues](#)

3. 1 ILLUMINA (Illumina HiSeq 2000) run: 14.6M spots, 3G bases, 1.3Gb downloads  
Accession: ERX288589

[HPA RNA-seq normal tissues](#)

4. 1 ILLUMINA (Illumina HiSeq 2000) run: 10.2M spots, 2.1G bases, 954.5Mb downloads  
Accession: ERX288504

Summary Send to:

Find related data  
Database: Select  
Find items

Recent activity  
Turn Off Clear

SRA Links for BioSample (Select 2 documents) (4) SRA

BioSample for BioProject (Select 231263) (95) BioSample

HPA RNA-seq normal tissues BioProject

SAMEA1968832 (1) BioSample

Sample from Homo sapiens biosample

See more...

**ERX288599: HPA RNA-seq normal tissues**

1 ILLUMINA (Illumina HiSeq 2000) run: 10.1M spots, 2G bases, 937.9Mb downloads

**Design:** HPA RNA-seq normal tissues**Submitted by:** SCIENCE FOR LIFE LABORATORY, STOCKHOLM, SWEDEN**Study:** HPA RNA-seq normal tissues[PRJEB4337](#) • [ERP003613](#) • [All experiments](#) • [All runs](#)[show Abstract](#)**Sample:** Homo sapiens; salivarygland\_6c[SAMEA2150585](#) • [ERS326959](#) • [All experiments](#) • [All runs](#)*Organism:* [Homo sapiens](#)**Library:***Name:* V240*Instrument:* Illumina HiSeq 2000*Strategy:* RNA-Seq*Source:* TRANSCRIPTOMIC*Selection:* cDNA*Layout:* PAIRED

*Construction protocol:* The use of human tissue samples was approved by the Uppsala Ethical Review Board (Reference #2011/473). Tissues samples, collected within the infrastructure of an established biobank, were embedded in Optimal Cutting Temperature (O.C.T.) compound and stored at -80C. A hematoxylin-eosin (HE) stained frozen section (4um) was prepared from each sample using a cryostat and the CryoJane Tape-Transfer System (Instrumedics, St. Louis, MO, USA). Each slide was examined by a pathologist to ensure proper tissue morphology. Three sections (10um) were cut from each frozen tissue block and collected into a tube for subsequent RNA extraction. The tissue was homogenized mechanically using a 3 mm steel grinding ball (VWR). Total RNA was extracted from cell lines and tissue samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturers instructions. The extracted RNA samples were analyzed using either an Experion automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA) with the standard-sensitivity RNA chip or an Agilent 2100 Bioanalyzer system (Agilent Biotechnologies, Palo Alto, USA) with the RNA 6000 Nano Labchip Kit. Only samples of high-quality RNA (RNA Integrity Number 7.5) were used in the following mRNA sample preparation for sequencing. Illumina Truseq RNA v2

**Spot descriptor:****Experiment attributes:***Experimental Factor: organism part:* salivary gland*Experimental Factor: individual:* V240**Runs:** 1 run, 10.1M spots, 2G bases, [937.9Mb](#)

Run	# of Spots	# of Bases	Size	Published
<a href="#">ERR315449</a>	10,096,395	2G	937.9Mb	2013-12-12

ID: 558501

- The table at the bottom of the record takes you to the **Run** (results) information for the experiment:

HPA RNA-seq normal tissues (ERR315449) [Change accession...](#)

**Metadata** Reads Download

Run	Spots	Bases	Size	GC content	Published	Access Type
ERR315449	10.1M	2.0Gbp	983.5M	50.4%	2013-12-12	public

Quality graph (bigger)

This run has 2 reads per spot

L=101, 100% L=101, 100%

Legend

Experiment	Library				
ERX288599					
Name	Platform	Strategy	Source	Selection	Layout
V240	Illumina	RNA-Seq	TRANSCRIPTOMIC	cDNA	PAIRED

Show design

Biosample	Sample Description	Organism
SAMEA2150585 (ERS326959)	Protocols: The use of human tissue samples was approved by the Uppsala Ethical Review Board (Reference #2011/473). Tissues samples, collected within the infrastructure of an established biobank, were embedded in Optimal Cutting Temperature (O.C.T.) compound and stored at -80C. A hematoxylin-eosin (HE) stained frozen section (4um) was prepared from each sample using a cryostat and the CryoJane Tape-Transfer System (Instrumedics, St. Louis, MO, USA). Each slide was examined by a pathologist to ensure proper tissue morphology. Three sections (10um) were cut from each frozen tissue block and collected into a tube for subsequent RNA extraction. The tissue was homogenized mechanically using a 3 mm steel grinding ball (VWR). Total RNA was extracted from cell lines and tissue samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturers instructions. The extracted RNA samples were analyzed using either an Exponent automated electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA) with the standard-sensitivity RNA chip or an Agilent 2100 Bioanalyzer system (Agilent Biotechnologies, Palo Alto, USA) with the RNA 6000 Nano Labchip Kit. Only samples of high-quality RNA (RNA Integrity Number 7.5) were used in the following mRNA sample preparation for sequencing. Illumina Truseq RNA v2	Homo sapiens

Bioproject	SRA Study	Title
PRJEB4337	ERP003613	HPA RNA-seq normal tissues

Show abstract

**Sequence Read Archive**

Main Browse Search Download Submit Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

Studies Samples Analyses **Run Browser** Run Selector Provisional SRA

HPA RNA-seq normal tissues (ERR315449) [Change accession...](#)

**Metadata** Reads Download

Filter:  Find Filtered Download [What does it do?](#)

[What can the filter be applied to?](#) This button will lead you to the page where you can subset the data using "Filter" field and download the results in FASTQ or FASTA format.

< 1 1 1009640 >

View:  biological reads  technical reads  quality scores [advanced options](#)

**Reads (separated)**

1. ERR315449.1 ERS326959  
name: HISQ:78:H0CCDADXX:1:1101:6506:1995  
member: default  
x: 6506, y: 1995

```
>gnl|SRA|ERR315449.1.1 HISQ:78:H0CCDADXX:1:1101:6506:1995 (Biological)
CTGAATACTGTCCATGGGGTAACTACTAATCAAATGCCCGCCTGTCGAATTTTC
AGGTTTTATGTAGATAAGTATTAGTTAACCTGTTGCAAAA
```

2. ERR315449.2 ERS326959  
name: HISQ:78:H0CCDADXX:1:1101:7495:1995  
member: default  
x: 7495, y: 1997

```
>gnl|SRA|ERR315449.1.2 HISQ:78:H0CCDADXX:1:1101:6506:1995 (Biological)
NTTAGGAAGAGCATCATTAATCACTGATCTCACCCCATAAAAGGATGATTTTGAGTT
TCAAGATTTTAGATTTACTAGAGATAGCATGTATCTATCAA
```

3. ERR315449.3 ERS326959  
name: HISQ:78:H0CCDADXX:1:1101:1314:201  
member: default

**Sequence Read Archive**

Main Browse Search Download Submit Documentation Software Trace Archive Trace Assembly Trace Home Trace BLAST

Studies Samples Analyses **Run Browser** Run Selector Provisional SRA

HPA RNA-seq normal tissues (ERR315449) [Change accession...](#)

**Metadata** Reads **Download**

Please:

Use [SRA Toolkit](#) tools to directly operate on SRA runs. Toolkit has capacity to find requested runs at NCBI and download (and cache) only the part you really need. For example quality scores represent a majority of data volume and you may not need them if you dump fasta only (versus fastq). Or if you are looking at particular gene you may not need the reads aligned to other regions or not aligned at all.

Use SRA Toolkit [prefetch](#) utility if you want to cache all data in advance (for example in case your processing cluster does not connect to internet). Read more at [Downloading SRA data using command line utilities](#).

Use SRA Run Selector to filter and download a list of SRA runs in the scope of [experiment](#), [sample](#) and [study](#).

[How can I get fastq format?](#) See [Converting SRA format data into FASTQ](#) in the [SRA Toolkit Documentation](#)

## 2. RNA-Seq experiments in GEO also are also registered in BioProject- and BioSample databases:

- Revisit the BRAF RNA-Seq experiment (GSE46805)
- The **Download data** link leads to the FTP site:

NCBI Resources How To

GEO DataSets GEO DataSets GSE46805[Accession]  
Create alert Advanced

Entry type Summary Sort by Default order Send to

DataSets (0)  
Series (1)  
Samples (2)  
Platforms (1)

Organism  
Customize ...

Study type  
Expression profiling by array  
Methylation profiling by array  
Customize ...

Author  
Customize ...

Attribute name  
tissue (0)  
strain (0)  
Customize ...

Publication dates  
30 days  
1 year  
Custom range...

Clear all  
Show additional filters

**Search results**  
Items: 4

[Expression data from Uninfected and BRAF600V induced melanocytes](#)  
1. (Submitter supplied) Melanocytes within benign human nevi are the paradigm for tumor suppressive senescent cells in a pre-malignant neoplasm. These cells typically contain mutations in either the BRAF or N-RAS oncogene and express markers of senescence, including p16. However, a nevus can contain 10s to 100s of thousands of clonal melanocytes and approximately 20-30% of melanoma are thought to arise in association with a pre-existing nevus. [more...](#)  
Organism: Homo sapiens  
Type: Expression profiling by high throughput sequencing  
Platform: GPL10999 2 Samples  
**Download data: GEO (BIGWIG), SRA SRP022259**  
Series Accession: [GSE46805](#) ID: 200046805  
[PubMed](#) [Full text in PMC](#) [Similar studies](#)

[Illumina Genome Analyzer Ix \(Homo sapiens\)](#)  
2. Organism: Homo sapiens  
679 Series 7336 Samples  
Download data: GEO  
Platform Accession: GPL10999 ID: 100010999

[BRAF](#)  
3. Organism: Homo sapiens  
Source name: BRAF  
Platform: GPL10999 Series: [GSE46805](#) GSE46818  
Download data: GEO (BIGWIG), SRA SRX276897  
Sample Accession: GSM1138581 ID: 301138581

[Uninfected](#)  
4. Organism: Homo sapiens  
Source name: Uninfected  
Platform: GPL10999 Series: [GSE46805](#) GSE46818  
Download data: GEO (BIGWIG), SRA SRX276896  
Sample Accession: GSM1138580 ID: 301138580

- The SRA link in the record takes you to the records on the SRA web:

senescence due to activated Wnt signaling. The observation that activation of Wnt signaling correlates inversely with nevus maturation, an indicator of clinical benignancy, supports the notion that persistent destabilization of senescence by Wnt signaling contributes to the malignant potential of nevi.

**Overall design** We used RNA-Seq to detail the global programme of gene expression in primary human melanocytes which were Uninfected and BRAF600V induced cells

**Contributor(s)** Pawlikowski JS, McBryan T, Adams PD

**Citation(s)** Pawlikowski JS, McBryan T, van Tuyn J, Drotar ME et al. Wnt signaling potentiates neovogenesis. *Proc Natl Acad Sci U S A* 2013 Oct 1;110(40):16009-14. PMID: [24043806](#)  
 Capell BC, Drake AM, Zhu J, Shah PP et al. MLL1 is essential for the senescence-associated secretory phenotype. *Genes Dev* 2016 Feb 1;30(3):321-36. PMID: [26833731](#)

**Submission date** May 09, 2013  
**Last update date** May 24, 2016  
**Contact name** Peter D. Adams  
**Organization name** University of Glasgow, Beatson Institute for Cancer Research  
**Street address** Switchback Rd, Bearsden  
**City** Glasgow  
**ZIP/Postal code** G61 1BD  
**Country** United Kingdom

**Platforms (1)** [GPL10999](#) Illumina Genome Analyzer Iix (Homo sapiens)

**Samples (2)** [GSM1138580](#) Uninfected  
[GSM1138581](#) BRAF

This SubSeries is part of SuperSeries:  
[GSE46818](#) Wnt-signaling potentiates neovogenesis.

**Relations**  
**BioProject** [PRJNA202399](#)  
**SRA** [SRP022259](#)

[See the data on Genome Data Viewer](#)

Download family	Format
<a href="#">SOFT formatted family file(s)</a>	SOFT <a href="#">?</a>
<a href="#">MINIML formatted family file(s)</a>	MINIML <a href="#">?</a>
<a href="#">Series Matrix File(s)</a>	TXT <a href="#">?</a>

Supplementary file	Size	Download	File type/resource
<a href="#">GSE46805_RAW.tar</a>	202.3 Mb	<a href="#">(http)(custom)</a>	TAR (of BIGWIG)

- From the SRA web you can send the results to BLAST:

NCBI Resources How To

SRA SRA SRP022259  
 Create alert Advanced

Access Public (2)  
 Source RNA (2)  
 Clear all  
 Show additional filters

Summary Send to: ▼

[Send results to Blast](#)

**Search results**  
**Items: 2**

- [GSM1138581: BRAF; Homo sapiens; RNA-Seq](#)  
 1. 1 ILLUMINA (Illumina Genome Analyzer Iix) run: 29.2M spots, 4.1G bases, 2.3Gb downloads  
 Accession: SRX276897
- [GSM1138580: Uninfected; Homo sapiens; RNA-Seq](#)  
 2. 1 ILLUMINA (Illumina Genome Analyzer Iix) run: 30.7M spots, 4.3G bases, 2.3Gb downloads  
 Accession: SRX276896

- The two experiments (**SRX276896** and **SRX276897**) are now listed as the database to which you can align your sequence of interest:

The screenshot shows the BLAST search interface for 'Sequence Read Archive Nucleotide BLAST'. The 'Choose Search Set' section is active, showing 'Sequences: 119,879,516' and two selected SRA experiments: SRX276897 and SRX276896. The 'Program Selection' section is set to 'Somewhat similar sequences (blastn)'. The 'Algorithm parameters' section shows 'Max target sequences' set to 5000.

- We will work with one experiment at a time, for example **SRX276897**:

The screenshot shows the BLAST search interface with the 'Choose Search Set' section active. The 'SRA Experiment set (SRX)' field now contains only 'SRX276897', which is highlighted in yellow. A tooltip is visible over the text, showing details: 'SRX276897 GSM1138581: BRAF; Homo sapiens; RNA-Seq (Homo sapiens taxid:9606; study:SRP022259; subm...'. The 'Program Selection' section remains set to 'Somewhat similar sequences (blastn)'. The 'Algorithm parameters' section shows 'Max target sequences' set to 5000.



- Our query sequence will be the genomic sequence of the TCN1 gene: **ref|NC\_000011.10 (From: 59852808 To: 59866568)**

**Sequence Read Archive Nucleotide BLAST**

blastn Reset page    Bookmark

BLASTN programs search SRA databases using a nucleotide query.

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) Clear    Query subrange

ref|NC\_000011.10 From: 59852808  
To: 59866568

Or, upload file Browse...    No file selected.

Job Title NC\_000011.Homo sapiens chromosome 11, GRCh38.p2...

Enter a descriptive title for your BLAST search

**Choose Search Set**

SRA Experiment set (SRX) Sequences: 58,388,984

SRX276897 +

Enter an SRA accession (experiment, study, or submission), title, the scientific name or tax id. Only 20 top suggestions will be shown.

**Program Selection**

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

**BLAST** Search database SRA using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

**Algorithm parameters** Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign    Restore default search parameter

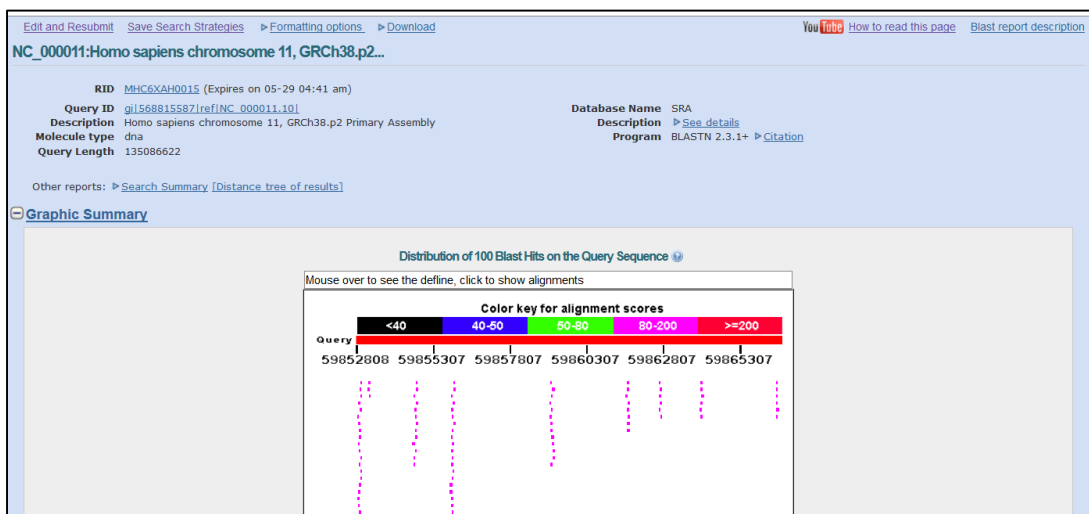
**General Parameters**

Max target sequences ♦ 5000

Select the maximum number of aligned sequences to display

- See the search results for SRX276897 BLAST:

Q: Can you identify the nine exons on the sequence? \_\_\_\_\_



- Re-run BLAST for SRX276896 BLAST:

Q: Can you identify the nine exons on the sequence? \_\_\_\_\_

blastn BLASTn programs search SRA databases using a nucleotide query.

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

ref|NC\_000011.10 From 59852808

To 59866568

Or, upload file Browse... No file selected.

**Job Title** NC\_000011:Homo sapiens chromosome 11, GRCh38.p2...  
Enter a descriptive title for your BLAST search

**Choose Search Set** Sequences: 58,388,984

**SRA Experiment set (SRX)** SRX276896 +

SRX276896 GSM1138580: Uninfected; Homo sapiens; RNA-Seq (Homo sapiens taxid:9606; study:SRP022259; ... down.

**Program Selection**

**Optimize for**

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

**BLAST** Search database SRA using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

