

# Data Curation Manual

## 1 Website

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

## 2 Filter selection

- By organism - `Homo sapiens`
- By experiment type - `rna assay`

## 3 Procedure

### 3.1 Search disease name by using the correct filter in ArrayExpress

### 3.2 For each Accession, check the following

#### 3.2.1 Check if this accession is about MicroRNA

The following two datasets are all examples of MicroRNA

- Huntington's Disease, E-GEOD-64977

EMBL-EBI ArrayExpress

Home Browse Submit Help About ArrayExpress

ArrayExpress > Browse > E-GEOD-64977

E-GEOD-64977 - miRNA-seq expression profiling of Huntington's Disease and neurologically normal human post-mortem prefrontal cortex (BA9) brain samples

Status Released on 12 March 2015, last updated on 21 March 2015

Organism Homo sapiens

Samples (64) [Click for detailed sample information and links to data](#)

Protocols (2) [Click for detailed protocol information](#)

Description

BACKGROUND MicroRNAs (miRNAs) are small non-coding RNAs that recognize sites of complementarity of target messenger RNAs, resulting in transcriptional regulation and translational repression of target genes. In Huntington's disease (HD), a neurodegenerative disease caused by a trinucleotide repeat expansion, miRNA dysregulation has been reported, which may impact gene expression and modify the progression and severity of HD. METHODS: We performed next-generation miRNA sequence analysis in prefrontal cortex (Brodmann Area 9) from 26 HD, 2 asymptomatic HD, and 36 control brains. Neuropathological information was available for all HD brains, including age at disease onset, CAG-repeat size, Vonsattel grade, and Hadzi-Vonsattel striatal and cortical scores, a continuous measure of the extent of neurodegeneration. Linear models were performed to examine the relationship of miRNA expression to these clinical features, and messenger RNA targets of associated miRNAs were tested for gene ontology term enrichment. RESULTS: We identified 75 miRNAs differentially expressed in HD brain (FDR q-value <0.05). Among the HD brains, nine miRNAs were significantly associated with Vonsattel grade of neuropathological involvement and three of these, miR-10b-5p, miR-10b-3p, and miR-302a-3p, significantly related to the Hadzi-Vonsattel striatal score (a continuous measure of striatal involvement) after adjustment for CAG length. Five miRNAs (miR-10b-5p, miR-196a-5p, miR-196b-5p, miR-10b-3p, and miR-106a-5p) were identified as having a significant relationship to CAG length-adjusted age of onset including miR-10b-5p, the mostly strongly over-expressed miRNA in HD cases. Although prefrontal cortex was the source of tissue profiled in these studies, the relationship of miR-10b-5p expression to striatal involvement in the disease was independent of cortical involvement. Correlation of miRNAs to the clinical features clustered by direction of effect and the gene targets of the observed miRNAs showed association to processes relating to nervous system development and transcriptional regulation. CONCLUSIONS: These results demonstrate that miRNA expression in cortical BA9 provides insight into striatal involvement and support a role for these miRNAs, particularly miR-10b-5p, in HD pathogenicity. The miRNAs identified in our studies of postmortem brain tissue may be detectable in peripheral fluids and thus warrant consideration as accessible biomarkers for disease stage, rate of progression, and other important clinical characteristics of HD. 26 Huntington's disease, 2 asymptomatic HD gene positive and 49 neurologically normal control prefrontal cortex samples

Experiment type RNA-seq of non coding RNA

Contacts [Richard H Myers <rmeyers@bu.edu>](#), Adam Labadorf, Andrew G Hoss, Jeanne Latourelle

MINSEQE

Exp. design Protocols Variables Processed Seq. reads

Files

Investigation description [E-GEOD-64977.idf.txt](#)

Sample and data relationship [E-GEOD-64977.sdrf.txt](#)

[Click to browse all available files](#)

Links

ENA - SRP052236, GEO - GSE64977

Send E-GEOD-64977 data to [GENOME SPACE](#)

- Encephalitis, E-GEOD-55069

EMBL-EBI

ArrayExpress

Home Browse Submit Help About ArrayExpress

ArrayExpress > Browse > E-GEOD-55069

**E-GEOD-55069 - MicroRNA-21 dysregulates the expression of MEF2C in neurons in monkey and human SIV/HIV neurological disease**

Status Released on 18 February 2014, last updated on 3 May 2014

Organism Homo sapiens

Samples (34) [Click for detailed sample information and links to data](#)

Array (1) A-GEOD-15829 - Exiqon miRCURY LNA microRNA array v.11.0 (miRBase v13)

Protocols (6) [Click for detailed protocol information](#)

Description MicroRNAs (miRNAs) have important roles in regulating a plethora of physiological and pathophysiological processes including neurodegeneration. In both human immunodeficiency virus (HIV)-associated dementia in humans and its monkey model simian immunodeficiency virus encephalitis (SIVE), we find miR-21, a miRNA largely known for its link to oncogenesis, to be significantly upregulated in the brain. In situ hybridization of the diseased brain sections revealed induction of miR-21 in neurons. miR-21 can be induced in neurons by prolonged N-methyl-D-aspartic acid receptor stimulation, an excitotoxic process active in HIV and other neurodegenerative diseases. Introduction of miR-21 into human neurons leads to pathological functional defects. Furthermore, we show that miR-21 specifically targets the mRNA of myocyte enhancer factor 2C (MEF2C), a transcription factor crucial for neuronal function, and reduces its expression. MEF2C is dramatically downregulated in neurons of HIV-associated dementia patients, as well as monkeys with SIVE. Together, this study elucidates a novel role for miR-21 in the brain, not only as a potential signature of neurological disease, but also as a crucial effector of HIV-induced neuronal dysfunction and neurodegeneration. Total cellular RNA was isolated from frozen (-80°C) brain specimens by using TRIzol (Invitrogen, Carlsbad, CA, USA) for the monkey samples, and RNazol (Biotech Laboratories, Houston, TX, USA) for the human samples, followed by column purification (miRNeasy, Qiagen, Valencia, CA, USA) as per manufacturer's instructions. For the monkeys, four samples were from uninfected animals and four from SIV-infected animals that developed simian AIDS with SIV encephalitis. For the human samples, six samples were from individuals who were HIV negative with no history of dementia or neurocognitive disability and showed no significant neuropathology. Five samples were from HIV positive individuals that were neurocognitively impaired (NNTC clinical rating >6)38 and a neuropathological diagnosis of HIV encephalitis. Clinical information is provided in the supplemental material. For microarray analysis, only three of the HAD specimens were sufficient for use, but all five were used for the qRT-PCR studies.

Experiment type transcription profiling by array

Contacts [Howard Fox <geo@ncbi.nlm.nih.gov>](#), Howard S Fox

MIAME

Platforms Protocols Variables Processed Raw

Files

Investigation description	<a href="#">E-GEOD-55069.idf.txt</a>
Sample and data relationship	<a href="#">E-GEOD-55069.sdrf.txt</a>
Raw data (1)	<a href="#">E-GEOD-55069.raw.1.zip</a>
Processed data (1)	<a href="#">E-GEOD-55069.processed.1.zip</a>
Array design	<a href="#">A-GEOD-15829.adf.txt</a>

[Click to browse all available files](#)

Links

[GEO - GSE55069](#)

[Send E-GEOD-55069 data to GENOMESPACE](#)

### 3.2.2 Check if it has enough replicates

- Creutzfeldt-Jakob Disease, E-GEOD-30643

Samples and Data < E-GEOD

https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-30643/samples/

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our Terms of Use.

EMBL-EBI

ArrayExpress

Home Browse Submit Help About ArrayExpress

Feedback Login

ArrayExpress > Browse > E-GEOD-30643 > Samples and Data

E-GEOD-30643 - Expression data from parietal cortex of G114V genetic Creutzfeldt-Jakob disease patient

2 rows

Source Name	Sample_description	Sample_source_name	age	genot	Links to Data
GSM759883 1	Gene expression data from parietal cortex of G114V gCJD patient brain	Human brain from a G114V gCJD patient(postmortem)	47 year	PRNP-C	Raw Processed
GSM759884 1	Gene expression data from commercial normal human parietal cortex pooled from four male/female	Commercial normal human brain total RNA(clontech)	35-89 year	PRNP-V	Raw Processed

Download Samples and Data table in Tab-delimited format

EMBL-EBI

Services

By topic  
By name (A-Z)  
Help & Support

Research

Overview  
Publications  
Research groups  
Postdocs & PhDs

Training

Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry

Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme

About us

Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs  
People & groups  
News  
Events  
Visit us  
Contact us

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
Copyright © EMBL-EBI 2016 | EBI is an outstation of the European Molecular Biology Laboratory | Terms of use

### 3.2.3 Check if it has disease vs. control

- Tuberous Sclerosis, E-GEOD-9715

It is hard to determine whether there is disease and control or not sometimes. The following dataset is about a disease called Tuberous Sclerosis

EMBL-EBI

ArrayExpress

Home Browse Submit Help About ArrayExpress

ArrayExpress > Browse > E-GEOD-9715

### E-GEOD-9715 - Transcription profiling of human tuberous sclerosis complex (TSC) Tumor Fibroblasts

Status	Submitted on 28 November 2007, released on 6 April 2008, last updated on 10 June 2011			
Organism	Homo sapiens			
Samples (11)	<a href="#">Click for detailed sample information and links to data</a>			
Array (1)	A-AFFY-33 - Affymetrix GeneChip Human Genome HG-U133A [HG-U133A]			
Protocols (5)	<a href="#">Click for detailed protocol information</a>			
Description	<p>Patients with tuberous sclerosis complex (TSC) develop hamartomas containing biallelic inactivating mutations in either TSC1 or TSC2, resulting in mammalian target of rapamycin (mTOR) activation. Hamartomas overgrow epithelial and mesenchymal cells in TSC skin. The pathogenetic mechanisms for these changes had not been investigated, and the existence or location of cells with biallelic mutations ("two-hit" cells) that resulted in mTOR activation was unclear. We compared TSC skin hamartomas (facial angiofibromas and periungual fibromas) to normal-appearing skin of the same patient, and observed more proliferation and mTOR activation in hamartoma epidermis. "Two-hit" cells were not detected in the epidermis. Fibroblast-like cells in the dermis, however, exhibited allelic deletion of TSC2, in both touch preparations of fresh tumor samples and cells grown from TSC skin tumors, suggesting that increased epidermal proliferation and mTOR activation were not caused by second-hit mutations in the keratinocytes but by mesenchymal-epithelial interactions. Gene expression arrays, used to identify potential paracrine factors released by mesenchymal cells, revealed more epi-regulin mRNA in fibroblast-like angiofibroma and periungual fibroma cells than in fibroblasts from normal-appearing skin of the same patient. Elevation of epi-regulin mRNA was confirmed using real-time PCR, and increased amounts of epi-regulin protein were demonstrated using immunoprecipitation and ELISA. Epi-regulin stimulated keratinocyte proliferation and phosphorylation of ribosomal protein S6 in vitro. These results suggest that hamartomatous TSC skin tumors are induced by paracrine factors released by "two-hit" cells in the dermis, and that proliferation with mTOR activation of the overlying epidermis is an effect of epi-regulin. Experiment Overall Design: The study is of case/control design with biological replication. Tumor (case) and normal (control) fibroblast cells were isolated from each of four patients (biological replicates).</p>			
Experiment types	transcription profiling by array, unknown experiment type			
Contact	Eric Billings			
MIAME	<div> <div>★</div> <div>—</div> <div>—</div> <div>★</div> <div>★</div> </div> <div>Platforms Protocols Variables Processed Raw</div>			
Files	<div> <div>Investigation description</div> <div>Sample and data relationship</div> <div>Raw data (1)</div> <div>Processed data (1)</div> <div>Array design</div> <div>R ExpressionSet</div> <div><a href="#">Click to browse all available files</a></div> </div> <div> <a href="#">E-GEOD-9715.idf.txt</a>  <a href="#">E-GEOD-9715.sdrrf.txt</a>  <a href="#">E-GEOD-9715.raw.1.zip</a>  <a href="#">E-GEOD-9715.processed.1.zip</a>  <a href="#">A-AFFY-33.adf.txt</a>  <a href="#">E-GEOD-9715.eSet.r</a> </div>			
Links	<a href="#">GEO - GSE9715</a> <a href="#">Send E-GEOD-9715 data to GENOME SPACE</a>			

EMBL-EBI

News  
Brochures  
Contact us  
Intranet

Services

By topic  
By name (A-Z)  
Help & Support

Research

Overview  
Publications  
Research groups  
Postdocs & PhDs

Training

Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry

Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme

About us

Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs  
People & groups  
News  
Events  
Visit us  
Contact us

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
Copyright © EMBL-EBI 2016 | EBI is an outstation of the European Molecular Biology Laboratory | Terms of use

If we check its samples like what we usually do, we will get the following

EMBL-EBI

ArrayExpress

Home Browse Submit Help About ArrayExpress

ArrayExpress > Browse > E-GEOD-9715 > Samples and Data\*

E-GEOD-9715 - Transcription profiling of human uberous sclerosis complex (TSC) Tumor Fibroblasts

11 rows

Sample Attributes		Links to Data	
Source Name	Organism	Raw	Processed Matrix
GSE9715GSM245372	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245373	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245374	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245375	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245376	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245377	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245378	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245379	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245380	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245381	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>
GSE9715GSM245382	Homo sapiens	<a href="#">Download</a>	<a href="#">Download</a>

Download Samples and Data table in Tab-delimited format

EMBL-EBI

Services

By topic  
By name (A-Z)  
Help & Support

Research

Overview  
Publications  
Research groups  
Postdocs & PhDs

Training

Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry

Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme

About us

Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs  
People & groups  
News  
Events  
Visit us  
Contact us

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
Copyright © EMBL-EBI 2016 | EBI is an outstation of the European Molecular Biology Laboratory | Terms of use

At this time, we need to check its GEO database, by click Links GEO - GSE9715 at the bottom of its main page

Summary

containing biallelic inactivating mutations in either TSC1 or TSC2, resulting in mammalian target of rapamycin (mTOR) activation. Hamartomas overgrew epithelial and mesenchymal cells in TSC skin. The pathogenic mechanisms for these changes had not been investigated, and the existence or location of cells with biallelic mutations ("two-hit" cells) that resulted in mTOR activation was unclear. We compared TSC skin hamartomas (facial angiofibromas and periungual fibromas) to normal-appearing skin of the same patient, and observed more proliferation and mTOR activation in hamartoma epidermis. "Two-hit" cells were not detected in the epidermis. Fibroblast-like cells in the dermis, however, exhibited allelic deletion of TSC2, in both touch preparations of fresh tumor samples and cells grown from TSC skin tumors, suggesting that increased epidermal proliferation and mTOR activation were not caused by second-hit mutations in the keratinocytes but by mesenchymal-epithelial interactions. Gene expression arrays, used to identify potential paracrine factors released by mesenchymal cells, revealed more epiregulin mRNA in fibroblast-like angiofibroma and periungual fibroma cells than in fibroblasts from normal-appearing skin of the same patient. Elevation of epiregulin mRNA was confirmed using real-time PCR, and increased amounts of epiregulin protein were demonstrated using immunoprecipitation and ELISA. Epiregulin stimulated keratinocyte proliferation and phosphorylation of ribosomal protein S6 in vitro. These results suggest that hamartomatous TSC skin tumors are induced by paracrine factors released by "two-hit" cells in the dermis, and that proliferation with mTOR activation of the overlying epidermis is an effect of epiregulin.

Keywords: Disease state analysis

Overall design

The study is of case/control design with biological replication. Tumor (case) and normal (control) fibroblast cells were isolated from each of four patients (biological replicates).

Contributor(s)

Darling TN

Citation(s)

Li S, Takeuchi F, Wang JA, Fan Q et al. Mesenchymal-epithelial interactions involving epiregulin in tuberous sclerosis complex hamartomas. *Proc Natl Acad Sci U S A* 2008 Mar 4;105(9):3539-44. PMID: 18292222

Submission date

Nov 28, 2007

Last update date

May 14, 2015

Contact name

Eric Billings

Organization name

NHLBI

Street address

10 Center Drive

City

Bethesda

State/province

MD

ZIP/Postal code

20874

Country

USA

Platforms (1)

GPL96 [HG-U133A] Affymetrix Human Genome U133A Array

Samples (11)

at Less...

GSM245372 fibroblast\_N1\_Patient12

GSM245373 fibroblast\_AF\_Patient12

GSM245374 fibroblast\_N1\_Patient2

GSM245375 fibroblast\_AF\_Patient2

GSM245376 fibroblast\_PF\_Patient2

GSM245377 fibroblast\_N1\_Patient4

GSM245378 fibroblast\_AF\_Patient4

GSM245379 fibroblast\_PF\_Patient4

GSM245380 fibroblast\_N1\_Patient10

GSM245381 fibroblast\_AF\_Patient10

GSM245382 fibroblast\_PF\_Patient10

Relations

Suppl. Files

GSM11488616

Now, we can know what their samples are come from 'N1', 'AF' and 'PF' patients. To figure out what 'N1', 'AF', 'PF' mean, we need to check it description again



The screenshot shows the ArrayExpress website interface. The main heading is "E-GEOD-9715 - Transcription profiling of human tuberous sclerosis complex (TSC) Tumor Fibroblasts". The description states: "Patients with tuberous sclerosis complex (TSC) develop hamartomas containing biallelic inactivating mutations in either TSC1 or TSC2, resulting in mammalian target of rapamycin (mTOR) activation. Hamartomas overgrow epithelial and mesenchymal cells in TSC skin. The pathogenetic mechanisms for these changes had not been investigated, and the existence or location of cells with biallelic mutations ("two-hit" cells) that resulted in mTOR activation was unclear. We compared TSC skin hamartomas (facial angiofibromas and periungual fibromas) to normal-appearing skin of the same patient, and observed more proliferation and mTOR activation in hamartoma epidermis. "Two-hit" cells were not detected in the epidermis. Fibroblast-like cells in the dermis, however, exhibited allelic deletion of TSC2, in both touch preparations of fresh tumor samples and cells grown from TSC skin tumors, suggesting that increased epidermal proliferation and mTOR activation were not caused by second-hit mutations in the keratinocytes but by mesenchymal-epithelial interactions. Gene expression arrays, used to identify potential paracrine factors released by mesenchymal cells, revealed more epiregulin mRNA in fibroblast-like angiofibroma and periungual fibroma cells than in fibroblasts from normal-appearing skin of the same patient. Elevation of epiregulin mRNA was confirmed using real-time PCR, and increased amounts of epiregulin protein were demonstrated using immunoprecipitation and ELISA. Epiregulin stimulated keratinocyte proliferation and phosphorylation of ribosomal protein S6 in vitro. These results suggest that hamartomatous TSC skin tumors are induced by paracrine factors released by "two-hit" cells in the dermis, and that proliferation with mTOR activation of the overlying epidermis is an effect of epiregulin. Experiment Overall Design: The study is of case/control design with biological replication. Tumor (case) and normal (control) fibroblast cells were isolated from each of four patients (biological replicates)."

Now, we know they are comparing facial angiofibromas(AF) and periungual fibromas(PF) to normal-appearing(N1), so it is relevant.

### 3.2.4 Check if it is biomarker

- Narcolepsy, E-GEOD-21592

The following dataset is about narcolepsy, which is a neurological disorder. But the dataset source is from blood



Samples and Data < E-GE/ X

https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-21592/samples/

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our Terms of Use.

EMBL-EBI ArrayExpress

Services Research Training About us

Home Browse Submit Help About ArrayExpress

Search  
Examples: E-MEXP-31, cancer, p53, Geovadis  
Advanced

Feedback Login

ArrayExpress > Browse > E-GEOD-21592 > Samples and Data

E-GEOD-21592 - Genome-wide gene expression profiling of human narcolepsy

20 rows

Source Name	Sample_source_name	Organism	disease status	sex	tissue	DISEASE STATUS	SEX	Links to Data
								Raw Processed
GSMS39022 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39023 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39024 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39025 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39026 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39027 1	blood	Homo sapiens	healthy control	female	whole blood	healthy control	female	<a href="#">↓</a> <a href="#">↓</a>
GSMS39028 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39029 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39030 1	blood	Homo sapiens	healthy control	male	whole blood	healthy control	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39031 1	blood	Homo sapiens	healthy control	female	whole blood	healthy control	female	<a href="#">↓</a> <a href="#">↓</a>
GSMS39032 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39033 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39034 1	blood	Homo sapiens	narcolepsy patient	female	whole blood	narcolepsy patient	female	<a href="#">↓</a> <a href="#">↓</a>
GSMS39035 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39036 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39037 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39038 1	blood	Homo sapiens	narcolepsy patient	female	whole blood	narcolepsy patient	female	<a href="#">↓</a> <a href="#">↓</a>
GSMS39039 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39040 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>
GSMS39041 1	blood	Homo sapiens	narcolepsy patient	male	whole blood	narcolepsy patient	male	<a href="#">↓</a> <a href="#">↓</a>

Download Samples and Data table in Tab-delimited format

EMBL-EBI

News  
Brochures  
Contact us  
Intranet

Services  
By topic  
By name (A-Z)  
Help & Support

Research  
Overview  
Publications  
Research groups  
Postdocs & PhDs

Training  
Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry  
Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme

About us  
Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs

So, it is may be a biomarker. Check the dataset description now.

The screenshot shows the ArrayExpress website interface. The main heading is 'E-GEOD-21592 - Genome-wide gene expression profiling of human narcolepsy'. Below this, there is a detailed description of the study, including its objectives, design, and results. The experiment type is listed as 'transcription profiling by array'. Contact information for Camilla Bernardini is provided. A table of files for download is also shown, including investigation descriptions, sample and data relationships, raw data, processed data, array design, and R expression sets. Links to GEO and GENOMESPACE are also present.

You can see the goal of this study is to find the **biomarker**. The disease signature must reflect the underlying biological basis of the disease. For biomarker study, the goal is not to investigate disease mechanisms. Thus, data sets collected for biomarker study can not be used for drug repositioning.

## 4 Do It Yourself

Check the following datasets:

- Huntington's Disease, E-MTAB-2206
- Dravet Syndrome, E-GEOD-46472
- Tuberous Sclerosis, E-GEOD-6002
- Huntington's Disease, E-GEOD-1751

## 5 Newly added

- Do not focus too much on confirming whether it is about mRNA. The dataset should be about mRNA by default. Just pay a attention for MicroRNA(miRNA).
- Check whether it is disease vs control:

acute megakaryoblastic leukemia, E-GEOD-16677

Based on the title of this dataset, it is study the Down Syndrome (DS)-AMKL and non-DS AMKL samples. There is no healthy sample.

E-GEOD-16677 - Gene expression profiling of Down Syndrome (DS)-AMKL and non-DS AMKL samples




18 rows								
Source Name	Sample description	Sample_source_name	Organism	cell description	Sample Attributes		Links to Data	
					cell line	cell type	Raw	Processed
GSM417985 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 1, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417986 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 2, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417987 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 3, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417988 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 4, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417989 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 5, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417990 1	Gene expression data from sorted primary leukemic cells.	DS-AMKL sample 6, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417991 1	Gene expression data from sorted primary leukemic cells.	Non-DS AMKL sample 1, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417992 1	Gene expression data from sorted primary leukemic cells.	Non-DS AMKL sample 2, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417993 1	Gene expression data from sorted primary leukemic cells.	Non-DS AMKL sample 3, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417994 1	Gene expression data from sorted primary leukemic cells.	Non-DS AMKL sample 4, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417995 1	Gene expression data from sorted primary leukemic cells.	Non-DS AMKL sample 5, sorted blasts	Homo sapiens	FACS sorted leukemic blasts from patients	megaka		<a href="#">↓</a>	<a href="#">↓</a>
GSM417996 1	Gene expression data from a non-DS AMKL cell line.	Non-DS AMKL cell line UT-7	Homo sapiens	AMKL cell line	UT-7		<a href="#">↓</a>	<a href="#">↓</a>
GSM417997 1	Gene expression data from a DS-AMKL cell line.	DS-AMKL cell line CMK	Homo sapiens	DS-AMKL cell line	CMK		<a href="#">↓</a>	<a href="#">↓</a>
GSM417998 1	Gene expression data from a non-DS AMKL cell line.	Non-DS AMKL cell line UT-7	Homo sapiens	AMKL cell line	UT-7		<a href="#">↓</a>	<a href="#">↓</a>
GSM417999 1	Gene expression data from a DS-AMKL cell line.	DS-AMKL cell line CMK	Homo sapiens	DS-AMKL cell line	CMK		<a href="#">↓</a>	<a href="#">↓</a>
GSM418000 1	Gene expression data from a non-DS AMKL cell line.	AMKL cell line Meg01	Homo sapiens	AMKL cell line	MEG01		<a href="#">↓</a>	<a href="#">↓</a>
GSM418001 1	Gene expression data from a non-DS AMKL cell line.	AMKL cell line M07	Homo sapiens	AMKL cell line	M07		<a href="#">↓</a>	<a href="#">↓</a>
GSM418002 1	Gene expression data from a erythroleukemia cell line.	Erythroleukemia cell line K562	Homo sapiens	Erythroleukemia cell line	K562		<a href="#">↓</a>	<a href="#">↓</a>

[Download Samples and Data table in Tab-delimited format](#)

- If you see a comprehensive dataset, mark as comprehensive.

## Acute myeloblastic leukemia with maturation, E-MTAB-3732









### E-MTAB-3732 - [A comprehensive human expression map](#)

Status	Submitted on 23 May 2015, released on 23 July 2015, last updated on 1 August 2015			
Organism	Homo sapiens			
Samples (27887)	<a href="#">Click for detailed sample information and links to data</a>			
Protocols (1)	<a href="#">Click for detailed protocol information</a>			
Description	A compiled human gene expression, ontology-annotated dataset from publicly available data including 27887 Affymetrix HG-U133Plus2 arrays, filtered for quality control. The dataset contains samples from healthy individuals, from individuals with disease (including cancer) and cell lines.			
Experiment types	transcription profiling by array, array specific design, cell type comparison design, disease state design, organism part comparison design			
Contact	 <a href="mailto:aurora@ebi.ac.uk">Aurora Torrente &lt;aurora@ebi.ac.uk&gt;</a>			
MIAME	<div><div>—</div><div>*</div><div>*</div><div>*</div><div>*</div></div> <div>Platforms   Protocols   Variables   Processed   Raw</div>			
Files	<div>Investigation description   <a href="#">E-MTAB-3732.idf.txt</a></div> <div>Sample and data relationship   <a href="#">E-MTAB-3732.sdrf.txt</a></div> <div>Processed data (1)   <a href="#">E-MTAB-3732.processed.1.zip</a></div> <div> <a href="#">Click to browse all available files</a></div>			
Links	<a href="#">Send E-MTAB-3732 data to</a>  GENOMESPACE			

- If the dataset is about methylation, it is irrelevant.

## Ulcerative colitis, E-GEOD-27899

# E-GEOD-27899 - Methylation profiling in Ulcerative colitis

Status	Released on 27 September 2012, last updated on 2 May 2014		
Organism	Homo sapiens		
Samples (40)	<a href="#">Click for detailed sample information and links to data</a>		
Arrays (2)	<a href="#">A-GEOD-8490 - Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)</a> <a href="#">A-GEOD-13269 - Nimblegen Human high-density promoter array CpG island methylation array</a>		
Protocols (12)	<a href="#">Click for detailed protocol information</a>		
Description	The series was designed to identify the different methylated single CpGs involved in the pathophysiology of ulcerative colitis. A cohort of n=20 monozygotic twins, discordant for ulcerative colitis was selected. Illumina and Nimblegen platforms were used.		
Experiment type	methylation profiling by array		
Contacts	 <a href="mailto:Zhe.Feng@ncbi.nlm.nih.gov">Zhe Feng &lt;Zhe.Feng@ncbi.nlm.nih.gov&gt;</a> , Robert Häsler		
MIAME	     Platforms   Protocols   Variables   Processed   Raw		
Files	Investigation description <a href="#">↓ E-GEOD-27899.idf.txt</a> Sample and data relationship <a href="#">↓ E-GEOD-27899.sdrf.txt</a> Raw data (2) <a href="#">↓ E-GEOD-27899.raw.1.zip</a> , <a href="#">↓ E-GEOD-27899.raw.2.zip</a> Processed data (1) <a href="#">↓ E-GEOD-27899.processed.1.zip</a> Array designs <a href="#">↓ A-GEOD-13269.adf.txt</a> , <a href="#">↓ A-GEOD-8490.adf.txt</a>  <a href="#">Click to browse all available files</a>		
Links	<a href="#">GEO - GSE27899</a> <a href="#">Send E-GEOD-27899 data to</a>  GENOMESPACE		