NCBI resources III: GEO and ftp site

Yanbin Yin

Spring 2013

Homework assignment 2

- Search "colon cancer" at GEO and find a data Series and perform a GEO2R analysis

- Write a report (in word or ppt) to include all the operations and screen shots

Due on Feb 12 (send by email or bring printed hard copy to class)

Office hour:
Tue, Thu and Fri 2-4pm, MO325A
Or email: yyin@niu.edu

Gene Expression Omnibus (GEO)

http://www.ncbi.nlm.nih.gov/geo/

GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community.

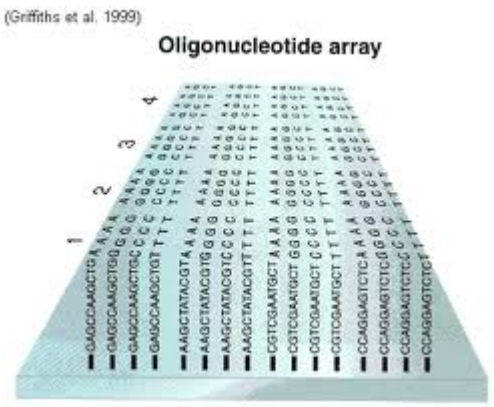
The three main goals of GEO are to:

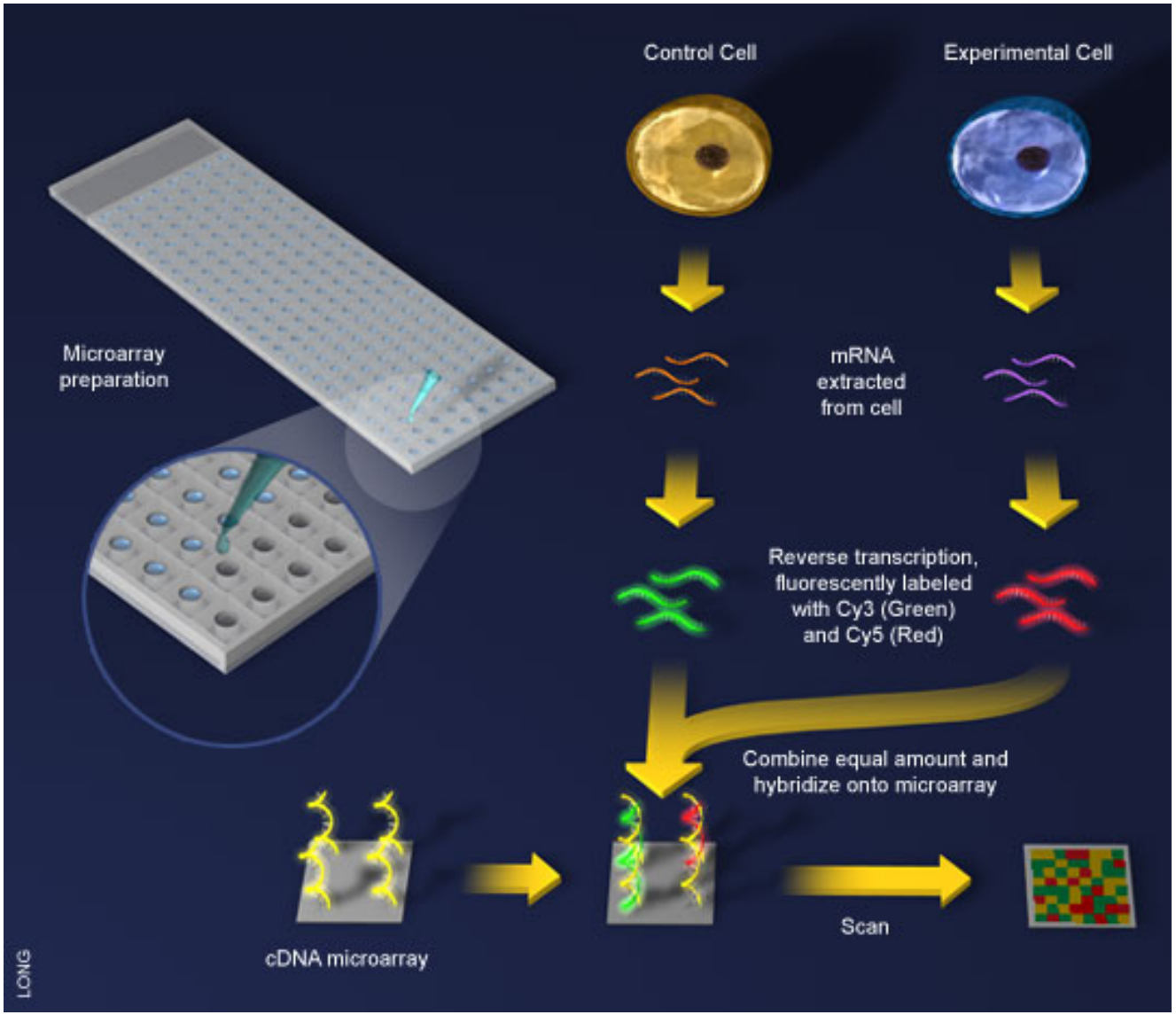
Provide a robust, versatile database in which to efficiently store high-throughput functional genomic data

Offer simple submission procedures and formats that support complete and well-annotated data deposits from the research community

Provide user-friendly mechanisms that allow users to query, locate, review and download studies and gene expression profiles of interest (Query and analysis)

Basic intro to microarray



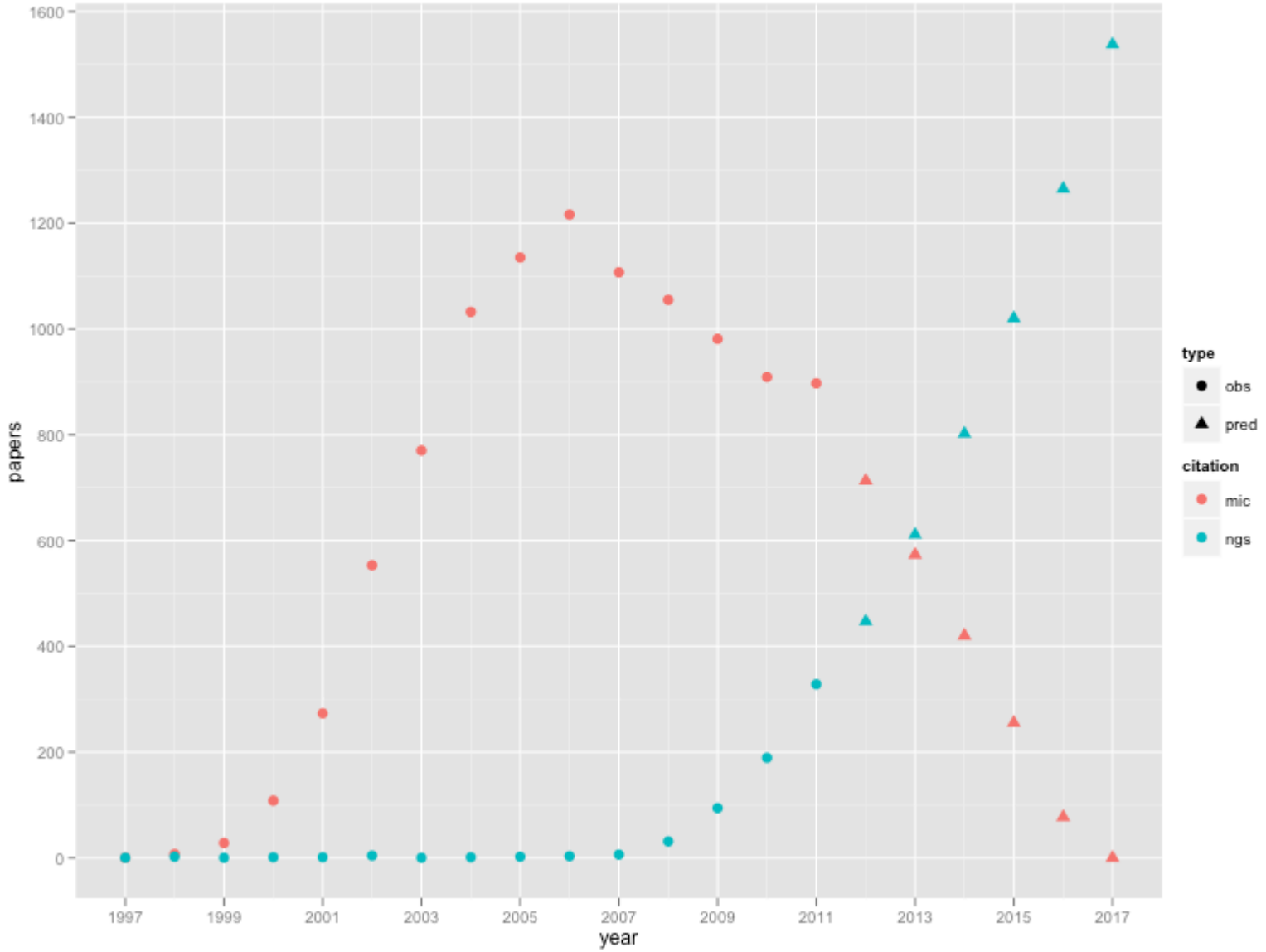


People are moving from microarray to high throughput sequencing

Table 1 | **Advantages of RNA-Seq compared with other transcriptomics methods**

Technology	Tiling microarray	cDNA or EST sequencing	RNA-Seq
Technology specifications			
Principle	Hybridization	Sanger sequencing	High-throughput sequencing
Resolution	From several to 100 bp	Single base	Single base
Throughput	High	Low	High
Reliance on genomic sequence	Yes	No	In some cases
Background noise	High	Low	Low
Application			
Simultaneously map transcribed regions and gene expression	Yes	Limited for gene expression	Yes
Dynamic range to quantify gene expression level	Up to a few-hundredfold	Not practical	>8,000-fold
Ability to distinguish different isoforms	Limited	Yes	Yes
Ability to distinguish allelic expression	Limited	Yes	Yes
Practical issues			
Required amount of RNA	High	High	Low
Cost for mapping transcriptomes of large genomes	High	High	Relatively low

When can we expect the last microarray paper?



http://jermdemo.blogspot.com/2012/01/when-can-we-expect-last-damn-microarray.html

What data does GEO have?

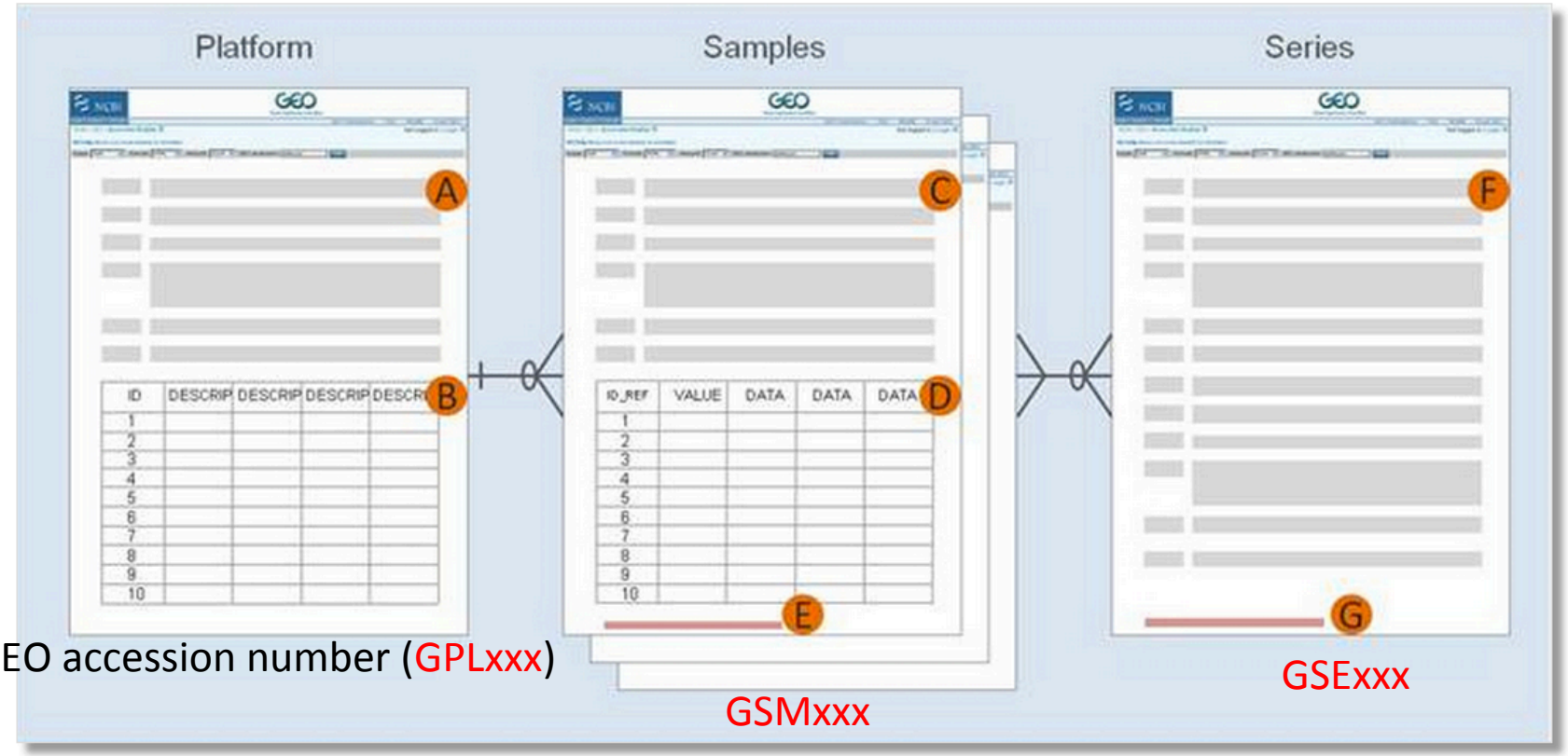
http://www.ncbi.nlm.nih.gov/geo/

- Submitter supplied: Platform, Sample, Series
- NCBI curated: DataSets and Profiles
- Tools: GEO BLAST and GEO2R

Public data	
Platforms	11,021
Samples	871,896
Series	35,642
DataSets	2,720

Omics data:

Genomics
Transcriptomics
Epigenomics
Proteomics
…



For almost all array data submissions, you will be asked to provide the following information:

A	Text description of the array or sequencer
B	Text tab-delimited table of the array template
C	Text description of the biological sample and protocols to which it was subjected
D	Text tab-delimited table of processed hybridization result or sequence counts
E	Raw data file, or processed sequence data file
F	Text description of the overall experiment
G	Tar archive of raw data files, or processed sequence data files

Series | Platforms | Samples | Organisms | History

Technology	Count
in situ oligonucleotide	4,199
spotted oligonucleotide	2,537
spotted DNA/cDNA	2,749
antibody	18
MS	16
SAGE NlaIII	67
SAGE Sau3A	4
SAGE RsaI	1
SARST	2
MPSS	17
RT-PCR	128
other	125
oligonucleotide beads	179
mixed spotted oligonucleotide/cDNA	14
spotted peptide or protein	46
high-throughput sequencing	918

Microarray

NGS

Series | Platforms | Samples | Organisms | History

Series type	Count
Expression profiling by array	27,297
Expression profiling by genome tiling array	486
Expression profiling by high throughput sequencing	1,117
Expression profiling by SAGE	231
Expression profiling by MPSS	19
Expression profiling by RT-PCR	115
Expression profiling by SNP array	11
Genome variation profiling by array	447
Genome variation profiling by genome tiling array	774
Genome variation profiling by high throughput sequencing	38
Genome variation profiling by SNP array	563
Genome binding/occupancy profiling by array	127
Genome binding/occupancy profiling by genome tiling array	1,603
Genome binding/occupancy profiling by high throughput sequencing	1,468
Genome binding/occupancy profiling by SNP array	9
Methylation profiling by array	262
Methylation profiling by genome tiling array	330
Methylation profiling by high throughput sequencing	212
Methylation profiling by SNP array	6
Protein profiling by protein array	80
Protein profiling by Mass Spec	4
SNP genotyping by SNP array	315
Other	449
Non-coding RNA profiling by array	1,053
Non-coding RNA profiling by genome tiling array	100
Non-coding RNA profiling by high throughput sequencing	713
Third-party reanalysis	57

Expression

Genome variation

DNA-binding

Methylation/
Epigenomics

Protein array

ncRNAs

Series | Platforms | **Samples** | Organisms | History

Sample type	Count
RNA	663,297
genomic	158,478
protein	5,934
SAGE	1,735
mixed	3,276
other	5,229
SARST	9
MPSS	207
SRA	33,731

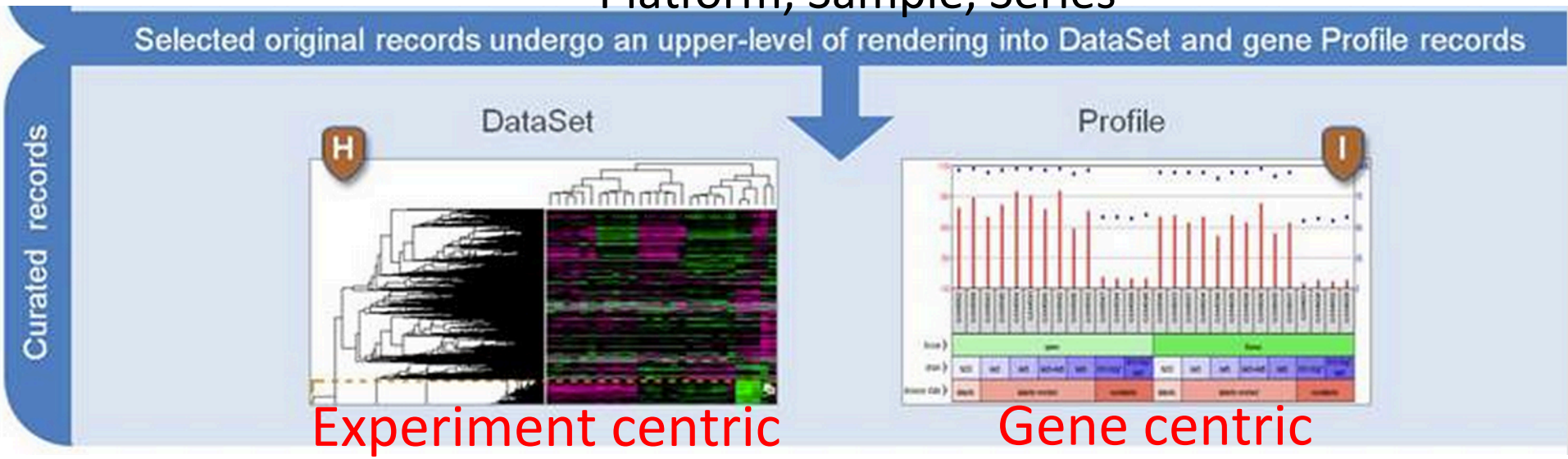
Series | Platforms | Samples | Organisms | History

See full list of organisms

Organism	Series	Platforms	Samples
Homo sapiens	13,290	3,763	487,496
Mus musculus	9,001	1,631	139,231
Rattus norvegicus	1,631	372	37,987
Saccharomyces cerevisiae	1,307	491	24,944
Arabidopsis thaliana	1,726	279	20,911
Drosophila melanogaster	1,603	269	15,913
Sus scrofa	263	72	6,609
Caenorhabditis elegans	703	159	5,384
Bos taurus	276	110	5,079
Glycine max	122	32	4,697
Zea mays	176	74	4,387
Escherichia coli	394	109	4,022
Oryza sativa	345	159	3,837
Gallus gallus	259	78	3,775
Macaca mulatta	160	27	2,563
Xenopus laevis	86	21	806

http://www.ncbi.nlm.nih.gov/geo/info/overview.html

Platform, Sample, Series



Data of a GEO Series are reassembled by GEO staff into GEO Dataset records (GDSxxx).

A DataSet represents a curated collection of biologically and statistically comparable GEO Samples and forms the basis of GEO's suite of data display and analysis tools.

Not all submitted data are suitable for DataSet assembly, so not all Series have corresponding DataSet record(s).

Profiles are derived from DataSets

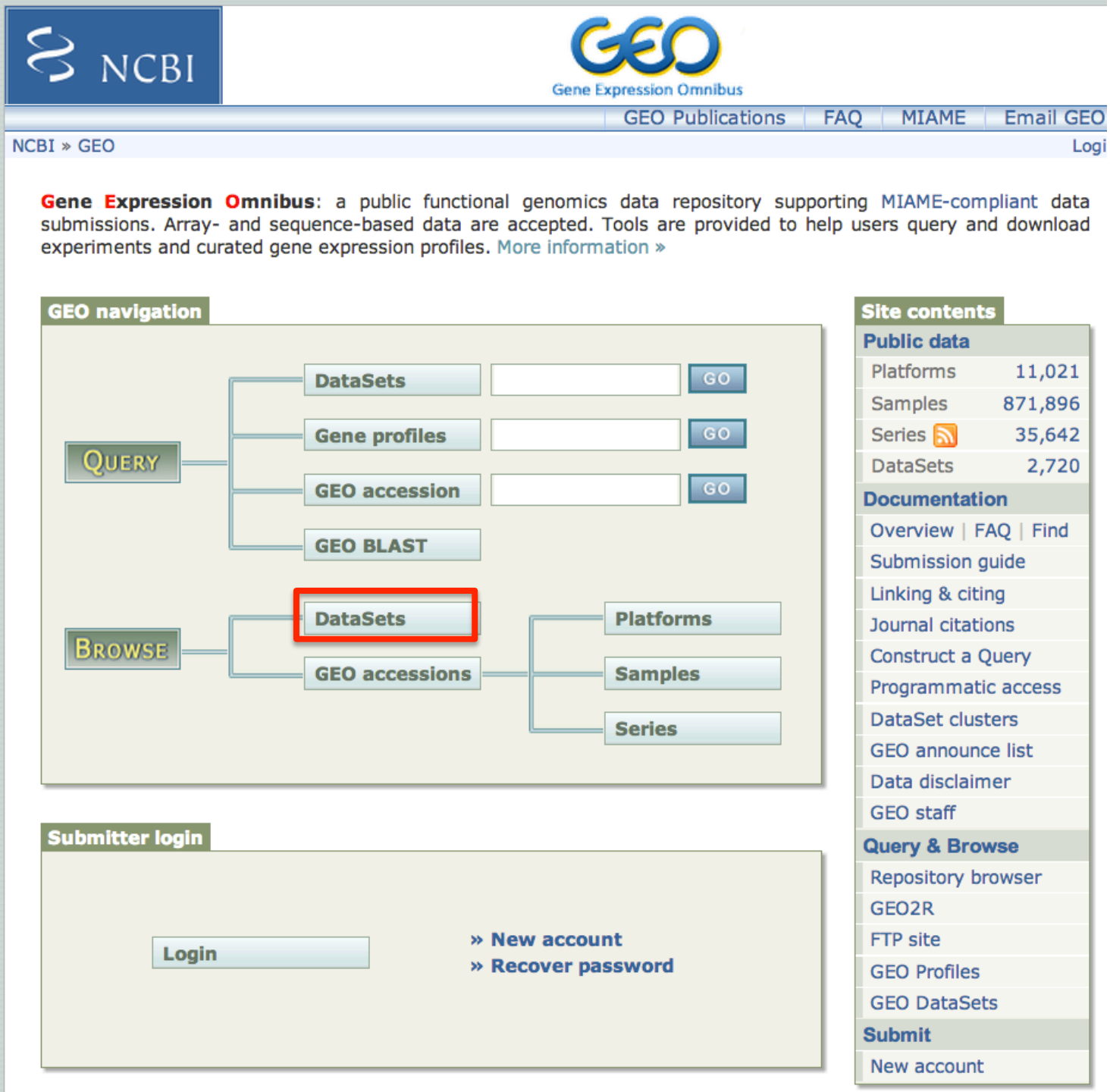
A Profile consists of the expression measurements for an individual gene across all Samples in a DataSet.

Total holdings

	Public	Unreleased	Total
Series	35,642	5,521	41,163
Platforms	11,021	450	11,471
Samples	871,896	140,563	1,012,459

Hands on exercise 1

GEO browse and query


NCBI
Gene Expression Omnibus
GEO Publications
FAQ
MIAME
Email GEO
NCBI » GEO
Logi
Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »
GEO navigation
QUERY
DataSets
GO
Gene profiles
GO
GEO accession
GO
GEO BLAST
BROWSE
DataSets
GEO accessions
Platforms
Samples
Series
Submitter login
Login
» New account
» Recover password
Site contents
Public data
Platforms 11,021
Samples 871,896
Series 35,642
DataSets 2,720
Documentation
Overview | FAQ | Find
Submission guide
Linking & citing
Journal citations
Construct a Query
Programmatic access
DataSet clusters
GEO announce list
Data disclaimer
GEO staff
Query & Browse
Repository browser
GEO2R
FTP site
GEO Profiles
GEO DataSets
Submit
New account

Try:
cancer
colon cancer
arabidopsis
ecoli

These are only DataSets

Search for [ ] Search Clear Show All Advanced Search Page size

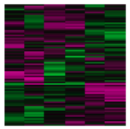
2720 DataSet records Page 1 of 136

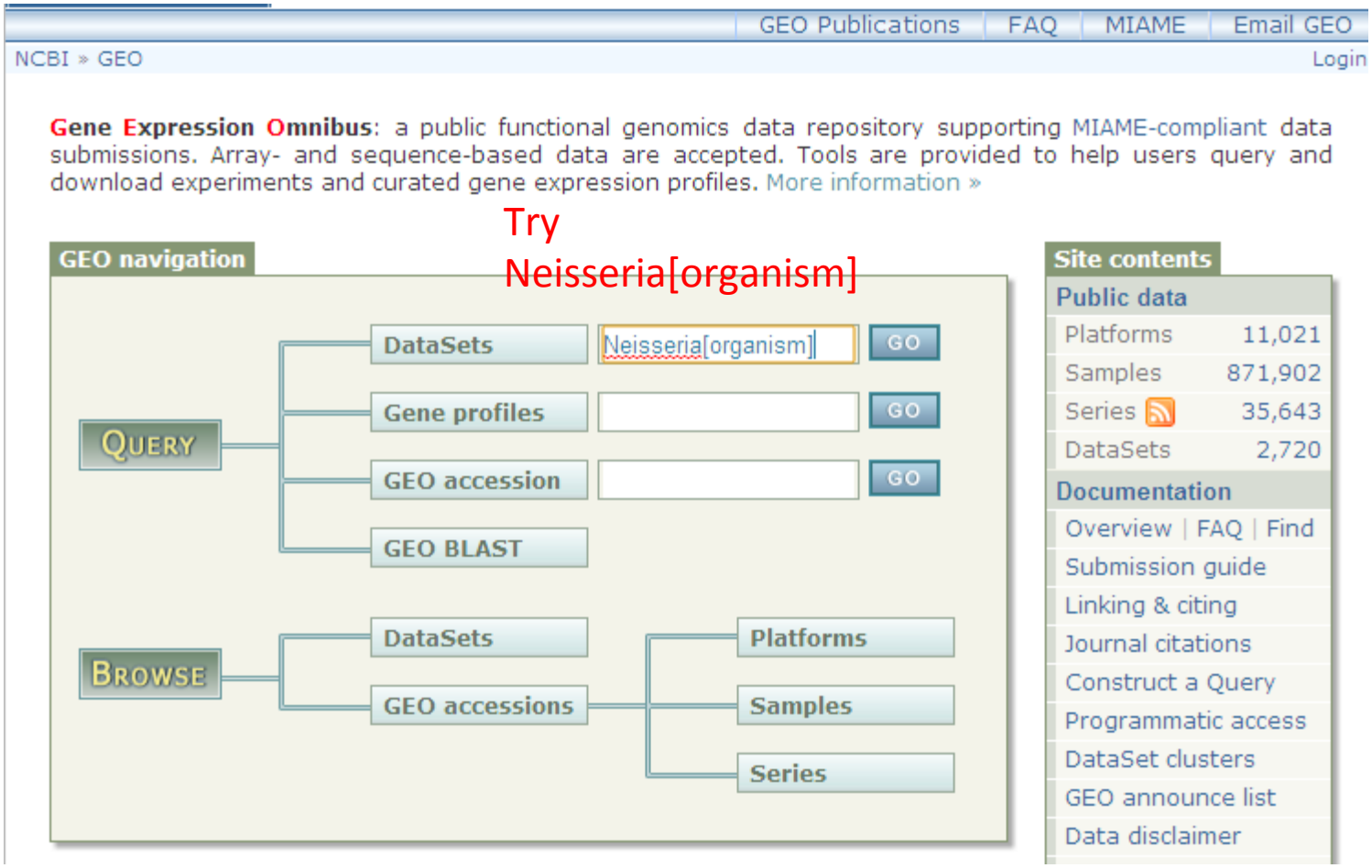
DataSet	Title	Organism(s)	Platform	Series	Sar
GDS3719	Transcription factor Ovo1 depletion effect on 12 hour post-fertilization embryos	*Danio rerio*	GPL1319	GSE21539	
GDS3718	Zinc finger Zbtb20 deficiency effect on the developing hippocampus	*Mus musculus*	GPL1261	GSE19513	
GDS3717	NOTCH antagonist SAHM1 effect on T-ALL cell lines	*Homo sapiens*	GPL570	GSE18198	
GDS3716	Breast cancer: histologically normal breast epithelium	*Homo sapiens*	GPL96	GSE20437	
GDS3715	Insulin effect on skeletal muscle	*Homo sapiens*	GPL91	GSE22309	
GDS3711	Asthmatic atopic epithelium	*Homo sapiens*	GPL96	GSE18965	
GDS3710	TGF-beta-induced epithelial-mesenchymal transition model	*Homo sapiens*	GPL570	GSE17708	
GDS3709	Cigarette smoke effect on the oral mucosa	*Homo sapiens*	GPL570	GSE17913	
GDS3707	Acute ethanol exposure: time course	*Drosophila melanogaster*	GPL1322	GSE18208	
GDS3706	Benzo[a]pyrene diol epoxide effect on lung WI-38 fibroblasts: dose-response	*Homo sapiens*	GPL570	GSE19510	

DataSet Record GDS3719: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	Transcription factor Ovo1 depletion effect on 12 hour post-fertilization embryos
Summary:	Analysis of wild-type embryos injected with Ovo1 morpholino antisense oligos. At 12 hpf, the Ovo1 morphants begin showing neural crest (NC) defects wherein a subset of NC cells aggregates in the dorsal midline above the neural tube. Results provide insight into the role of Ovo1 in NC migration.
Organism:	*Danio rerio*
Platform:	GPL1319: [Zebrafish] Affymetrix Zebrafish Genome Array

Cluster Analysis




GEO Publications
FAQ
MIAME
Email GEO
NCBI » GEO
Login
Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »
Try
Neisseria[organism]
GEO navigation
QUERY
DataSets
Neisseria[organism]
GO
Gene profiles
GO
GEO accession
GO
GEO BLAST
BROWSE
DataSets
GEO accessions
Platforms
Samples
Series
Site contents
Public data
Platforms 11,021
Samples 871,902
Series 35,643
DataSets 2,720
Documentation
Overview | FAQ | Find
Submission guide
Linking & citing
Journal citations
Construct a Query
Programmatic access
DataSet clusters
GEO announce list
Data disclaimer

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

Send to:

Results: 1 to 20 of 684

<< First < Prev Page 1 of 35 Next > Last >>

1. Zur regulon of **Neisseria** meningitidis

(Submitter supplied) The zur regulon in **Neisseria** meningitidis was elucidated in the strain MC58 using a zur knockout strain and conditions which activate Zur (zinc supplementation in the medium)

Organism: **Neisseria** meningitidis
Type: Expression profiling by array
Platform: GPL8787 11 Samples
Download data: GEO (GPR, TXT)
Series Accession: GSE38033 ID: 200038033
PubMed Analyze with GEO2R

2. MpeR regulation in **Neisseria** gonorrhoeae (F19) through an iron-responsive mechanism

(Submitter supplied) Previous studies have shown that the MpeR transcriptional regulator produced by **Neisseria** gonorrhoeae represses expression of mtrF, which encodes a putative inner membrane protein that works with the MtrC-MtrD-MtrE efflux pump to allow gonococci to resist high levels of multiple hydrophobic antimicrobials. Regulation of mpeR has been reported to occur by an iron-dependent mechanism involving Fur (Ferric uptake regulator). more...

Organism: **Neisseria** gonorrhoeae
Type: Expression profiling by array
Platform: GPL7218 12 Samples
Download data: GEO (CEL, TXT)
Series Accession: GSE32717 ID: 200032717
PubMed Full text in PMC Similar studies Analyze with GEO2R

3. Comparative genomic hybridization study of **Neisseria** meningitidis

(Submitter supplied) PFGRC has developed a cost effective alternative to complete genome sequencing in order to study the genetic differences between closely related species and/or strains. The comparative genomics approach combines Gene Discovery (GD) and Comparative Genomic Hybridization (CGH) techniques, resulting in the design and production of species microarrays that represent the diversity of a species beyond just

Filter your results:

All (684)
DataSets (0)
Platforms (23)
Samples (640)
Series (21)

Manage

Top Organisms [Tree]

Neisseria meningitidis *(595)*
Neisseria gonorrhoeae *(87)*
Neisseria gonorrhoeae FA 1090 *(25)*
Neisseria meningitidis MC58 *(20)*
Neisseria gonorrhoeae F62 *(6)*
More

Find related data

Database: Select

Find items

Search details

"Neisseria"[Organism]


NCBI
Gene Expression Omnibus
GEO Publications
FAQ
MIAME
Email GEO
NCBI » GEO
Logi
Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »
GEO navigation
QUERY
DataSets
GO
Gene profiles
GO
GEO accession
GO
GEO BLAST
BROWSE
DataSets
GEO accessions
Platforms
Samples
Series
Submitter login
Login
» New account
» Recover password
Site contents
Public data
Platforms 11,021
Samples 871,896
Series 35,642
DataSets 2,720
Documentation
Overview | FAQ | Find
Submission guide
Linking & citing
Journal citations
Construct a Query
Programmatic access
DataSet clusters
GEO announce list
Data disclaimer
GEO staff
Query & Browse
Repository browser
GEO2R
FTP site
GEO Profiles
GEO DataSets
Submit
New account

Querying GEO DataSets and GEO Profiles

- Quick examples
- How to construct queries
- Tables of query fields and examples

term [field] OPERATOR term [field]

Quick examples

GEO DataSets | GEO Profiles

This database stores original submitter-supplied study descriptions, as well as curated gene expression DataSets. DataSets form the basis of GEO's advanced data display and analysis tools, including gene expression profile charts and clusters.

Search Examples:

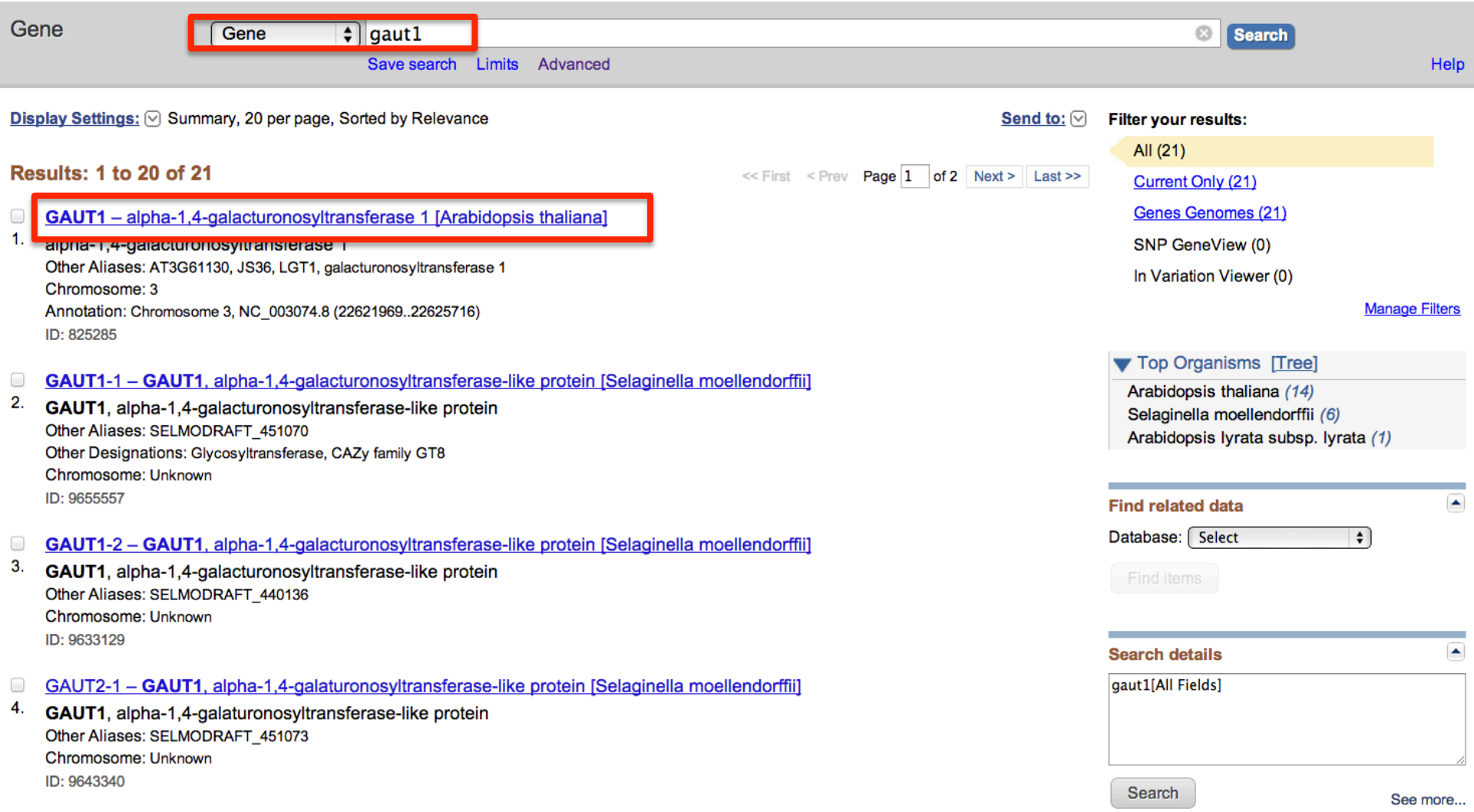
Search by...	Search text
Free text	smoking cancer
Keywords and species	(smok* OR diet) AND (mammals[organism] NOT human[organism])
Studies in the **NIH Roadmap Epigenomics project**	"roadmap epigenomics"[Project]
Study type	"expression profiling by high throughput sequencing"[DataSet Type]
Studies with between 100 and 500 samples	100:500[Number of Samples]
Studies with CEL files	"cel"[Supplementary Files]
DataSets that have 'age' as an experimental variable	"age"[Subset Variable Type]
Author	smith a[Author]
Published between January and June 2007	2007/01:2007/06[Publication Date]
Platform accession	GPL570
	"gds pubmed"[Filter]

cbi.nlm.nih.gov/geo/info/qqtutorial.html#profiles-table

Hands on exercise 2

GEO gene profiles

Search for a gene: GAUT1



Display Settings: Full Report

Send to:

GAUT1 alpha-1,4-galacturonosyltransferase 1 [*Arabidopsis thaliana*]

Gene ID: 825285, updated on 6-Jan-2013

Summary

Gene symbol	GAUT1
Gene description	alpha-1,4-galacturonosyltransferase 1
Primary source	TAIR:AT3G61130
Locus tag	AT3G61130
Gene type	protein coding
RNA name	alpha-1,4-galacturonosyltransferase 1
RefSeq status	REVIEWED
Organism	Arabidopsis thaliana (ecotype: Columbia)
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis
Also known as	galacturonosyltransferase 1; GAUT1; JS36; LGT1
Summary	Encodes a protein with putative galacturonosyltransferase activity.

Genomic context

Location: chromosome: 3

Sequence: Chromosome: 3; NC_003074.8 (22621969..22625716)

See GAUT1 in MapViewer

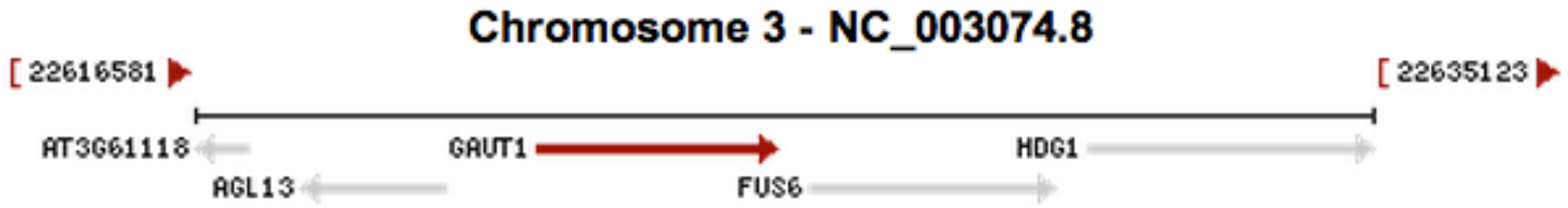


Table of contents

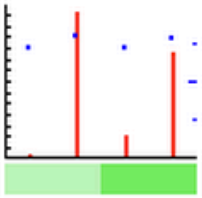
Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Interactions
General gene info
General protein info
Reference sequences
Related sequences
Additional links

Related information

BioProjects
BioSystems
Conserved Domains
EST
Full text in PMC
Genome
GEO Profiles
HomoloGene

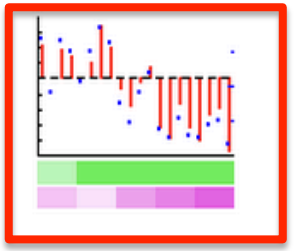
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors

15. GAUT1 - Alternative oxidase anti-sense silencing effect on leaves



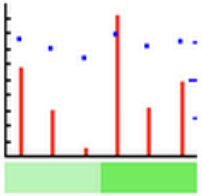
Annotation: GAUT1, GAUT1 (GALACTURONOSYLTRANSFERASE 1); polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups
Organism: Arabidopsis thaliana
Reporter: GPL198, 251308_at (ID_REF), GDS1532, At3g61130 (ORF)
DataSet type: Expression profiling by array, count, 4 samples
ID: 15716008
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors

16. GAUT1 - Stem development



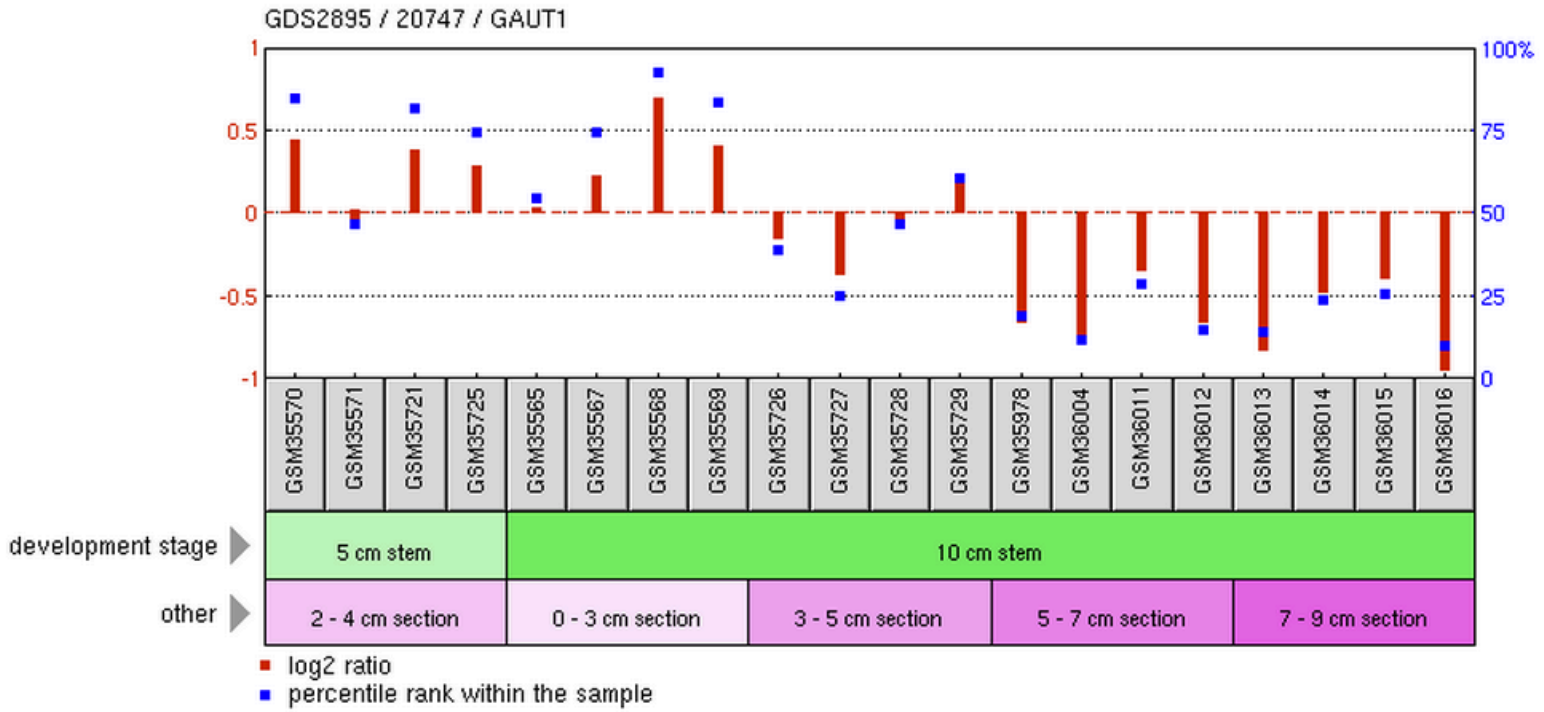
Annotation: GAUT1, GAUT1 (GALACTURONOSYLTRANSFERASE 1); polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups
Organism: Arabidopsis thaliana
Reporter: GPL1713, 20747 (ID_REF) GDS2895, AY039515, At3g61130 (ORF)
DataSet type: Expression profiling by array, log2 ratio, 20 samples
ID: 43520847
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

17. GAUT1 - MicroRNA miR159a overexpression effect on flower



Annotation: GAUT1, GAUT1 (GALACTURONOSYLTRANSFERASE 1); polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups
Organism: Arabidopsis thaliana
Reporter: GPL198, 251308_at (ID_REF), GDS2063, At3g61130 (ORF)
DataSet type: Expression profiling by array, count, 6 samples
ID: 24707408

Profile GDS2895 / 20747 / GAUT1
Title Stem development
Organism Arabidopsis thaliana



Graph caption help

Sample	Value	Rank
GSM35570	0.456367	85
GSM35571	-0.00232472	47
GSM35721	0.388293	82
GSM35725	0.297085	75
GSM35565	0.0420401	55
GSM35567	0.23765	75
GSM35568	0.7	93
GSM35569	0.421382	84

Profile neighbors: what are the co-expressed genes sharing similar expression profiles?

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

Send to:

Filters: Manage Filters

Results: 1 to 20 of 200

<< First < Prev Page 1 of 10 Next > Last >>

GAUT1 - Stem development

1. Annotation: GAUT1, GAUT1 (GALACTURONOSYLTRANSFERASE 1); polygalacturonate 4-alpha-galacturonosyltransferase/ transferase, transferring glycosyl groups
Organism: Arabidopsis thaliana
Reporter: GPL1713, 20747 (ID_REF), GDS2895, AY039515, At3g61130 (ORF)
DataSet type: Expression profiling by array, log2 ratio, 20 samples
ID: 43520847
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

Stem development

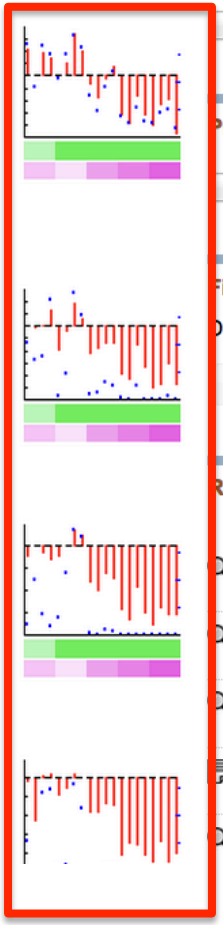
2. Annotation: Arabidopsis thaliana chromosome 1 BAC F3C3 genomic sequence, complete sequence
Organism: Arabidopsis thaliana
Reporter: GPL1713, 28681 (ID_REF), GDS2895, AC084165, At1g32080 (ORF)
DataSet type: Expression profiling by array, log2 ratio, 20 samples
ID: 43528781
GEO DataSets Profile neighbors Sequence neighbors

Stem development

3. Annotation: Arabidopsis thaliana DNA chromosome 4, contig fragment No. 43
Organism: Arabidopsis thaliana
Reporter: GPL1713, 24698 (ID_REF), GDS2895, AL161543, At4g16180 (ORF)
DataSet type: Expression profiling by array, log2 ratio, 20 samples
ID: 43524798
GEO DataSets Profile neighbors Sequence neighbors

Stem development

4. Annotation: Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21H1
Organism: Arabidopsis thaliana
Reporter: GPL1713, 19369 (ID_REF), GDS2895, AB020742, At5g67180 (ORF)



Profile data

Download profile data

Profile pathways

Find pathways

Find related data

Database: Select

Find items

Recent activity

Turn Off

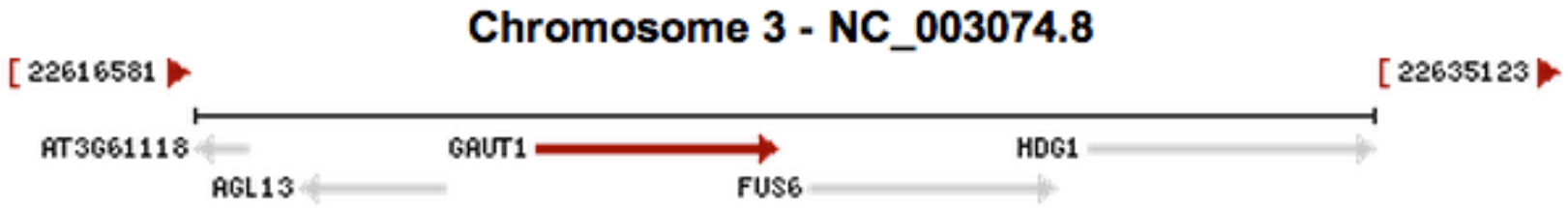
Profile neighbors for GEO Profiles (Sel 43520847) (200)

GEO Profiles for Gene (Select 825285)

gaut1 (21)

GAUT1 [Arabidopsis thaliana]

Gene Links for GEO Profiles (Select



Chromosome neighbors:
are neighboring genes co-expressed?

Results: 20

1. AT3G60880 - Stem development
   Annotation: AT3G60880, dihydrodipicolinate synthase 1 (DHDPS1) (DHDPS) (DHPS1)
   Organism: Arabidopsis thaliana
   Reporter: GPL1713, 24307 (ID_REF), GDS2895, X72971, At3g60880 (ORF)
   DataSet type: Expression profiling by array, log2 ratio, 20 samples
   ID: 43524407
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

2. AtPP2-A13 - Stem development
   Annotation: AtPP2-A13, AtPP2-A13 (Arabidopsis thaliana phloem protein 2-A13); carbohydrate binding
   Organism: Arabidopsis thaliana
   Reporter: GPL1713, 22586 (ID_REF), GDS2895, AY034967, At3g61060 (ORF)
   DataSet type: Expression profiling by array, log2 ratio, 20 samples
   ID: 43522686
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

3. FUS6 - Stem development
   Annotation: FUS6, FUS6 (FUSCA 6)
   Organism: Arabidopsis thaliana
   Reporter: GPL1713, 21904 (ID_REF), GDS2895, AF360295, At3g61140 (ORF)
   DataSet type: Expression profiling by array, log2 ratio, 20 samples
   ID: 43522004
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

4. SYP73 - Stem development
   Annotation: SYP73, SYP73 (SYNTAXIN OF PLANTS 73); protein transporter
   Organism: Arabidopsis thaliana
   Reporter: GPL1713, 21301 (ID_REF), GDS2895, AF355759, At3g61450 (ORF)
   DataSet type: Expression profiling by array, log2 ratio, 20 samples
   ID: 43521401
   GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors

GAUT1 - Stem development

Profile data

Download profile data

Profile pathways

Find pathways

Find related data

Database: Select

Find items

Recent activity

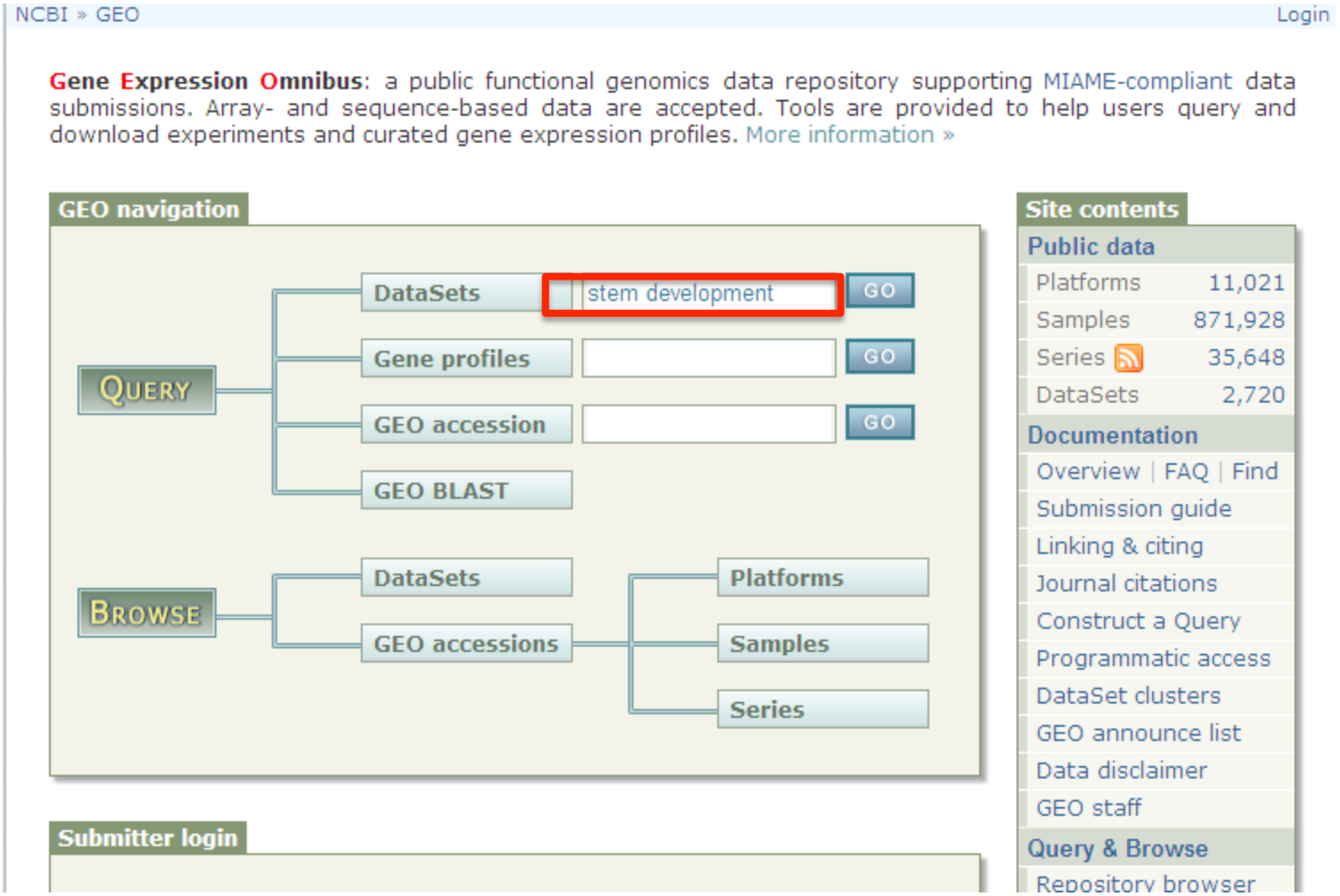
Turn Off

- Chromosome neighbors for GEO Profiles (Select 43520847) (20)
- Similar studies for GEO DataSets (Select 200013043) (20)
- (poplar stem) AND "Populus trichocarpa" [porgn] (59)
- poplar stem (319)
- arabidopsis stem (467)

Hands on exercise 3

GEO DataSets analysis tool

Search:
stem development



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

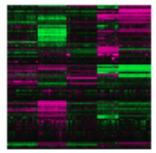
Send to:

Results: 1 to 20 of 2062

<< First < Prev Page 1 of 104 Next > Last >>

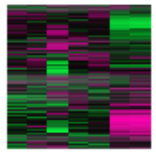
☐ Atherosclerotic Coronary Artery Disease: circulating mononuclear cell types

1. Analysis of various mononuclear cells from patients with severe triple-vessel coronary artery disease (CAD). CD34+ **stem** cells, CD4+ T-helper cells, CD14+ resting monocytes, LPS-stimulated monocytes, and macrophages were examined. Results provide insight into the pathophysiology of atherosclerosis.
Organism: Homo sapiens
Type: Expression profiling by array, transformed count, 5 cell type, 2 disease state, 31 individual sets
Platform: GPL6255 Series: GSE9820 153 Samples
Download data: GEO
DataSet Accession: GDS3690 ID: 3690
PubMed Similar studies GEO Profiles Analyze DataSet

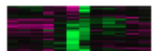


☐ Calreticulin deficiency effect on embryonic **stem** cell line

2. Analysis of D3 embryonic **stem** cells lacking calreticulin (CRT). CRT is a calcium binding protein.
Organism: Mus musculus
Type: Expression profiling by array, count, 2 genotype/variation sets
Platform: GPL1261 Series: GSE13805 7 Samples
Download data: GEO (CEL)
DataSet Accession: GDS3680 ID: 3680
PubMed Full text in PMC Similar studies GEO Profiles Analyze DataSet



☐ Fasting effect on the liver



Filter your results:

All (2062)
DataSets (39)
Platforms (36)
Samples (1289)
Series (698)

Top Organisms [Tree]
Mus musculus *(1227)*
Homo sapiens *(326)*
Hordeum vulgare *(169)*
Vitis vinifera *(104)*
Arabidopsis thaliana *(65)*
More...

Find related data

Database: Select

Find items

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

Send to:

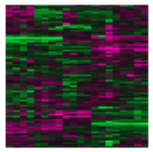
Results: 1 to 20 of 65

<< First < Prev Page 1 of 4 Next > Last >>

1. **Stem development**

Analysis of sections from 5 and 10 cm long ecotype Ler bolting stems. Different stages of vascular and interfascicular fiber differentiation can be identified along the axis of bolting stems. Results provide insight into the molecular mechanisms controlling this pattern of **development**.

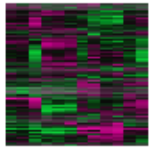
Organism: **Arabidopsis thaliana**
Type: Expression profiling by array, log2 ratio, 2 **development** stage, 5 other sets
Platform: GPL1713 Series: GSE2000 20 Samples
Download data: GEO
DataSet Accession: GDS2895 ID: 2895
PubMed Similar studies GEO Profiles Analyze DataSet



2. Auxin response transcription factor ARF6 and ARF8 mutations effect on flower **development**

Analysis of mutant flowers heterozygous for arf6 but homozygous for arf8, and mutants double homozygous for both arf6 and arf8. Mutants examined at various stages of **development**. ARF6 and ARF8 are auxin response factors that coordinate the transition from immature to mature fertile flowers.

Organism: **Arabidopsis thaliana**
Type: Expression profiling by array, count, 4 **development** stage, 3 genotype/variation, 2 tissue sets
Platform: GPL198 Series: GSE2848 12 Samples
Download data: GEO (CEL)
DataSet Accession: GDS2114 ID: 2114



Filter your results:

All (65)
DataSets (2)
Platforms (2)
Samples (43)
Series (18)

Find related data

Database: Select

Find items

Search details

```
(("plant stems"[MeSH Te
"microscopy, electron,
transmission"[MeSH Term
stem[All Fields]) AND (
```

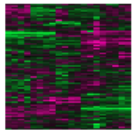
Search for GDS2895[ACCN] Search Clear Show All Advanced Search

DataSet Record GDS2895: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	Stem development
Summary:	Analysis of sections from 5 and 10 cm long ecotype Ler bolting stems. Different stages of vascular and interfascicular fiber differentiation can be identified along the axis of bolting stems. Results provide insight into the molecular mechanisms controlling this pattern of development.
Organism:	*Arabidopsis thaliana*
Platform:	GPL1713: GBC_FOAR03_0001
Citation:	Ehlting J, Mattheus N, Aeschliman DS, Li E et al. Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant J* 2005 Jun;42(5):618-40. PMID: 15918878

Reference Series:	GSE2000	**Sample count:**	20
Value type:	log2 ratio	**Series published:**	2005/11/18

Cluster Analysis



Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

Data Analysis Tools

- Find genes
- **Compare 2 sets of samples**
- Cluster heatmaps
- Experiment design and value distribution

Step 1: Select test and significance level

Two-tailed t-test (A vs B) Significance level: 0.100

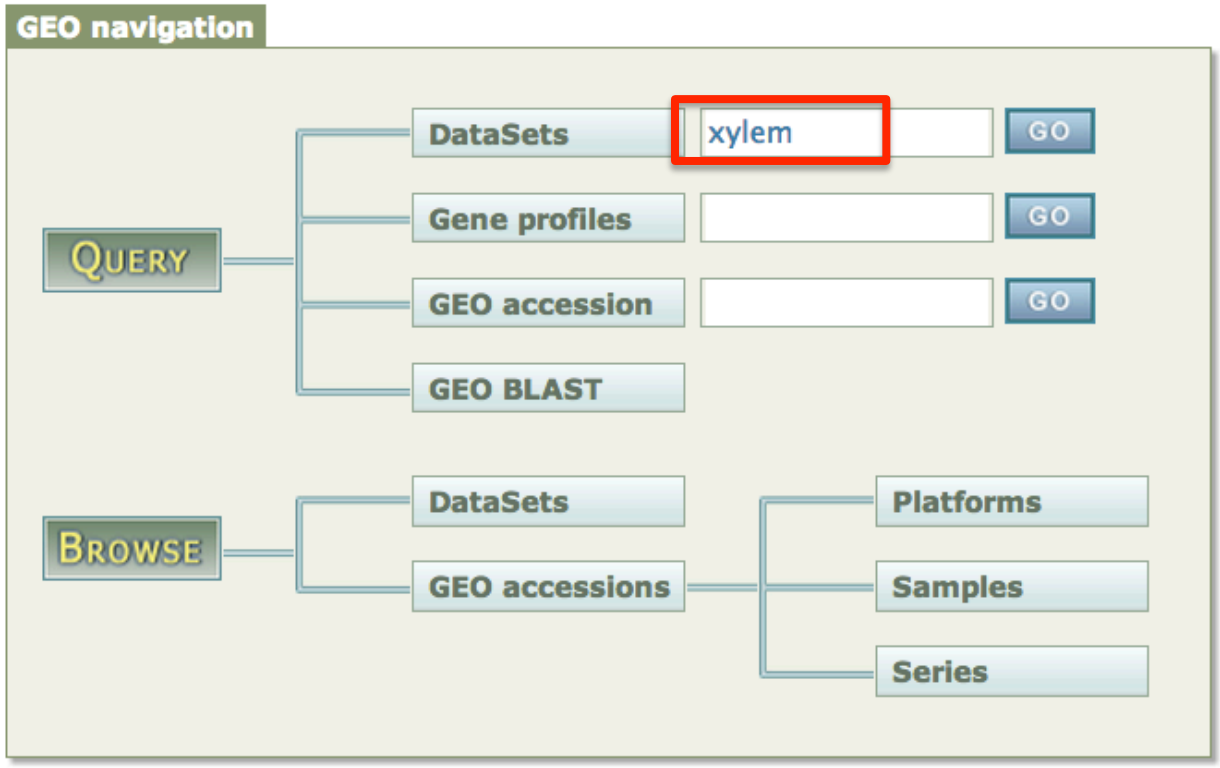
Step 2: Select which Samples to put in Group A and Group B

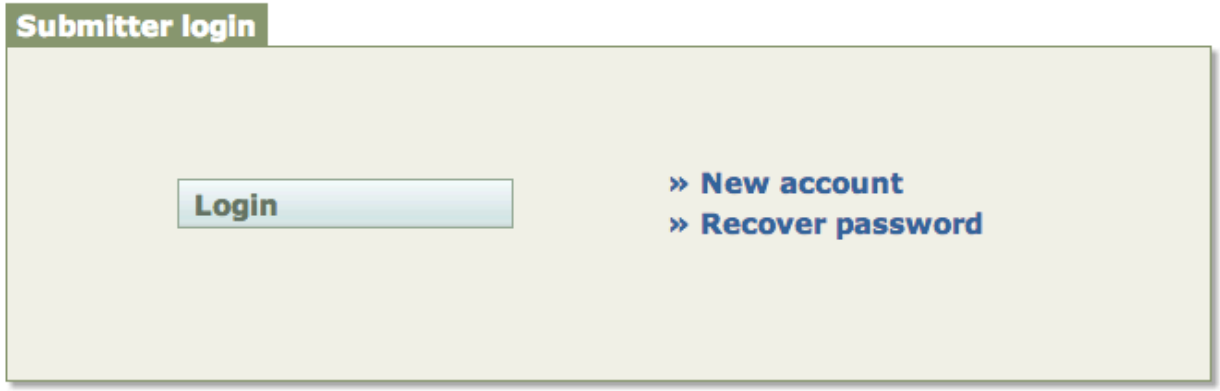
Step 3: Query Group A vs. B

GEO2R: differentially expressed genes

http://www.youtube.com/watch?v=EUPmGWS8ik0

Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »





Site contents

Public data

Platforms	11,019
Samples	872,327
Series	35,668
DataSets	2,720

Documentation

- Overview | FAQ | Find
- Submission guide
- Linking & citing
- Journal citations
- Construct a Query
- Programmatic access
- DataSet clusters
- GEO announce list
- Data disclaimer
- GEO staff

Query & Browse

- Repository browser
- GEO2R
- FTP site
- GEO Profiles
- GEO DataSets

Submit

- New account

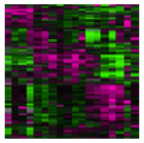
Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Results: 1 to 20 of 997

<< First < Prev Page 1 of 50 Next > Last >>

1. Adventitious root development



Analysis of hypocotyls treated with auxin to induce adventitious root formation. Hypocotyls were harvested at day 0, day 3 (cell expansion), day 6 (when root primordia are formed), day 9 (when root meristems are formed), day 12 (when roots are fully developed), and day 33 (root elongation).

Organism: Pinus contorta; Pinus taeda
Type: Expression profiling by array, count, 6 time sets
Platform: GPL1014 Series: GSE1261 18 Samples
Download data: GEO
DataSet Accession: GDS966 ID: 966
PubMed Full text in PMC Similar studies GEO Profiles Analyze DataSet

2. xyleme-Molecular bases of acclimation and adaptation to water deficit in poplar.

(Submitter supplied) affy_popsec_orleans_poplar - xyleme - This project aims to identify candidate genes for water deficit acclimation and/or adaptation in a tree species: poplar. Due to compelling evidence that transcriptional regulation plays a major role in regulating many biological processes, we will look for genes and gene expression networks related to drought stress. We intend to analyse the transcriptome in two poplar genotypes of contrasted tolerance to water deficit, at various stages and intensities of stress and simultaneously in whole **xylem** and cambial zone from young trees. more...

Organism: Populus sp.; Populus x canadensis
Type: Expression profiling by array
Platform: GPL4359 32 Samples
Download data: GEO (CEL)
Series Accession: GSE17220 ID: 200017220
Analyze with GEO2R

Filter your results:

All (997)
DataSets (1)
Platforms (11)
Samples (918)
Series (67)

Manage Filt

Top Organisms [Tree]

Eucalyptus grandis x Eucalyptus urophylla *(208)*
Populus trichocarpa x Populus deltoides *(183)*
Picea glauca *(118)*
Eucalyptus grandis *(102)*
Arabidopsis thaliana *(75)*
More...

Find related data

Database: Select

Find items

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

Send to:

Results: 1 to 20 of 75

<< First < Prev Page 1 of 4 Next > Last >>

Finding direct target genes of VND7

1. (Submitter supplied) The **Arabidopsis thaliana** NAC domain transcription factor, VASCULAR-RELATED NAC-DOMAIN7 (VND7), acts as a key regulator of **xylem** vessel differentiation. In order to identify direct target genes of VND7, we performed global transcriptome analysis using Arabidopsis transgenic lines in which VND7 activity could be induced post-translationally. This analysis identified 63 putative direct target genes of VND7, which encode a broad range of proteins, such as transcription factors, IRREGULAR **XYLEM** proteins and proteolytic enzymes, known to be closely associated with **xylem** vessel formation. more...

Organism: **Arabidopsis thaliana**
Type: Expression profiling by array
Platform: GPL198 20 Samples
Download data: GEO (CEL)
Series Accession: GSE24[illegible] ID: 200024[illegible]
PubMed Similar studies Analyze with GEO2R

Expression data from Arabidopsis suspension cells overexpressing VND6 and SND1

2. (Submitter supplied) **Xylem** consists of three types of cells: vessel cells, also referred to as tracheary elements (TEs), parenchyma cells, and fiber cells. TE differentiation includes two essential processes, programmed cell death (PCD) and secondary cell wall formation. These two processes are tightly coupled. However, little is known about the molecular mechanism of their gene regulation. Here, we show that VASCULAR-RELATED NAC-DOMAIN 6 (VND6), a master regulator of TEs, regulates these processes in a coordinated manner. more...

Organism: **Arabidopsis thaliana**
Type: Expression profiling by array
Platform: GPL198 18 Samples
Download data: GEO (CEL, CHP)
Series Accession: GSE20586 ID: 200020586
PubMed Full text in PMC Similar studies Analyze with GEO2R

Cell signalling by microRNA165/6 directs gene dose dependent root cell fate

3. (Submitter supplied) A key question in developmental biology is how cells exchange positional information for proper patterning during organ

Filter your results:

All (75)
DataSets (0)
Platforms (0)
Samples (60)
Series (15)

Manag

Find related data

Database: Select

Find items

Search details

xylem[All Fields] AND "Arabidopsis thaliana [porgn]

Search

Se

Recent activity

Turn O

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 GSE24169 Set Finding direct target genes of VND7

Samples

Define groups

Enter a group name: List

Cancel selection

1 (5 samples)

2 (5 samples)

3 (5 samples)

4 (5 samples)

Selected **20** out of **20** samples

Columns Set

Group	Accession	Title	Source name
1	GSM594701	10d seedlin GR treated with DEX plus CHX, rep1	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hr
1	GSM594702	10d seedlin GR treated with DEX plus CHX, rep2	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hr
1	GSM594703	10d seedlin GR treated with DEX plus CHX, rep3	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hr
1	GSM594704	10d seedlin GR treated with DEX plus CHX, rep4	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hr
1	GSM594705	10d seedlin GR treated with DEX plus CHX, rep5	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hr
2	GSM594706	10d seedlings expressing 35S:VND7-VP16-GR treated with CHX, rep1	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 6hrs
2	GSM594707	10d seedlings expressing 35S:VND7-VP16-GR treated with CHX, rep2	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 6hrs
2	GSM594708	10d seedlings expressing 35S:VND7-VP16-GR treated with CHX, rep3	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 6hrs
2	GSM594709	10d seedlings expressing 35S:VND7-VP16-GR treated with CHX, rep4	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 6hrs
2	GSM594710	10d seedlings expressing 35S:VND7-VP16-GR treated with CHX, rep5	10d whole seedlings expressing 35S:VND7-VP16-GR treated with 10mM CHX for 6hrs
3	GSM594711	10d seedlings harboring empty vector treated with DEX plus CHX, rep1	10d whole seedlings harboring empty vector treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hrs
3	GSM594712	10d seedlings harboring empty vector treated with DEX plus CHX, rep2	10d whole seedlings harboring empty vector treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hrs
3	GSM594713	10d seedlings harboring empty vector treated with DEX plus CHX, rep3	10d whole seedlings harboring empty vector treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hrs
3	GSM594714	10d seedlings harboring empty vector treated with DEX plus CHX, rep4	10d whole seedlings harboring empty vector treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hrs
3	GSM594715	10d seedlings harboring empty vector treated with DEX plus CHX, rep5	10d whole seedlings harboring empty vector treated with 10mM CHX for 2hrs followed by 10mM DEX for 4hrs
4	GSM594716	10d seedlings harboring empty vector treated with CHX, rep1	10d whole seedlings harboring empty vector treated with 10mM CHX for 6hrs
4	GSM594717	10d seedlings harboring empty vector treated with CHX, rep2	10d whole seedlings harboring empty vector treated with 10mM CHX for 6hrs
4	GSM594718	10d seedlings harboring empty vector treated with CHX, rep3	10d whole seedlings harboring empty vector treated with 10mM CHX for 6hrs

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 GSE24169 Set Finding direct target genes of VND7

▸ **Samples** ▸ Define groups Selected **20** out of **20** samples

GEO2R | Value distribution | Options | Profile graph | R script

▸ **Quick start**

Recalculate if you changed any options. Save all results Select columns

ID	adj.P.Val	P.Value	F	Gene.symbol	Gene.title
260173_at	2.03e-13	8.91e-18	552.7	VND7	VND7 (VASCULAR RELATED NA…
253191_at	5.61e-10	4.97e-14	206.19	XCP1	XCP1 (XYLEM CYSTEINE PEPTI…
265174_s_at	5.61e-10	7.39e-14	196.93	AT1G23460	polygalacturonase
260333_at	6.97e-10	1.22e-13	185.74	AT1G70500	polygalacturonase, putative / pect…
263629_at	9.15e-10	2.01e-13	175.36	AT2G04850	auxin-responsive protein-related
265672_at	2.96e-09	7.81e-13	149.72	AT2G31980	cysteine proteinase inhibitor-related
249173_at	1.34e-08	4.13e-12	123.14	AT5G43000	hypothetical protein
250322_at	2.10e-08	7.37e-12	115.01	MYB46	MYB46 (MYB DOMAIN PROTEIN…
249518_at	1.06e-07	4.18e-11	93.55	AT5G38610	invertase/pectin methylesterase i…
256803_at	9.35e-07	4.11e-10	70.99	CYP705A33	CYP705A33; electron carrier/ he…
252439_at	1.51e-06	7.31e-10	66.16	AT3G47400	pectinesterase family protein
247590_at	1.80e-06	9.55e-10	64.02	AT5G60720	hypothetical protein
249375_at	1.80e-06	1.05e-09	63.26	AGP24	AGP24
262796_at	1.80e-06	1.17e-09	62.42	XCP2	XCP2 (xylem cysteine peptidase …
247522_at	1.80e-06	1.19e-09	62.34	AT5G61340	hypothetical protein
262838_at	2.38e-06	1.67e-09	59.74	AT1G14960	major latex protein-related / MLP-…
262657_at	3.10e-06	2.32e-09	57.37	AT1G14210	ribonuclease T2 family protein
250231_at	3.90e-06	3.26e-09	54.98	LEP	LEP (LEAFY PETIOLE); DNA bin…

ftp

FTP stands for File Transfer Protocol.
HTTP stands for Hyper Text Transfer Protocol.

When ftp appears in a URL it means that the user is connecting to a file server and not a Web server and that some form of file transfer is going to take place.

When http appears in a URL it means that the user is connecting to a Web server and not a file server. The files are transferred but not downloaded, therefore not copied into the memory of the receiving device.

http://wiki.answers.com/Q/What_is_the_difference_between_FTP_and_HTTP

ftp server of NCBI

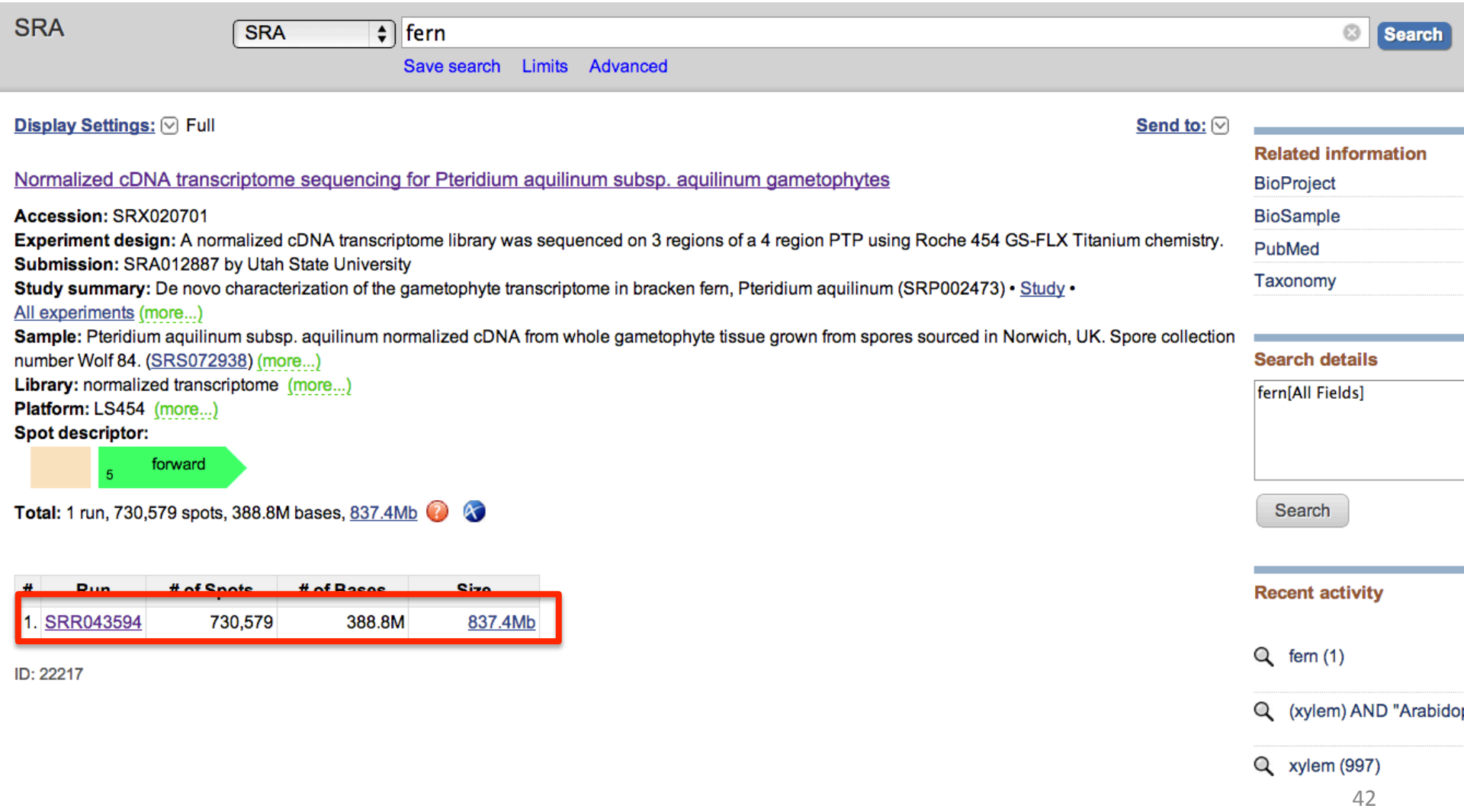
F

- Frequency-weighted Link (FLink)
- FTP: BLAST Databases
- FTP: CDD
- FTP: dbGAP Open-Access Data
- FTP: dbMHC Data
- FTP: FASTA BLAST Databases
- FTP: GenBank
- FTP: Gene
- FTP: Gene Expression Omnibus (GEO) Profiles and Datasets
- FTP: Genome
- FTP: Genome Mapping Data
- FTP: Genome Markers (UniSTS)
- FTP: GenPept
- FTP: HomoloGene
- FTP: NCBI Field Guide Manual
- FTP: NCBI Structure Course Materials
- FTP: NCBI Taxonomy
- FTP: Protein Clusters
- FTP: PubChem
- FTP: RefSeq
- FTP: Sequence Read Archive (SRA) Download Facility
- FTP: Site
- FTP: SKY/M-Fish and CGH Data
- FTP: SNP
- FTP: Structure (MMDB)
- FTP: Trace Archive
- FTP: UniGene
- FTP: UniVec
- FTP: Whole Genome Shotgun Sequences

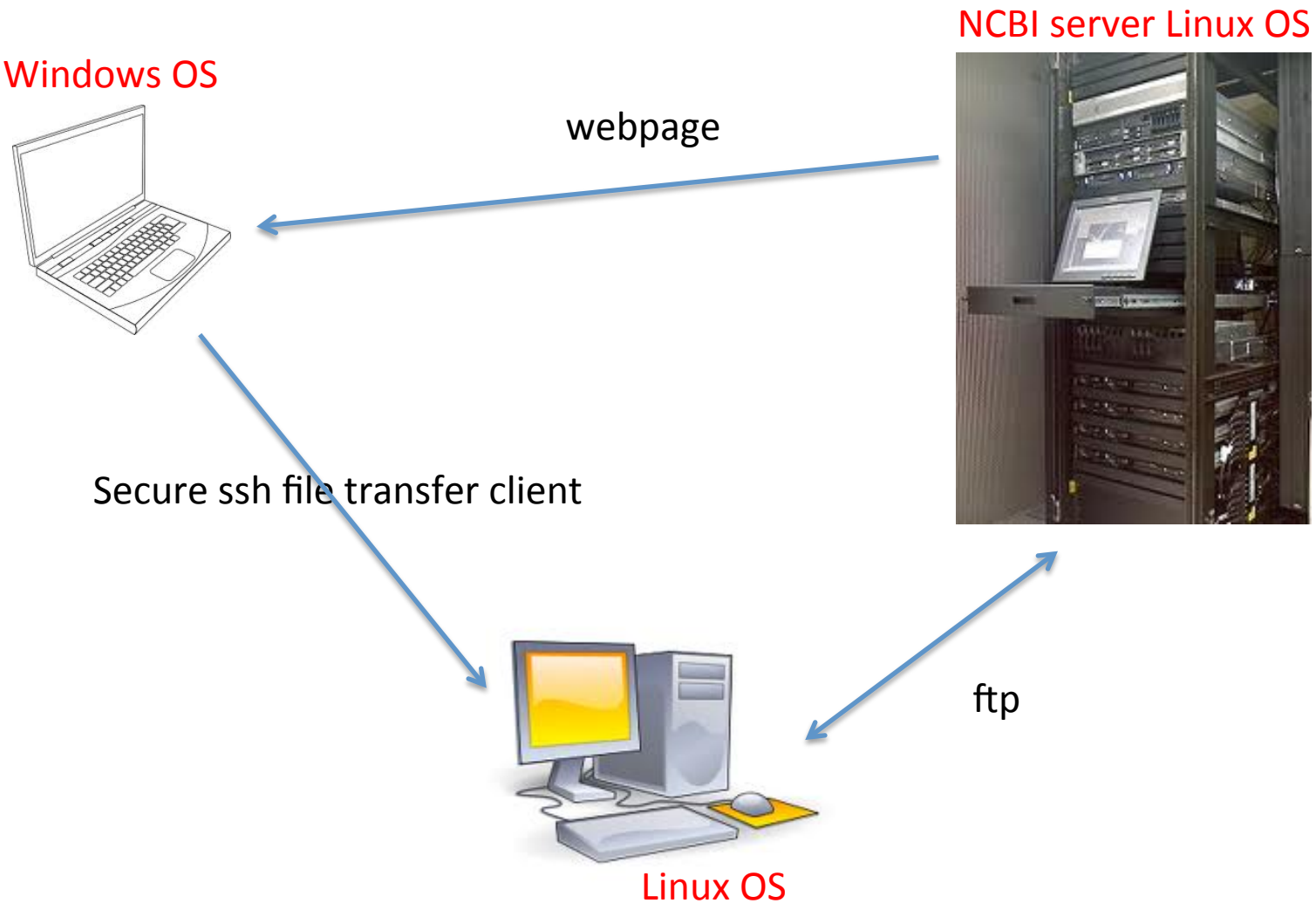
ftp resources

- Refseq genomes, proteins, mRNAs
- Microbial genomes
- Plant genomes
- Fungal genomes
- Blast database folder
- Sra reads
- Geo datasets

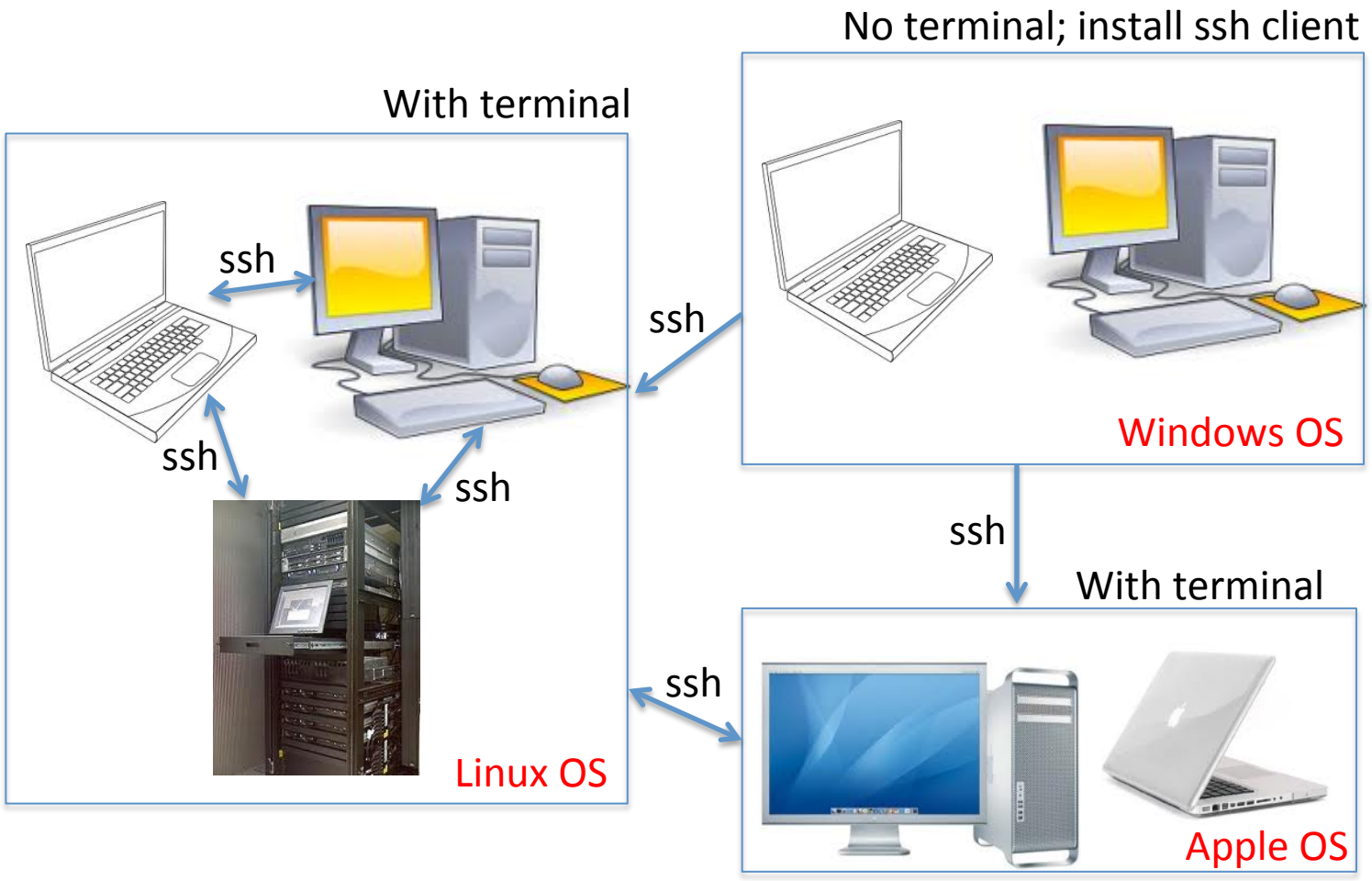
http download example



ftp connections



How to connect different machines using terminals?



Hands on exercise:
Linux terminal access to NCBI ftp sites

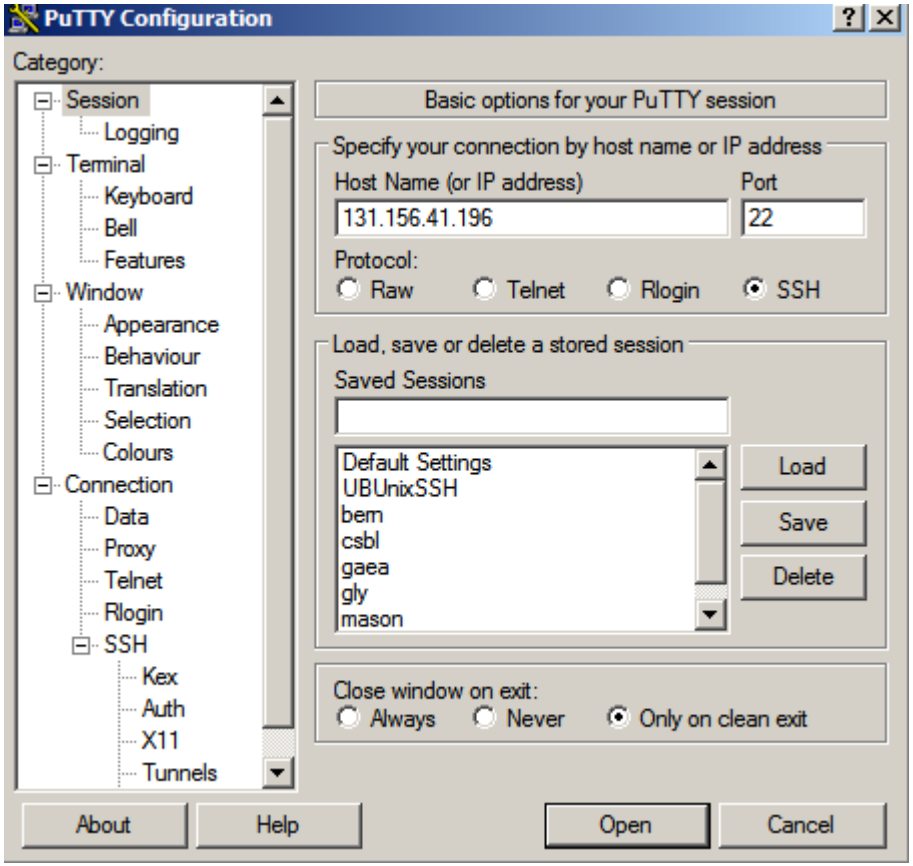
login info

IP 131.156.41.196

Accountstudent ID (e.g. z1003529)

Pswd student ID (e.g. z1003529)

z117479
z1245176
z1608050
z1576493
z1598039
z1559435
z1660438
z1003529
z1678230



PuTTY

```
yyin@glu:~$ ssh z117479@glu
z117479@glu's password:
Welcome to Ubuntu 12.04.1 LTS (GNU/Linux 3.2.0-35-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

66 packages can be updated.
25 updates are security updates.

Last login: Mon Jan 28 17:00:38 2013 from localhost
z117479@glu:~$ ls
examples.desktop
z117479@glu:~$ lftp ftp.ncbi.nih.gov
lftp ftp.ncbi.nih.gov:~>
```

```
lftp ftp.ncbi.nih.gov:/> ls
dr-xr-xr-x   3 ftp      anonymous     4096 May 27  2011 1000genomes
-r--r--r--   1 ftp      anonymous 10738466816 Dec  6 14:25 10GB
-r--r--r--   1 ftp      anonymous 1074790400 Dec  6 14:25 1GB
-r--r--r--   1 ftp      anonymous     1868 Dec  5 21:45 README.ftp
lr--r--r--   1 ftp      anonymous       29 Dec  5 22:08 asn1-converters -> toolbox/ncbi_tools/converters
dr-xr-xr-x  12 ftp      anonymous   184320 Dec 14 23:00 bigwig
dr-xr-xr-x   4 ftp      anonymous     4096 Jan 28 16:10 bioproject
dr-xr-xr-x   2 ftp      anonymous     4096 Jan 28 09:53 biosample
dr-xr-xr-x  10 ftp      anonymous     4096 May 24  2012 blast
dr-xr-xr-x   3 ftp      anonymous     4096 Sep 13  2004 cgap
dr-xr-xr-x   4 ftp      anonymous     4096 May 25  2011 cn3d
dr-xr-xr-x  30 ftp      anonymous     4096 Jan 14 20:20 dbgap
dr-xr-xr-x  11 ftp      anonymous     4096 Jun  4  2006 entrez
dr-xr-xr-x   7 ftp      anonymous     4096 Oct 13  2011 epigenomics
dr-xr-xr-x   6 ftp      anonymous     4096 Aug  4  2006 fa2htgs
-r--r--r--   1 ftp      anonymous     3262 Dec  5 21:45 favicon.ico
dr-xr-xr-x  12 ftp      anonymous    65536 Dec 21 18:02 genbank
dr-xr-xr-x   6 ftp      anonymous     4096 Dec 13 19:24 gene
dr-xr-xr-x  97 ftp      anonymous     4096 Jan  9 22:54 genomes
dr-xr-xr-x 1073741824 ftp      anonymous        0 Jan 28 20:06 geo
dr-xr-xr-x  25 ftp      anonymous     4096 Sep 20  2011 hapmap
dr-xr-xr-x  13 ftp      anonymous     4096 Jan 27  2012 mmdb
dr-xr-xr-x   6 ftp      anonymous    49152 Dec 21 17:05 ncbi-asn1
dr-xr-xr-x 164 ftp      anonymous     8192 Dec  2 17:49 pub
dr-xr-xr-x  11 ftp      anonymous     4096 Nov 29 13:17 pubchem
dr-xr-xr-x   2 ftp      anonymous     4096 Jan 28 09:03 pubmed
dr-xr-xr-x  15 ftp      anonymous     4096 Jan 14 18:59 refseq
dr-xr-xr-x  57 ftp      anonymous     4096 Aug 20  2008 repository
-r--r--r--   1 ftp      anonymous       26 Dec  5 21:45 robots.txt
dr-xr-xr-x   5 ftp      anonymous     4096 Nov 14 14:07 sequin
dr-xr-xr-x   9 ftp      anonymous     4096 May 24  2010 sky-cgh
dr-xr-xr-x  17 ftp      anonymous     4096 Jan 18 18:23 snp
dr-xr-xr-x  13 ftp      anonymous     4096 May 16  2012 sra
dr-xr-xr-x   2 ftp      anonymous     4096 Sep 29  2004 tech-reports
dr-xr-xr-x  12 ftp      anonymous     4096 Jun 29  2011 toolbox
dr-xr-xr-x   5 ftp      anonymous     4096 Apr 24  2009 tpa
dr-xr-xr-x   4 ftp      anonymous     4096 Sep 13 15:46 variation
lftp ftp.ncbi.nih.gov:/>
```

ls command:

list files and folders

All sequenced bacterial genomes

http://www.ncbi.nlm.nih.gov/genome/browse/

Genome Information by organism

Search by organism | Clear | Download Reports from FTP site

Overview [9348] | Eukaryotes [2391] | Prokaryotes [16577] | Viruses [3501]

Statuses: All | Complete | Scaffolds or contigs | SRA or Traces | No Data

First | Previous | Shown: 1 - 100 out of 2555 items | Next | Last | Download selected records

Organism/Name	BioProject	Group	SubGroup	Size (Mb)	GC%	Chromosomes		Plasmids		WGS	Scaffolds	Gene	Protein	Release Date	Modify Date	Status
		-- All Prokaryotes --	-- All Prokaryc			RefSeq	INSDC	RefSeq	INSDC							
Acaryochloris marina MBIC11017	PRJNA58167 PRJNA12997	Cyanobacteria	Chroococcales	8.36	46.99	NC_009925.1	CP000828.1	NC_009926.1 NC_009929.1 NC_009928.1 NC_009930.1 NC_009927.1 NC_009931.1 NC_009933.1 NC_009932.1 NC_009934.1	CP000838.1 CP000841.1 CP000840.1 CP000842.1 CP000839.1 CP000843.1 CP000845.1 CP000844.1 CP000846.1	-	-	8571	8383	2007/10/16	2008/02/21	Complete
Acetobacter pasteurianus IFO 3283-01	PRJNA59279 PRJDA31129	Proteobacteria	Alphaproteobacteria	3.34	53.07	NC_013209.1	AP011121.1	NC_013211.1 NC_013214.1 NC_013215.1 NC_013213.1 NC_013212.1 NC_013210.1	AP011123.1 AP011126.1 AP011127.1 AP011125.1 AP011124.1 AP011122.1	-	-	3121	3049	2009/08/27	2009/10/02	Complete
Acetobacter pasteurianus IFO 3283-01-42C	PRJNA158377 PRJDA31141	Proteobacteria	Alphaproteobacteria	3.25	53.16	NC_017150.1	AP011163.1	NC_017106.1 NC_017107.1 NC_017152.1 NC_017104.1 NC_017105.1 NC_017151.1	AP011167.1 AP011168.1 AP011169.1 AP011164.1 AP011165.1 AP011166.1	-	-	3050	2984	2009/08/27	2012/06/18	Complete
Acetobacter pasteurianus IFO 3283-03	PRJNA158373 PRJDA31131	Proteobacteria	Alphaproteobacteria	3.34	53.07	NC_017100.1	AP011128.1	NC_017101.1 NC_017142.1 NC_017109.1 NC_017118.1 NC_017119.1	AP011129.1 AP011132.1 AP011134.1 AP011130.1 AP011131.1	-	-	3120	3048	2009/08/27	2012/06/18	Complete

Where bacterial genomes are in the ftp site?

```
lftp ftp.ncbi.nih.gov:/> cd genomes/
lftp ftp.ncbi.nih.gov:/genomes> cd Bacteria
lftp ftp.ncbi.nih.gov:/genomes/Bacteria>
```

The end of the page after ls

```
dr-xr-xr-x   2 ftp      anonymous     4096 Dec  6  2010 _Nostoc_azollae__0708_uid49725
-r--r--r--   1 ftp      anonymous 455294518 Jan 28 08:08 all.GeneMark.tar.gz
-r--r--r--   1 ftp      anonymous 102841288 Jan 28 08:14 all.Glimmer3.tar.gz
-r--r--r--   1 ftp      anonymous 227990948 Jan 28 12:02 all.Prodigal.tar.gz
-r--r--r--   1 ftp      anonymous 4532882964 Jan 28 08:42 all.asn.tar.gz
-r--r--r--   1 ftp      anonymous 1515030808 Jan 28 09:03 all.faa.tar.gz
-r--r--r--   1 ftp      anonymous 2190975020 Jan 28 09:20 all.ffn.tar.gz
-r--r--r--   1 ftp      anonymous 2416170041 Jan 28 09:55 all.fna.tar.gz
-r--r--r--   1 ftp      anonymous  9095708 Jan 28 10:29 all.frn.tar.gz
-r--r--r--   1 ftp      anonymous 6720488727 Jan 28 10:45 all.gbk.tar.gz
-r--r--r--   1 ftp      anonymous 548255240 Jan 28 11:37 all.gff.tar.gz
-r--r--r--   1 ftp      anonymous 182962489 Jan 28 11:41 all.ptt.tar.gz
-r--r--r--   1 ftp      anonymous  2600313 Jan 28 11:43 all.rnt.tar.gz
-r--r--r--   1 ftp      anonymous   357580 Jan 28 11:44 all.rpt.tar.gz
-r--r--r--   1 ftp      anonymous 4107354627 Jan 28 11:53 all.val.tar.gz
dr-xr-xr-x   2 ftp      anonymous     4096 Sep 25 04:11 alpha_proteobacterium_HIMB59_uid175778
dr-xr-xr-x   2 ftp      anonymous     4096 Sep 25 04:12 alpha_proteobacterium_HIMB5_uid175779
dr-xr-xr-x   2 ftp      anonymous     4096 Dec 20 05:32 bacterium_BT_1_uid184079
dr-xr-xr-x   2 ftp      anonymous     4096 Dec  6  2010 cyanobacterium_UCYN_A_uid43697
dr-xr-xr-x   2 ftp      anonymous     4096 Dec  6  2010 gamma_proteobacterium_HdN1_uid51635
dr-xr-xr-x   2 ftp      anonymous     4096 Sep  1  2011 halophilic_archaeon_DL31_uid72619
dr-xr-xr-x   2 ftp      anonymous     4096 Aug 20 04:10 secondary_endosymbiont_of_Ctenarytaina_eucalypti_uid172737
dr-xr-xr-x   2 ftp      anonymous     4096 Nov  2 16:10 secondary_endosymbiont_of_Heteropsylla_cubana_Thao2000_uid172738
dr-xr-xr-x   2 ftp      anonymous     4096 Jan 28  2011 uncultured_Termite_group_1_bacterium_phylotype_Rs_D17_uid59059
lftp ftp.ncbi.nih.gov:/genomes/Bacteria>
```

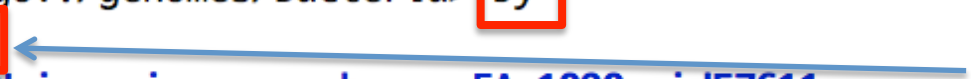
cd ne

Then press tab key to auto-complete or list

```
lftp ftp.ncbi.nih.gov:/genomes/Bacteria> cd Ne
Neisseria_gonorrhoeae_FA_1090_uid57611/          Neisseria_meningitidis_M01_240355_uid162075/
Neisseria_gonorrhoeae_NCCP11945_uid59191/        Neisseria_meningitidis_M04_240196_uid162081/
Neisseria_gonorrhoeae_TCDC_NG08107_uid161097/    Neisseria_meningitidis_MC58_uid57817/
Neisseria_lactamica_020_06_uid60851/             Neisseria_meningitidis_NZ_05_33_uid162077/
Neisseria_meningitidis_053442_uid58587/          Neisseria_meningitidis_WUE_2594_uid162093/
Neisseria_meningitidis_8013_uid161967/           Neisseria_meningitidis_Z2491_uid57819/
Neisseria_meningitidis_FAM18_uid57825/           Neisseria_meningitidis_alpha14_uid61649/
Neisseria_meningitidis_G2136_uid162085/          Neisseria_meningitidis_alpha710_uid161971/
Neisseria_meningitidis_H44_76_uid162083/         Neorickettsia_risticii_Illinois_uid58889/
Neisseria_meningitidis_M01_240149_uid162079/     Neorickettsia_sennetsu_Miyayama_uid57965/
lftp ftp.ncbi.nih.gov:/genomes/Bacteria> cd Ne
```

Download the folder using mirror

```
lftp ftp.ncbi.nih.gov:/genomes/Bacteria> mirror Neisseria_gonorrhoeae_FA_1090_uid57611/
Total: 1 directory, 15 files, 0 symlinks
New: 15 files, 0 symlinks
23930081 bytes transferred in 5 seconds (4.36M/s)
lftp ftp.ncbi.nih.gov:/genomes/Bacteria> by
z117479@glu:~$ ls
examples.desktop  Neisseria_gonorrhoeae_FA_1090_uid57611
z117479@glu:~$
```



You just left NCBI ftp site and are listing files in glu

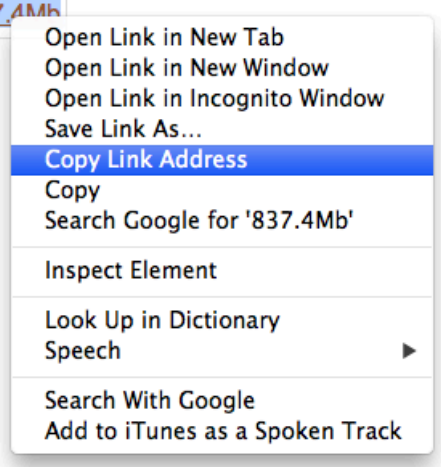
Normalized cDNA transcriptome sequencing for Pteridium aquilinum subsp. aquilinum gametophytes

Accession: SRX020701
Experiment design: A normalized cDNA transcriptome library was sequenced on 3 regions of a 4 region PTP using Roche 454 GS-FLX Titanium chemistry.
Submission: SRA012887 by Utah State University
Study summary: De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum (SRP002473) • Study • All experiments (more...)
Sample: Pteridium aquilinum subsp. aquilinum normalized cDNA from whole gametophyte tissue grown from spores sourced in Norwich, UK. Spore collection number Wolf 84. (SRS072938) (more...)
Library: normalized transcriptome (more...)
Platform: LS454 (more...)
Spot descriptor:

Total: 1 run, 730,579 spots, 388.8M bases, 837.4Mb

#	Run	# of Spots	# of Bases	Size
1.	SRR043594	730,579	388.8M	837.4Mb

ID: 22217



lftp and paste the address

```
z117479@glu:~$ lftp ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR043/SRR043594
cd ok, cwd=/sra/sra-instant/reads/ByRun/sra/SRR/SRR043/SRR043594
lftp ftp-trace.ncbi.nlm.nih.gov:/sra/sra-instant/reads/ByRun/sra/SRR/SRR043/SRR043594> ls
-r--r--r--   1 ftp      anonymous 878100441 Jan 19  2012 SRR043594.sra
lftp ftp-trace.ncbi.nlm.nih.gov:/sra/sra-instant/reads/ByRun/sra/SRR/SRR043/SRR043594>
```

To download the data

```
lftp ftp-trace.ncbi.nlm.nih.gov:/sra/sra-instant/reads/ByRun/sra/SRR/SRR043/SRR043594> get SRR043594.sra
Interrupt
```

To stop before finish, ctrl+c

Next lecture: `EBI resources I`