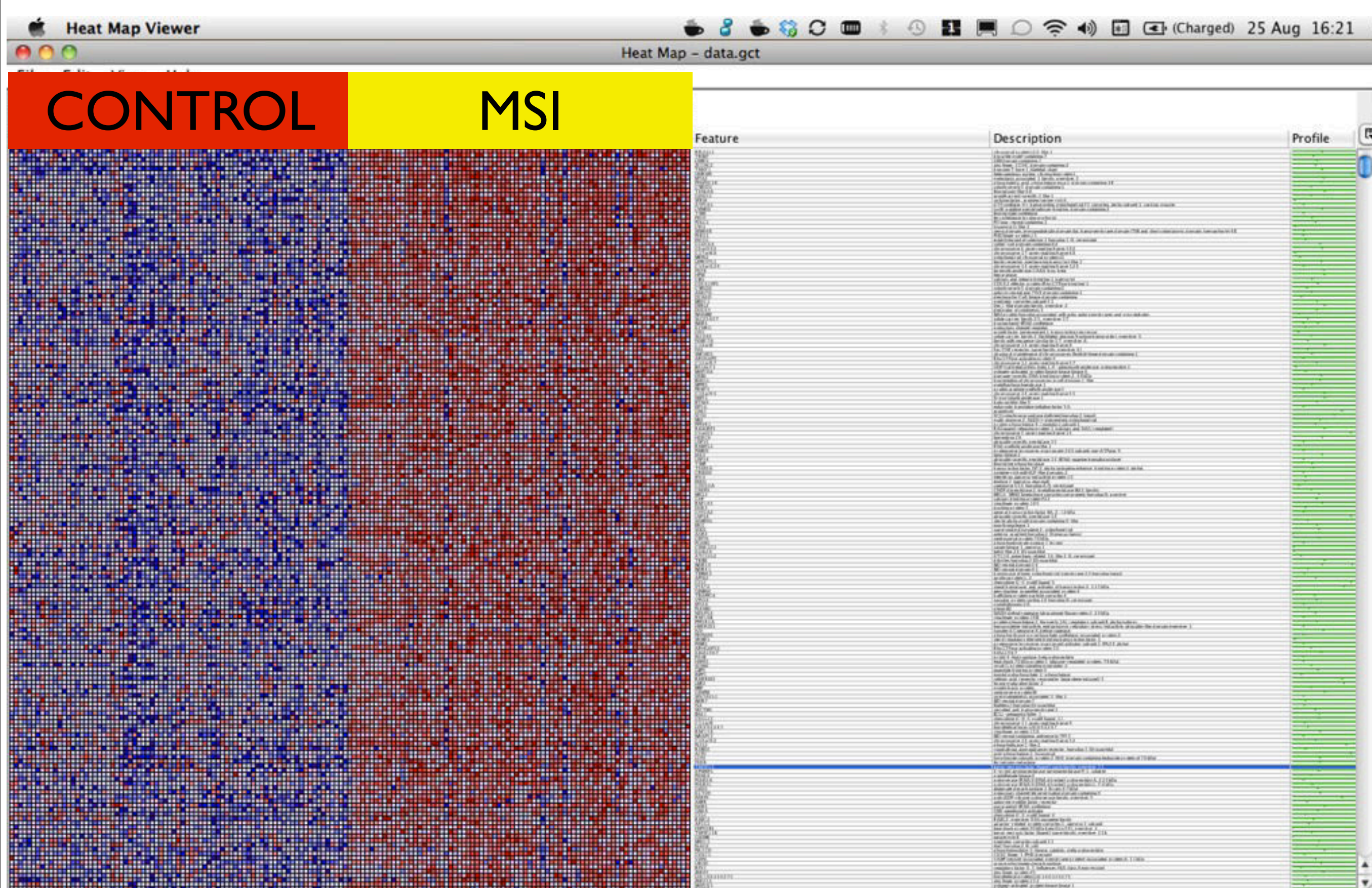


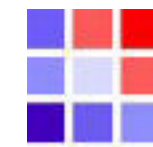
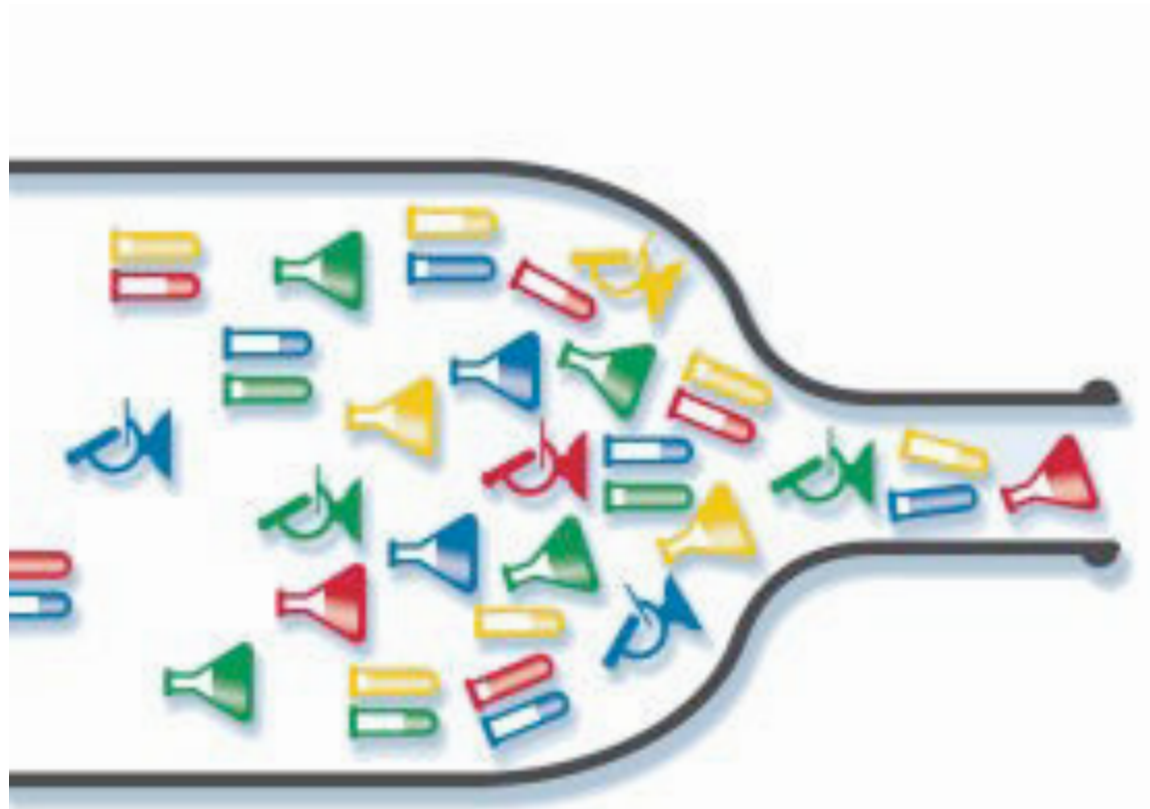
Clinical Annotations and gene expression data to find biomarkers



Clinical Annotations and formatting import bottleneck

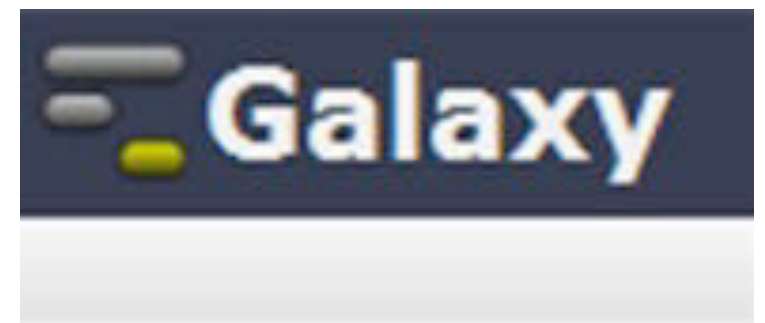


Gene Expression Omnibus



GenePattern

A platform for integrative genomics



Gene Expression Omnibus data is unstructured

The screenshot shows the NCBI GEO website interface. At the top left is the NCBI logo, and at the top right is the GEO logo (Gene Expression Omnibus). Below the logos are navigation links for HOME, SEARCH, and SITE MAP. The main content area displays the breadcrumb path: NCBI > GEO > Accession Display. A help message states: "GEO help: Mouse over screen elements for information." Below this is a search bar with the following controls: Scope: Self, Format: HTML, Amount: Quick, and GEO accession: GSM610544. The main content area is titled "Sample GSM610544" and includes a link "Query DataSets for GSM610544". The sample details are as follows:

Status	Public on Oct 19, 2010
Title	colorectal cancer cell line C80
Sample type	RNA
Source name	colorectal cancer cell line, replication error negative
Organism	Homo sapiens
Characteristics	cell line: colorectal cancer cell line C80 genotype/variation: Replication error negative (RER-/MSI-)
Growth protocol	DMEM 10% FCS
Extracted molecule	total RNA
Extraction protocol	samples were extracted using Qiagen Rneasy kit.
Label	biotin
Label protocol	20 micrograms RNA sent to Molecular biology core facility of the Paterson Institute for Cancer Research.

Gene Expression Omnibus data is unstructured

The image shows a screenshot of the NCBI Gene Expression Omnibus (GEO) website. The top navigation bar includes the NCBI logo and the GEO logo (Gene Expression Omnibus). Below the navigation bar, there are links for HOME, SEARCH, and SITE. The main content area displays a hierarchical view of data for GSE24795-GPL570. The hierarchy is as follows:

- GSE24795-GPL570
 - Cell Line (30)
 - Disease (30)
 - colorectal cancer (30)
 - Genotype/Variation (30)
 - replication error negative (RER-/MSI-) (16)
 - replication error positive (RER+/MSI+) (14)

On the left side of the screenshot, there is a sidebar with the following information:

NCBI > GEO > Acc

GEO help: Mouse over

Scope: Self

Sample GSM610

Status

Title

Sample type

Source name

Organism

Characteristics

Growth protocol

Extracted molecule

Extraction protocol: samples were extracted using Qiagen Rneasy kit.

Label: biotin























Label protocol: 20 micrograms RNA sent to Molecular biology core facility of the Paterson Institute for Cancer Research.

<http://insilico.ulb.ac.be>

INSILICO home Browse genomic datasets

Search query eg: lung cancer Selected datasets: 0 Page 1 of 35 Datasets displayed 1 - 50 of

Send to GenePattern Download in R format

Title	Platform	#Samples	Public
<input type="checkbox"/>  Osteosarcoma TE85 cell tissue culture study	HG-U133A	10	
<input type="checkbox"/>  Decreased Expression of Intelectin 1 in The Human Airway Epithelium of Smokers Com...	HG-U133_Plus_2	87	
<input type="checkbox"/>  mRNA expression profiles in human cell lines	HG-U133_Plus_2	16	
<input type="checkbox"/>  Key Regulatory Molecules of Cartilage Destruction in Rheumatoid Arthritis: An in vitro S...	HG-U133A	6	
<input type="checkbox"/>  High resolution gene expression profiling for simultaneous analysis of RNA synthesis, ...	HG-U133_Plus_2	9	
<input type="checkbox"/>  Expression data from PBMC treated with rabbit anti-thymocyte globulin (rATG) or horse...	HG-U133A_2	33	
<input type="checkbox"/>  Genomic Counter-Stress Changes Induced by Mind-Body Practice	HG-U133_Plus_2	72	
<input type="checkbox"/>  Identification of genes responsive to mild hyperthermia in human leukemia U937 cells	HG-U133A	4	
<input type="checkbox"/>  Breast cancer-associated fibroblasts confer AKT1-mediated epigenetic silencing of Cyst...	HG-U133_Plus_2	6	
<input type="checkbox"/>  Effects of tobacco smoke on gene expression and cellular pathways in a cellular model ...	HG-U133_Plus_2	60	
<input type="checkbox"/>  Gene Expression in MCF10A cells through Differentiation on Transwells	HG-U133_Plus_2	13	

Filters

DataSets

- GEO datasets
- InSilico datasets

Sharing

- Public
- Shared
- Owned

Platforms

- HG-U133A
- HG-U133B
- HG-U133A_2
- HG-U133_Plus_2

Normalization

- All Datasets
- FRMA Datasets

Options

Download options

- normalized
- original

InSilico DB - Galaxy integration

Galaxy Analyze Data Workflow Shared Data Help User

INSILICO home Browse genomic datasets Login or Register About us

Colorectal cancer Selected datasets: 1 colin molter 2011-05-26

Send to Galaxy Send to GenePattern Download in R format

Title	Platform	#Samples	Public
<input type="checkbox"/> GEO Genome-wide maps of tran...	Illumina Geno...	4	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Sex-specific and lineage-sp...	Illumina Geno...	12	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO miRNA expression data fro...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO BI Human Reference Epige...	Illumina Geno...	227	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO RNA-Seq of melanoma sho...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Dynamic transcriptomes du...	Illumina Geno...	6	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Mutational screening of lin...	Illumina Geno...	21	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Comparative transcriptomi...	Illumina Geno...	30	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Integrative model of genom...	Illumina Geno...	16	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> GEO Alternative expression anal...	Illumina Geno...	2	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Tissue-type specific estrog...	Illumina Geno...	12	<input checked="" type="checkbox"/>

Filters

DataSets

- GEO datasets
- InSilico datasets

Sharing

- Public
- Shared
- Owned

Platforms

- HG-U133A
- HG-U133B
- HG-U133A_2
- HG-U133_Plus_2
- Illumina G Analyzer II

Normalization

- All Datasets
- FRMA Datasets

History

- 45: BED-to-GFF on d
- 42: InSilico DB
- 37: InSilico DB
- 9: InSilico DB
- 6: Convert on data 1
- 4: UCSC Main on Human knownGene (genome)
- 2: Gene BED To Exon/Intron BED on data
- 1: GSM486704 4 Primary C K27me3.CD 61.bed

InSilico DB - Galaxy integration

The screenshot shows the Galaxy web interface with the InSilico DB integration. The search results for 'Colorectal cancer' are displayed in a table. The 'Illumina G Analyzer II' filter is highlighted in the left sidebar.

Galaxy Analyze Data Workflow Shared Data Help User

INSILICO home Browse genomic datasets Login or Register About us

Colorectal cancer Selected datasets: 1 colin molter 2011-05-26

Send to Galaxy Send to GenePattern Download in R format

Title	Platform	#Samples	Public
<input type="checkbox"/> GEO Genome-wide maps of tran...	Illumina Geno...	4	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Sex-specific and lineage-sp...	Illumina Geno...	12	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO miRNA expression data fro...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO BI Human Reference Epige...	Illumina Geno...	227	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO RNA-Seq of melanoma sho...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Dynamic transcriptomes du...	Illumina Geno...	6	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Mutational screening of lin...	Illumina Geno...	21	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Comparative transcriptomi...	Illumina Geno...	30	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Integrative model of genom...	Illumina Geno...	16	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> GEO Alternative expression anal...	Illumina Geno...	2	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Tissue-type specific estrog...	Illumina Geno...	12	<input checked="" type="checkbox"/>

Filters

DataSets

- GEO datasets
- InSilico datasets

Sharing

- Public
- Shared
- Owned

Platforms

- HG-U133A
- HG-U133B
- HG-U133A_2
- HG-U133_Plus_2
- Illumina G Analyzer II

Normalization

- All Datasets
- FRMA Datasets

History

- 45: BED-to-GFF on d
- 42: InSilico DB
- 37: InSilico DB
- 9: InSilico DB
- 6: Convert on data 1
- 4: UCSC Main on Hur knownGene (genome)
- 2: Gene BED To Exon/Intron BED on data
- 1: GSM486704 4 Primary C K27me3.CD 61.bed

InSilico DB - Galaxy integration

Galaxy Analyze Data Workflow Shared Data Help User

INSILICO home Browse genomic datasets Login or Register About us

Colorectal cancer Selected datasets: 1 colin molter 2011-05-26

Send to Galaxy Send to GenePattern Download in R format

Title	Platform	#Samples	Public
<input type="checkbox"/> GEO Genome-wide maps of tran...	Illumina Geno...	4	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Sex-specific and lineage-sp...	Illumina Geno...	12	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO miRNA expression data fro...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO BI Human Reference Epige...	Illumina Geno...	227	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO RNA-Seq of melanoma sho...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Dynamic transcriptomes du...	Illumina Geno...	6	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Mutational screening of lin...	Illumina Geno...	21	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Comparative transcriptomi...	Illumina Geno...	30	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Integrative model of genom...	Illumina Geno...	16	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> GEO Alternative expression anal...	Illumina Geno...	2	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Tissue-type specific estrog...	Illumina Geno...	12	<input checked="" type="checkbox"/>

Filters

DataSets

- GEO datasets
- InSilico datasets

Sharing

- Public
- Shared
- Owned

Platforms

- HG-U133A
- HG-U133B
- HG-U133A_2
- HG-U133_Plus_2
- Illumina G Analyzer II

Normalization

- All Datasets
- FRMA Datasets

History

- 45: BED-to-GFF on d
- 42: InSilico DB
- 37: InSilico DB
- 9: InSilico DB
- 6: Convert on data 1
- 4: UCSC Main on Hur knownGene (genome)
- 2: Gene BED To Exon/Intron BED on data
- 1: GSM486704 4 Primary C K27me3.CD 61.bed

InSilico DB - Galaxy integration

Analyze Data Workflow Shared Data Help User

Options home Browse genomic datasets Login or Register About us

Colorectal cancer Selected datasets: 1 colin molter 2011-05-26

Send to Galaxy Send to GenePattern Download in R format

Filters

DataSets

- GEO datasets
- InSilico datasets

Sharing

- Public
- Shared
- Owned

Platforms

- HG-U133A
- HG-U133B
- HG-U133A_2
- HG-U133_Plus_2
- Illumina G Analyzer II

Normalization

- All Datasets
- FRMA Datasets

Title	Platform	#Samples	Public
<input type="checkbox"/> GEO Genome-wide maps of tran...	Illumina Geno...	4	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Sex-specific and lineage-sp...	Illumina Geno...	12	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO miRNA expression data fro...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO BI Human Reference Epige...	Illumina Geno...	227	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO RNA-Seq of melanoma sho...	Illumina Geno...	14	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Dynamic transcriptomes du...	Illumina Geno...	6	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Mutational screening of lin...	Illumina Geno...	21	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Comparative transcriptomi...	Illumina Geno...	30	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Integrative model of genom...	Illumina Geno...	16	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> GEO Alternative expression anal...	Illumina Geno...	2	<input checked="" type="checkbox"/>
<input type="checkbox"/> GEO Tissue-type specific estrog...	Illumina Geno...	12	<input checked="" type="checkbox"/>

History Options

- 45: BED- to-GFF on data 1
- 42: InSilico DB
- 37: InSilico DB
- 9: InSilico DB
- 6: Convert on data 1
- 4: UCSC Main on Human: knownGene (genome)
- 2: Gene BED To Exon/Intron/Codon BED on data 1
- 1: GSM486704_BI.CD34_4_Primary_Cells.H3_K27me3.CD34_396_61.bed