生醫資訊導論

Teacher: Prof. Kun-Mao Chao
TA: Chia-Jung Chang and Wu-Lung Yang

Some Useful Websites

- http://www.ncbi.nlm.nih.gov/ NCBI
- http://genome.ucsc.edu/ UCSC
- http://www.ensembl.org/index.html ENSEMBL
- http://hapmap.ncbi.nlm.nih.gov/ HapMap
- http://ctdbase.org/ CTD
- http://geneticassociationdb.nih.gov/ GAD
- http://www.informatics.jax.org/phenotypes.shtml MGI
- http://david.abcc.ncifcrf.gov/ DAVID
- http://www.genome.jp/kegg/pathway.html KEGG

Some Tutorials

- http://manuals.bioinformatics.ucr.edu/
- http://bioinf.wehi.edu.au/affylmGUI/R/library/affylmGUI/doc/estrogen/estrogen.html
- http://statlab.nchc.org.tw/rnotes/

NCBI

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases

Popular Resources

PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI Announcements

DELTA BLAST - more sensitive protein searching

30 Apr 201

Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) runs a fast RPS-

NCBI

Entrez, The Life Sciences Search Engine

HOME | SEARCH | SITE MAP | PubMed | All Databases | Human Genome | GenBank | Map Viewer | BLAST

Search across databases HLA GO Clear Help

- Result counts displayed in gray indicate one or more terms not found

88570		**PubMed:** biomedical literature citations and abstracts	920		**Books:** online books
48410		**PubMed Central:** free, full text journal articles	610		**OMIM:** online Mendelian Inheritance in Man
5		**Site Search:** NCBI web and FTP sites			

65025		**Nucleotide:** Core subset of nucleotide sequence records	148		**dbGaP:** genotype and phenotype
6689		**EST:** Expressed Sequence Tag records	410		**UniGene:** gene-oriented clusters of transcript sequences
114		**GSS:** Genome Survey Sequence records	2		**CDD:** conserved protein domain database
48706		**Protein:** sequence database	1		**Clone:** integrated data for clone resources

48706	**Protein:** sequence database	1	**Clone:** integrated data for clone resources
34	**Genome:** whole genome sequences	316	**UniSTS:** markers and mapping data
497	**Structure:** three-dimensional macromolecular structures	651	**PopSet:** population study data sets
none	**Taxonomy:** organisms in GenBank	47551	**GEO Profiles:** expression and molecular abundance profiles
5	**SNP:** short genetic variations	524	**GEO DataSets:** experimental sets of GEO data
2679	**dbVar:** Genomic structural variation	none	**Epigenomics:** Epigenetic maps and data sets
6816	**Gene:** gene-centered information	246	**PubChem BioAssay:** bioactivity screens of chemical substances
14	**SRA:** Sequence Read Archive	7	**PubChem Compound:** unique small molecule chemical structures
1594	**BioSystems:** Pathways and systems of interacting molecules	1038	**PubChem Substance:** deposited chemical substance records
36	**HomoloGene:** eukaryotic homology groups	none	**Protein Clusters:** a collection of related protein sequences
2493	**Probe:** sequence-specific reagents	none	**OMIA:** online Mendelian Inheritance in Animals
75	**BioProject:** aggregated biological research project data	246	**BioSample:** biological material descriptions

292	**NLM Catalog:** catalog of books, journals, and audiovisuals in the NLM collections	571	**MeSH:** detailed information about NLM's controlled vocabulary

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

Send to:

Results: 1 to 20 of 48410

<< First < Prev Page 1 of 2421 Next > Last >>

1. Nomenclature for factors of the **HLA** system, 2010

S G E Marsh, E D Albert, W F Bodmer, R E Bontrop, B Dupont, H A Erlich, M Fernández-Viña, D E Geraghty, R Holdsworth, C K Hurley, M Lau, K W Lee, B Mach, M Maiers, W R Mayr, C R Müller, P Parham, E W Petersdorf, T Sasazuki, J L Strominger, A Svejgaard, P I Terasaki, J M Tiercy, J Trowsdale

Tissue Antigens. 2010 April; 75(4): 291–455. doi: 10.1111/j.1399-0039.2010.01466.x

PMCID: PMC2848993

Full Text PDF–1.2M

2. Implications of the polymorphism of **HLA**-G on its function, regulation, evolution and disease association

Eduardo A. Donadi, Erick C. Castelli, Antonio Arnaiz-Villena, Michel Roger, Diego Rey, Philippe Moreau

Cell Mol Life Sci. 2011 February; 68(3): 369–395. Published online 2010 November 24. doi: 10.1007/s00018-010-0580-7

PMCID: PMC3021195

Abstract Full Text PDF–1.3M

3. Epistasis among ***HLA***-*DRB1*, ***HLA***-*DQA1*, and ***HLA***-*DQB1* loci determines multiple sclerosis susceptibility

Matthew R. Lincoln, Sreeram V. Ramagopalan, Michael J. Chao, Blanca M. Herrera, Gabriele C. DeLuca, Sarah-Michelle Orton, David A. Dyment, A Dessa Sadovnick, George C. Ebers

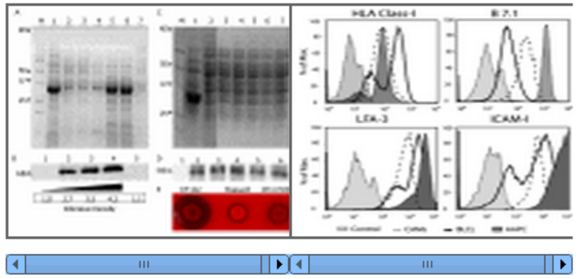
Filter your results:

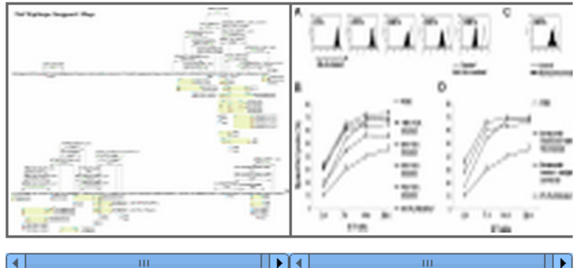
All (48410)

NIH grants (16052)

Manage Filters

PMC Images search for *HLA*





+142800

MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS I, A; HLA-A

Alternative titles; symbols

HLA-A HISTOCOMPATIBILITY TYPE

Other entities represented in this entry:

MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS I, H PSEUDOGENE, INCLUDED; HLA-H, INCLUDED

MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS I, J PSEUDOGENE, INCLUDED; HLA-J, INCLUDED

HGNC Approved Gene Symbol: HLA-A

Cytogenetic location: 6p22.1 ***Genomic coordinates (GRCh37): 6:29,910,246 - 29,913,660*** (from NCBI)

Gene Phenotype Relationships

Location	Phenotype	Phenotype MIM number
6p22.1	{Hypersensitivity syndrome, carbamazepine-induced, susceptibility to}	608579

Clinical Synopsis

TEXT

Description

The human major histocompatibility complex (MHC) has been divided into 3 regions on chromosome 6p21.3: class II (centromeric), class III, and class I (telomeric), with extended class I and class II regions on either side. The MHC encodes highly polymorphic proteins, many of which are associated with the immune system. The products of classical polymorphic class I genes, human leukocyte antigen-A (HLA-A), HLA-B (142830), and HLA-C (142840), interact with T-cell receptor (TCR; see 186880) molecules, as well as killer immunoglobulin-like receptors (KIRs; see 604936) expressed on natural killer cells and some T cells (review by Trowsdale, 2001).

Evidence from amino acid sequences suggests an evolutionary relatedness of transplantation antigens, immunoglobulins and beta-2-microglobins (Tragardh et al., 1979). Both the class I MHC antigens (A, B, and C) and the class II antigens DR and DC1 are polymorphic 2-chain cell surface glycoproteins; they are recognized by different subsets of T cells and have different functions, tissue distributions, and structures. The light chain of class I antigens is beta-2-microglobulin (B2M; 109700), which is coded by chromosome 15. The heavy chain, coded by chromosome 6, has a molecular mass of 44,000 and is made up of 3 N-terminal extracellular domains of 90 amino acids each, a small hydrophobic membrane-spanning segment and a small hydrophilic intracellular C-terminal domain. The 2 N-terminal domains are polymorphic, bear the carbohydrate and have no sequence homology with immunoglobulin. The third domain, closest to the membrane, and the 11.6-kD B2M light chain are highly conserved and have strong sequence homology with immunoglobulin.

- Table of Contents – +142800

External Links:

- Genome
- DNA
- Protein
- Gene Info
- Clinical Resources
- Variation
- Animal Models
- Cell Lines
- Cellular Pathways

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

Send to:

Results: 1 to 20 of 524

<< First < Prev Page 1 of 27 Next > Last >>

1: GDS3417 record**: Untreated juvenile dermatomyositis muscle biopsies** [*Homo sapiens*]

GEO Profiles, Links

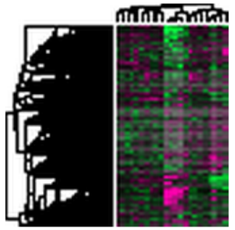
Summary: Analysis of skeletal muscle biopsies from untreated girls with active symptoms of juvenile dermatomyositis (JDM) less than 2 months or greater than 2 months. Results provide insight into the impact of the duration of chronic inflammation on gene expression in muscle of untreated children with JDM.
Parent Platform: GPL96
Reference Series: GSE11971

Type: Expression profiling by array, count

Subsets: 3 disease state sets.

Samples: 23

GSM303210: Muscle JDM (YJD-JDM-6UA-s2)
GSM303211: Muscle JDM (YJD-JDM-8UA-s2)
GSM303212: Muscle JDM (YJD-JDM-9UA-S2)
GSM303195: Muscle JDM (YJD-JDM-16UA-s2)
GSM303198: Muscle JDM (YJD-JDM-1UA-s2)
GSM303202: Muscle JDM (YJD-JDM-2UA-s2)



2: GDS3258 record**: Monocyte-derived macrophage response to decoy receptor 3** [*Homo sapiens*]

GEO Profiles, Links

Filter your results:

All (524)
DataSets (5)
Platforms (2)
Samples (441)
Series (76)

Manag

Top Organisms [Tree]

Homo sapiens *(471)*
Mus musculus *(25)*
Rattus norvegicus *(19)*
Staphylococcus aureus *(11)*
Bacteroides thetaiotaomicron *(7)*
More...

Find related data

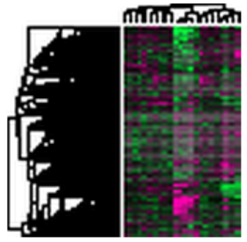
Database: Select

Find items

DataSet Record GDS3417: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	Untreated juvenile dermatomyositis muscle biopsies		
Summary:	Analysis of skeletal muscle biopsies from untreated girls with active symptoms of juvenile dermatomyositis (JDM) less than 2 months or greater than 2 months. Results provide insight into the impact of the duration of chronic inflammation on gene expression in muscle of untreated children with JDM.		
Organism:	*Homo sapiens*		
Platform:	GPL96: [HG-U133A] Affymetrix Human Genome U133A Array		
Citation:	Chen YW, Shi R, Geraci N, Shrestha S et al. Duration of chronic inflammation alters gene expression in muscle from untreated girls with juvenile dermatomyositis. *BMC Immunol* 2008 Jul 31;9:43. PMID: 18671865		
Reference Series:	GSE11971	**Sample count:**	23
Value type:	count	**Series published:**	2008/07/18

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

Find gene name or symbol: Go

Find genes that are up/down for this condition(s): disease state Go

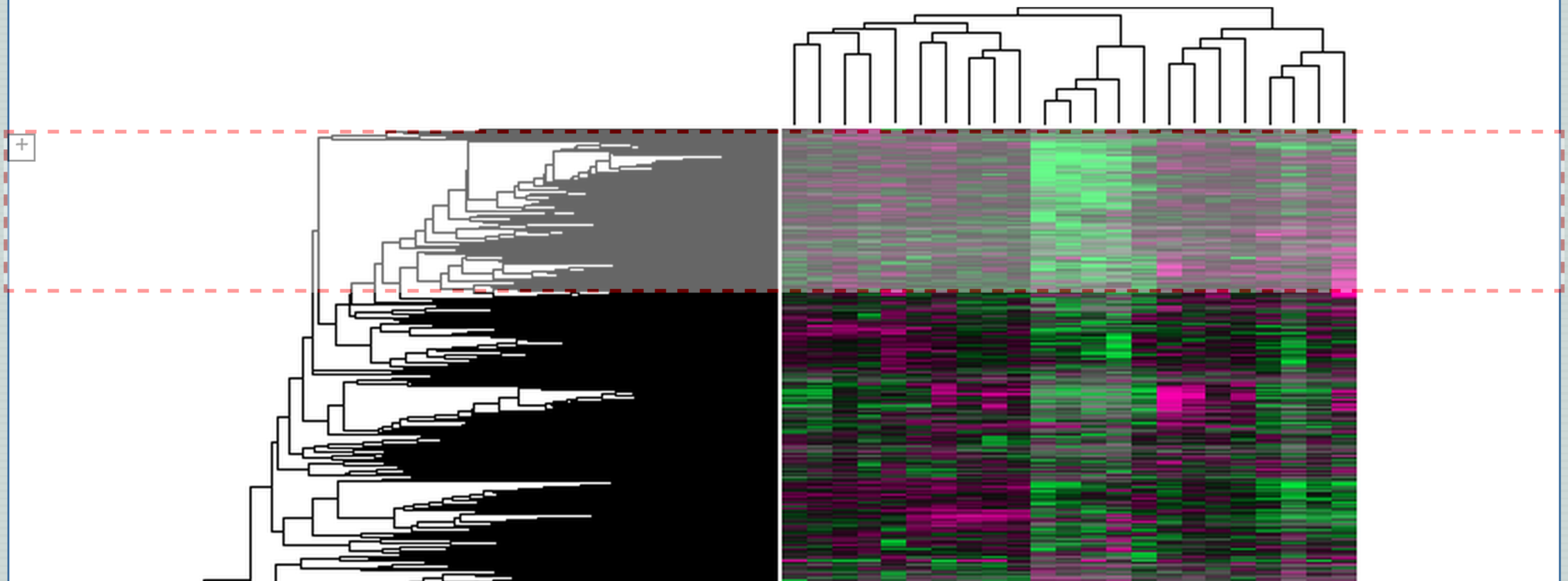
Selected profiles

Download | Plot values | View in Entrez | Stack up

« How To

GDS3417 **Untreated juvenile dermatomyositis muscle biopsies [Homo sapiens]**

Clustering: Uncentered Correlation UPGMA Colors: High Low Full image: 12870 x 23 spots

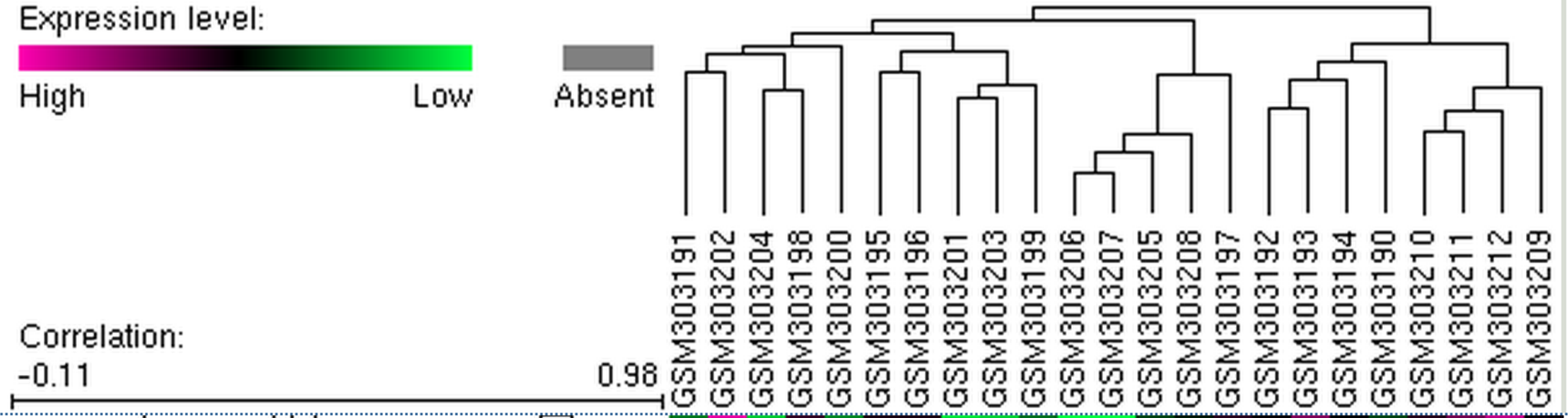


- Click on heat map to start selection. Drag/resize box to cover region of interest.
- Double click on active selection or click "Stack up" to zoom in picked region(s).

« How To

GDS3417 **Untreated juvenile dermatomyositis muscle biopsies [Homo sapiens]**

Clustering: Uncentered Correlation UPGMA **Colors:** High Low Full image: **12870** x **23** spots [Reset]



Gene list
(searchable)

HRC
SRPK3
LARGE
HIF1AN
NKX3-1
PSMC1
ZHX3
FHOD1
AA401963
CEP192

Ensembl

Search: All species for Go

e.g. **BRCA2** or **rat X:100000..200000** or **coronary heart disease**

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Click on a link below to go to the species' home page.

Popular genomes (Log in to customize this list)

Human
GRCh37

Mouse
NCBIM37

Zebrafish
Zv9

All genomes

-- Select a species --

View full list of all Ensembl species

Other species are available in Ensembl *Pre!* and EnsemblGenomes

New to Ensembl?

Did you know you can:

- **Learn how to use Ensembl**
  with our video tutorials and walk-throughs
- **Add custom tracks**
  using our new Control Panel
- **Upload and analyse your data**
  and save it to your Ensembl account
- **Search for a DNA or protein sequence**
  using BLAST or BLAT
- **Fetch only the data you want**
  from our public database, using the Perl API
- **Download our databases via FTP**
  in FASTA, MySQL and other formats
- **Mine Ensembl with BioMart**
  and export sequences or tables in text, html, or Excel format

Still got questions? Try our FAQs or glossary

Did you know...?

If you want to learn more about using the browser, host a workshop!

What's New in Release 66 (February 2012)

- New species: Coelacanth
- View patches aligned to reference sequence (Human)
- Region Report - new data export tool

Important Notice

We now use Blat as our default DNA search. This will make your query faster.

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

```
>test
AAAGAAAAAAAGAAAAATCCA
TGCATATGATACATCAGTTAACAAGGCACTGGTGAAATTAATTTTAAGTA
TTATTGTCTCTTTGTGTTTTTGGTCTCAGAAAAGTTACGATTTCCCTTAG
TTCCTTAGGGCAGAGAGAATCTTCAATCACTGAAGTCAGGAGACACACAT
```

Or Upload a file containing one or more FASTA sequences

Choose File No file chosen

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

Retrieve

- (•) dna queries
- ( ) peptide queries

No query sequences have been entered

Select the databases to search against

Select species:
Use 'ctrl' key to select multiple species

Gallus_gallus
Gasterosteus_aculeatus
Gorilla_gorilla
Homo_sapiens

- (•) dna database — RNA_NC
- ( ) peptide database — PEP_ALL

Select the Search Tool

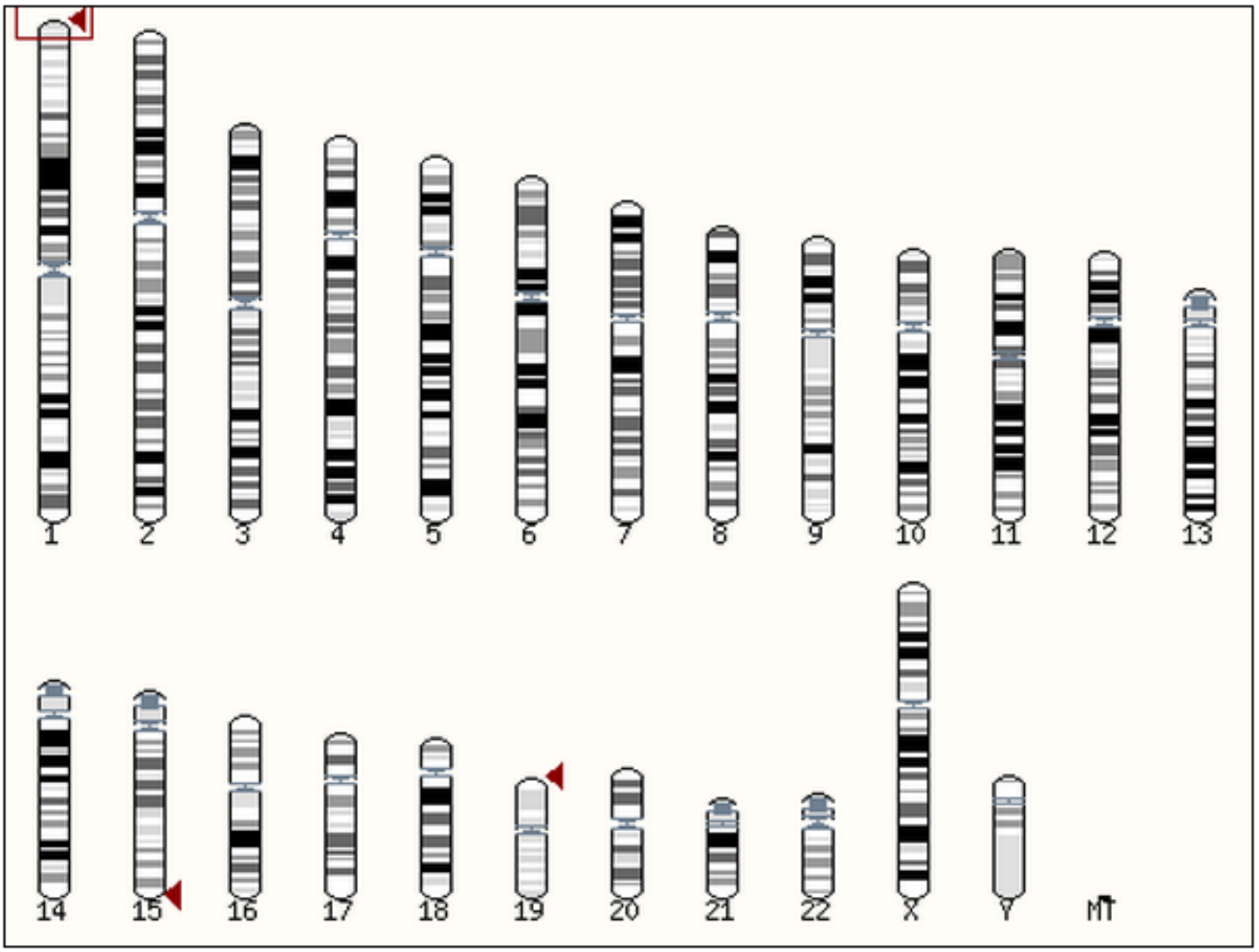
BLASTN
TBLASTX

configure ▶ RUN ▶

Search sensitivity:
Optimise search parameters to find the following alignments

Near-exact matches

Alignment Locations vs. Karyotype (click arrow to hide)



Alignment Locations vs. Query (click arrow to hide)

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

refresh

Query	Subject	Chromosome	Supercontig	Clone	Contig	Lrg	Stats	Sort By
off	_off_	_off_	_off_	_off_	_off_	_off_	_off_	<P-val
Name	Name	Name	Name	Name	Name	Name	Score	>P-val
Start	Start	Start	Start	Start	Start	Start	E-val	<%ID
End	End	End	End	End	End	End	P-val	>%ID

Links	Query			Chromosome				Stats			
	Start	*End*	*Ori*	*Name*	*Start*	*End*	*Ori*	*Score*	*E-val*	*%ID*	*Length*
[A] [G] [C]	1	571	+	Chr:1	49518	50088	+	2827	0.0e+00	100.00	571
[A] [G] [C]	1	571	+	Chr:19	91130	91700	+	2827	0.0e+00	100.00	571
[A] [G] [C]	1	571	-	Chr:15	102481063	102481629	+	2780	0.0e+00	99.30	571

Dataset

Homo sapiens genes (GRCh37.p6)

Filters

GO Evidence code : IEA

Attributes

Ensembl Gene ID
Ensembl Transcript ID
Gene Start (bp)
Gene End (bp)
Strand
GO Term Accession
PUBMED ID

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

- (•) **Features**
- ( ) **Structures**
- ( ) **Transcript Event**
- ( ) **Homologs**
- ( ) **Variation**
- ( ) **Sequences**

⊟ GENE:

Ensembl

- ☑ Ensembl Gene ID
- ☑ Ensembl Transcript ID
- ☐ Ensembl Protein ID
- ☐ Description
- ☐ Chromosome Name
- ☑ Gene Start (bp)
- ☑ Gene End (bp)
- ☑ Strand
- ☐ Band
- ☐ Transcript Start (bp)
- ☐ Transcript End (bp)
- ☐ Associated Gene Name
- ☐ Associated Transcript Name
- ☐ Associated Gene DB
- ☐ Associated Transcript DB
- ☐ Transcript count
- ☐ % GC content
- ☐ Gene Biotype
- ☐ Transcript Biotype
- ☐ Source
- ☐ Status (gene)
- ☐ Status (transcript)

⊟ EXTERNAL:

GO

- ☑ GO Term Accession
- ☐ GO Term Name
- ☐ GO Term Definition
- ☐ GO Term Evidence Code
- ☐ GO domain

GOSlim GOA

- ☐ GOSlim GOA Accession(s)
- ☐ GOSlim GOA Description

New | Count | Results | URL | XML | Perl | Help

Dataset

Homo sapiens genes (GRCh37.p6)

Filters

GO Evidence code : IEA

Attributes

Ensembl Gene ID
Ensembl Transcript ID
Gene Start (bp)
Gene End (bp)
Strand
GO Term Accession
PUBMED ID

Dataset

[None Selected]

Export all results to Compressed web file (notify by email) CSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Gene Start (bp)	Gene End (bp)	Strand	GO Term Accession	PUBMED ID
ENSG00000211814	ENST00000390462	22689792	22690371	1	GO:0005515	8188290
ENSG00000211815	ENST00000390463	22694641	22695149	1	GO:0005515	8188290
ENSG00000211816	ENST00000390464	22739851	22740446	1	GO:0005515	8188290
ENSG00000211817	ENST00000390465	22748988	22749631	1	GO:0005515	8188290
ENSG00000211818	ENST00000390466	22771939	22772438	1	GO:0005515	8188290
ENSG00000211819	ENST00000390467	22782922	22783351	1	GO:0005515	8412327
ENSG00000211820	ENST00000390468	22788620	22789123	1	GO:0005515	8188290
ENSG00000211821	ENST00000390469	22891362	22892033	1	GO:0005515	2526321
ENSG00000211829	ENST00000390477	22931924	22934779	1	GO:0016020	
ENSG00000211829	ENST00000390477	22931924	22934779	1	GO:0016021	

Human (Homo sapiens)

Search for: Go

e.g. BRCA2 or 6:133017695-133161157 or osteoarthritis

Description

Assembly

This site provides a data set based on the February 2009 *Homo sapiens* high coverage assembly GRCh37 (GCA_000001405.6) from the Genome Reference Consortium. This assembly is used by UCSC to create their hg19 database. The data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model of exonerate.

This release of the assembly has the following properties:

- 27478 contigs.
- contig length total 3.2 Gb.
- chromosome length total 3.1 Gb.

It also includes nine haplotypic regions, mainly in the MHC region of chromosome 6.

As the GRC maintains and improves the assembly, patches are being introduced. Patch release six (GRCh37.p6) was included in Ensembl release 66. Currently, assembly patches are of two types:

- Novel patch: new sequences that add alternative sequence at a loci and will remain as haplotypes in the next major assembly release by GRC
- Fix patch: sequences that correct the reference sequence and will replace the given region of the reference assembly at the next major assembly release by GRC

To convert your old data from Human assembly NCBI36 to GRCh37, click on 'Manage your data' on any human page and select 'Assembly converter' from the left-hand menu.

A preliminary assembly of the Neanderthal (*Homo sapiens neanderthalensis*) genome is available via the Neanderthal Genome Browser, an Ensembl-powered project based at the Max Planck Institute.
The genome assembly represented here corresponds to GenBank Assembly ID GCA_000005045.6

Download Human genome sequence (FASTA)

Previous assemblies

NCBI36 (Release 54, May2009) ▾ Go to archive

Annotation

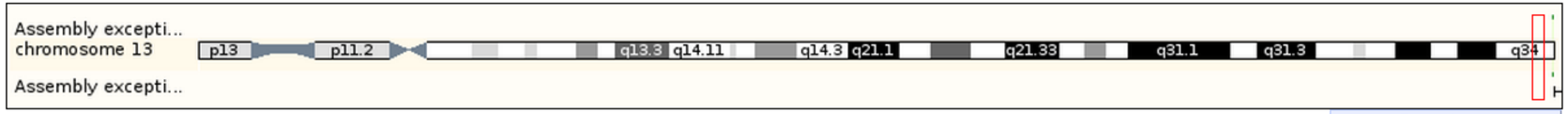
The Ensembl human gene annotations have been updated using Ensembl's automatic annotation pipeline. The updated annotation incorporates new protein and cDNA sequences which have become publicly available since the last GRCh37 genebuild (March 2009).

In release 66 (February 2012), we continue to display a joint gene set based on the merge between the automatic annotation from Ensembl and the manually curated annotation from Havana. This refined gene set corresponds to GENCODE release 11. The Consensus Coding Seque
(CCDS) identifiers have also been mapped to the annotations. More information about the CCDS project.

Updated manual annotation from Havana is merged into the Ensembl annotation every release. Transcripts from the two annotation sources are merged if they share the same internal exon-intron boundaries (i.e. have identical splicing pattern) with slight differences in the terminal exo
allowed. Importantly, all Havana transcripts are included in the final Ensembl/Havana merged (GENCODE) gene set. In this release, 23171 Ensembl gene models and 45484 Havana genes were merged together to create the final set of 56478 genes.

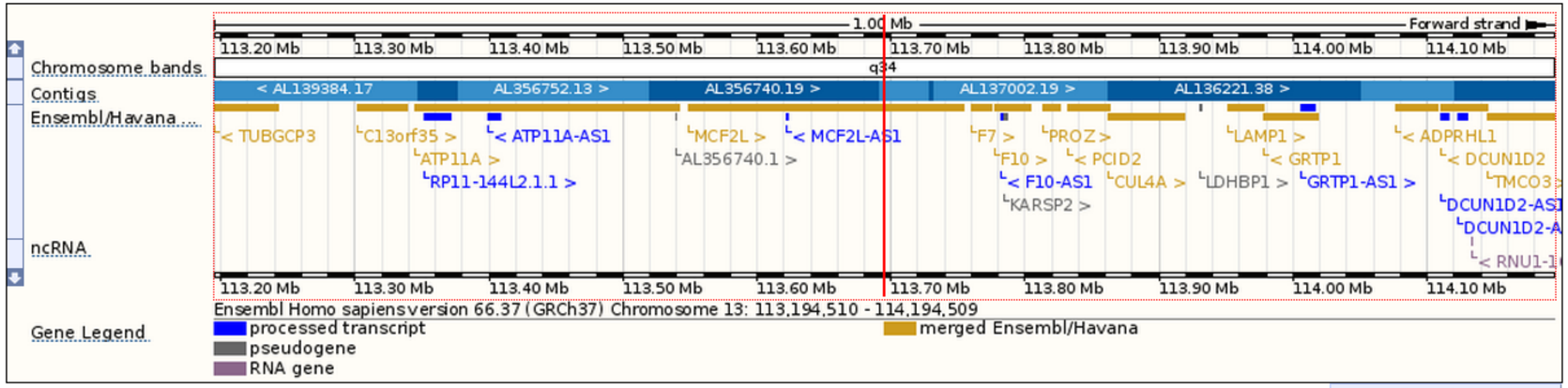
- Detailed information on genebuild (PDF)

Chromosome 13: 113,694,009-113,695,009



Export Image

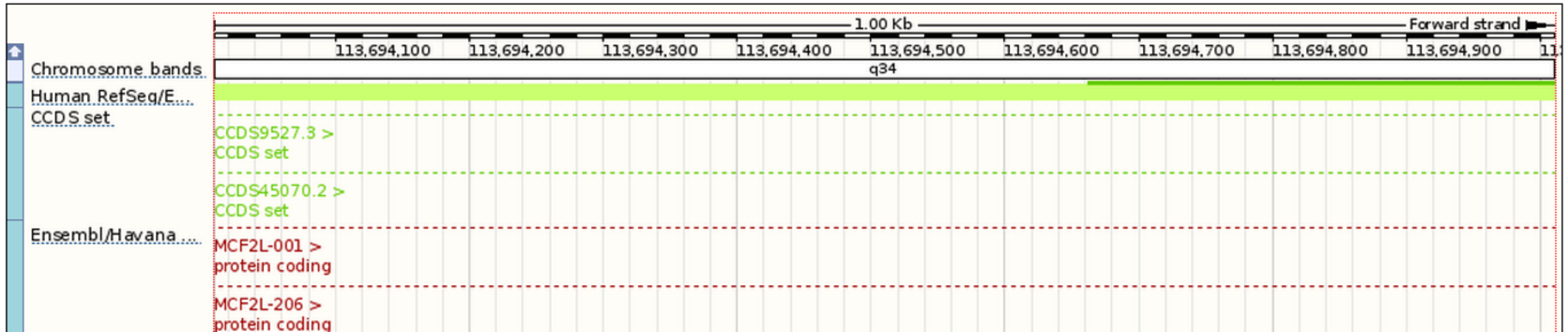
Region in detail help



Export Image

Location: 13:113694009-113695009 Go

Gene: Go



R: Microarray

```
>
> v1<-1
> v2<-as.integer(2)
> v3<-TRUE
> v4<-"abc"
> typeof(v1)
[1] "double"
> typeof(v2)
[1] "integer"
> typeof(v3)
[1] "logical"
> typeof(v4)
[1] "character"
>
```

```
>
> c(1:10)
 [1]  1  2  3  4  5  6  7  8  9 10
> c(6,3,6,1:4,2:3)
[1] 6 3 6 1 2 3 4 2 3
> runif(10)
 [1] 0.15281516 0.53352757 0.30039152 0.69891784 0.35381320 0.81001772
 [7] 0.90448145 0.89338562 0.07840998 0.46073054
> x<-c(6,3,1,5,7,5)
> sort(x)
[1] 1 3 5 5 6 7
> sum(x)
[1] 27
> prod(x)
[1] 3150
> range(x)
[1] 1 7
> x<3
[1] FALSE FALSE  TRUE FALSE FALSE FALSE
> which(x<3)
[1] 3
> x[x<3]
[1] 1
> which(x<5)
[1] 2 3
> x(x<5)
錯誤: 沒有這個函數 "x"
> x[x<5]
[1] 3 1
>
```

```
> mode(x)
[1] "numeric"
> l1<-list(a=3,b=c(4:7),c=c("Abb","bbb","ddd"))
> l1
$a
[1] 3

$b
[1] 4 5 6 7

$c
[1] "Abb" "bbb" "ddd"

> mode(l1)
[1] "list"
> typeof(l1)
[1] "list"
> l1$a
[1] 3
> l1$b
[1] 4 5 6 7
>
```

```
>
> m1<-matrix(c(1,2,3,4,5,6),3,2)
> m1
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
> colnames(m1)=c("c1","c2")
> m1
     c1 c2
[1,]  1  4
[2,]  2  5
[3,]  3  6
> rownames(m1)=c("r1","r2","r3")
> m1
   c1 c2
r1  1  4
r2  2  5
r3  3  6
>
```

File Edit View Workspace Run Data Analysis Plots Distributions Windows Settings Help

Open Create Save Run selection Interrupt running command

Workspace Files

All Non-Functions Functions

Show All Environments

Show Hidden Objects

Name	Label	Type	Class
package:base			
Autoloads			
package:methods			
package:datasets			
package:utils			
package:grDevices			
package:graphics			
package:stats			
package:rkward			
.GlobalEnv			
mycmap3			data.fra
mycmap2			data.fra
mycmap			data.fra
hg			data.fra
cmap2			data.fra
cmap			data.fra
columnna...		Factor	factor
Vendor		Factor	factor
Vehicle		Factor	factor
Scanner		Factor	factor
Sample		Number	integer
Concentra...		Number	numeri
Cmap		Factor	factor
Cell		Factor	factor
Batch		Factor	factor
Array.1		Factor	factor

Update

rkward_welcome | read.table.html | read.table.html | data.frame.html | cmap

	1	2	3	4	5	6	7	8	9	10	#New Variable#
Name	**Sample**	**Cmap**	**Concentrati...**	**Cell**	**Array.1**	**Scanner**	**Vehicle**	**Vendor**	**Batch**	**columnname**	
Label											
Type	Number	Factor	Number	Factor	Factor	Factor	Factor	Factor	Factor	Factor	
Format											
Levels		0173570-00...		HL60#,#MC...	HG-U133A#,...	Axon ImageX...	DMSO#,#eth...	Asinex#,#As...	1#,#1000#,...	s1#,#s1000...	
1	1	metformin	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s1	
2	2	metformin	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s2	
3	3	metformin	1e-07	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s3	
4	4	metformin	0.001	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s4	
21	21	phenformin	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s21	
22	22	phenyl bigua...	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s22	
23	23	valproic acid	0.001	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s23	
61	61	metformin	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s61	
121	121	estradiol	1e-08	MCF7	HG-U133A	HP GeneArra...	DMSO	Sigma-Aldrich	5	s121	
122	122	alpha-estradiol	1e-08	MCF7	HG-U133A	HP GeneArra...	DMSO	Sigma-Aldrich	5	s122	
123	123	dexamethas...	1e-06	MCF7	HG-U133A	HP GeneArra...	DMSO	Sigma-Aldrich	5	s123	
124	124	mesalazine	0.0001	MCF7	HG-U133A	HP GeneArra...	DMSO	Sigma-Aldrich	5	s124	
141	141	chlorpropami...	0.0001	MCF7	HG-U133A	HP GeneArra...	DMSO	Sigma-Aldrich	6	s141	

```
>
> cmap<-read.delim("biodata/cmap/pvca-rscript/b2cmap_instance.csv",header=TRUE,sep="\t",row.names=1)
> dim(cmap)
[1] 6092   10
> cmap[1,5]
[1] HG-U133A
Levels: HG-U133A HT_HG-U133A HT_HG-U133A_EA
> cmap[5,1]
[1] 21
> mycmap<-cmap[,c(2:3,9)]
> dim(mycmap)
[1] 6092    3
> mycmap2<-cmap[c(3:6),]
> dim(mycmap2)
[1]  4 10
> mycmap3<-cmap[c(3:6),c(2:3,9)]
> dim(mycmap3)
[1] 4 3
>
```

	1	2	3
Name	**Cmap**	**Concentrati...**	**Batch**
Label			
Type	Factor	Number	Factor
Format			
Levels	0173570-00...		1#,#1000#,...
1	metformin	1e-05	1
2	metformin	1e-05	1
3	metformin	1e-07	1
4	metformin	0.001	1
21	phenformin	1e-05	2
22	phenyl bigua...	1e-05	2
23	valproic acid	0.001	2
61	metformin	1e-05	2
121	estradiol	1e-08	5
122	alpha-estradiol	1e-08	5
123	dexamethas...	1e-06	5
124	mesalazine	0.0001	5
141	chlorpropami...	0.0001	6

	1	2	3	4	5	6	7	8	9	10	#New Variable#
Name	**Sample**	**Cmap**	**Concentrati...**	**Cell**	**Array.1**	**Scanner**	**Vehicle**	**Vendor**	**Batch**	**columnname**	
Label											
Type	Number	Factor	Number	Factor	Factor	Factor	Factor	Factor	Factor	Factor	
Format											
Levels		0173570-00...		HL60#,#MC...	HG-U133A#,...	Axon ImageX...	DMSO#,#eth...	Asinex#,#As...	1#,#1000#,...	s1#,#s1000...	
3	3	metformin	1e-07	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s3	
4	4	metformin	0.001	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	1	s4	
21	21	phenformin	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s21	
22	22	phenyl bigua...	1e-05	MCF7	HG-U133A	HP GeneArra...	medium	Sigma-Aldrich	2	s22	

	1	2	3	#New Variable#
Name	**Cmap**	**Concentrati...**	**Batch**	
Label				
Type	Factor	Number	Factor	
Format				
Levels	0173570-00...		1#,#1000#,...	
3	metformin	1e-07	1	
4	metformin	0.001	1	
21	phenformin	1e-05	2	
22	phenyl bigua...	1e-05	2	

[Unnamed Workspace] – mycm

File Edit View Workspace Run Data Analysis Plots Distributions Windows Settings Help

Open Create Save Run selection Interrupt running command

Workspace Files

All Non-Functions Functions

Show All Environments

Show Hidden Objects

Name	Label	Type	Class
package:base			
Autoloads			
package:methods			
package:datasets			
package:utils			
package:grDevices			
package:graphics			
package:stats			
package:rkward			
.GlobalEnv			
mycmap4			data.fra
Concentra...		Number	numeri
Cmap		Factor	factor
Batch		Factor	factor
mycmap3			data.fra
mycmap2			data.fra
mycmap			data.fra
Concentra...		Number	numeri
Cmap		Factor	factor
Batch		Factor	factor
hg			data.fra
cmap2			data.fra
cmap			data.fra

Update

rkward_welcome | read.table.html | read.table.html | data.frame

	1	2	3	#New Variable#
Name	**Cmap**	**Concentrati...**	**Batch**	
Label				
Type	Factor	Number	Factor	
Format				
Levels	0173570-00...		1#,#1000#,...	
1	metformin	1e-05	1	
2	metformin	1e-05	1	
3	metformin	1e-07	1	
4	metformin	0.001	1	
61	metformin	1e-05	2	
1694	metformin	2.42e-05	627	
1816	metformin	2.42e-05	628	
1858	metformin	2.42e-05	629	
5068	metformin	2.42e-05	718	
5487	metformin	2.42e-05	737	

```
>
> cmap[["Cmap"]][4]
[1] metformin
1309 Levels: 0173570-0000 0175029-0000 0179445-0000 0198306-0000 ... zuclopenthixol
> cmap[["Cmap"]][5]
[1] phenformin
1309 Levels: 0173570-0000 0175029-0000 0179445-0000 0198306-0000 ... zuclopenthixol
> which(cmap[["Cmap"]]=="metformin")
 [1]    1    2    3    4    8  925 1046 1088 3930 4249
> mycmap4<-cmap[cmap[["Cmap"]]=="metformin",c(2,3,9)]
> mycmap4[,c("Cmap","Batch")]
          Cmap Batch
1    metformin     1
2    metformin     1
3    metformin     1
4    metformin     1
61   metformin     2
1694 metformin   627
1816 metformin   628
1858 metformin   629
5068 metformin   718
5487 metformin   737
>
```

```
> unique(cmap$Cell)
[1] MCF7   HL60   ssMCF7 PC3    SKMEL5
Levels: HL60 MCF7 PC3 SKMEL5 ssMCF7
> c1<-unique(cmap[cmap$Cell=="MCF7",2])
> c2<-unique(cmap[cmap$Cell=="HL60",2])
> c3<-unique(cmap[cmap$Cell=="ssMCF7",2])
> c4<-unique(cmap[cmap$Cell=="PC3",2])
> c5<-unique(cmap[cmap$Cell=="SKMEL5",2])
> length(c1)
[1] 1294
> length(c2)
[1] 1078
> length(c3)
[1] 16
> length(c4)
[1] 1182
> length(c5)
[1] 16
> length(unique(cmap$C
cmap[["Concentration"]]            cmap[["Cmap"]]                  cmap[["Cell"]]
> length(unique(cmap[["Cmap"]]))
[1] 1309
>
```

File Edit View Workspace Run Data Analysis Plots Distributions Windows Settings Help

Open Create Save Run selection Interrupt running command

Workspace Files

All Non-Functions Functions

Show All Environments

Show Hidden Objects

Name	Label	Type
package:base		
Autoloads		
package:methods		
package:datasets		
package:utils		
package:grDevices		
package:graphics		
package:stats		
package:rkward		
package:Biobase		
package:affy		
package:AnnotationDbi		
package:hgu133acdf		
.GlobalEnv		
x		
X6146151114...		Number
X6146151114...		Factor
X6146151114...		Number

Update

rkward_welcome | x

	1	2	3	4	5	6
Name	**X61461511...**	**X61461511...**	**X61461511...**	**X61461511...**	**X61461511...**	**X61461511...**
Label						
Type	Number	Factor	Number	Number	Factor	Number
Format						
Levels		A#,#M#,#P			A#,#M#,#P	
1007_s_a	7.75619582...	A	7.76175916...	0.23455651...	A	7.63209988...
1053_at	8.04026676...	P	7.60805950...	0.00080466...	P	8.18401420...
117_at	6.31622052...	P	6.45869139...	0.03133563...	P	6.33528244...
121_at	8.11570091...	P	8.05893416...	0.01309178...	P	8.17648677...
1255_g_	4.43889246...	A	4.44377345...	0.26746255...	A	4.44060560...
1294_at	6.91313351...	P	6.80276038...	0.01493651...	P	6.91510208...

```
> library(affy)
> list.celfiles(path="~/biodata/cmap/",full.names=TRUE)
[1] "/home/john/biodata/cmap//614615111406.A02.CEL"
[2] "/home/john/biodata/cmap//614615111406.A03.CEL"
[3] "/home/john/biodata/cmap//614615111406.A04.CEL"
[4] "/home/john/biodata/cmap//614615111406.A05.CEL"
[5] "/home/john/biodata/cmap//614615111406.A12.CEL"
> mydata<-ReadAffy(filenames=list.celfiles(path="~/biodata/cmap/",full.names=TRUE))
> eset <- rma(mydata)
Background correcting
Normalizing
Calculating Expression
> eset_PMA <- mas5calls(mydata)
Getting probe level data...
Computing p-values
Making P/M/A Calls
> x <- data.frame(exprs(eset), exprs(eset_PMA), assayDataElement(eset_PMA, "se.exprs"))
> x <- x[,sort(names(x))]
>
```

```
> eset
ExpressionSet (storageMode: lockedEnvironment)
assayData: 22283 features, 5 samples
  element names: exprs
protocolData
  sampleNames: 614615111406.A02.CEL 614615111406.A03.CEL ...
    614615111406.A12.CEL (5 total)
  varLabels: ScanDate
  varMetadata: labelDescription
phenoData
  sampleNames: 614615111406.A02.CEL 614615111406.A03.CEL ...
    614615111406.A12.CEL (5 total)
  varLabels: sample
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
Annotation: hgu133a
> pData(eset)
                     sample
614615111406.A02.CEL      1
614615111406.A03.CEL      2
614615111406.A04.CEL      3
614615111406.A05.CEL      4
614615111406.A12.CEL      5
> mydata
AffyBatch object
size of arrays=712x712 features (11 kb)
cdf=HG-U133A (22283 affyids)
number of samples=5
number of genes=22283
annotation=hgu133a
notes=
>
```

hgu133a.db

Affymetrix Human Genome U133 Set annotation data (chip hgu133a)

Bioconductor version: Release (2.10)

Affymetrix Human Genome U133 Set annotation data (chip hgu133a) assembled using data from public repositories

Author: Marc Carlson, Seth Falcon, Herve Pages, Nianhua Li

Maintainer: Biocore Data Team <biocannotation at lists.fhcrc.org>

To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("hgu133a.db")
```

To cite this package in a publication, start R and enter:

```
citation("hgu133a.db")
```

All Non-Functions Functions

Show All Environments

Show Hidden Objects

Name	Label	Type
x		
probleGene		
probeset		String
probeGene		
mydata		Unknc
mapCdfName		
eset_PMA		Unknc
eset		Unknc
SYMBOL		String

Update

rkward_welcome x Documentation for package 'hgu133a.db'

hgu133a.db	Bioconductor annotation data package
hgu133aACCNUM	Map Manufacturer identifiers to Accession Numbers
hgu133aALIAS2PROBE	Map between Common Gene Symbol Identifiers and Manufacturer Identifiers
hgu133aCHR	Map Manufacturer IDs to Chromosomes
hgu133aCHRLENGTHS	A named vector for the length of each of the chromosomes
hgu133aCHRLOC	Map Manufacturer IDs to Chromosomal Location
hgu133aENSEMBL	Map Ensembl gene accession numbers with Entrez Gene identifiers
hgu133aENTREZID	Map between Manufacturer Identifiers and Entrez Gene
hgu133aENZYME	Map between Manufacturer IDs and Enzyme Commission (EC) Numbers
hgu133aENZYME2PROBE	Map between Enzyme Commission Numbers and Manufacturer Identifiers
hgu133aGENENAME	Map between Manufacturer IDs and Genes
hgu133aGO	Map between Manufacturer IDs and Gene Ontology (GO)
hgu133aGO2ALLPROBES	Map between Gene Ontology (GO) Identifiers and all Manufacturer Identifiers in the subtree
hgu133aGO2PROBE	Map between Gene Ontology (GO) and Manufacturer Identifiers

```
>
>
>
>
> library(hgu133a.db)
> library(help=hgu133a.db)
> hgu133a()
Quality control information for hgu133a:


This package has the following mappings:

hgu133aACCNUM has 22283 mapped keys (of 22283 keys)
hgu133aALIAS2PROBE has 54726 mapped keys (of 110701 keys)
hgu133aCHR has 20380 mapped keys (of 22283 keys)
hgu133aCHRLENGTHS has 93 mapped keys (of 93 keys)
hgu133aCHRLOC has 20163 mapped keys (of 22283 keys)
hgu133aCHRLOCEND has 20163 mapped keys (of 22283 keys)
hgu133aENSEMBL has 19792 mapped keys (of 22283 keys)
hgu133aENSEMBL2PROBE has 13125 mapped keys (of 20087 keys)
hgu133aENTREZID has 20387 mapped keys (of 22283 keys)
```

```
> contents(hgu133aSYMBOL)[1:5]
$`1007_s_at`
[1] "DDR1"

$`1053_at`
[1] "RFC2"

$`117_at`
[1] "HSPA6"

$`121_at`
[1] "PAX8"

$`1255_g_at`
[1] "GUCA1A"

> probeset<-rownames(exprs(eset))
> probeset[1:5]
[1] "1007_s_at" "1053_at"   "117_at"    "121_at"    "1255_g_at"
>
```

RKWar
well as

Getti

Due to
introdu

The se
the abo

affylmGUI – Untitled

File RNA Targets Normalization Linear Model Evaluate TopTable Plot Help

Choose working directory

Choose working directory (containing input files):

/home/acb/old/biodata/cmap/tmp

Browse

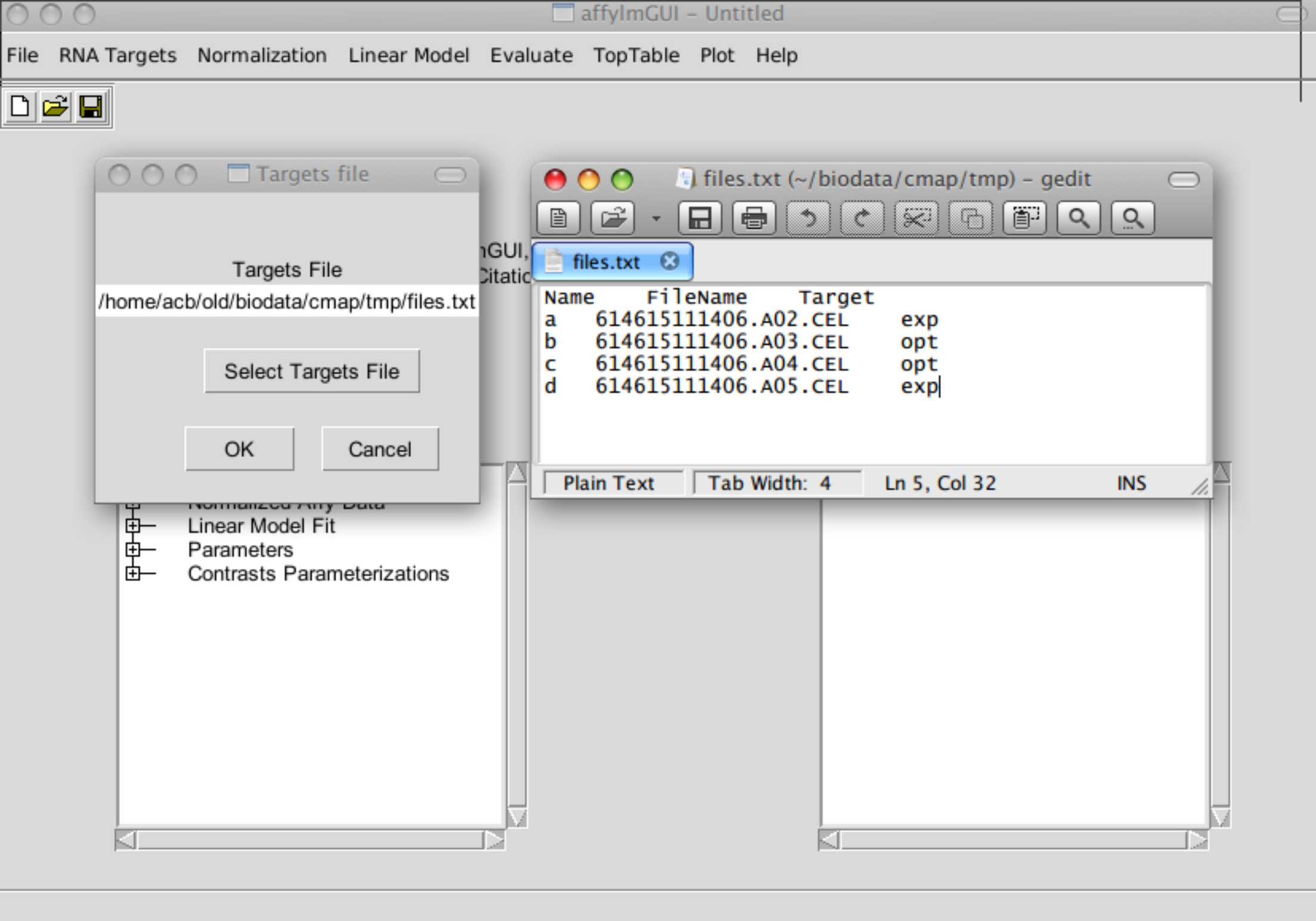
OK Cancel

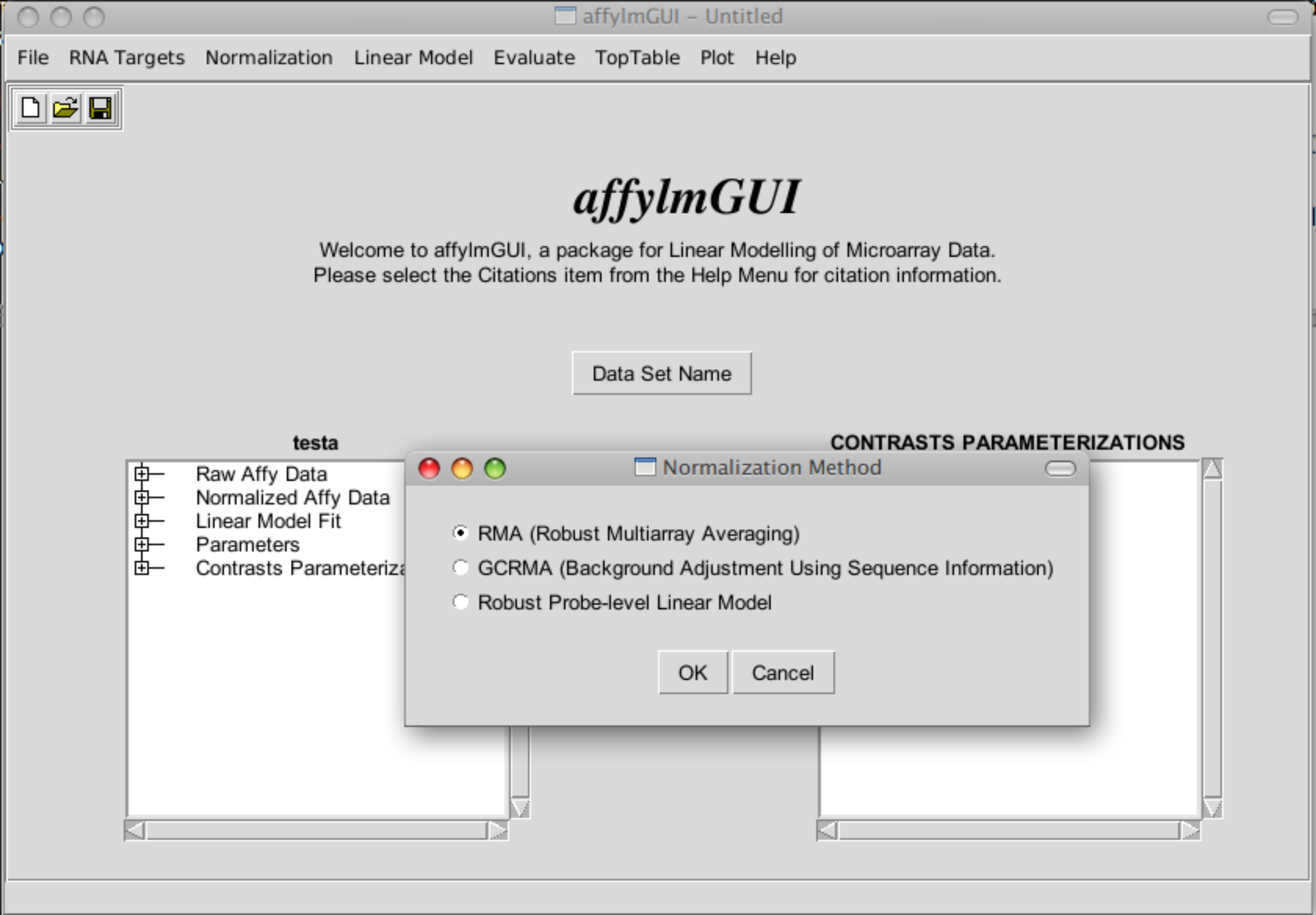
Untitled

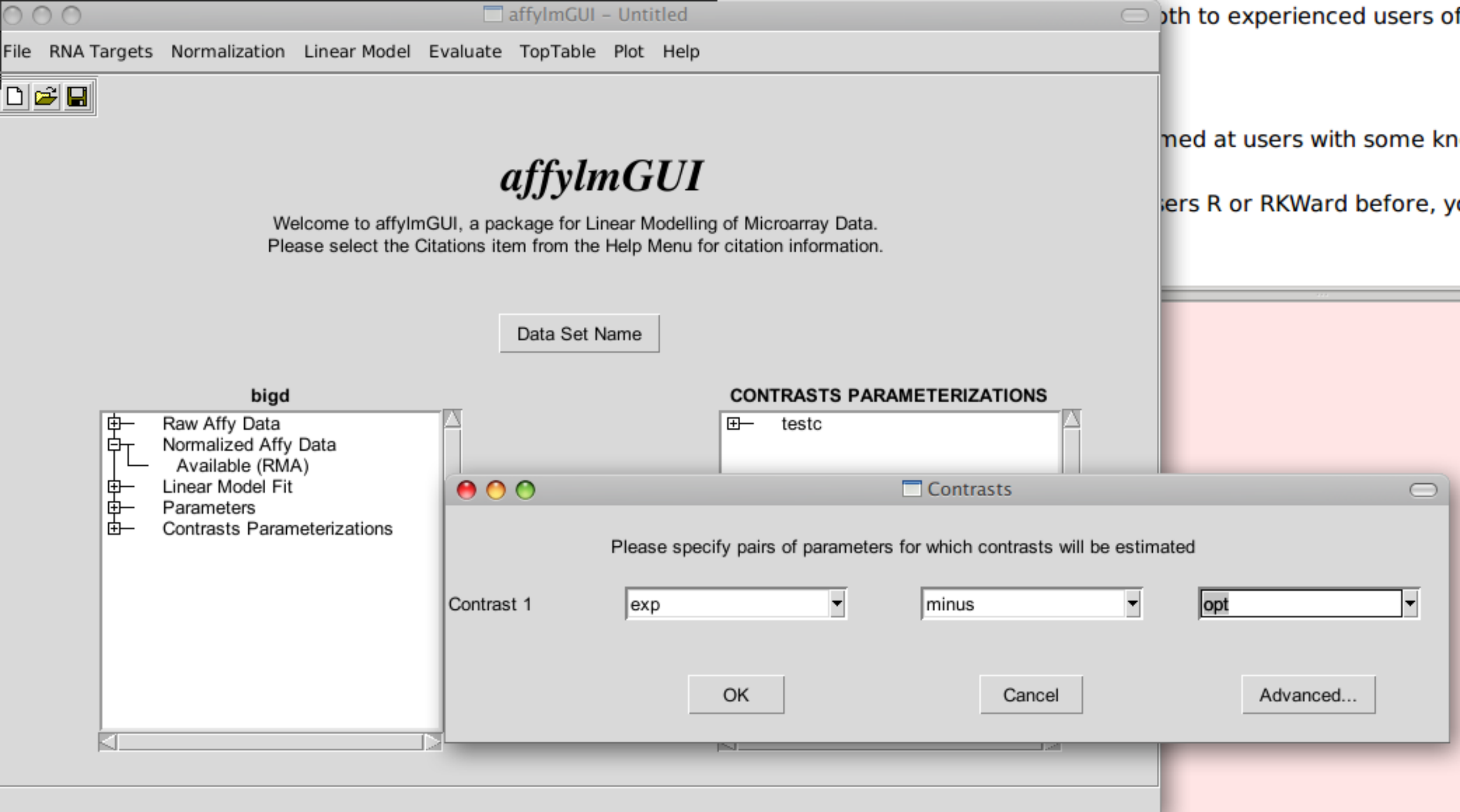
- Raw Affy Data
- Normalized Affy Data
- Linear Model Fit
- Parameters
- Contrasts Parameterizations

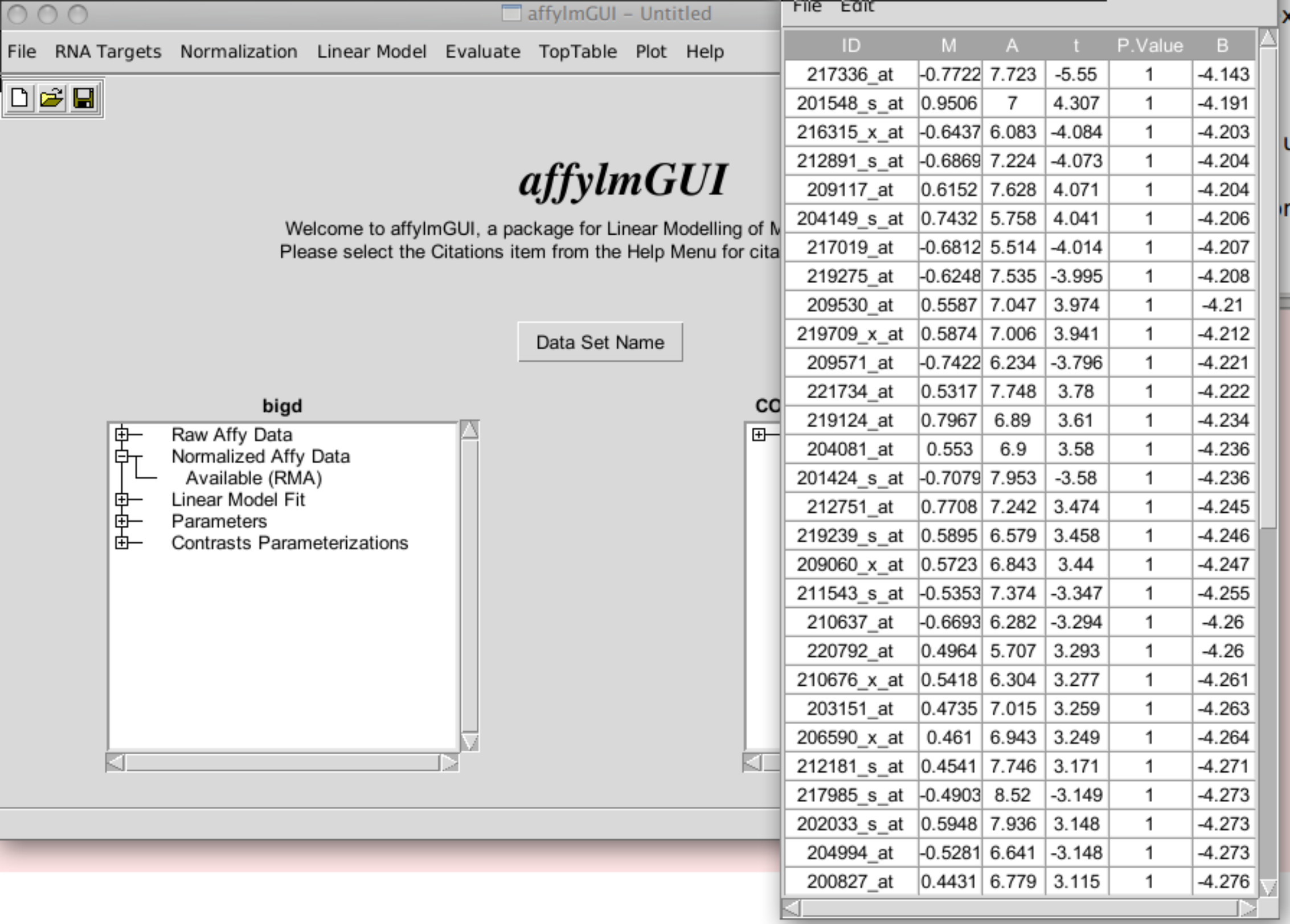
CONTRASTS PARAMETERIZATIONS

```
> library(affylmGUI)
> affylmGUI()

Searching for user-defined affylmGUI commands in /home/john/R/i486-pc-linux-gnu-library/2.15/affylmGUI/etc ...
>
```


affylmGUI - Untitled
File
RNA Targets
Normalization
Linear Model
Evaluate
TopTable
Plot
Help
Targets file
Targets File
/home/acb/old/biodata/cmap/tmp/files.txt
Select Targets File
OK
Cancel
files.txt (~/biodata/cmap/tmp) - gedit
files.txt
Name FileName Target
a 614615111406.A02.CEL exp
b 614615111406.A03.CEL opt
c 614615111406.A04.CEL opt
d 614615111406.A05.CEL exp
Plain Text
Tab Width: 4
Ln 5, Col 32
INS
Linear Model Fit
Parameters
Contrasts Parameterizations


affylmGUI – Untitled
File
RNA Targets
Normalization
Linear Model
Evaluate
TopTable
Plot
Help
affylmGUI
Welcome to affylmGUI, a package for Linear Modelling of Microarray Data.
Please select the Citations item from the Help Menu for citation information.
Data Set Name
testa
CONTRASTS PARAMETERIZATIONS
Raw Affy Data
Normalized Affy Data
Linear Model Fit
Parameters
Contrasts Parameteriza
Normalization Method
RMA (Robust Multiarray Averaging)
GCRMA (Background Adjustment Using Sequence Information)
Robust Probe-level Linear Model
OK
Cancel


affylmGUI – Untitled
File
RNA Targets
Normalization
Linear Model
Evaluate
TopTable
Plot
Help
affylmGUI
Welcome to affylmGUI, a package for Linear Modelling of Microarray Data.
Please select the Citations item from the Help Menu for citation information.
Data Set Name
bigd
Raw Affy Data
Normalized Affy Data
Available (RMA)
Linear Model Fit
Parameters
Contrasts Parameterizations
CONTRASTS PARAMETERIZATIONS
testc
Contrasts
Please specify pairs of parameters for which contrasts will be estimated
Contrast 1
exp
minus
opt
OK
Cancel
Advanced...

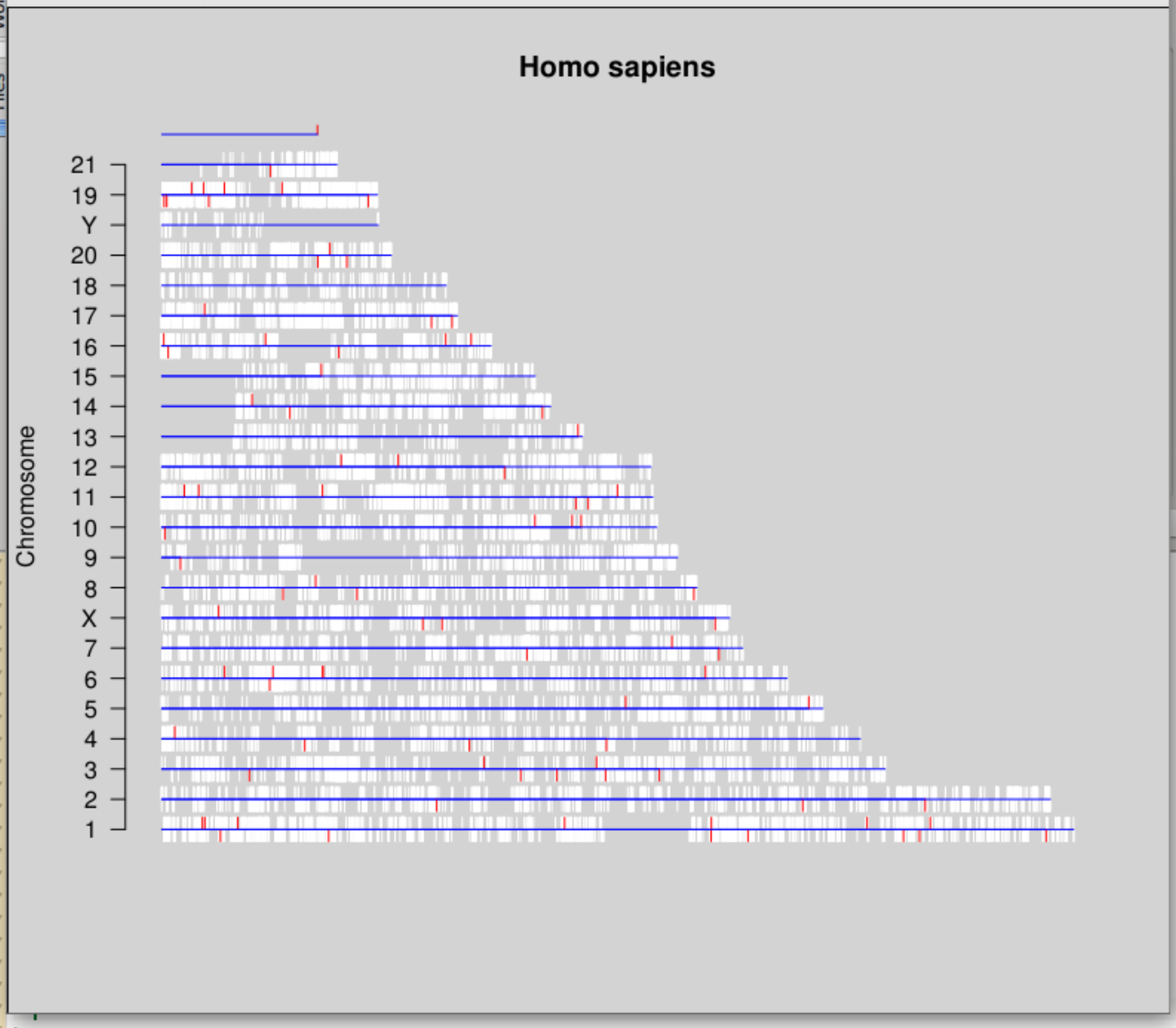


ID	M	A	t	P.Value	B
217336_at	-0.7722	7.723	-5.55	1	-4.143
201548_s_at	0.9506	7	4.307	1	-4.191
216315_x_at	-0.6437	6.083	-4.084	1	-4.203
212891_s_at	-0.6869	7.224	-4.073	1	-4.204
209117_at	0.6152	7.628	4.071	1	-4.204
204149_s_at	0.7432	5.758	4.041	1	-4.206
217019_at	-0.6812	5.514	-4.014	1	-4.207
219275_at	-0.6248	7.535	-3.995	1	-4.208
209530_at	0.5587	7.047	3.974	1	-4.21
219709_x_at	0.5874	7.006	3.941	1	-4.212
209571_at	-0.7422	6.234	-3.796	1	-4.221
221734_at	0.5317	7.748	3.78	1	-4.222
219124_at	0.7967	6.89	3.61	1	-4.234
204081_at	0.553	6.9	3.58	1	-4.236
201424_s_at	-0.7079	7.953	-3.58	1	-4.236
212751_at	0.7708	7.242	3.474	1	-4.245
219239_s_at	0.5895	6.579	3.458	1	-4.246
209060_x_at	0.5723	6.843	3.44	1	-4.247
211543_s_at	-0.5353	7.374	-3.347	1	-4.255
210637_at	-0.6693	6.282	-3.294	1	-4.26
220792_at	0.4964	5.707	3.293	1	-4.26
210676_x_at	0.5418	6.304	3.277	1	-4.261
203151_at	0.4735	7.015	3.259	1	-4.263
206590_x_at	0.461	6.943	3.249	1	-4.264
212181_s_at	0.4541	7.746	3.171	1	-4.271
217985_s_at	-0.4903	8.52	-3.149	1	-4.273
202033_s_at	0.5948	7.936	3.148	1	-4.273
204994_at	-0.5281	6.641	-3.148	1	-4.273
200827_at	0.4431	6.779	3.115	1	-4.276

R Graphics: Device 2 (ACTIVE) – RKWard

Device Edit View History Window Settings Help

Copy to output | Go to plot | Previous plot | Next plot | Remove this plot



rkward_welcome | toptablef | Ou

	1	2	3
Name	X	ID	M
Label			
Type	Number	Factor	Number
Format			
Levels		200054_at#,...	

1	16701	217336_at	-0.772243
2	1076	201548_s_at	0.950628
3	15687	216315_x_at	-0.643748
4	12276	212891_s_at	-0.686907
5	8611	209117_at	0.615163
6	3676	204149_s_at	0.743227

```
>
>
>
> library(annotate)
> library(geneplotter)
> library("hgu133a.db")
> newChrom <- buildChromLocation("hgu133a.db")
> cPlot(newChrom,c(1:21,'X','Y'))
> cColor(as.character(toptablef[["ID"]]),"red",newChrom)
```

```
> ?unlist
> library(GOstats); library(GO.db); library(ath1121501.db); library(annotate)
Loading required package: Biobase
Loading required package: BiocGenerics

Attaching package: 'BiocGenerics'

The following object(s) are masked from 'package:stats':

    xtabs

The following object(s) are masked from 'package:base':

    anyDuplicated, cbind, colnames, duplicated, eval, Filter, Find,
    get, intersect, lapply, Map, mapply, mget, order, paste, pmax,
    pmax.int, pmin, pmin.int, Position, rbind, Reduce, rep.int,
    rownames, sapply, setdiff, table, tapply, union, unique

Welcome to Bioconductor

    Vignettes contain introductory material; view with
    'browseVignettes()'. To cite Bioconductor, see
    'citation("Biobase")', and for packages 'citation("pkgname")'.

Loading required package: Category
Loading required package: AnnotationDbi
Loading required package: graph
Loading required package: DBI

Loading required package: org.At.tair.db


> GOTERM$"GO:0003700"
GOID: GO:0003700
Term: sequence-specific DNA binding transcription factor activity
Ontology: MF
Definition: Interacting selectively and non-covalently with a specific
    DNA sequence in order to modulate transcription. The transcription
    factor may or may not also interact selectively with a protein or
    macromolecular complex.
Synonym: transcription factor activity
Synonym: GO:0000130
Secondary: GO:0000130
> GOTERM$"GO:0003700"@Ontology
[1] "MF"
> GOTERM$"GO:0003700"@Term
[1] "sequence-specific DNA binding transcription factor activity"
> zz <- eapply(GOTERM, function(x) x@Ontology)
>
```

Upload List Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

213809_x_at
203266_s_at
202912_at
215818_at

Clear

Or

B:Choose From a File

Choose File No file chosen

Multi-List File

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID

Step 3: List Type

Gene List
Background

Step 4: Submit List

Submit List

Analysis Wizard

Tell us how you like the tool
Contact us for questions

Step 1. Successfully submitted gene list

Current Gene List: List_1
Current Background: Human Genome U133A Array

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

Functional Annotation Tool

- Functional Annotation Clustering
- Functional Annotation Chart
- Functional Annotation Table

Gene Functional Classification Tool

Gene ID Conversion Tool

Gene Name Batch Viewer

Upload List
Background

Population Manager

Select a background Help

HT Human Genome U133A
Homo sapiens
Human Genome U133A Arra

Select List to:

Use Rename

Affymetrix 3' IVT Backgrounds H

- Human Genome U133A 2 Array
- Human Genome U133A Array
- Human Genome U133B Array
- Human Genome U95A Array
- Human Genome U95Av2 Array

Affymetrix Exon Backgrounds He

- HuEx-1_0-st-v2
- HuGene-1_0-st-v1
- MoEx-1_0-st-v1
- MoGene-1_0-st-v1
- RaEx-1_0-st-v1

Affymetrix SNP Backgrounds Help

Open Create Save Cut Copy Paste Paste inside selection Paste

Workspace Files

All Non-Functions Functions

Show All Environments

Show Hidden Objects

Name	Label	Type	Class
+ package:annotate			
+ package:org.Hs.eg.db			
+ package:hgu133a.db			

Update

top

Name X
Label
1

```
> hgu133a.db
ChipDb object:
| DBSCHEMAVERSION: 2.1
| Db type: ChipDb
| Supporting package: AnnotationDbi
| DBSCHEMA: HUMANCHIP_DB
| ORGANISM: Homo sapiens
| SPECIES: Human
| MANUFACTURER: Affymetrix
| CHIPNAME: Human Genome U133 Set
| MANUFACTURERURL: http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133
| EGSOURCEDATE: 2012-Mar7
| EGSOURCENAME: Entrez Gene
| EGSOURCEURL: ftp://ftp.ncbi.nlm.nih.gov/gene/DATA
| CENTRALID: ENTREZID
| TAXID: 9606
| GOSOURCENAME: Gene Ontology
| GOSOURCEURL: ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/
| GOSOURCEDATE: 20120303
```

Command log R Console Help search

Ready.

Step 1. Successfully submitted gene list

Current Gene List: List_1

Current Background: Human Genome U133A Array

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

- Functional Annotation Tool
  - Functional Annotation Clustering
  - Functional Annotation Chart
  - Functional Annotation Table
- Gene Functional Classification Tool
- Gene ID Conversion Tool
- Gene Name Batch Viewer

Annotation Summary Results

Help and Tool Manual

Current Gene List: List_1 **99 DAVID IDs**

Current Background: Human Genome U133A Array **Check Defaults** ☑ Clear All

- ⊞ **Disease** (1 selected)
- ⊞ **Functional_Categories** (3 selected)
- ⊞ **Gene_Ontology** (3 selected)
- ⊞ **General Annotations** (0 selected)
- ⊞ **Literature** (0 selected)
- ⊞ **Main_Accessions** (0 selected)
- ⊞ **Pathways** (3 selected)
- ⊞ **Protein_Domains** (3 selected)
- ⊞ **Protein_Interactions** (0 selected)
- ⊞ **Tissue_Expression** (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

Functional Annotation Table

Gene_Ontology (3 selected)

GOTERM_BP_1	74.7%	74	Chart	
GOTERM_BP_2	73.7%	73	Chart	
GOTERM_BP_3	70.7%	70	Chart	
GOTERM_BP_4	68.7%	68	Chart	
GOTERM_BP_5	64.6%	64	Chart	
GOTERM_BP_ALL	74.7%	74	Chart	
☑ GOTERM_BP_FAT	70.7%	70	Chart	
GOTERM_CC_1	79.8%	79	Chart	
GOTERM_CC_2	78.8%	78	Chart	
GOTERM_CC_3	78.8%	78	Chart	
GOTERM_CC_4	78.8%	78	Chart	
GOTERM_CC_5	78.8%	78	Chart	
GOTERM_CC_ALL	79.8%	79	Chart	
☑ GOTERM_CC_FAT	59.6%	59	Chart	
GOTERM_MF_1	75.8%	75	Chart	
GOTERM_MF_2	74.7%	74	Chart	
GOTERM_MF_3	63.6%	63	Chart	
GOTERM_MF_4	60.6%	60	Chart	
GOTERM_MF_5	49.5%	49	Chart	
GOTERM_MF_ALL	75.8%	75	Chart	
☑ GOTERM_MF_FAT	64.6%	64	Chart	
PANTHER_BP_ALL	58.6%	58	Chart	
PANTHER_MF_ALL	67.7%	67	Chart	

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of Immunopathogenesis and Bioinformatics (LIB); National Institute of Allergies and Infectious Dise

david.abcc.ncifcrf.gov/chartReport.jsp?annot=39

99 DAVID IDs

Options

Rerun Using Options | Create Sublist

6 chart records

Download

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjam
	GOTERM_MF_FAT	calmodulin binding	RT		5	5.1	1.1E-2	8.7E-1
	GOTERM_MF_FAT	specific transcriptional repressor activity	RT		3	3.0	1.7E-2	8.0E-1
	GOTERM_MF_FAT	transcription repressor activity	RT		6	6.1	3.5E-2	9.0E-1
	GOTERM_MF_FAT	DNA-dependent ATPase activity	RT		3	3.0	4.8E-2	9.1E-1
	GOTERM_MF_FAT	DNA binding	RT		17	17.2	7.0E-2	9.4E-1
	GOTERM_MF_FAT	magnesium ion binding	RT		6	6.1	8.0E-2	9.3E-1

72 gene(s) from your list are not in the output.

General Annotations (0 selected)

Literature (0 selected)

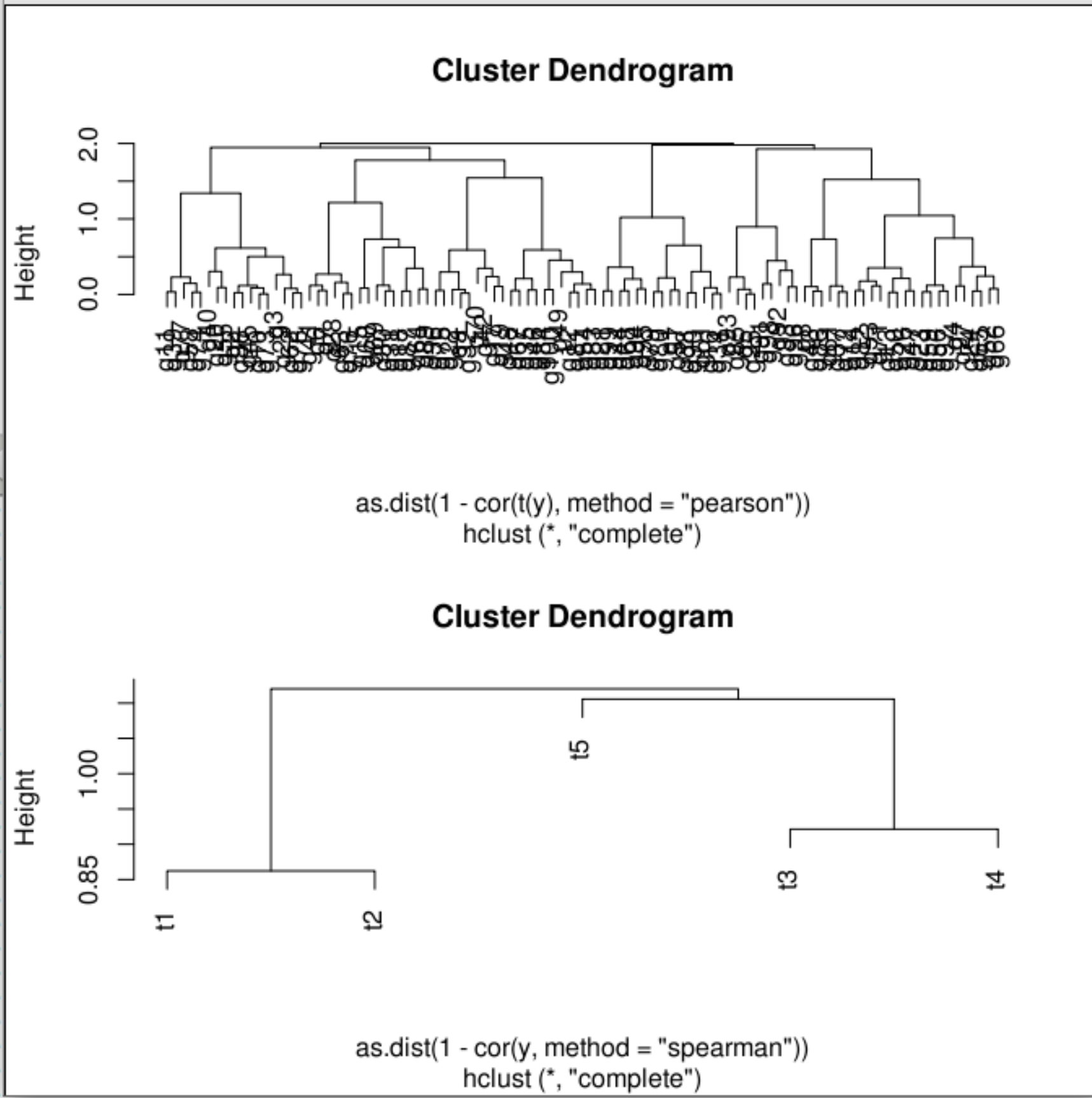
	t1	t2	t3	t4	t5
g1	-1.04520178	-0.380935801	0.32253285	0.188196643	0.28007615
g2	0.30260081	-0.970551319	2.59950534	1.176315379	0.18996084
g3	0.15975841	-0.478759036	1.73153639	1.061661475	1.57966492
g4	-0.78985782	0.081254590	-0.53945136	0.234886833	0.77129062
g5	-0.07311381	-0.375386205	-0.40928292	-1.308708764	-0.63216358
g6	-2.34589634	0.840200643	0.66167230	-1.007925101	-1.21183069
g7	-0.87402211	0.733826049	1.25435683	-0.294792532	-0.63150717
g8	1.37970856	-0.586133357	-0.92211485	-0.652446545	-0.04455500
g9	-0.38065076	0.303324485	-1.08968012	0.418261480	-1.33007870
g10	-0.18620830	-0.841703730	0.03012982	-0.278655199	-0.64582184
g11	1.14214019	0.053108651	0.23298126	0.899382537	-0.80270697
g12	1.44513788	-0.558154629	2.04270902	0.321967760	-1.54073249
g13	0.99803579	-0.776420713	-0.43827247	1.243851533	1.35663368
g14	0.96289504	-0.007877564	-1.09371899	1.340719288	1.60252259
g15	-1.32178711	1.120730238	0.37916397	1.248446497	-0.46664231
g16	-0.56333368	-1.093202573	-0.24292916	1.186015185	1.16267207
g17	-0.44602650	1.497993750	-1.32988839	0.180666341	-0.66531597
g18	-0.59273243	0.809950819	0.91060528	0.828761280	-1.71046485
g19	0.83650966	2.229876451	-0.81521291	0.081180504	0.20237456
g20	1.20476925	-1.801373447	0.48730660	1.276596164	0.25094667
g21	-1.91517457	-0.611499880	-0.12086295	-1.251266330	0.88628046
g22	0.18428016	1.182773003	0.39024382	0.638688662	1.05100182

```
> hr <- hclust(as.dist(1-cor(t(y), method="pearson")), method="complete")
> hr

Call:
hclust(d = as.dist(1 - cor(t(y), method = "pearson")), method = "complete")

Cluster method   : complete
Number of objects: 100

> hc <- hclust(as.dist(1-cor(y, method="spearman")), method="complete")
> hc

Call:
hclust(d = as.dist(1 - cor(y, method = "spearman")), method = "complete")

Cluster method   : complete
Number of objects: 5

> str(hc)
List of 7
 $ merge      : int [1:4, 1:2] -1 -3 -5 1 -2 -4 2 3
 $ height     : num [1:4] 0.862 0.921 1.106 1.12
 $ order      : int [1:5] 1 2 5 3 4
 $ labels     : chr [1:5] "t1" "t2" "t3" "t4" ...
 $ method     : chr "complete"
 $ call       : language hclust(d = as.dist(1 - cor(y, method = "spearman")), method = "complete")
 $ dist.method: NULL
 - attr(*, "class")= chr "hclust"
>
```

R Graphics: Device 2 (ACTIVE) - RKWard

Device Edit View History Window Settings Help

Copy to output | Go to plot | Previous plot | Next plot | Remove this plot

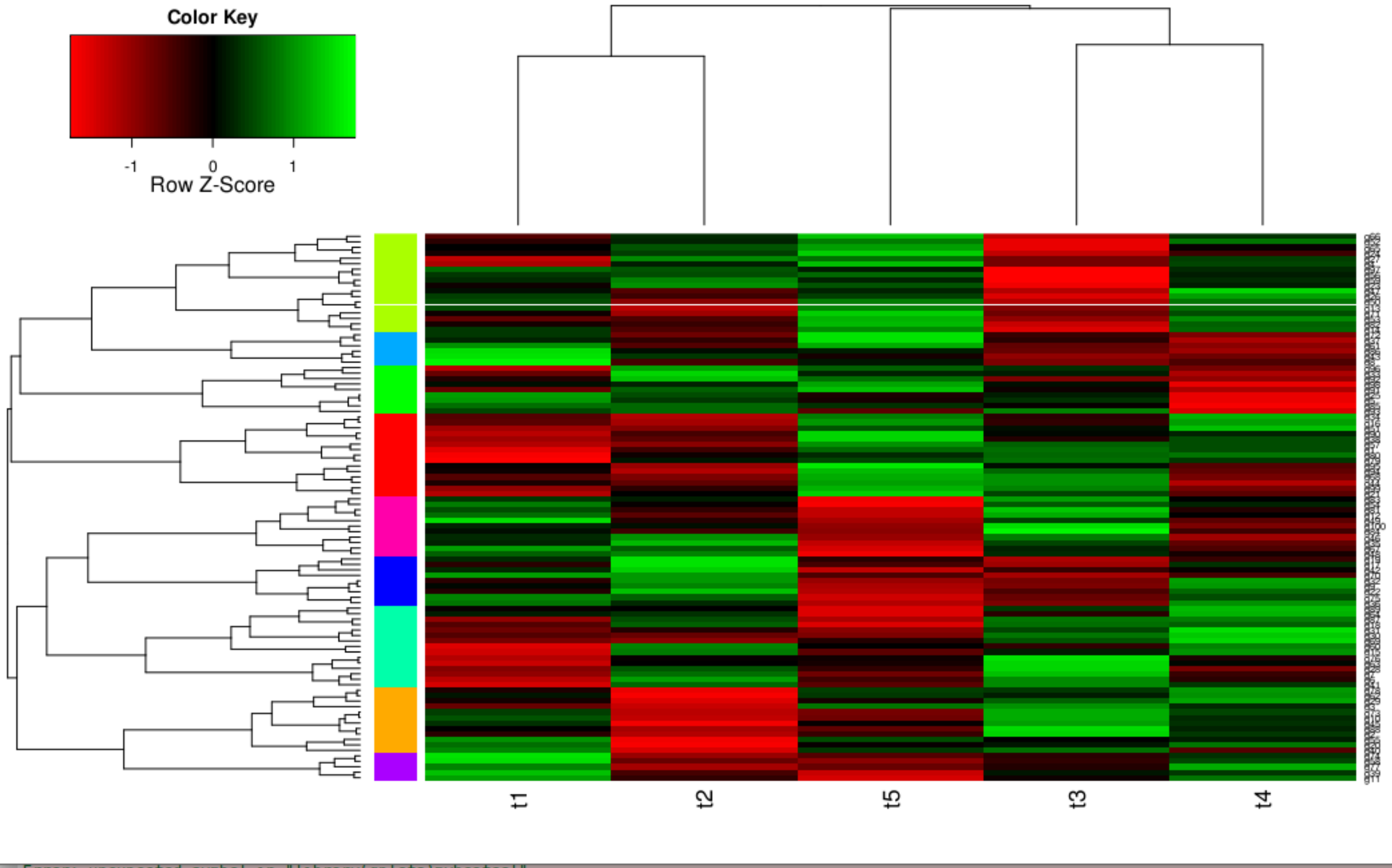



```
par(mfrow = c(2, 1))
plot(hr)
plot(hc)
```

```
> mycl <- cutree(hr, h=max(hr$height)/1.5)
> mycl
  g1   g2   g3   g4   g5   g6   g7   g8   g9  g10  g11  g12  g13  g14  g15  g16
   1    2    2    3    4    5    5    6    7    2    8    9    3    3    5    1
 g17  g18  g19  g20  g21  g22  g23  g24  g25  g26  g27  g28  g29  g30  g31  g32
   7    5    7    2    1    7    3    3    4    3    3    5    2    5    5    7
 g33  g34  g35  g36  g37  g38  g39  g40  g41  g42  g43  g44  g45  g46  g47  g48
   4    1    9    7    6    1    8    2    5    7    6    1    2    9    3    9
 g49  g50  g51  g52  g53  g54  g55  g56  g57  g58  g59  g60  g61  g62  g63  g64
   9    3    1    3    3    9    2    3    1    8    3    5    6    2    5    5
 g65  g66  g67  g68  g69  g70  g71  g72  g73  g74  g75  g76  g77  g78  g79  g80
   3    3    9    1    5    7    3    6    2    8    7    5    8    2    1    1
 g81  g82  g83  g84  g85  g86  g87  g88  g89  g90  g91  g92  g93  g94  g95  g96
   9    3    9    9    4    6    5    2    5    1    4    4    4    1    1    4
 g97  g98  g99 g100
   3    4    1    9
> ?rainbow
> mycolhc <- rainbow(length(unique(mycl)))
> mycolhc <- mycolhc[as.vector(mycl)]
> mycolhc
  [1] "#FF0000FF" "#FFAA00FF" "#FFAA00FF" "#AAFF00FF" "#00FF00FF" "#00FFAAFF"
  [7] "#00FFAAFF" "#00AAFFFF" "#0000FFFF" "#FFAA00FF" "#AA00FFFF" "#FF00AAFF"
 [13] "#AAFF00FF" "#AAFF00FF" "#00FFAAFF" "#FF0000FF" "#0000FFFF" "#00FFAAFF"
 [19] "#0000FFFF" "#FFAA00FF" "#FF0000FF" "#0000FFFF" "#AAFF00FF" "#AAFF00FF"
 [25] "#00FF00FF" "#AAFF00FF" "#AAFF00FF" "#00FFAAFF" "#FFAA00FF" "#00FFAAFF"
 [31] "#00FFAAFF" "#0000FFFF" "#00FF00FF" "#FF0000FF" "#FF00AAFF" "#0000FFFF"
 [37] "#00AAFFFF" "#FF0000FF" "#AA00FFFF" "#FFAA00FF" "#00FFAAFF" "#0000FFFF"
 [43] "#00AAFFFF" "#FF0000FF" "#FFAA00FF" "#FF00AAFF" "#AAFF00FF" "#FF00AAFF"
 [49] "#FF00AAFF" "#AAFF00FF" "#FF0000FF" "#AAFF00FF" "#AAFF00FF" "#FF00AAFF"
 [55] "#FFAA00FF" "#AAFF00FF" "#FF0000FF" "#AA00FFFF" "#AAFF00FF" "#00FFAAFF"
 [61] "#00AAFFFF" "#FFAA00FF" "#00FFAAFF" "#00FFAAFF" "#AAFF00FF" "#AAFF00FF"
 [67] "#FF00AAFF" "#FF0000FF" "#00FFAAFF" "#0000FFFF" "#AAFF00FF" "#00AAFFFF"
 [73] "#FFAA00FF" "#AA00FFFF" "#0000FFFF" "#00FFAAFF" "#AA00FFFF" "#FFAA00FF"
 [79] "#FF0000FF" "#FF0000FF" "#FF00AAFF" "#AAFF00FF" "#FF00AAFF" "#FF00AAFF"
 [85] "#00FF00FF" "#00AAFFFF" "#00FFAAFF" "#FFAA00FF" "#00FFAAFF" "#FF0000FF"
 [91] "#00FF00FF" "#00FF00FF" "#00FF00FF" "#FF0000FF" "#FF0000FF" "#00FF00FF"
 [97] "#AAFF00FF" "#00FF00FF" "#FF0000FF" "#FF00AAFF"
>
```

Device Edit View History Window Settings Help

Copy to output | Go to plot | Previous plot | Next plot | Remove this plot

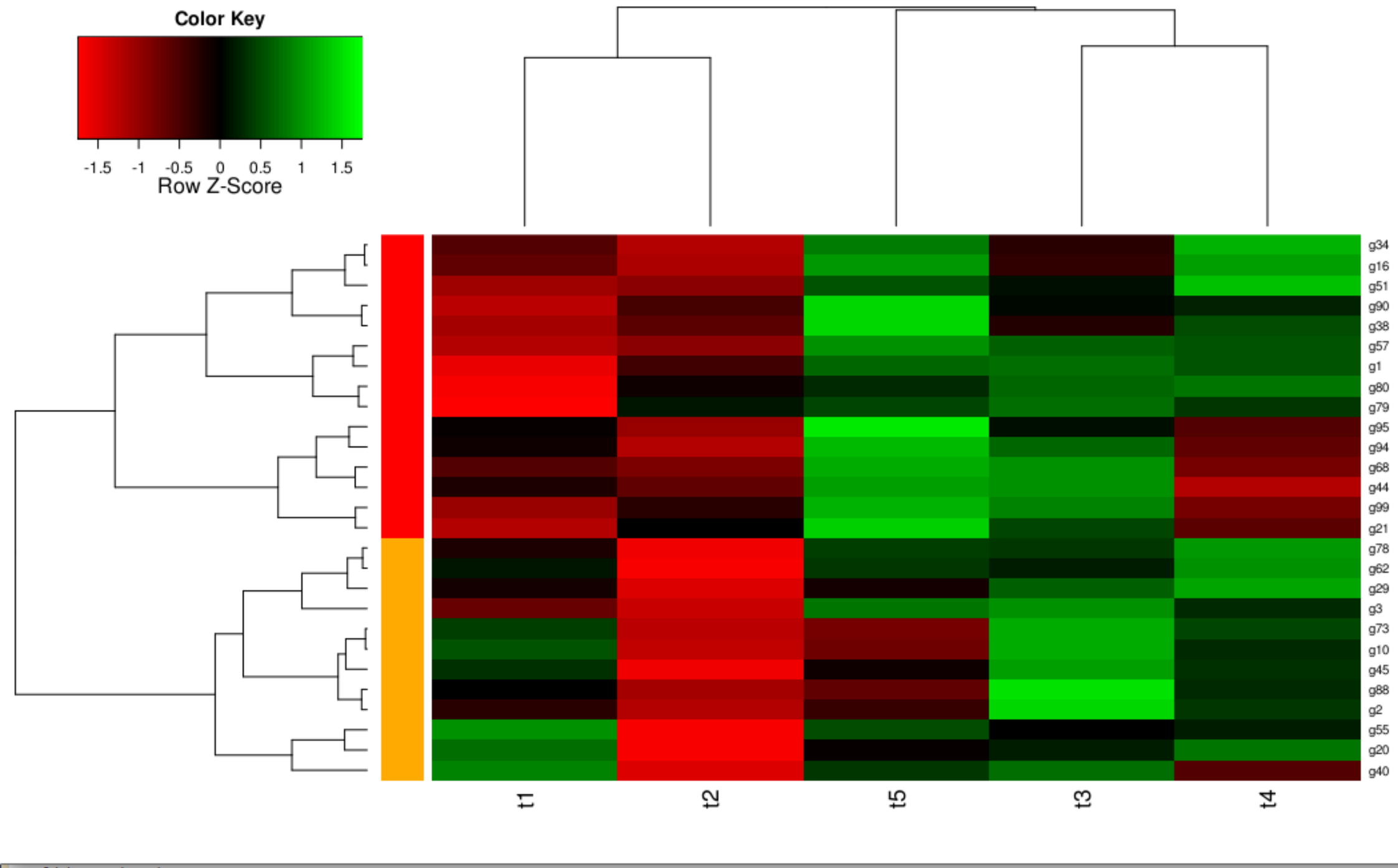



```
Error: unexpected symbol in "library(gplots)myheatcol"
> library(gplots)
> myheatcol <- redgreen(75)
> myheatcol
 [1] "#FF0000" "#F80000" "#F10000" "#EA0000" "#E30000" "#DD0000" "#D60000"
 [8] "#CF0000" "#C80000" "#C10000" "#BA0000" "#B30000" "#AC0000" "#A50000"
[15] "#9F0000" "#980000" "#910000" "#8A0000" "#830000" "#7C0000" "#750000"
[22] "#6E0000" "#670000" "#600000" "#5A0000" "#530000" "#4C0000" "#450000"
[29] "#3E0000" "#370000" "#300000" "#290000" "#220000" "#1C0000" "#150000"
[36] "#0E0000" "#070000" "#000000" "#000700" "#000E00" "#001500" "#001C00"
[43] "#002200" "#002900" "#003000" "#003700" "#003E00" "#004500" "#004C00"
[50] "#005300" "#005A00" "#006000" "#006700" "#006E00" "#007500" "#007C00"
[57] "#008300" "#008A00" "#009100" "#009800" "#009F00" "#00A500" "#00AC00"
[64] "#00B300" "#00BA00" "#00C100" "#00C800" "#00CF00" "#00D600" "#00DD00"
[71] "#00E300" "#00EA00" "#00F100" "#00F800" "#00FF00"
> heatmap.2(y, Rowv=as.dendrogram(hr), Colv=as.dendrogram(hc), col=myheatcol, scale="row", density.info="none", trace="none", RowSideColors=mycolhc) # C
>
```

R Graphics: Device 2 (ACTIVE) – RKWard

Device Edit View History Window Settings Help

Copy to output | Go to plot | Previous plot | Next plot | Remove this plot



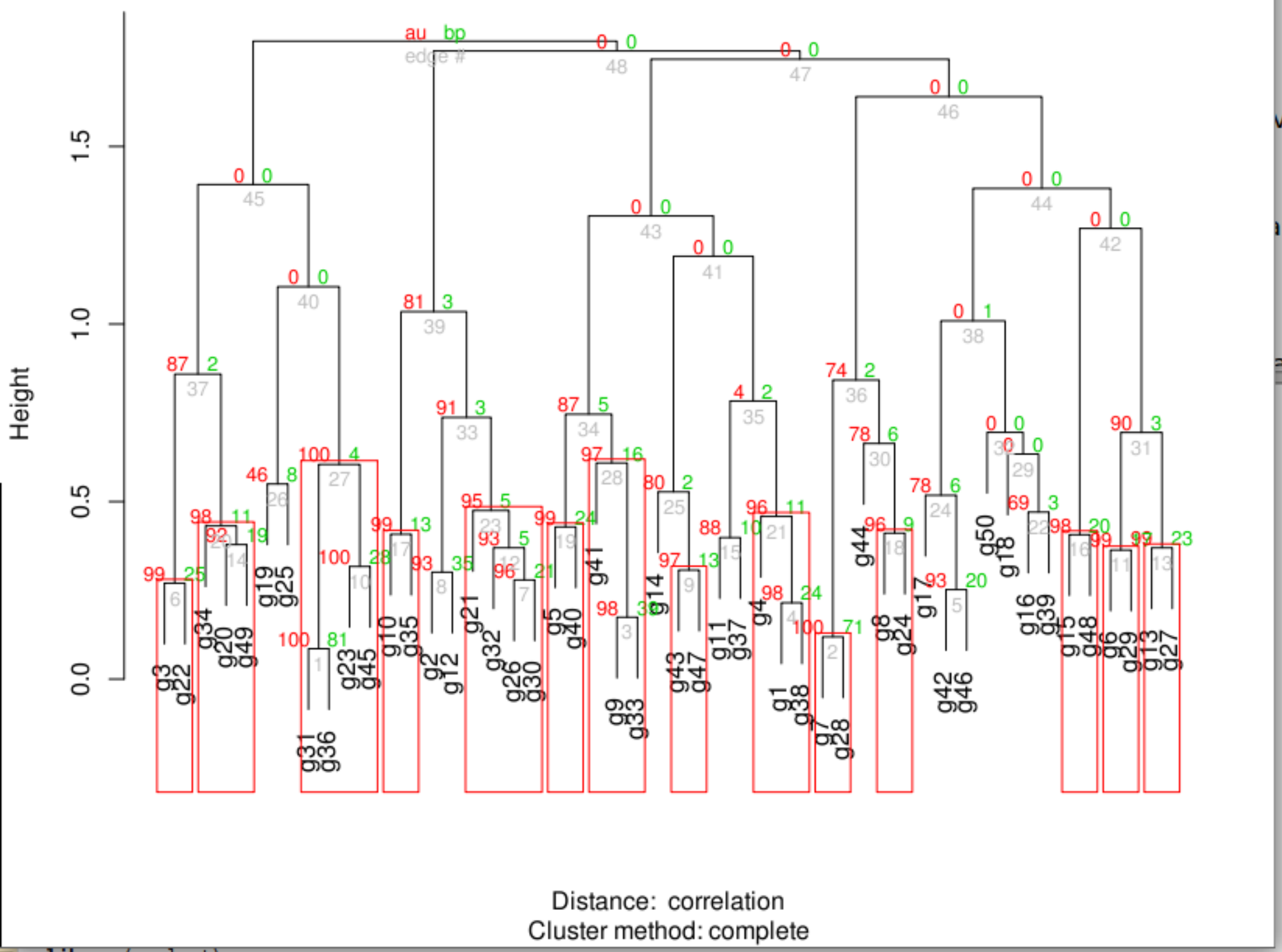
Object Viewer: y

colors.

start = 0, e

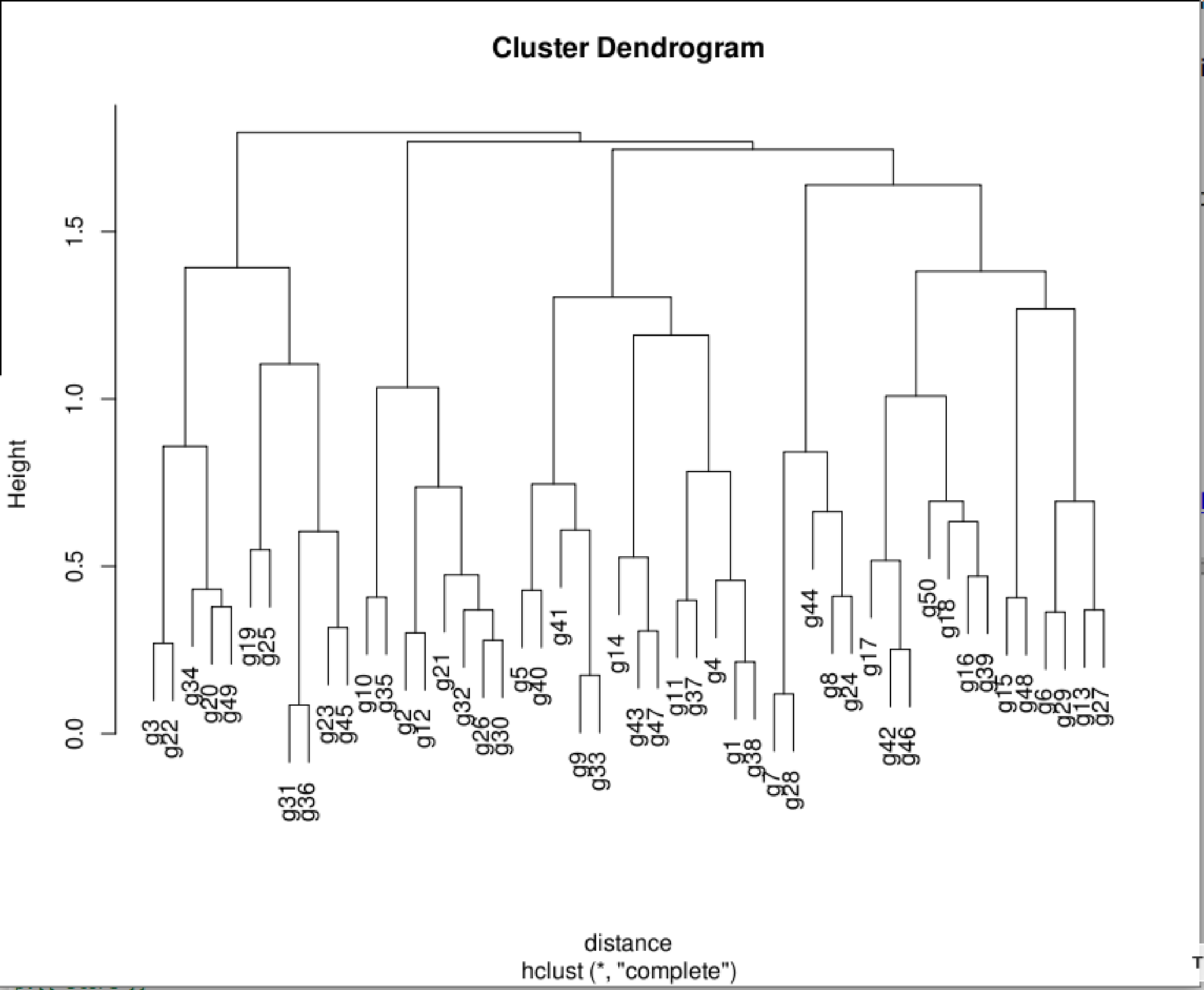
1)

```
> clid <- c(1,2)
> ysub <- y[names(mycl[mycl%in%clid]),]
> mycl[mycl%in%clid]
 g1  g2  g3 g10 g16 g20 g21 g29 g34 g38 g40 g44 g45 g51 g55 g57 g62 g68 g73 g78
  1   2   2   2   1   2   1   2   1   1   2   1   2   1   2   1   2   1   2   2
g79 g80 g88 g90 g94 g95 g99
  1   1   2   1   1   1   1
> names(mycl[mycl%in%clid])
 [1] "g1"  "g2"  "g3"  "g10" "g16" "g20" "g21" "g29" "g34" "g38" "g40" "g44"
[13] "g45" "g51" "g55" "g57" "g62" "g68" "g73" "g78" "g79" "g80" "g88" "g90"
[25] "g94" "g95" "g99"
> hrsub <- hclust(as.dist(1-cor(t(ysub), method="pearson")), method="complete")
> heatmap.2(ysub, Rowv=as.dendrogram(hrsub), Colv=as.dendrogram(hc), col=myheatcol, scale="row", density.info="none", trace="none", RowSideColors=mycolhc[mycl%in%clid])
>
```

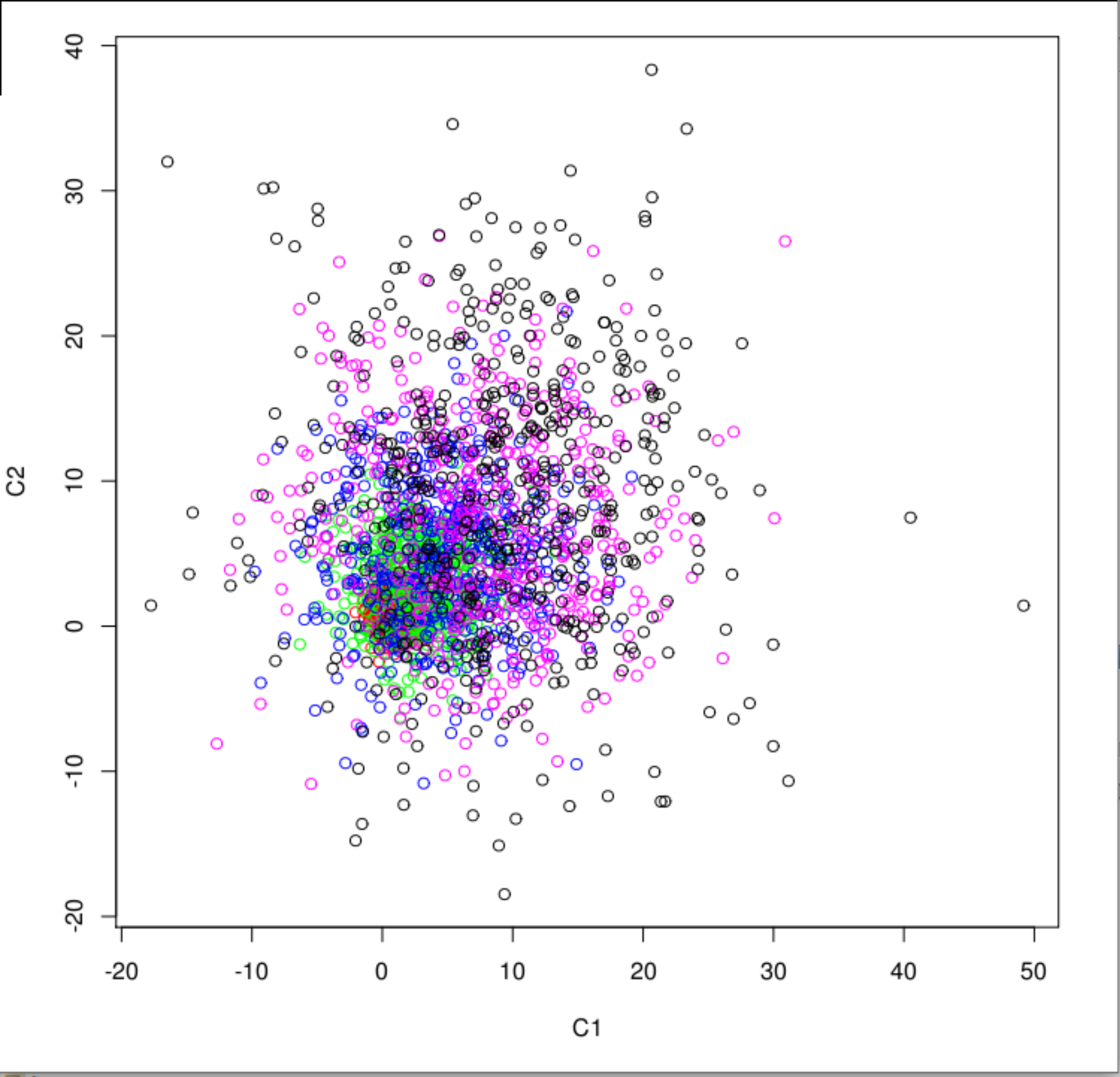



```
> library(pvclust)
> pv <- pvclust(scale(t(y)), method.dist="correlation", method.hclust="complete", nboot=100)
Bootstrap (r = 0.5)... Done.
Bootstrap (r = 0.6)... Done.
Bootstrap (r = 0.7)... Done.
Bootstrap (r = 0.8)... Done.
Bootstrap (r = 0.9)... Done.
Bootstrap (r = 1.0)... Done.
Bootstrap (r = 1.1)... Done.
Bootstrap (r = 1.2)... Done.
Bootstrap (r = 1.3)... Done.
Bootstrap (r = 1.4)... Done.
> plot(pv)
> pvrect(pv, alpha=0.95)
>
```

```
> pvpick(pv, alpha=0.95, pv="au", type="geq", max.only=TRUE)
$clusters
$clusters[[1]]
[1] "g7"  "g28"

$clusters[[2]]
[1] "g3"  "g22"

$clusters[[3]]
[1] "g43" "g47"

$clusters[[4]]
[1] "g6"  "g29"

$clusters[[5]]
[1] "g13" "g27"

$clusters[[6]]
[1] "g15" "g48"
```




```
> plot(pv[["hclust"]])
```

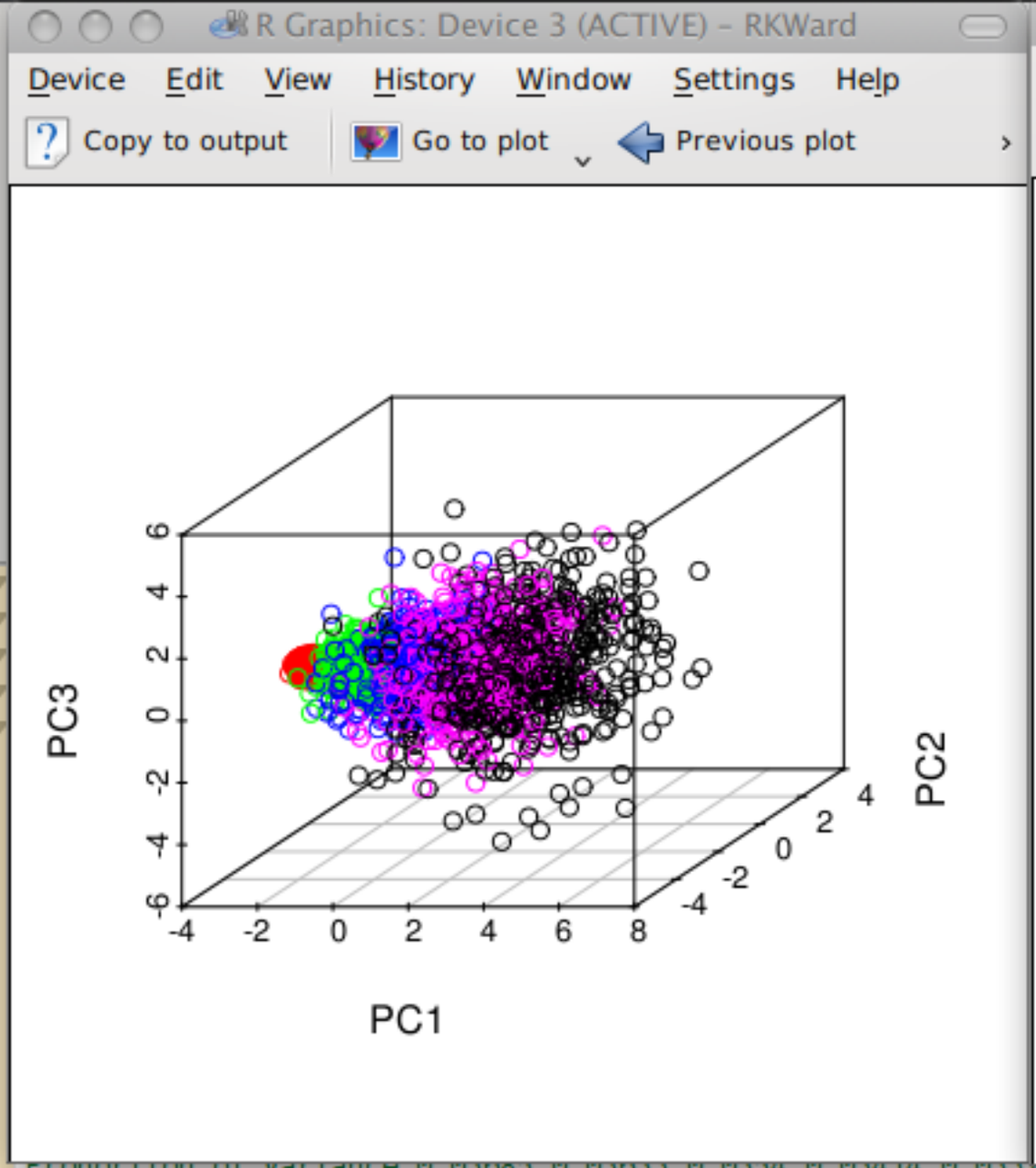


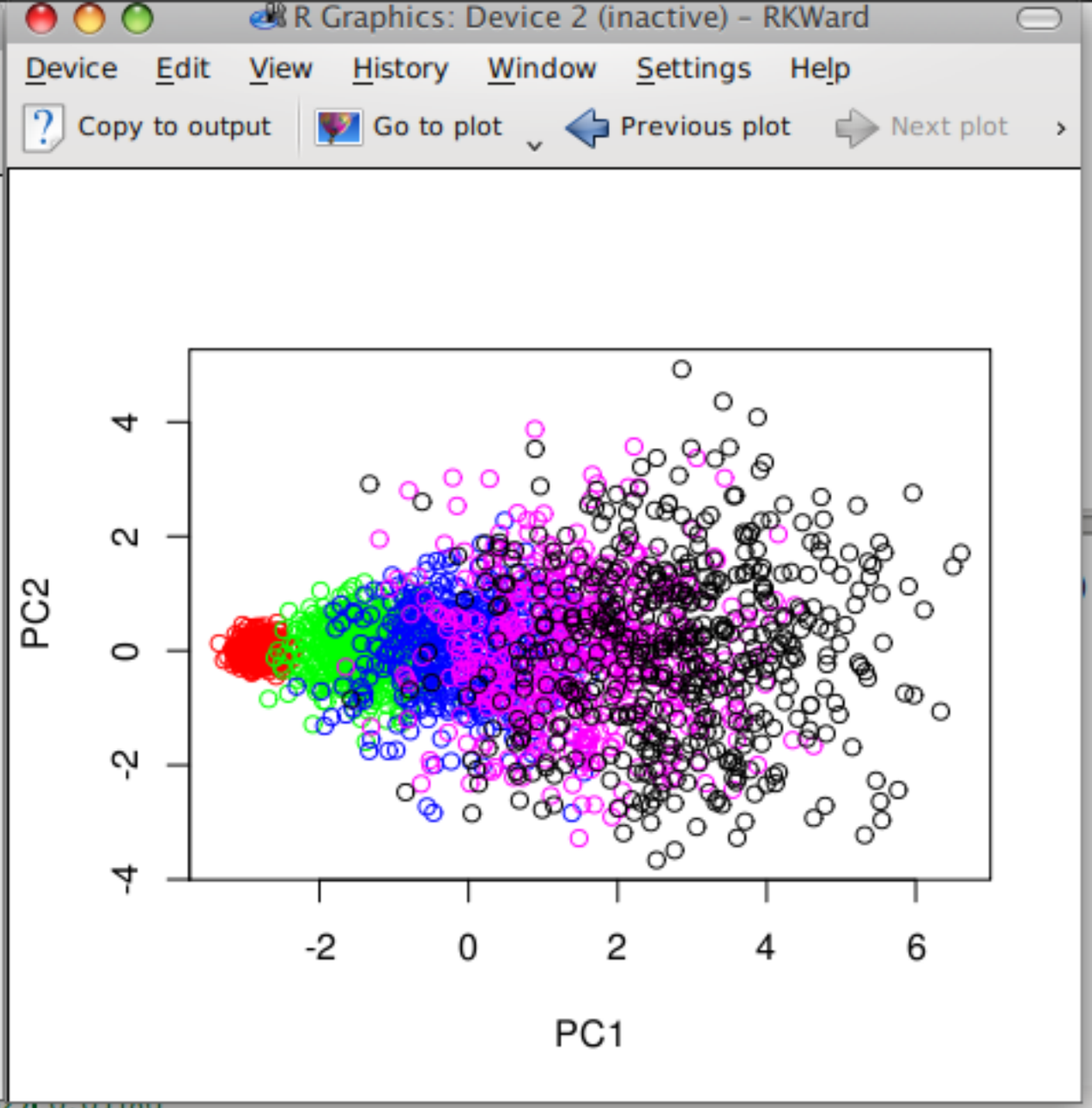
summary (x)

Ready

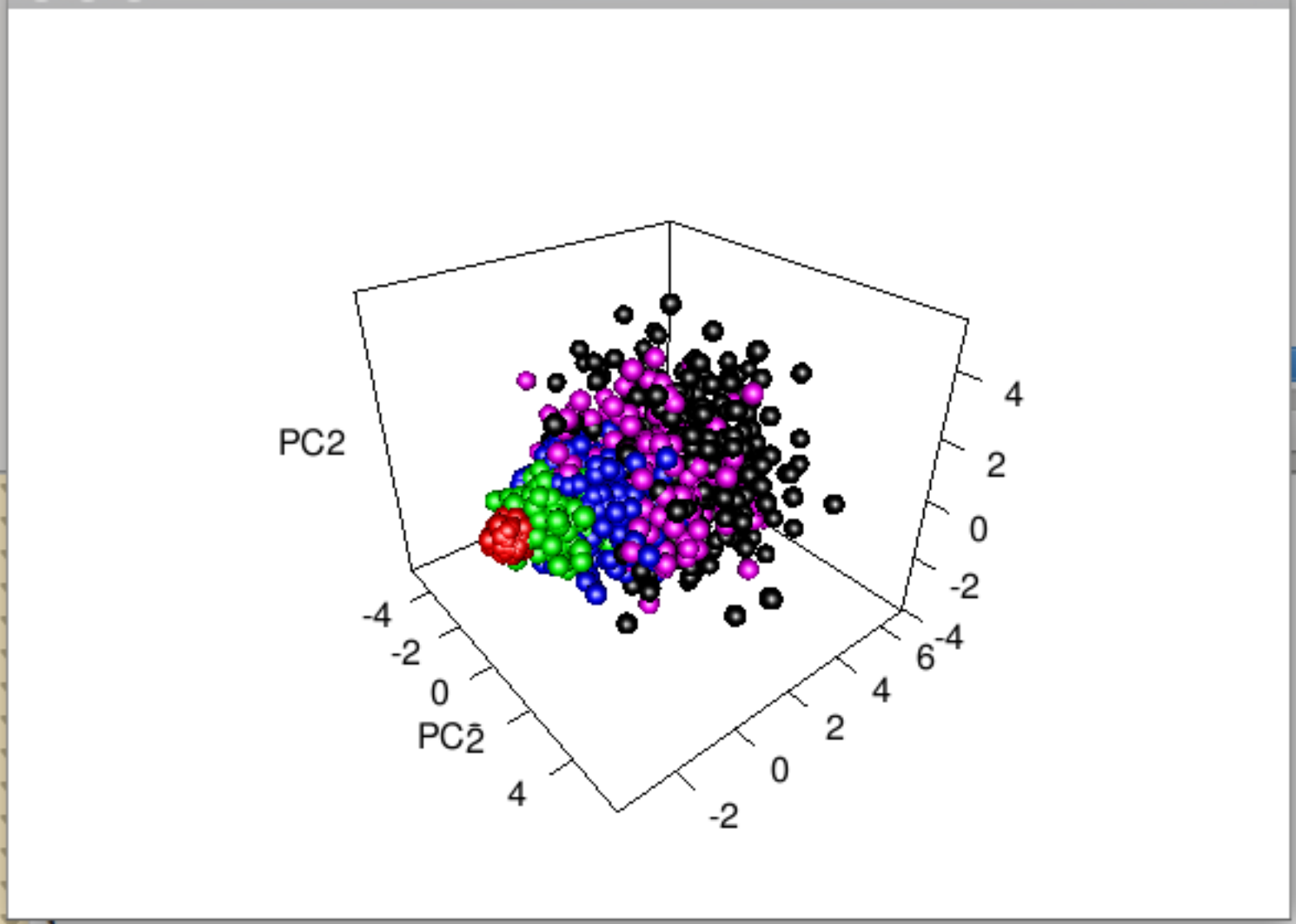
```
       C1
Min.   :-16.
1st Qu.:  0.
Median :  3.
Mean   :  4.
3rd Qu.:  8.
Max.   : 36.
       C5
Min.   :-20.
1st Qu.:  0.
Median :  3.
Mean   :  5.
3rd Qu.:  8.
Max.   : 38.
       C9
Min.   :-19.
1st Qu.:  0.
Median :  3.
Mean   :  4.
3rd Qu.:  7.
Max.   : 39.
      C13
Min.   :-19.
1st Qu.:  0.
Median :  3.
Mean   :  5.
3rd Qu.:  8.
Max.   : 31.
      C17
Min.   :-17.
1st Qu.:  0.
```

```
>
> z1 <- rnorm(10000, mean=1, sd=1)
> z2 <- rnorm(10000, mean=3, sd=3)
> z3 <- rnorm(10000, mean=5, sd=5)
> z4 <- rnorm(10000, mean=7, sd=7)
> z5 <- rnorm(10000, mean=9, sd=9)
> mydata <- matrix(c(z1, z2, z3, z4, z5), 2500, 20, byrow=T, dimnames=list(paste("R", 1:2500, sep=""), paste("C", 1:20, sep="")))
> plot(mydata,col=mycolors[sort(rep(1:5, 500))])
>
```






```
Proportion of Variance 0.03683 0.03633 0.0354 0.03414 0.03327 0.03186
Cumulative Proportion  0.82900 0.86533 0.9007 0.93487 0.96814 1.00000
> plot(pca$x,col=mycolors[sort(rep(1:5, 500))])
> x11()
> library(scatterplot3d)
> scatterplot3d(pca$x[,1:3],color=mycolors[sort(rep(1:5, 500))])
>
```




```
> library(scatterplot3d)
> rgl.open(); offset <- 50; par3d(windowRect=c(offset, offset, 640+offset, 640+offset))
> rgl.open();
> spheres3d(pca$x[,1], pca$x[,2], pca$x[,3], radius=0.3, color=mycolors[sort(rep(1:5, 500))], alpha=1, shininess=20)
> axes3d(col='black')
> title3d("", "", "PC1", "PC2", "PC3", col='black')
> bg3d("white")
> rgl.viewpoint(theta=45, phi=30, fov=60, zoom=1)
```

R: KEGG

KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps and update history) representing our knowledge on the molecular interaction and reaction networks for:

0. Global Map
1. Metabolism
Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

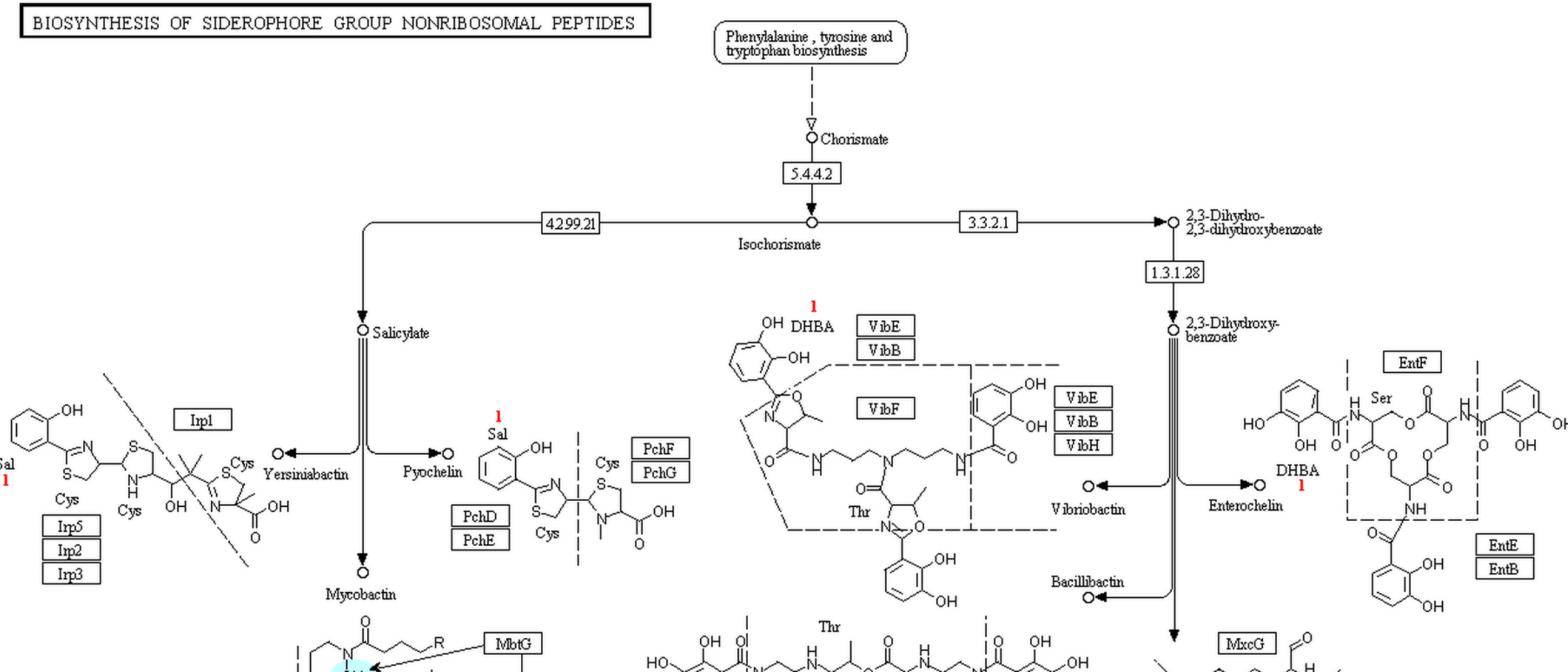
7. Drug Development

Pathway Mapping

KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool
- Color Pathway - selected pathway map coloring tool

0. Global Map


BIOSYNTHESIS OF SIDEROPHORE GROUP NONRIBOSOMAL PEPTIDES
Phenylalanine , tyrosine and tryptophan biosynthesis
Chorismate
5.4.4.2
4.2.99.21
Isochorismate
3.3.2.1
2,3-Dihydro-2,3-dihydroxybenzoate
1.3.1.28
2,3-Dihydroxy-benzoate
Salicylate
DHBA
VibE
VibB
VibF
VibH
Irp1
Sal
Cys
Irp5
Irp2
Irp3
Yersiniabactin
Pyochelin
PchF
PchG
PchD
PchE
Mycobactin
Thr
Vibriobactin
Enterochelin
EntF
Ser
EntE
EntB
Bacillibactin
MbtG
MxcG
R

```
> library(graphite)
> kegg[[1]]
"ABC transporters" pathway from KEGG
Number of nodes     = 44
Number of edges     = 0
Type of identifiers = native
Retrieved on        = 2011-05-05
> kegg[[2]]
"Acute myeloid leukemia" pathway from KEGG
Number of nodes     = 58
Number of edges     = 167
Type of identifiers = native
Retrieved on        = 2011-05-05
> length(kegg)
[1] 232
> biocarta[[1]]
"acetylation and deacetylation of rela in nucleus" pathway from BioCarta
Number of nodes     = 6
Number of edges     = 9
Type of identifiers = native
Retrieved on        = 2011-05-12
> length(biocarta)
[1] 254
> reactome[[1]]
"2-LTR circle formation" pathway from Reactome
Number of nodes     = 142
Number of edges     = 8485
Type of identifiers = native
Retrieved on        = 2011-05-12
> length(reactome)
[1] 1070
> nci[[1]]
"ALK1 pathway" pathway from NCI
Number of nodes     = 312
Number of edges     = 890
Type of identifiers = native
Retrieved on        = 2011-05-12
> length(nci)
[1] 177
>
```

```
>
> grep("Neuro",names(kegg))
[1] 129 130
> p<-kegg[[129]]
> p
"Neuroactive ligand-receptor interaction" pathway from KEGG
Number of nodes      = 272
Number of edges      = 46
Type of identifiers = native
Retrieved on         = 2011-05-05
> nodes(p)[1:10]
 [1] "EntrezGene:7442"   "EntrezGene:1141"   "EntrezGene:3363"
 [4] "EntrezGene:116443" "EntrezGene:8973"   "EntrezGene:2557"
 [7] "EntrezGene:6751"   "EntrezGene:3350"   "EntrezGene:5734"
[10] "EntrezGene:1136"
> edges(p)[1:10]
Error in `[.data.frame`(edges(p), 1:10) : undefined columns selected
Calls: [ -> [.data.frame
> edges(p)[1:10,]
                   src              dest direction       type
20649 EntrezGene:1081  EntrezGene:2492  directed activation
20650 EntrezGene:1081  EntrezGene:3973  directed activation
20651 EntrezGene:1081  EntrezGene:7253  directed activation
20652 EntrezGene:1442  EntrezGene:2690  directed activation
20653 EntrezGene:1511  EntrezGene:2149  directed activation
20654 EntrezGene:1511  EntrezGene:2150  directed activation
20655 EntrezGene:1511  EntrezGene:2151  directed activation
20656 EntrezGene:1511 EntrezGene:56288  directed activation
20657 EntrezGene:1511  EntrezGene:9002  directed activation
20658 EntrezGene:2147  EntrezGene:2149  directed activation
> psym<-convertIdentifiers(p,"symbol")
> nodes(psym)[1:10]
 [1] "TRPV1"  "CHRNB2" "HTR7"   "GRIN3A" "CHRNA6" "GABRA4" "SSTR1"  "HTR1A"
 [9] "PTGER4" "CHRNA3"
> edges(psym)[1:10,]
    src  dest direction       type
1   CGA  FSHR  directed activation
2   CGA LHCGR  directed activation
3   CGA  TSHR  directed activation
4  CSH1   GHR  directed activation
5  CTSG   F2R  directed activation
6  CTSG F2RL1  directed activation
7  CTSG F2RL2  directed activation
8  CTSG PARD3  directed activation
9  CTSG F2RL3  directed activation
10   F2   F2R  directed activation
>
```

UCSC: Galaxy

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, yo may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page.

clade: Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks **track:** UCSC Genes add custom tracks track hubs

table: knownGene describe table schema

region: genome ENCODE Pilot regions position chr7:97756624-97910747 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

output format: BED - browser extensible data Send output to Galaxy GREAT

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

get output summary/statistics

To reset **all** user cart settings (including custom tracks), click here.

Using the Table Browser

This section provides brief line-by-line descriptions of the Table Browser controls. For more information on using this program, see the Table Browser User's Guide.

- **clade:** Specifies which clade the organism is in.
- **genome:** Specifies which organism data to use.
- **assembly:** Specifies which version of the organism's genome sequence to use.
- **group:** Selects the type of tracks to be displayed in the *track* list. The options correspond to the track groupings shown in the Genome Browser. Select 'All Tracks' for an alphabetical list of all available tracks in all groups. Select 'All Tables' to see all tables including those not associated with a track.
- **database:** (with "All Tables" group option) Determines which database should be used for options in table menu.
- **track:** Selects the annotation track data to work with. This list displays all tracks belonging to the group specified in the *group* list.
- **table:** Selects the SQL table data to use. This list shows all tables associated with the track specified in the *track* list.
- **describe table schema:** Displays schema information for the tables associated with the selected track.
- **region:** Restricts the query to a particular chromosome or region. Select *genome* to apply the query to the entire genome or *ENCODE* to examine only the ENCODE Pilot regions. To limit the query to a specific position, type a chromosome name, e.g. *chrX*, or a chromosome coordinate range, such as chrX:100000-200000, or a gene name or other id in the text box.
- **lookup:** Press this button after typing in a gene name or other id in the position text box to look up the chromosome position

- EpiGRAPH server

Send Data

- Perform genome analysis and prediction with EpiGRAPH
- Encode Submission
- Encode Data Submission

ENCODE Tools

- Gencode Partition an interval file
- Random Intervals create a random set of intervals

Lift-Over

- Convert genome coordinates between assemblies and genomes

Text Manipulation

- Add column to an existing dataset
- Compute an expression on every row
- Concatenate datasets tail-to-head
- Condense consecutive characters
- Convert delimiters to TAB
- Merge Columns together
- Create single interval as a new dataset
- Cut columns from a table
- Change Case of selected columns
- Paste two files side by side
- Remove beginning of a file
- Select random lines from a file
- Select first lines from a dataset
- Select last lines from a dataset
- Trim leading or trailing characters
- Line/Word/Character count of a dataset

Gencode Partition (version 1.0.0)

File to Partition:

107: Filter on data 106

Execute

For detailed information about partitioning, click here.

Datasets are partitioned according to the protocol below:

A partition scheme has been defined that is similar to what has previously been done with TARs/TRANSFRAGs such that any feature can be classified as falling into one of the following 6 categories:

1. **Coding** -- coding exons defined from the GENCODE experimentally verified coding set (coding in any transcript)
2. **5UTR** -- 5' UTR exons defined from the GENCODE experimentally verified coding set (5' UTR in some transcript but never coding in any other)
3. **3UTR** -- 3' UTR exons defined from the GENCODE experimentally verified coding set (3' UTR in some transcript but never coding in any other)
4. **Intronic Proximal** -- intronic and no more than 5kb away from an exon.
5. **Intergenic Proximal** -- between genes and no more than 5kb away from an exon.
6. **Intronic Distal** -- intronic and greater than 5kb away from an exon.
7. **Intergenic Distal** -- between genes and greater than 5kb away from an exon.

Note: Features overlapping more than one partition will take the identity of the lower-numbered partition.

Citation

If you use this tool, please cite Blankenberg D, Taylor J, Schenck I, He J, Zhang Y, Ghent M, Veeraraghavan N, Albert I, Miller W, Makova KD, Hardison RC, Nekrutenko A. A framework for collaborative analysis of ENCODE data: making la
scale analyses biologist-friendly. Genome Res. 2007 Jun;17(6):960-4.

on a dataset

Convert Formats

- BED-to-GFF converter
- FASTA-to-Tabular converter
- GFF-to-BED converter
- Maf to BED Converts a MAF formatted file to the BED format
- MAF to Interval Converts a MAF formatted file to the Interval format
- MAF to FASTA Converts a MAF formatted file to FASTA format
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTQ to FASTA converter
- SFF converter
- Wig-to-bigWig converter
- BED-to-bigBed converter

FASTA manipulation

- Compute sequence length
- Filter sequences by length
- Concatenate FASTA alignment by species
- FASTA-to-Tabular converter
- Tabular-to-FASTA converts tabular file to FASTA format
- FASTA Width formatter
- RNA/DNA converter
- Collapse sequences

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an expression
- Filter on ambiguities in

BED-to-GFF (version 2.0.0)

Convert this query:

107: Filter on data 106

Execute

What it does

This tool converts data from BED format to GFF format (scroll down for format description).

Example

The following data in BED format:

```
chr28   346187  388197  BC114771        0       +       346187  388197  0       9       144,81,115,63,155,96,134,105,112,       0,24095,26190,31006,32131,33534,36994,41793,41898,
```

Will be converted to GFF (**note** that the start coordinate is incremented by 1):

```
##gff-version 2
##bed_to_gff_converter.py

chr28   bed2gff mRNA    346188  388197  0       +       .       mRNA BC114771;
chr28   bed2gff exon    346188  346331  0       +       .       exon BC114771;
chr28   bed2gff exon    370283  370363  0       +       .       exon BC114771;
chr28   bed2gff exon    372378  372492  0       +       .       exon BC114771;
chr28   bed2gff exon    377194  377256  0       +       .       exon BC114771;
chr28   bed2gff exon    378319  378473  0       +       .       exon BC114771;
chr28   bed2gff exon    379722  379817  0       +       .       exon BC114771;
chr28   bed2gff exon    383182  383315  0       +       .       exon BC114771;
chr28   bed2gff exon    387981  388085  0       +       .       exon BC114771;
chr28   bed2gff exon    388086  388197  0       +       .       exon BC114771;
```

About formats

BED format Browser Extensible Data format was designed at UCSC for displaying data tracks in the Genome Browser. It has three required fields and several additional optional ones:

The first three BED fields (required) are:

1. chrom - The name of the chromosome (e.g. chr1, chrY_random).
2. chromStart - The starting position in the chromosome. (The first base in a chromosome is numbered 0.)
3. chromEnd - The ending position in the chromosome, plus 1 (i.e., a half-open interval).

The additional BED fields (optional) are:

4. name - The name of the BED line.
5. score - A score between 0 and 1000.
6. strand - Defines the strand - either '+' or '-'.
7. thickStart - The starting position where the feature is drawn thickly at the Genome Browser.
8. thickEnd - The ending position where the feature is drawn thickly at the Genome Browser.
9. reserved - This should always be set to zero.
10. blockCount - The number of blocks (exons) in the BED line.

This dataset is large and only the first megabyte is shown below.
Show all | Save

@SRR002319.1 FC3003UAAXX_R1:1:1866:151 length=41
CAGGAAACTGAATAAATAAAATCCATAGAACACACAAACAA
+SRR002319.1 FC3003UAAXX_R1:1:1866:151 length=41
H>II:=HFIICFIBBCH9742I530696+—04,10–(–.,
@SRR002319.2 FC3003UAAXX_R1:1:1857:215 length=41
TGTATAATGTGTAATTCAGCTTAAACTCAAAAGATATCATG
+SRR002319.2 FC3003UAAXX_R1:1:1857:215 length=41
II>>IA>IIAI7CAAA.0I9I#7B.145,,,*5.,0./(0
@SRR002319.3 FC3003UAAXX_R1:1:1832:215 length=41
TTTTCTCTCTTACTAATTGCTTCCCAAATGCCATTACTTCT
+SRR002319.3 FC3003UAAXX_R1:1:1832:215 length=41
III=II–III<7DH28ID<5=F8151&,7+,,*9/1+++%=
@SRR002319.4 FC3003UAAXX_R1:1:1:1737:50 length=41
AAATCACTAACTCATCACAATTGCTCATTAAACCAGTCCAA
+SRR002319.4 FC3003UAAXX_R1:1:1:1737:50 length=41
CIIIIIIII@7IIII8;96.?##0;/6,6,1..0–$&–8)
@SRR002319.5 FC3003UAAXX_R1:1:1:1777:107 length=41
CATTCCTTGTGGTTCGATTCCTTTTCACTCTAGTCCATTCC
+SRR002319.5 FC3003UAAXX_R1:1:1:1777:107 length=41
DIIII>III?II>=5I,?!33"5?;),+:'5*//$*)07{&
@SRR002319.6 FC3003UAAXX_R1:1:1:793:170 length=41
TTCCTAAGTGATTCCAATATGCAAGCAGATTTTAGAACCAC
+SRR002319.6 FC3003UAAXX_R1:1:1:793:170 length=41
GI;IFI:IBICFI,>0<C800)002'/1+40.5(}*+){}&
@SRR002319.7 FC3003UAAXX_R1:1:1:923:232 length=41
ATGCAGAAAGAGGGTTTAAGTATATTAAAATGTCCTAAGTT
+SRR002319.7 FC3003UAAXX_R1:1:1:923:232 length=41
IIII=I8:/I4IIICII)4<.–I5>@520/30+/&5+)2*+
@SRR002319.8 FC3003UAAXX_R1:1:1:792:244 length=41
TAGCTTGTTTTCGTTTCTCCCTCAGCCATTCCTGTGGCATT
+SRR002319.8 FC3003UAAXX_R1:1:1:792:244 length=41
I8I<II:IIEI8@2II2>..)I.+–&**72&'/{+*–&".,
@SRR002319.9 FC3003UAAXX_R1:1:1:932:144 length=41
AAAAATCTCAAAGGAAACTAATAACAAACTGTGGAGCTCCT
+SRR002319.9 FC3003UAAXX_R1:1:1:932:144 length=41
9H?5BAB@D?8AII65I;I=:)3559/..&6'>.///{–'#+
@SRR002319.10 FC3003UAAXX_R1:1:1819:240 length=41
TTCAATCAATCAATCAATAAATAGTGCTATGTTAACAGACT
+SRR002319.10 FC3003UAAXX_R1:1:1819:240 length=41
II24IAC*0E1@766*212,*9+'%*.–)4**+.0')).+)
@SRR002319.11 FC3003UAAXX_R1:1:1:955:134 length=41
TGTATAGTTTGGTTACTTTGATATTTTATGTACACTATAGA
+SRR002319.11 FC3003UAAXX_R1:1:1:955:134 length=41
II3?I;I>III0>I62IIGI;I4?BI76:+41)/&.)4–&%
@SRR002319.12 FC3003UAAXX_R1:1:1:951:214 length=41
CTAATTTTTGTATTTTTAGTAGAGACAAGGTTTAACCATCT
+SRR002319.12 FC3003UAAXX_R1:1:1:951:214 length=41
I.22H1IIFI55II@IF5.?:+0+,,4–1*'–4&–)&{3'.
@SRR002319.13 FC3003UAAXX_R1:1:1:734:235 length=41

12: FASTQ Groomer on data 5

11: FASTQ Groomer on data 7
1.8 Gb
format: fastqsanger, database: hg19
Info: Groomed 9722922 illumina reads into sanger reads.
Based upon quality and sequence, the input data is valid for: sanger
Input ASCII range: '!'(33) – 'I'(73)
Input decimal range: –31 – 9

FC3003UAAXX_R1:2:1:1214:755 length=4
TCATGTGTCTGTTGTCTGCATAAATGTC
FC3003UAAXX_R1:2:1:1214:755 length=4
!*"**!*&!!*!#!!!!!!!!!!!!!!!
FC3003UAAXX_R1:2:1:1185:1814 length=
ATCACAAGGACAGTACTAAGAGGATGCA

9: SRR003962_1.fastq
1.7 Gb
format: fastq, database: hg19
Info: uploaded fastq file

@SRR003962.1 FC3003UAAXX_R1:4:1:356
GTATATATTTAGTATTGTAATCATTTCCTATTGGA
+SRR003962.1 FC3003UAAXX_R1:4:1:356
IIIIIIIIIIIIIIIIIIIIIDBIII:CI8II.&<.
@SRR003962.2 FC3003UAAXX_R1:4:1:278
GTTAGACAAAGAGGAGGGCAAGGGCTGAAAAACCA

8: SRR003961_2.fastq
7: SRR003960_1.fastq
5: SRR003960_2.fastq
4: SRR003961_1.fastq

FASTQ Groomer (version 1.0.4)

File to groom:
3: SRR002319_2.fastq

Input FASTQ quality scores type:
Illumina 1.3-1.7

Advanced Options:
Hide Advanced Options

Execute

What it does

This tool offers several conversions options relating to the FASTQ format.

When using *Basic* options, the output will be *sanger* formatted or *cssanger* formatted (when the input is Color Space Sanger).

When converting, if a quality score falls outside of the target score range, it will be coerced to the closest available value (i.e. the minimum or maximum).

When converting between Solexa and the other formats, quality scores are mapped between Solexa and PHRED scales using the equations found in Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2009 Dec 16.

When converting between color space (csSanger) and base/sequence space (Sanger, Illumina, Solexa) formats, adapter bases are lost or gained; if gained, the base 'G' is used as the adapter. You cannot convert a color space read to base space if there is no adapter present in the color space sequence. Any masked or ambiguous nucleotides in base space will be converted to 'N's when determining color space encoding.

Quality Score Comparison

```
 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
 ..........................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
 ..........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
 |                         |    |        |                              |                     |
33                        59   64       73                            104                   126

S - Sanger        Phred+33,  93 values  (0, 93) (0 to 60 expected in raw reads)
I - Illumina 1.3 Phred+64,  62 values  (0, 62) (0 to 40 expected in raw reads)
X - Solexa        Solexa+64, 67 values  (-5, 62) (-5 to 40 expected in raw reads)
```

Diagram adapted from http://en.wikipedia.org/wiki/FASTQ_format

Output from Illumina 1.8+ pipelines are Sanger encoded.

Citation

If you use this tool, please cite Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A; Galaxy Team. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010 Jul 15;26(14):1783-5.

15: FASTQ Groomer on data 1

14: FASTQ Groomer on data 8

13: FASTQ Groomer on data 4
1.7 Gb
format: fastqsanger, database: hg19
Info: Groomed 8995012 illumina reads into sanger reads.
Based upon quality and sequence, the input data is valid for: sanger
Input ASCII range: '!'(33) - 'I'(73)
Input decimal range: -31 - 9

```
FC3003UAAXX_R1:3:1:1811:296 length=
TCACTCATGATTTGGCTCTTGTTTGTCTT
FC3003UAAXX_R1:3:1:1811:296 length=
*(*!*!!**!!!!*!!!!"!!!!#!!!!!
FC3003UAAXX_R1:3:1:1768:260 length=
GGTTTTTGTGGTTTGGTGTTTTATATTTA
```

12: FASTQ Groomer on data 5

11: FASTQ Groomer on data 7
1.8 Gb
format: fastqsanger, database: hg19
Info: Groomed 9722922 illumina reads into sanger reads.
Based upon quality and sequence, the input data is valid for: sanger
Input ASCII range: '!'(33) - 'I'(73)
Input decimal range: -31 - 9

```
FC3003UAAXX_R1:2:1:1214:755 length=4
TCATGTGTCTGTTGTCTGCATAAATGTC
FC3003UAAXX_R1:2:1:1214:755 length=4
!*"**!*&!!*!#!!!!!!!!!!!!!!!
```

This dataset is large and only the first megabyte is shown below.
Show all | Save

```
@SRR003961.1 FC3003UAAXX_R1:3:1:1811:296 length=41
GTGAATGGGAGTTCACTCATGATTTGGCTCTTGTTTGTCTT
+SRR003961.1 FC3003UAAXX_R1:3:1:1811:296 length=41
********!****(*!*!!**!!!!*!!!!"!!!!#!!!!!
@SRR003961.2 FC3003UAAXX_R1:3:1:1768:260 length=41
GTTTTCTTCTAGGGTTTTTGTGGTTTGGTGTTTTATATTTA
+SRR003961.2 FC3003UAAXX_R1:3:1:1768:260 length=41
*!**********!*****$%!!!**!!!!*****!*$***!
@SRR003961.3 FC3003UAAXX_R1:3:1:230:380 length=41
GAAGATATTTCCTTTTTCAGCATAGGCCCCAAGGAGCTCAA
+SRR003961.3 FC3003UAAXX_R1:3:1:230:380 length=41
**********************************************
@SRR003961.4 FC3003UAAXX_R1:3:1:1818:333 length=41
GAATGGAATGGAATGCAGTGGAATGGAATCAAACCTAGTGC
+SRR003961.4 FC3003UAAXX_R1:3:1:1818:333 length=41
*&"***!!***!!*!!*!!(!!*!*!!(!!!!!!!!!!!
@SRR003961.5 FC3003UAAXX_R1:3:1:1709:379 length=41
GAATAATTGCTAATGGGTATCACAGTTCCTTTTAGGGTGAT
+SRR003961.5 FC3003UAAXX_R1:3:1:1709:379 length=41
**"*!*******#****!!$!!!!!!!!"!*!!!!!!*
@SRR003961.6 FC3003UAAXX_R1:3:1:1386:237 length=41
GTTGTAAGCAGGTGTGTTCATGTCACTAAGTTTGGCCATTA
+SRR003961.6 FC3003UAAXX_R1:3:1:1386:237 length=41
****!*!*!*!!!!**!)!!&!!!!!!!!!!%!!!!!!!!
@SRR003961.7 FC3003UAAXX_R1:3:1:1377:242 length=41
GTCTCACATGCCCTCTCACCATAGGTTTTTTCCAACAACCA
+SRR003961.7 FC3003UAAXX_R1:3:1:1377:242 length=41
*!"******!&!****!*!!*!*!**!!!!"(!!!!!!!!!
@SRR003961.8 FC3003UAAXX_R1:3:1:1858:302 length=41
GAAGGTCAGACCATTGTGTAACACATTTACTGTTTTATTCA
+SRR003961.8 FC3003UAAXX_R1:3:1:1858:302 length=41
*!****!*(*********!**!*!$!!*)*!!!!!!!!!!!
@SRR003961.9 FC3003UAAXX_R1:3:1:1756:325 length=41
GGAAATGAAGGCGACACTTAGAGAAATACAAACTGCACTAG
+SRR003961.9 FC3003UAAXX_R1:3:1:1756:325 length=41
**&*$**!**%$#!(!!*'!!!!!!'!!#!!!!!!!!!!
@SRR003961.10 FC3003UAAXX_R1:3:1:1694:356 length=41
GATTTTTGGTAGTGACAGCGTTTCACTGTATTGGCCAGGAT
+SRR003961.10 FC3003UAAXX_R1:3:1:1694:356 length=41
********!*%**!**#!!!!*'!!!!!!!&*!!!!!!!!!
@SRR003961.11 FC3003UAAXX_R1:3:1:1899:350 length=41
GGTTTGCTAGACTTTAAATCCTAGAACACGGCATAAGAGGA
+SRR003961.11 FC3003UAAXX_R1:3:1:1899:350 length=41
**!**************!*!**!!!!!!!!!!!!!!!!!
```

Tools
DATA
- Quality format converter (ASCII-Numeric)
- Compute quality statistics
- Draw quality score boxplot
- Draw nucleotides distribution chart
- FASTQ to FASTA converter
- Filter by quality
- Remove sequencing artifacts
- Barcode Splitter
- Clip adapter sequences
- Collapse sequences
- Rename sequences
- Reverse-Complement
- Trim sequences

FASTQ QC
- Fastqc: Fastqc QC using FastQC from Babraham

NGS: Mapping
ILLUMINA
- Map with Bowtie for Illumina
- Map with BWA for Illumina

ROCHE-454
- Lastz map short reads against reference sequence
- Megablast compare short reads against htgs, nt, and wgs databases
- Parse blast XML output

AB-SOLID
- Map with Bowtie for SOLiD
- Map with BWA for SOLiD

NGS: SAM Tools
NGS: Indel Analysis
NGS: Peak Calling

Map with Bowtie for Illumina (version 1.1.2)

Will you select a reference genome from your history or use a built-in index?:
Use a built-in index
Built-ins were indexed using default options

Select a reference genome:
Human (Homo sapiens): hg19 Full
if your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?:
Single-end

FASTQ file:
40: FASTQ Groomer on data 35
Must have ASCII encoded quality scores

Bowtie settings to use:
Commonly used
For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:
Bowtie produces SAM with several lines of header information by default

Execute

What it does

Bowtie is a short read aligner designed to be ultrafast and memory-efficient. It is developed by Ben Langmead and Cole Trapnell. Please cite: Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

Know what you are doing

There is no such thing (yet) as an automated gearshift in short read mapping. It is all like stick-shift driving in San Francisco. In other words = running this tool with default parameters will probably not give you meaningful results. A way to deal with this is to **understand** the parameters by carefully reading the documentation and experimenting. Fortunately, Galaxy makes experimenting easy.

Input formats

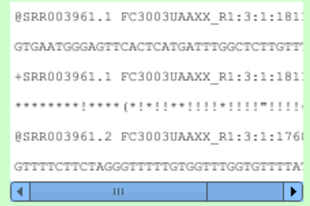
Bowtie accepts files in Sanger FASTQ format. Use the FASTQ Groomer to prepare your files.

A Note on Built-in Reference Genomes

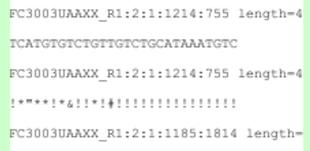
The default variant for all genomes is "Full", defined as all primary chromosomes (or scaffolds/contigs) including mitochondrial plus associated unmapped, plasmid, and other segments. When only one version of a genome is available in this tool, it represents the default "Full" variant. Some genomes will have more than one variant available. The "Canonical Male" or sometimes simply "Canonical" variant contains the primary chromosomes for a genome. For example a human "Canonical" variant contains chr1-chr22, chrX, chrY, and chrM. The "Canonical Female" variant contains the primary chromosomes excluding chrY.

Outputs

History
15: FASTQ Groomer on data 1
14: FASTQ Groomer on data 8
13: FASTQ Groomer on data 4
1.7 Gb
format: fastqsanger, database: hg19
Info: Groomed 8995012 illumina reads into sanger reads.
Based upon quality and sequence, the input data is valid for: sanger
Input ASCII range: '!'(33) - 'I'(73)
Input decimal range: -31 - 9



12: FASTQ Groomer on data 5
11: FASTQ Groomer on data 7
1.8 Gb
format: fastqsanger, database: hg19
Info: Groomed 9722922 illumina reads into sanger reads.
Based upon quality and sequence, the input data is valid for: sanger
Input ASCII range: '!'(33) - 'I'(73)
Input decimal range: -31 - 9



Tools

against htgs, nt, and wgs databases

- Parse blast XML output

AB-SOLID

- Map with Bowtie for SOLiD
- Map with BWA for SOLiD

NGS: SAM Tools

- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files
- rmdup remove PCR duplicates

SAM-to-BAM (version 1.1.2)

Choose the source for the reference list:

Locally cached

SAM File to Convert:

17: Map with Bowtie f..apped reads

Execute

What it does

This tool uses the SAMTools toolkit to produce an indexed BAM file based on a sorted input SAM file.

Citation

For the underlying tool, please cite Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, A
Bioinformatics. 2009 Aug 15;25(16):2078-9.

Tools

against htgs, nt, and wgs databases

- Parse blast XML output

AB-SOLID

- Map with Bowtie for SOLiD
- Map with BWA for SOLiD

NGS: SAM Tools

- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files
- rmdup remove PCR duplicates
- MPileup SNP and indel caller

NGS: Indel Analysis

Merge BAM Files (version 1.1.2)

Name for the output merged bam file:

This name will appear in your history so use it to remember what the new file in your history contains

Merge all component bam file headers into the merged bam file:

☑

Control the MERGE_SEQUENCE_DICTIONARIES flag for Picard MergeSamFiles. Default (true) correctly propagates read groups and other important metadata

First file:

101: Concatenate datas..nd data 100

with file:

101: Concatenate datas..nd data 100

Need to add more files? Use controls below.

Input Files

Input Files 1

Add file:

101: Concatenate datas..nd data 100

Remove Input Files 1

Add new Input Files

Execute

What it does

This tool uses the Picard merge command to merge any number of BAM files together into one BAM file while preserving the BAM metadata such as read groups

Tools

against htgs, nt, and wgs databases
- Parse blast XML output

AB-SOLID
- Map with Bowtie for SOLiD
- Map with BWA for SOLiD

NGS: SAM Tools
- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files
- rmdup remove PCR duplicates
- MPileup SNP and indel caller

NGS: Indel Analysis
NGS: Peak Calling
NGS: RNA Analysis
NGS: Picard (beta)

RGENETICS
SNP/WGA: Data; Filters
SNP/WGA: QC; LD; Plots
SNP/WGA: Statistical Models
NGS: GATK Tools (beta)
NGS: Variant Detection

Workflows
- map and filter to bam
- Workflow constructed from

Generate pileup (version 1.1.1)

Will you select a reference genome from your history or use a built-in index?:

Use a built-in index

Select the BAM file to generate the pileup file for:

57: Merge BAM Files o.. merged BAM

Whether or not to print the mapping quality as the last column:

Do not print the mapping quality as the last col

Makes the output easier to parse, but is space inefficient

Whether or not to print only output pileup lines containing indels:

Print all lines

Where to cap mapping quality:

60

Call consensus according to MAQ model?:

No

Execute

What it does

Uses SAMTools' pileup command to produce a pileup dataset from a provided BAM dataset. It generates two types of pileup datasets depending on the specified options. If *Call consensus according to MAQ model?* option is set to N tool produces simple pileup. If the option is set to **Yes**, a ten column pileup dataset with consensus is generated. Both types of datasets are briefly summarized below.

Types of pileup datasets

The description of pileup format below is largely based on information that can be found on SAMTools Pileup documentation page. The 6- and 10-column variants are described below.

Six column pileup:

```
   1    2  3  4        5        6
-------------------------------------
chrM  412  A  2        .,       II
chrM  413  G  4       ..t,      IIIH
chrM  414  C  4       ...a      III2
chrM  415  C  4       TTTt      III7
```

where:

```
 Column  Definition
-------- ------------------------------
      1 Chromosome
      2 Position (1-based)
      3 Reference base at that position
      4 Coverage (# reads aligning over that position)
      5 Bases within reads where (see Galaxy wiki for more info)
      6 Quality values (phred33 scale, see Galaxy wiki for more)
```

Ten column pileup

What it does

Uses SAMTools' pileup command to produce a pileup dataset from a provided BAM dataset. It generates two types of pileup datasets depending on the specified options. If *Call consensus accor* ool produces simple pileup. If the option is set to **Yes**, a ten column pileup dataset with consensus is generated. Both types of datasets are briefly summarized below.

Types of pileup datasets

The description of pileup format below is largely based on information that can be found on SAMTools Pileup documentation page. The 6- and 10-column variants are described below.

Six column pileup:

```
  1    2  3  4       5        6
-------------------------------
chrM  412  A  2       .,       II
chrM  413  G  4      ..t,     IIIH
chrM  414  C  4      ...a     III2
chrM  415  C  4      TTTt     III7
```

where:

Column	Definition
1	Chromosome
2	Position (1-based)
3	Reference base at that position
4	Coverage (# reads aligning over that position)
5	Bases within reads where (see Galaxy wiki for more info)
6	Quality values (phred33 scale, see Galaxy wiki for more)

Ten column pileup

The ten-column (consensus) pileup incorporates additional consensus information generated with *-c* option of *samtools pileup* command:

```
  1    2  3  4   5   6   7   8       9        10
-------------------------------------------------
chrM  412  A  A  75   0  25  2       .,       II
chrM  413  G  G  72   0  25  4      ..t,     IIIH
chrM  414  C  C  75   0  25  4      ...a     III2
chrM  415  C  T  75  75  25  4      TTTt     III7
```

where:

Column	Definition
1	Chromosome
2	Position (1-based)
3	Reference base at that position
4	Consensus bases
5	Consensus quality
6	SNP quality
7	Maximum mapping quality
8	Coverage (# reads aligning over that position)
9	Bases within reads where (see Galaxy wiki for more info)
10	Quality values (phred33 scale, see Galaxy wiki for more)

---

chr10	60070	T	N	0	0	0	5	.....	!!!!!
chr10	60071	T	N	0	0	0	6	.....^~.	!!!!!!
chr10	60072	T	N	0	0	0	6	.....	!!!!!!
chr10	60073	G	N	0	0	0	6	.....	!!!!!!
chr10	60074	G	N	0	0	0	6	TT....	!!!!!!
chr10	60075	T	N	0	0	0	6	.$.....	!!!!!!
chr10	60076	G	N	0	0	0	5	.....	!!!!!
chr10	60077	C	N	0	0	0	5	.....	!!!!!
chr10	60078	T	N	0	0	0	5	.....	!!!!!
chr10	60079	C	N	0	0	0	5	.....	!!!!!
chr10	60080	T	N	0	0	0	5	.....	!!!!!
chr10	60081	T	N	0	0	0	5	.....	!!!!!
chr10	60082	T	N	0	0	0	5	C$.C..	!!!!!
chr10	60083	A	N	0	0	0	4	.$...	!!!!
chr10	60084	T	N	0	0	0	3	...	!!!
chr10	60085	T	N	0	0	0	3	...	!!!
chr10	60086	T	N	0	0	0	3	...	!!!
chr10	60087	T	N	0	0	0	3	.$..	!!!
chr10	60088	G	N	0	0	0	3	..^~.	!!!
chr10	60089	C	N	0	0	0	3	...	!!!
chr10	60090	G	N	0	0	0	3	.A.	!!&
chr10	60091	T	N	0	0	0	3	...	!!*
chr10	60092	A	N	0	0	0	3	.T.	!!*
chr10	60093	T	N	0	0	0	3	...	!!*
chr10	60094	T	N	0	0	0	4	...^~.	!!**
chr10	60095	T	N	0	0	0	4	....	!!**
chr10	60096	A	N	0	0	0	5	....^~.	!!**!
chr10	60097	A	N	0	0	0	5	.$....	!!**!
chr10	60098	A	N	0	0	0	4	....	!**!
chr10	60099	A	N	0	0	0	4	....	!)*!
chr10	60100	C	N	0	0	0	4	..AG	!**!
chr10	60101	T	N	0	0	0	4	....	!**!
chr10	60102	A	N	0	0	0	4	....	!!*!
chr10	60103	T	N	0	0	0	4	....	!!*!
chr10	60104	T	N	0	0	0	4	....	!!*!
chr10	60105	A	N	0	0	0	4	....	!!*!

Pileup-to-Interval (version 1.0.0)

Choose a pileup file to condense::

58: Generate pileup o..rted pileup

which contains::

Pileup with six columns (simple)

See "Types of pileup datasets" below for examples

Do not report bases with coverage less than::

3

Execute

What is does

Reduces the size of a results set by taking a pileup file and producing a condensed version showing consecutive sequences of bases meeting coverage criteria. The tool works on six and ten column pileup formats produced w
pileup command. You also can specify columns for the input file manually. The tool assumes that the pileup dataset was produced by *samtools pileup* command (although you can override this by setting column assignments r

---

Types of pileup datasets

The description of pileup format below is largely based on information that can be found on SAMTools documentation page. The 6- and 10-column variants are described below.

Six column pileup:

```
   1    2  3  4        5         6
-------------------------------------
chrM  412  A  2         .,        II
chrM  413  G  4       ..t,      IIIH
chrM  414  C  4       ...a      III2
chrM  415  C  4       TTTt      III7
```

where:

```
Column  Definition
------  ----------
     1  Chromosome
     2  Position (1-based)
     3  Reference base at that position
     4  Coverage (# reads aligning over that position)
     5  Bases within reads where (see Galaxy wiki for more info)
     6  Quality values (phred33 scale, see Galaxy wiki for more)
```

Ten column pileup

The ten-column pileup incorporates additional consensus information generated with *-c* option of *samtools pileup* command:

```
   1    2  3  4   5   6   7   8        9         10
-------------------------------------------------
chrM  412  A  A  75   0  25  2         .,        II
chrM  413  G  G  72   0  25  4       ..t,      IIIH
chrM  414  C  C  75   0  25  4       ...a      III2
chrM  415  C  T  75  75  25  4       TTTt      III7
```

The output format

The output file condenses the information in the pileup file

Given the following input with minimum coverage set to 3:

```
   1    2  3  4        5         6
-------------------------------------
chr1  112  G  3     ..Ta      III6
chr1  113  T  2     aT..      III5
chr1  114  A  5     ,,,.      IIH2
chr1  115  C  4      ,,,      III
chrM  412  A  2        .,       II
chrM  413  G  4      ..t,      IIIH
chrM  414  C  4      ...a      III2
chrM  415  C  4     TTTt       III7
chrM  490  T  3        a         I
```

the following would be the output:

```
   1    2    3  4
-----------------
chr1  111  112  G
chr1  113  115  AC
chrM  412  415  GCC
chrM  489  490  T
```

where:

Column	Definition
1	Chromosome
2	Starting position (0-based)
3	Ending position (1-based)
4	Sequence of bases

NGS: SAM Tools

- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files
- rmdup remove PCR duplicates
- MPileup SNP and indel caller

NGS: Indel Analysis

- Filter Indels for SAM
- Extract indels from SAM
- Indel Analysis

NGS: Peak Calling

NGS: RNA Analysis

NGS: Picard (beta)

RGENETICS

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

SNP/WGA: Statistical Models

NGS: GATK Tools (beta)

NGS: Variant Detection

Workflows

- map and filter to bam
- Workflow constructed from history '8 human'

Indel Analysis (version 1.0.0)

Select sam file to analyze:

42: Map with Bowtie f..apped reads

Frequency threshold:

0.015

Cutoff

Execute

What it does

Given an input sam file, this tool provides analysis of the indels. It filters out matches that do not meet the frequency threshold. The way this frequency of occurence is calculated is different for deletions and insertions. The CIGAR string's "M" can indicate an exact match or a mismatch. For SAM containing the following bits of information (assuming the reference "ACTGCTCGAT"):

CHROM	POS	CIGAR	SEQ
ref	3	2M1I3M	TACTTC
ref	1	2M1D3M	ACGCT
ref	4	4M2I3M	GTTCAAGAT
ref	2	2M2D3M	CTCCG
ref	1	3M1D4M	AACCTGG
ref	6	3M1I2M	TTCAAT
ref	5	3M1I3M	CTCTGTT
ref	7	4M	CTAT
ref	5	5M	CGCTA
ref	3	2M1D2M	TGCC

The following totals would be calculated (this is an intermediate step and not output):

POS	BASE	NUMREADS	DELPROPCALC	DELPROP	INSPROPSTARTCALC	INSSTARTPROP	INSPROPENDCALC	INSENDPROP
1	A	2	2/2	1.00	---	---	---	---
2	A	1	1/3	0.33	---	---	---	---
	C	2	2/3	0.67	---	---	---	---
3	C	1	1/5	0.20	---	---	---	---
	T	3	3/5	0.60	---	---	---	---
	-	1	1/5	0.20	---	---	---	---
4	A	1	1/6	0.17	---	---	---	---
	G	3	3/6	0.50	---	---	---	---
	-	1	1/6	0.17	---	---	---	---
	--	1	1/6	0.17	---	---	---	---
5	C	4	4/7	0.57	---	---	---	---
	T	2	2/7	0.29	---	---	---	---
	-	1	1/7	0.14	---	---	---	---
	+C	1	---	---	1/7	0.14	1/9	0.11
6	C	2	2/9	0.22	---	---	---	---
	G	1	1/9	0.11	---	---	---	---
	T	6	6/9	0.67	---	---	---	---
7	C	7	7/9	0.78	---	---	---	---
	G	1	1/9	0.11	---	---	---	---


```
GTCACAGGGTCTTGATGCTGTGGTCTTCATCTGCA
GCAACTGCTGGCCTGTGCCAGGGTGCAAGCTGAGC
TCCTGTGGAGAGGAGCCATGCCTAGAGTGGGATGG
```

85: UCSC Main on Human: knownGene (genome)

77,614 regions

format: bed, database: hg19

display at UCSC main

view in GeneTrack

display at Ensembl Current

display at RViewer main

```
                                    12
109,1189,           0,739,1347,
127,1007,           0,721,1529,
52,1189,            0,772,1347,
69,147,159,         0,607,1433,2244,
519,                0,375,
159,198,136,456,    0,811,1062,1437,181
```

58: Generate pileup on data 57: converted pileup

~1,276,188,655 lines

format: tabular, database: hg19

Info: Samtools Version: 0.1.12 (r862)

Converted BAM to pileup

```
1     2     3 4 5 6 7 8 9   10
chr10 60006 C N 0 0 0 1 ^~. !
chr10 60007 C N 0 0 0 1 .   !
chr10 60008 T N 0 0 0 1 .   !
chr10 60009 T N 0 0 0 1 .   !
chr10 60010 G N 0 0 0 1 .   !
chr10 60011 A N 0 0 0 1 .   !
```

57: Merge BAM Files on data 46, data 56, and others:

Tools
- MPileup SNP and indel caller

NGS: Indel Analysis
- Filter Indels for SAM
- Extract indels from SAM
- Indel Analysis

NGS: Peak Calling
- MACS Model-based Analysis of ChIP-Seq
- SICER Statistical approach for the Identification of ChIP-Enriched Regions
- GeneTrack indexer on a BED file
- Peak predictor on GeneTrack index

NGS: RNA Analysis

RNA-SEQ
- Cufflinks transcript assembly and FPKM (RPKM) estimates for RNA-Seq data
- Cuffcompare compare assembled transcripts to a reference annotation and track Cufflinks transcripts across multiple experiments
- Cuffmerge merge together several Cufflinks assemblies
- Cuffdiff find significant changes in transcript expression, splicing, and promoter use

FILTERING
- Filter Combined Transcripts using tracking file
- Tophat for Illumina Find splice junctions using RNA-seq data

NGS: Picard (beta)

RGENETICS

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

Cufflinks (version 0.0.5)

SAM or BAM file of aligned RNA-Seq reads:

101: Concatenate datas..nd data 100

Max Intron Length:

300000

Min Isoform Fraction:

0.1

Pre MRNA Fraction:

0.15

Perform quartile normalization:

No

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Use Reference Annotation:

No

Perform Bias Correction:

No

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Set Parameters for Paired-end Reads? (not recommended):

No

Execute

Cufflinks Overview

Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples. It accepts aligned RNA-Seq reads and assembles the alignments into a parsimonious set of transcripts. Cufflinks then estimates the relative abundances of these transcripts based on how many reads support each one. Please cite: Trapnell C, Williams BA, Pertea G, Mortazavi AM, Kwan G, van Baren MJ, Salzberg SL, Wold B, Pachter L. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nature Biotechnology doi:10.1038/nbt.1621

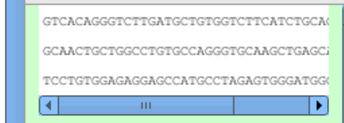
Know what you are doing

There is no such thing (yet) as an automated gearshift in expression analysis. It is all like stick-shift driving in San Francisco. In other words, running this tool with default parameters will probably not give you meaningful results. A way to deal with this is to **understand** the parameters by carefully reading the documentation and experimenting. Fortunately, Galaxy makes experimenting easy.

Input formats

Cufflinks takes a text file of SAM alignments as input. The RNA-Seq read mapper TopHat produces output in this format, and is recommended for use with Cufflinks. However Cufflinks will accept SAM alignments generated by any read mapper. Here's an example of an alignment Cufflinks will accept:

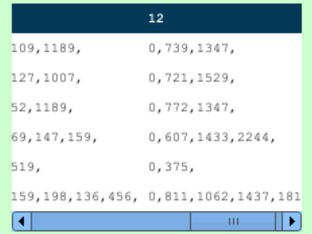
History



85: UCSC Main on Human: knownGene (genome)

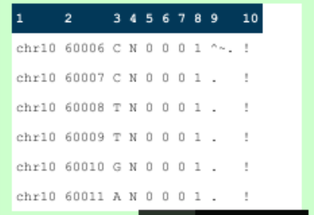
77,614 regions
format: bed, database: hg19

display at UCSC main
view in GeneTrack
display at Ensembl Current
display at RViewer main



58: Generate pileup on data 57: converted pileup

~1,276,188,655 lines
format: tabular, database: hg19
Info: Samtools Version: 0.1.12 (r862)
Converted BAM to pileup



Display data in browser

chr1	Cufflinks	transcript	11354	17137	1000	.	.	gene_id "CUFF.3"; transcript_id "CUFF.3.1"; FPKM "5674141.4810736002"; frac "1.000000"; conf_lo "4509832.215824"; conf_hi "6838450.746324"; cov "1.39583
chr1	Cufflinks	exon	11354	17137	1000	.	.	gene_id "CUFF.3"; transcript_id "CUFF.3.1"; exon_number "1"; FPKM "5674141.4810736002"; frac "1.000000"; conf_lo "4509832.215824"; conf_hi "6838450.7463
chr1	Cufflinks	transcript	23453	26584	1000	.	.	gene_id "CUFF.12"; transcript_id "CUFF.12.1"; FPKM "8990942.0815893132"; frac "1.000000"; conf_lo "6967822.050805"; conf_hi "11014062.112374"; cov "2.21
chr1	Cufflinks	exon	23453	26584	1000	.	.	gene_id "CUFF.12"; transcript_id "CUFF.12.1"; exon_number "1"; FPKM "8990942.0815893132"; frac "1.000000"; conf_lo "6967822.050805"; conf_hi "11014062.1
chr1	Cufflinks	transcript	18530	20705	1000	.	.	gene_id "CUFF.2"; transcript_id "CUFF.2.1"; FPKM "8785830.4311732948"; frac "1.000000"; conf_lo "6349079.499087"; conf_hi "11222581.363260"; cov "2.1613
chr1	Cufflinks	exon	18530	20705	1000	.	.	gene_id "CUFF.2"; transcript_id "CUFF.2.1"; exon_number "1"; FPKM "8785830.4311732948"; frac "1.000000"; conf_lo "6349079.499087"; conf_hi "11222581.363
chr1	Cufflinks	transcript	26692	28915	1000	.	.	gene_id "CUFF.15"; transcript_id "CUFF.15.1"; FPKM "10061654.9449292421"; frac "1.000000"; conf_lo "7485129.132709"; conf_hi "12638180.757150"; cov "2.4
chr1	Cufflinks	exon	26692	28915	1000	.	.	gene_id "CUFF.15"; transcript_id "CUFF.15.1"; exon_number "1"; FPKM "10061654.9449292421"; frac "1.000000"; conf_lo "7485129.132709"; conf_hi "12638180.
chr1	Cufflinks	transcript	29885	31055	1000	.	.	gene_id "CUFF.4"; transcript_id "CUFF.4.1"; FPKM "8954336.8441765849"; frac "1.000000"; conf_lo "5442156.974966"; conf_hi "12466516.713387"; cov "2.2027
chr1	Cufflinks	exon	29885	31055	1000	.	.	gene_id "CUFF.4"; transcript_id "CUFF.4.1"; exon_number "1"; FPKM "8954336.8441765849"; frac "1.000000"; conf_lo "5442156.974966"; conf_hi "12466516.713
chr1	Cufflinks	transcript	31292	35854	1000	.	.	gene_id "CUFF.8"; transcript_id "CUFF.8.1"; FPKM "10933038.5500250086"; frac "1.000000"; conf_lo "9104505.323513"; conf_hi "12761571.776537"; cov "2.689
chr1	Cufflinks	exon	31292	35854	1000	.	.	gene_id "CUFF.8"; transcript_id "CUFF.8.1"; exon_number "1"; FPKM "10933038.5500250086"; frac "1.000000"; conf_lo "9104505.323513"; conf_hi "12761571.77
chr1	Cufflinks	transcript	21062	23398	1000	.	.	gene_id "CUFF.1"; transcript_id "CUFF.1.1"; FPKM "5623575.7029602556"; frac "1.000000"; conf_lo "3749050.468640"; conf_hi "7498100.937280"; cov "1.38340
chr1	Cufflinks	exon	21062	23398	1000	.	.	gene_id "CUFF.1"; transcript_id "CUFF.1.1"; exon_number "1"; FPKM "5623575.7029602556"; frac "1.000000"; conf_lo "3749050.468640"; conf_hi "7498100.9372
chr1	Cufflinks	transcript	10001	10471	1000	.	.	gene_id "CUFF.25"; transcript_id "CUFF.25.1"; FPKM "1398669370.7421174049"; frac "1.000000"; conf_lo "1315231751.350605"; conf_hi "1482106990.133630"
chr1	Cufflinks	exon	10001	10471	1000	.	.	gene_id "CUFF.25"; transcript_id "CUFF.25.1"; exon_number "1"; FPKM "1398669370.7421174049"; frac "1.000000"; conf_lo "1315231751.350605"; conf_hi "1482
chr1	Cufflinks	transcript	35927	36799	1000	.	.	gene_id "CUFF.5"; transcript_id "CUFF.5.1"; FPKM "7961700.4630200835"; frac "1.000000"; conf_lo "3980850.231510"; conf_hi "11942550.694530"; cov "1.9585
chr1	Cufflinks	exon	35927	36799	1000	.	.	gene_id "CUFF.5"; transcript_id "CUFF.5.1"; exon_number "1"; FPKM "7961700.4630200835"; frac "1.000000"; conf_lo "3980850.231510"; conf_hi "11942550.694
chr1	Cufflinks	transcript	36925	38285	1000	.	.	gene_id "CUFF.6"; transcript_id "CUFF.6.1"; FPKM "6333448.6850877833"; frac "1.000000"; conf_lo "3632857.091763"; conf_hi "9034040.278412"; cov "1.55802
chr1	Cufflinks	exon	36925	38285	1000	.	.	gene_id "CUFF.6"; transcript_id "CUFF.6.1"; exon_number "1"; FPKM "6333448.6850877833"; frac "1.000000"; conf_lo "3632857.091763"; conf_hi "9034040.2784
chr1	Cufflinks	transcript	38425	42891	1000	.	.	gene_id "CUFF.9"; transcript_id "CUFF.9.1"; FPKM "6254094.0159574756"; frac "1.000000"; conf_lo "4855636.080222"; conf_hi "7652551.951693"; cov "1.53850
chr1	Cufflinks	exon	38425	42891	1000	.	.	gene_id "CUFF.9"; transcript_id "CUFF.9.1"; exon_number "1"; FPKM "6254094.0159574756"; frac "1.000000"; conf_lo "4855636.080222"; conf_hi "7652551.9516
chr1	Cufflinks	transcript	43008	44856	1000	.	.	gene_id "CUFF.10"; transcript_id "CUFF.10.1"; FPKM "7493487.3332633190"; frac "1.000000"; conf_lo "5029643.875315"; conf_hi "9957330.791211"; cov "1.843
chr1	Cufflinks	exon	43008	44856	1000	.	.	gene_id "CUFF.10"; transcript_id "CUFF.10.1"; exon_number "1"; FPKM "7493487.3332633190"; frac "1.000000"; conf_lo "5029643.875315"; conf_hi "9957330.79
chr1	Cufflinks	transcript	50336	51639	1000	.	.	gene_id "CUFF.16"; transcript_id "CUFF.16.1"; FPKM "6661376.0323455902"; frac "1.000000"; conf_lo "3820955.748325"; conf_hi "9501796.316366"; cov "1.638
chr1	Cufflinks	exon	50336	51639	1000	.	.	gene_id "CUFF.16"; transcript_id "CUFF.16.1"; exon_number "1"; FPKM "6661376.0323455902"; frac "1.000000"; conf_lo "3820955.748325"; conf_hi "9501796.31
chr1	Cufflinks	transcript	52657	53438	1000	.	.	gene_id "CUFF.7"; transcript_id "CUFF.7.1"; FPKM "7485801.6916323258"; frac "1.000000"; conf_lo "3333426.024381"; conf_hi "11638177.358884"; cov "1.8415
chr1	Cufflinks	exon	52657	53438	1000	.	.	gene_id "CUFF.7"; transcript_id "CUFF.7.1"; exon_number "1"; FPKM "7485801.6916323258"; frac "1.000000"; conf_lo "3333426.024381"; conf_hi "11638177.358
chr1	Cufflinks	transcript	44928	48364	1000	.	.	gene_id "CUFF.11"; transcript_id "CUFF.11.1"; FPKM "5669155.7336361185"; frac "1.000000"; conf_lo "4140298.127276"; conf_hi "7198013.339996"; cov "1.394
chr1	Cufflinks	exon	44928	48364	1000	.	.	gene_id "CUFF.11"; transcript_id "CUFF.11.1"; exon_number "1"; FPKM "5669155.7336361185"; frac "1.000000"; conf_lo "4140298.127276"; conf_hi "7198013.33
chr1	Cufflinks	transcript	48415	50221	1000	.	.	gene_id "CUFF.14"; transcript_id "CUFF.14.1"; FPKM "6442735.2780004917"; frac "1.000000"; conf_lo "4128436.447033"; conf_hi "8757034.108968"; cov "1.584
chr1	Cufflinks	exon	48415	50221	1000	.	.	gene_id "CUFF.14"; transcript_id "CUFF.14.1"; exon_number "1"; FPKM "6442735.2780004917"; frac "1.000000"; conf_lo "4128436.447033"; conf_hi "8757034.10
chr1	Cufflinks	transcript	53801	54735	1000	.	.	gene_id "CUFF.13"; transcript_id "CUFF.13.1"; FPKM "7742684.3858391885"; frac "1.000000"; conf_lo "3986930.791882"; conf_hi "11498437.979796"; cov "1.90
chr1	Cufflinks	exon	53801	54735	1000	.	.	gene_id "CUFF.13"; transcript_id "CUFF.13.1"; exon_number "1"; FPKM "7742684.3858391885"; frac "1.000000"; conf_lo "3986930.791882"; conf_hi "11498437.9
chr1	Cufflinks	transcript	54874	58344	1000	.	.	gene_id "CUFF.17"; transcript_id "CUFF.17.1"; FPKM "6426197.0510835936"; frac "1.000000"; conf_lo "4806947.596506"; conf_hi "8045446.505661"; cov "1.580
chr1	Cufflinks	exon	54874	58344	1000	.	.	gene_id "CUFF.17"; transcript_id "CUFF.17.1"; exon_number "1"; FPKM "6426197.0510835936"; frac "1.000000"; conf_lo "4806947.596506"; conf_hi "8045446.50
chr1	Cufflinks	transcript	62815	63894	1000	.	.	gene_id "CUFF.19"; transcript_id "CUFF.19.1"; FPKM "4561660.6328496579"; frac "1.000000"; conf_lo "1927984.638522"; conf_hi "7195336.627177"; cov "1.122
chr1	Cufflinks	exon	62815	63894	1000	.	.	gene_id "CUFF.19"; transcript_id "CUFF.19.1"; exon_number "1"; FPKM "4561660.6328496579"; frac "1.000000"; conf_lo "1927984.638522"; conf_hi "7195336.62
chr1	Cufflinks	transcript	63964	66062	1000	.	.	gene_id "CUFF.20"; transcript_id "CUFF.20.1"; FPKM "5802074.4218300879"; frac "1.000000"; conf_lo "3782051.391765"; conf_hi "7822097.451895"; cov "1.427
chr1	Cufflinks	exon	63964	66062	1000	.	.	gene_id "CUFF.20"; transcript_id "CUFF.20.1"; exon_number "1"; FPKM "5802074.4218300879"; frac "1.000000"; conf_lo "3782051.391765"; conf_hi "7822097.45
chr1	Cufflinks	transcript	66695	67725	1000	.	.	gene_id "CUFF.21"; transcript_id "CUFF.21.1"; FPKM "6844844.8388877828"; frac "1.000000"; conf_lo "3524607.396335"; conf_hi "10165082.281441"; cov "1.68

History

chr1 14362 16765 input_line_4
chr1 16857 17751 input_line_5
chr1 15795 18061 input_line_6

104: Cufflinks on data 101: assembled transcripts

~260,000 lines
format: gtf, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa

display at UCSC main
display at Ensembl Current

1.Seqname	2.Source	3.Feature	4.St
chr1	Cufflinks	transcript	1135
chr1	Cufflinks	exon	1135
chr1	Cufflinks	transcript	2345
chr1	Cufflinks	exon	2345
chr1	Cufflinks	transcript	1853
chr1	Cufflinks	exon	1853

103: Cufflinks on data 101: transcript expression

~130,000 lines
format: tabular, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa

1		
tracking_id	class_code	nea

tracking_id class_code nearest_ref_id gene_id gene_short_name tss_id locus length coverage FPKM FPKM_conf_lo FPKM_conf_hi FPKM_status
CUFF.3.1 - - CUFF.3 - - chr1:11353-17137 5784 1.39584 5.67414e+06 4.50983e+06 6.83845e+06 OK
CUFF.12.1 - - CUFF.12 - - chr1:23452-26584 3132 2.21177 8.99094e+06 6.96782e+06 1.10141e+07 OK
CUFF.2.1 - - CUFF.2 - - chr1:18529-20705 2176 2.16131 8.78583e+06 6.34908e+06 1.12226e+07 OK
CUFF.15.1 - - CUFF.15 - - chr1:26691-28915 2224 2.47517 1.00617e+07 7.48513e+06 1.26382e+07 OK
CUFF.4.1 - - CUFF.4 - - chr1:29884-31055 1171 2.20277 8.95434e+06 5.44216e+06 1.24665e+07 OK
CUFF.8.1 - - CUFF.8 - - chr1:31291-35854 4563 2.68953 1.0933e+07 9.10451e+06 1.27616e+07 OK
CUFF.1.1 - - CUFF.1 - - chr1:21061-23398 2337 1.3834 5.62358e+06 3.74905e+06 7.4981e+06 OK
CUFF.25.1 - - CUFF.25 - - chr1:10000-10471 471 344.073 1.39867e+09 1.31523e+09 1.48211e+09 OK
CUFF.5.1 - - CUFF.5 - - chr1:35926-36799 873 1.95858 7.9617e+06 3.98085e+06 1.19426e+07 OK
CUFF.6.1 - - CUFF.6 - - chr1:36924-38285 1361 1.55803 6.33345e+06 3.63286e+06 9.03404e+06 OK
CUFF.9.1 - - CUFF.9 - - chr1:38424-42891 4467 1.53851 6.25409e+06 4.85564e+06 7.65255e+06 OK
CUFF.10.1 - - CUFF.10 - - chr1:43007-44856 1849 1.8434 7.49349e+06 5.02964e+06 9.95733e+06 OK
CUFF.16.1 - - CUFF.16 - - chr1:50335-51639 1304 1.6387 6.66138e+06 3.82096e+06 9.5018e+06 OK
CUFF.7.1 - - CUFF.7 - - chr1:52656-53438 782 1.84151 7.4858e+06 3.33343e+06 1.16382e+07 OK
CUFF.11.1 - - CUFF.11 - - chr1:44927-48364 3437 1.39461 5.66916e+06 4.1403e+06 7.19801e+06 OK
CUFF.14.1 - - CUFF.14 - - chr1:48414-50221 1807 1.58491 6.44274e+06 4.12844e+06 8.75703e+06 OK
CUFF.13.1 - - CUFF.13 - - chr1:53800-54735 935 1.9047 7.74268e+06 3.98693e+06 1.14984e+07 OK
CUFF.17.1 - - CUFF.17 - - chr1:54873-58344 3471 1.58084 6.4262e+06 4.80695e+06 8.04545e+06 OK
CUFF.19.1 - - CUFF.19 - - chr1:62814-63894 1080 1.12217 4.56166e+06 1.92798e+06 7.19534e+06 OK
CUFF.20.1 - - CUFF.20 - - chr1:63963-66062 2099 1.42731 5.80207e+06 3.78205e+06 7.8221e+06 OK
CUFF.21.1 - - CUFF.21 - - chr1:66694-67725 1031 1.68383 6.84484e+06 3.52461e+06 1.01651e+07 OK
CUFF.22.1 - - CUFF.22 - - chr1:67777-69778 2001 1.36828 5.56213e+06 3.53112e+06 7.59313e+06 OK
CUFF.18.1 - - CUFF.18 - - chr1:58450-61939 3489 1.57218 6.39099e+06 4.78062e+06 8.00137e+06 OK
CUFF.29.1 - - CUFF.29 - - chr1:69835-71779 1944 1.31888 5.36129e+06 3.33491e+06 7.38767e+06 OK
CUFF.23.1 - - CUFF.23 - - chr1:79289-80921 1632 4.7058 1.91293e+07 1.49043e+07 2.33542e+07 OK
CUFF.27.1 - - CUFF.27 - - chr1:81081-83838 2757 3.24292 1.31826e+07 1.05592e+07 1.5806e+07 OK
CUFF.28.1 - - CUFF.28 - - chr1:83989-87313 3324 2.23335 9.07865e+06 7.10922e+06 1.10481e+07 OK
CUFF.24.1 - - CUFF.24 - - chr1:87746-88937 1191 1.41111 5.73623e+06 2.95375e+06 8.5187e+06 OK
CUFF.26.1 - - CUFF.26 - - chr1:89066-90884 1818 6.80425 2.76595e+07 2.28807e+07 3.24384e+07 OK
CUFF.30.1 - - CUFF.30 - - chr1:71904-78568 6664 11.3084 4.5969e+07 4.2889e+07 4.90491e+07 OK
CUFF.31.1 - - CUFF.31 - - chr1:91084-120988 29904 4.23516 1.72161e+07 1.6337e+07 1.80952e+07 OK
CUFF.41.1 - - CUFF.41 - - chr1:176053-177408 1355 4.2713 1.7363e+07 1.28799e+07 2.18461e+07 OK
CUFF.37.1 - - CUFF.37 - - chr1:121127-175952 54825 3.62397 1.47316e+07 1.41319e+07 1.53313e+07 OK
CUFF.43.1 - - CUFF.43 - - chr1:227417-267296 39879 4.59231 1.86679e+07 1.78759e+07 1.946e+07 OK
CUFF.38.1 - - CUFF.38 - - chr1:317719-342194 24475 3.98989 1.6219e+07 1.52751e+07 1.7163e+07 OK
CUFF.32.1 - - CUFF.32 - - chr1:342335-354967 12632 4.78981 1.94708e+07 1.80255e+07 2.0916e+07 OK
CUFF.33.1 - - CUFF.33 - - chr1:387195-387999 804 2.31996 9.43071e+06 4.85615e+06 1.40053e+07 OK
CUFF.34.1 - - CUFF.34 - - chr1:394312-396504 2192 1.77286 7.20676e+06 5.00872e+06 9.4048e+06 OK
CUFF.35.1 - - CUFF.35 - - chr1:388067-394261 6194 2.05312 8.34603e+06 6.98313e+06 9.70893e+06 OK
CUFF.50.1 - - CUFF.50 - - chr1:396573-409814 13241 2.956 1.20163e+07 1.09077e+07 1.31248e+07 OK
CUFF.36.1 - - CUFF.36 - - chr1:424784-427729 2945 1.8243 7.41585e+06 5.51685e+06 9.31486e+06 OK
CUFF.64.1 - - CUFF.64 - - chr1:355018-387052 32034 3.66581 1.49017e+07 1.41116e+07 1.56917e+07 OK

History

104: Cufflinks on data
101: assembled transcripts
~260,000 lines
format: gtf, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa
display at UCSC main
display at Ensembl Current

1.Seqname	2.Source	3.Feature	4.St
chr1	Cufflinks	transcript	1135
chr1	Cufflinks	exon	1135
chr1	Cufflinks	transcript	2345
chr1	Cufflinks	exon	2345
chr1	Cufflinks	transcript	1853
chr1	Cufflinks	exon	1853

103: Cufflinks on data
101: transcript expression
~130,000 lines
format: tabular, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa

cus	length	coverage	FPKM
84	1.39584	5.67414e+06	4.5098
32	2.21177	8.99094e+06	6.9678
76	2.16131	8.78583e+06	6.3490
24	2.47517	1.00617e+07	7.4851
71	2.20277	8.95434e+06	5.4421

tracking_id class_code nearest_ref_id gene_id gene_short_name tss_id locus length coverage FPKM FPKM_conf_lo FPKM_conf_hi FPKM_status
CUFF.3 - - CUFF.3 - - chr1:11353-17137 - - 5.67414e+06 4.50983e+06 6.83845e+06 OK
CUFF.12 - - CUFF.12 - - chr1:23452-26584 - - 8.99094e+06 6.96782e+06 1.10141e+07 OK
CUFF.2 - - CUFF.2 - - chr1:18529-20705 - - 8.78583e+06 6.34908e+06 1.12226e+07 OK
CUFF.15 - - CUFF.15 - - chr1:26691-28915 - - 1.00617e+07 7.48513e+06 1.26382e+07 OK
CUFF.4 - - CUFF.4 - - chr1:29884-31055 - - 8.95434e+06 5.44216e+06 1.24665e+07 OK
CUFF.8 - - CUFF.8 - - chr1:31291-35854 - - 1.0933e+07 9.10451e+06 1.27616e+07 OK
CUFF.1 - - CUFF.1 - - chr1:21061-23398 - - 5.62358e+06 3.74905e+06 7.4981e+06 OK
CUFF.25 - - CUFF.25 - - chr1:10000-10471 - - 1.39867e+09 1.31523e+09 1.48211e+09 OK
CUFF.5 - - CUFF.5 - - chr1:35926-36799 - - 7.9617e+06 3.98085e+06 1.19426e+07 OK
CUFF.6 - - CUFF.6 - - chr1:36924-38285 - - 6.33345e+06 3.63286e+06 9.03404e+06 OK
CUFF.9 - - CUFF.9 - - chr1:38424-42891 - - 6.25409e+06 4.85564e+06 7.65255e+06 OK
CUFF.10 - - CUFF.10 - - chr1:43007-44856 - - 7.49349e+06 5.02964e+06 9.95733e+06 OK
CUFF.16 - - CUFF.16 - - chr1:50335-51639 - - 6.66138e+06 3.82096e+06 9.5018e+06 OK
CUFF.7 - - CUFF.7 - - chr1:52656-53438 - - 7.4858e+06 3.33343e+06 1.16382e+07 OK
CUFF.11 - - CUFF.11 - - chr1:44927-48364 - - 5.66916e+06 4.1403e+06 7.19801e+06 OK
CUFF.14 - - CUFF.14 - - chr1:48414-50221 - - 6.44274e+06 4.12844e+06 8.75703e+06 OK
CUFF.13 - - CUFF.13 - - chr1:53800-54735 - - 7.74268e+06 3.98693e+06 1.14984e+07 OK
CUFF.17 - - CUFF.17 - - chr1:54873-58344 - - 6.4262e+06 4.80695e+06 8.04545e+06 OK
CUFF.19 - - CUFF.19 - - chr1:62814-63894 - - 4.56166e+06 1.92798e+06 7.19534e+06 OK
CUFF.20 - - CUFF.20 - - chr1:63963-66062 - - 5.80207e+06 3.78205e+06 7.8221e+06 OK
CUFF.21 - - CUFF.21 - - chr1:66694-67725 - - 6.84484e+06 3.52461e+06 1.01651e+07 OK
CUFF.22 - - CUFF.22 - - chr1:67777-69778 - - 5.56213e+06 3.53112e+06 7.59313e+06 OK
CUFF.18 - - CUFF.18 - - chr1:58450-61939 - - 6.39099e+06 4.78062e+06 8.00137e+06 OK
CUFF.29 - - CUFF.29 - - chr1:69835-71779 - - 5.36129e+06 3.33491e+06 7.38767e+06 OK
CUFF.23 - - CUFF.23 - - chr1:79289-80921 - - 1.91293e+07 1.49043e+07 2.33542e+07 OK
CUFF.27 - - CUFF.27 - - chr1:81081-83838 - - 1.31826e+07 1.05592e+07 1.5806e+07 OK
CUFF.28 - - CUFF.28 - - chr1:83989-87313 - - 9.07865e+06 7.10922e+06 1.10481e+07 OK
CUFF.24 - - CUFF.24 - - chr1:87746-88937 - - 5.73623e+06 2.95375e+06 8.5187e+06 OK
CUFF.26 - - CUFF.26 - - chr1:89066-90884 - - 2.76595e+07 2.28807e+07 3.24384e+07 OK
CUFF.30 - - CUFF.30 - - chr1:71904-78568 - - 4.5969e+07 4.2889e+07 4.90491e+07 OK
CUFF.31 - - CUFF.31 - - chr1:91084-120988 - - 1.72161e+07 1.6337e+07 1.80952e+07 OK
CUFF.41 - - CUFF.41 - - chr1:176053-177408 - - 1.7363e+07 1.28799e+07 2.18461e+07 OK
CUFF.37 - - CUFF.37 - - chr1:121127-175952 - - 1.47316e+07 1.41319e+07 1.53313e+07 OK
CUFF.43 - - CUFF.43 - - chr1:227417-267296 - - 1.86679e+07 1.78759e+07 1.946e+07 OK
CUFF.38 - - CUFF.38 - - chr1:317719-342194 - - 1.6219e+07 1.52751e+07 1.7163e+07 OK
CUFF.32 - - CUFF.32 - - chr1:342335-354967 - - 1.94708e+07 1.80255e+07 2.0916e+07 OK
CUFF.33 - - CUFF.33 - - chr1:387195-387999 - - 9.43071e+06 4.85615e+06 1.40053e+07 OK
CUFF.34 - - CUFF.34 - - chr1:394312-396504 - - 7.20676e+06 5.00872e+06 9.4048e+06 OK
CUFF.35 - - CUFF.35 - - chr1:388067-394261 - - 8.34603e+06 6.98313e+06 9.70893e+06 OK
CUFF.50 - - CUFF.50 - - chr1:396573-409814 - - 1.20163e+07 1.09077e+07 1.31248e+07 OK
CUFF.36 - - CUFF.36 - - chr1:424784-427729 - - 7.41585e+06 5.51685e+06 9.31486e+06 OK
CUFF.64 - - CUFF.64 - - chr1:355018-387052 - - 1.49017e+07 1.41116e+07 1.56917e+07 OK
CUFF.40 - - CUFF.40 - - chr1:438756-440439 - - 4.27964e+06 2.31601e+06 6.24328e+06 OK

History

103: Cufflinks on data 101: transcript expression
~130,000 lines
format: tabular, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa

gth	coverage	FPKM	FPKM_co
9584	5.67414e+06	4.50983e+06	
1177	8.99094e+06	6.96782e+06	
6131	8.78583e+06	6.34908e+06	
7517	1.00617e+07	7.48513e+06	
0277	8.95434e+06	5.44216e+06	

102: Cufflinks on data 101: gene expression
~130,000 lines
format: tabular, database: hg19
Info: cufflinks v1.3.0
cufflinks -q --no-update-check -I 300000 -F 0.100000 -j 0.150000 -p 8 -N -b /galaxy/data/hg19/sam_index/hg19.fa

length	coverage	FPKM	FP
5.67414e+06	4.50983e+06	6.	
8.99094e+06	6.96782e+06	1.	
8.78583e+06	6.34908e+06	1.	
1.00617e+07	7.48513e+06	1.	
8.95434e+06	5.44216e+06	1.	

101: Concatenate datasets on data 99 and data 100
~83,000,000 lines

Your workflows

Create new workflow | Upload or import workflow

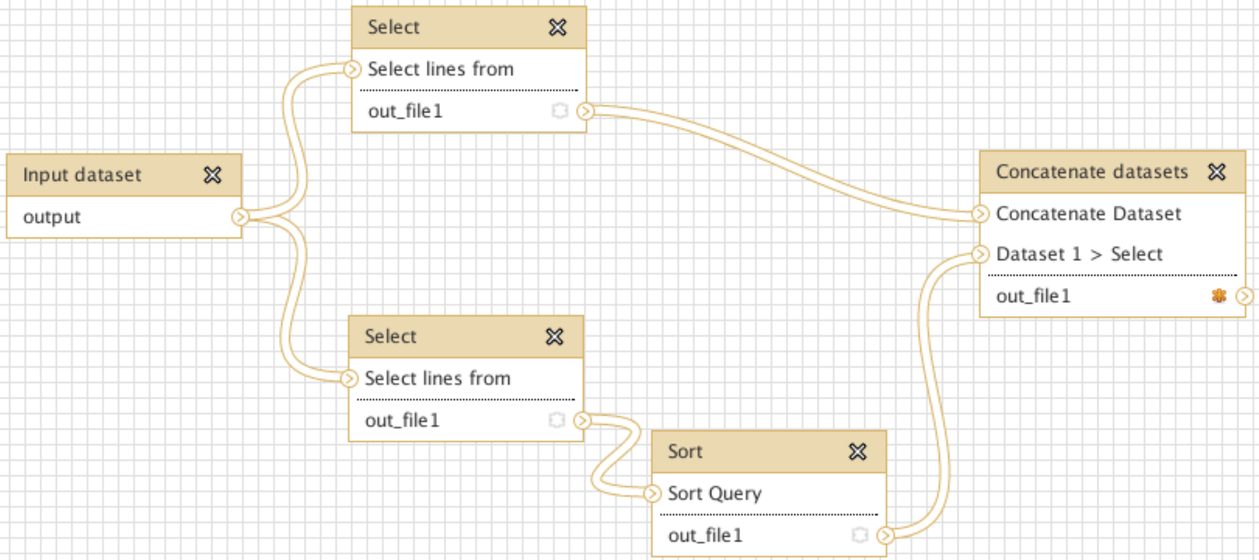
Name	# of Steps
imported: imported: Make Ensembl GTF compatible with Cufflinks	7
imported: Bristol workflow to get sorted unique proper pair mapped reads	11
imported: Sort SAM file for Cufflinks	5
imported: metagenomic analysis	16
pileup and filter	4
map and filter to bam	3
filter pileup and cut sort	3
Workflow constructed from history '8 human'	4
filter and bam	2
Clone of 'map and filter to bam'	3
Single hg19 map Workflow	3
imported: Sort SAM file for Cufflinks	5

Workflows shared with you by others

No workflows have been shared with you.

Other options

Configure your workflow menu



Tool: Concatenate datasets

Concatenate Dataset
Data input 'input1' (data)

Datasets:

Dataset 1

Select
Data input 'input2' (data)

Remove Dataset 1

Add new Dataset

Edit Step Actions

Rename Dataset
out_file1
Create

Add actions to this step; actions are applied when this workflow step completes.

Edit Step Attributes

Annotation / Notes:

Concantenate headers and sorted SAM entries.

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

WARNING: Be careful not to concatenate datasets of different kinds (e.g., sequences with intervals). This tool does not check if the datasets being concatenated are in the same format.

What it does

Concatenates datasets

Example

Concatenating Dataset:

```
chrX  151087187  151087355  A  0
chrX  151572400  151572481  B  0
```

Running workflow "imported: Sort SAM file for Cufflinks"

Expand All | Collapse

Cufflinks requires that SAM files be sorted by chromosome and position. This workflow performs the sorting necessary for Cufflinks.

Step 1: Input dataset

Input Dataset

42: Map with Bowtie f..apped reads

type to filter

Step 2: Select

Remove comment lines.

Select lines from
Output dataset 'output' from step 1

that
NOT Matching

the pattern
^@

Actions:
Hide this dataset.
Hide this dataset.
Hide this dataset.
Hide this dataset.
Hide this dataset.
Hide this dataset.
Hide this dataset.

Step 3: Select

Select SAM headers.

Select lines from
Output dataset 'output' from step 1

that
Matching

the pattern
^@

Actions:
Hide this dataset.
Hide this dataset.
Hide this dataset.
Hide this dataset.

History

NM18507 521.2 Gb

116: Cufflinks on data 101 and data 113: assembled transcripts

115: Cufflinks on data 101 and data 113: transcript expression

114: Cufflinks on data 101 and data 113: gene expression

113: Paste on data 109 and data 112

112: Cut on data 111

111: Merge Columns on data 110

110: Add column on data 108

109: Cut on data 85

108: Cut on data 85

107: Filter on data 106
1,394 regions
format: bed, database: hg19
Info: Filtering with c7!='no_overlap', kept 1.80% of 77614 valid lines (77614 total lines).
display at UCSC main
view in GeneTrack
display at Ensembl Current
display at RViewer main

1.Chrom	2.Start	3.End	4.Name
chr1	147574323	148005511	input_1
chr1	148005402	148011788	input_1
chr1	148005402	148023669	input_1
chr1	148003642	148025863	input_1
chr1	147930760	148176401	input_1
chr1	148201753	148202536	input_1


History
NM18507
116: C
101 and d
transcript
115: C
101 and d
expressio
114: C
101 and d
expressio
113: Pa
109 and d
112: C
111: M
on data 1
110: A
data 108
HISTORY LISTS
Saved Histories
Histories Shared with Me
CURRENT HISTORY
Create New
Clone
Copy Datasets
Share or Publish
Extract Workflow
Dataset Security
Show Deleted Datasets
Show Hidden Datasets
Purge Deleted Datasets
Show Structure
Export to File
Delete
Delete Permanently
OTHER ACTIONS
Import from File

Some datasets still queued or running were ignored

Workflow name

Workflow constructed from history 'NM18507'

Create Workflow | Check all | Uncheck all

Tool	History items created
Upload File *This tool cannot be used in workflows*	**1: SRR002319_1.fastq** Treat as input dataset
	3: SRR002319_2.fastq Treat as input dataset
	4: SRR003961_1.fastq Treat as input dataset
	5: SRR003960_2.fastq Treat as input dataset
Upload File *This tool cannot be used in workflows*	**7: SRR003960_1.fastq** Treat as input dataset
Upload File *This tool cannot be used in workflows*	**8: SRR003961_2.fastq** Treat as input dataset
Upload File *This tool cannot be used in workflows*	**9: SRR003962_1.fastq** Treat as input dataset
FASTQ Groomer Include "FASTQ Groomer" in workflow	**11: FASTQ Groomer on data 7**
FASTQ Groomer Include "FASTQ Groomer" in workflow	**12: FASTQ Groomer on data 5**
FASTQ Groomer Include "FASTQ Groomer" in workflow	**13: FASTQ Groomer on data 4**

Saved Histories

search history names and tags

Advanced Search

Name	Datasets		Tags	Sharing	Size on Disk	Created	Last Updated ↑	Status
NM18507	55	9	0 Tags		521.2 Gb	Jan 28, 2011	37 minutes ago	current history
YH	17		0 Tags		193.6 Gb	Feb 18, 2011	Feb 18, 2011	
Unnamed history	196	5	0 Tags		459.4 Gb	Dec 27, 2010	Feb 17, 2011	
Sort SAM file for Cufflinks workflow results	9		0 Tags		51.1 Gb	Jan 26, 2011	Jan 27, 2011	
chimpanzee	8		0 Tags		178.6 Gb	Jan 18, 2011	Jan 20, 2011	
test	9	1	0 Tags		127.3 Mb	Jan 18, 2011	Jan 20, 2011	
fa2	26		0 Tags		8.2 Gb	Jan 19, 2011	Jan 19, 2011	
fa	30		0 Tags		7.6 Gb	Jan 19, 2011	Jan 19, 2011	

For 0 selected histories: Rename | Delete | Delete Permanently | Undelete

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

History

NM18507 521.2 Gb

116: Cufflinks on data 101 and data 113: assembled transcripts

115: Cufflinks on data 101 and data 113: transcript expression

114: Cufflinks on data 101 and data 113: gene expression

113: Paste on data 109 and data 112

112: Cut on data 111

111: Merge Columns on data 110

110: Add column on data 108

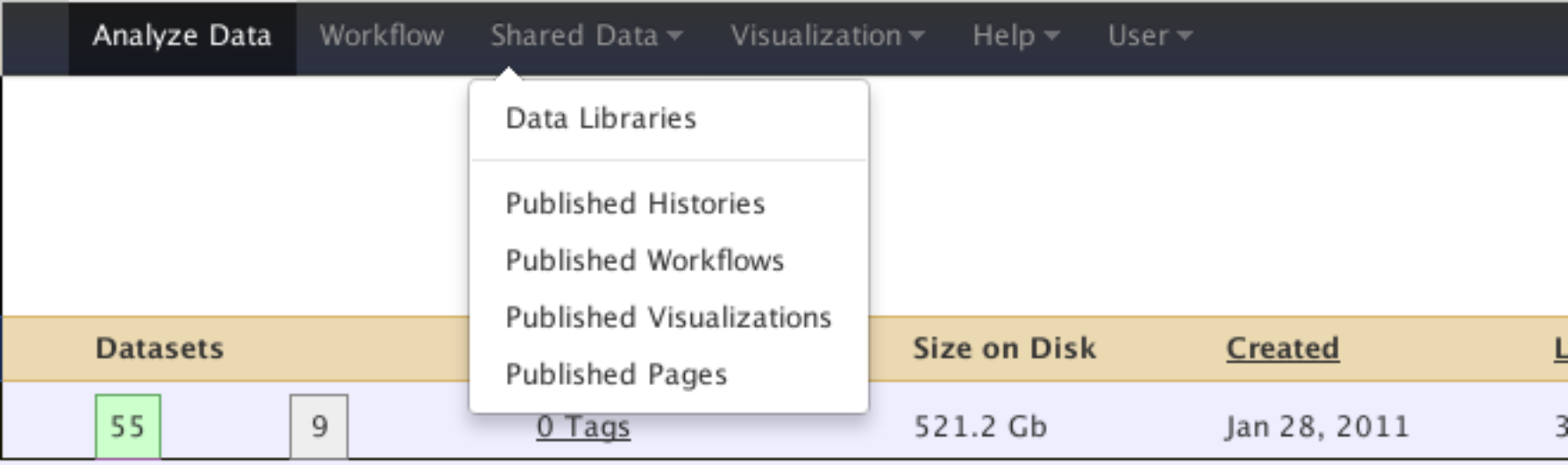
109: Cut on data 85

108: Cut on data 85

107: Filter on data 106

1,394 regions
format: bed, database: hg19
Info: Filtering with c7!='no_overlap', kept 1.80% of 77614 valid lines (77614 total lines).

display at UCSC main
view in GeneTrack
display at Ensembl Current
display at RViewer main


Analyze Data
Workflow
Shared Data
Visualization
Help
User
Data Libraries
Published Histories
Published Workflows
Published Visualizations
Published Pages
Datasets
Size on Disk
Created
55
9
0 Tags
521.2 Gb
Jan 28, 2011

Published Workflows

search name, annotation, owner, and t.

Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
imported: imported: RNASeq workflow		sjyap			~ 4 hours ago
imported: RNASeq workflow		benpass			Apr 25, 2012
RNASeq workflow		fluidigmngs			Apr 16, 2012
mt analysis 0.017 strand-specific (fastq single) from TopHat BAM		aun1			Apr 03, 2012
imported: Workflow from UCSC genes and symbols		saad-uconn			Mar 28, 2012
Prep pgSnp file to run SIFT	This adds the reference allele to homozygous SNPs in a pgSnp file for use in SIFT.	Belinda			Mar 26, 2012
Workflow 1 (PLUS version): A faire faire par les etudiants		yann-lbbe			Mar 26, 2012
Q64~Eigentable, Groups, Tail signatures, Score distributions'		leemsilver			Mar 26, 2012
Q64=AAAA ACTUAL SCORES, Eigentable, Tail signatures, Score distributions		leemsilver			Mar 24, 2012
Sureselect Filter BLAT		odhardy			Mar 16, 2012
imported: Tophat - Cuffdiff (paired-end, fastq)		muehlsch12			Mar 14, 2012
imported: metagenomic analysis	Generic workflow for performing a metagenomic analysis on NGS data.	yong27			Mar 13, 2012
imported: Make Ensembl GTF compatible with Cufflinks	Converts an Ensembl gene annotation file so that it can be used with Cufflinks/compare/diff.	guzhi100			Mar 07, 2012
LNE Workflod	Workflow genome collaboration	josephcarter			Mar 02, 2012
mapping porcine small RNA by bowtie		wanbo			Feb 26, 2012
bwa-version-analysis		aun1			Jan 31, 2012
imported: metagenomic analysis	Generic workflow for performing a metagenomic analysis on NGS data.	cristiane			Jan 25, 2012
MACS		nanleng			Jan 19, 2012
Constructed Workflow		james			Jan 19, 2012
Clone of 'Avinash Workflow - Nov 14th 2011' shared by 'avinash.banala01@gmail.com'		abhishekreddy			Dec 01, 2011
WF'Metagenomics'		koozyn			Nov 27, 2011
Workflow constructed from history 'Clone of 'imported: workshop data' (active items only)'		arnisut			Nov 24, 2011
cshl-workflow		cartman			Nov 18, 2011
mt analysis 0.01 strand-specific (fastq double)		aun1			Nov 18, 2011
rama proj		rakhi			Nov 16, 2011
Avinash Workflow - Nov 14th 2011		nash			Nov 16, 2011
linear regression analysis of Genetic Association Studies of Complex Diseases and Disorders using snps and exons related to them		salendra			Nov 16, 2011
Workflow constructed from history 'Test1'		mahe			Nov 15, 2011

Thanks for your listening