

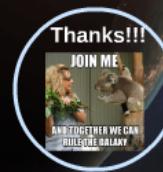
RNA-Seq Data Analysis Using Galaxy with Dolphin

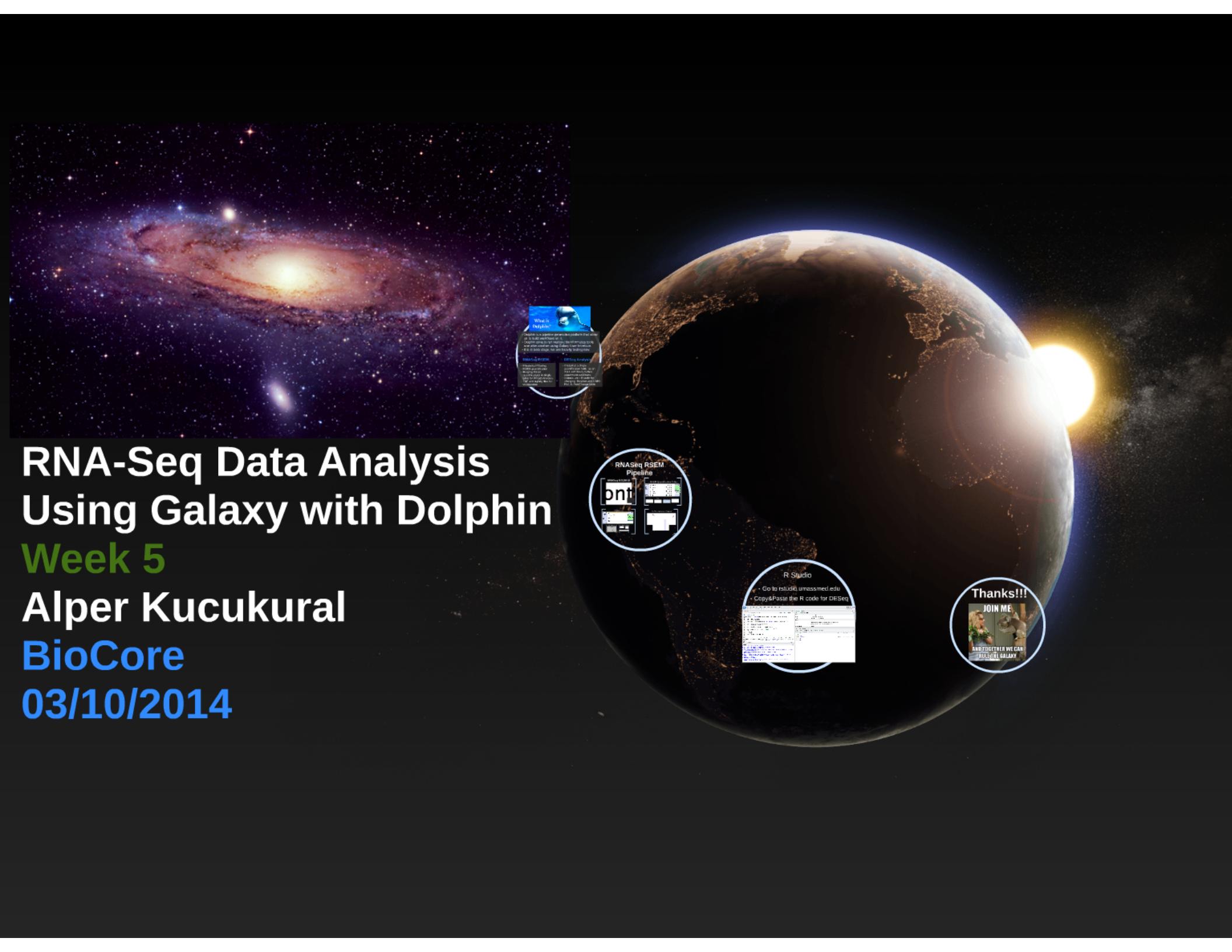
Week 5

Alper Kucukural

BioCore

03/10/2014





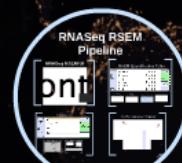
RNA-Seq Data Analysis Using Galaxy with Dolphin

Week 5

Alper Kucukural

BioCore

03/10/2014



What is Dolphin?

- Dolphin is a pipeline generation platform that allow us to build workflows on it.
- Dolphin aims to run multiple bioinformatics tools one after another using Galaxy User Interface.
- It is in beta stage, we are heavily testing now.

RNASeq RSEM

- Ribosomal Filtering.
- RSEM quantification
- Merging library quantifications in single table for DESeq Analysis.
- TDF and bigWig files for visualization

DESeq Analysis

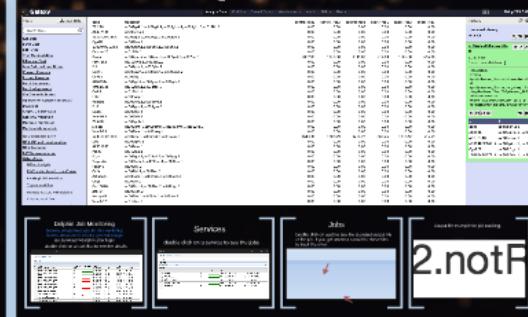
- It requires a single quantification table as an input and library names, experiment conditions.
- Outputs are i. R code for changing the plots and ii. MA-Plot, iii. FoldChange table

RNASeq RSEM Pipeline

RNASeq RSEM UI



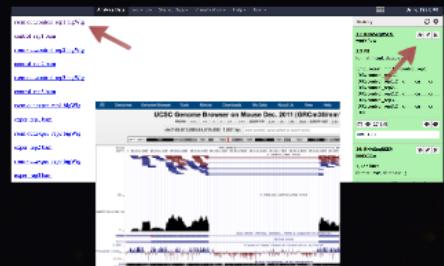
RSEM Quantification Table



DESeq Analysis



UCSC Genome Browser



RNASeq RSEM UI

Galaxy

Analyze Data Workflow Shared Data Visualization

Tools

[Get Genomic Scores](#)

[Operate on Genomic Intervals](#)

[Statistics](#)

[Graph/Display Data](#)

[Regional Variation](#)

[Multiple regression](#)

[Multivariate Analysis](#)

NGS TOOLBOX BETA

[NGS: QC and manipulation](#)

[NGS: Mapping](#)

[FASTA manipulation](#)

[UMassTools](#)

[DESeq Analysis](#)

[FileTransfer from/to the Cluster](#)

[RNASeqRSEM workflow](#)

[Tophat workflow](#)

[Multiple FastQC with dolphin](#)

[ChipSeq workflow](#)

[MIRZA miRNA target search for each miRNA sequence](#)

[MIRZA miRNA target search with dolphin](#)

NGS: SAM Tools

NGS: GATK Tools (beta)

[Phenotype Association](#)

Workflows

- [All workflows](#)

RNASeqRSEM workflow (version 1.0.0)

Select a genome: MouseTest

Select a genome build: mm10

Source for Input Parameter File: Free Text

Input Directory(Full path in the cluster): /share/data/umw_biocore/genome_data/mousetest/mm10

Input Parameters (See below for example):

```
control_rep1 DIR/control_rep1.1 DIR/control_rep1.2  
control_rep2 DIR/control_rep2.1 DIR/control_rep2.2  
control_rep3 DIR/control_rep3.1 DIR/control_rep3.2  
exper_rep1 DIR/exper_rep1.1 DIR/exper_rep1.2  
exper_rep2 DIR/exper_rep2.1 DIR/exper_rep2.2  
exper_rep3 DIR/exper_rep3.1 DIR/exper_rep3.2
```

RSEM Parameters:

Please refer to parameters for rsem-calculate-expression

Output Directory: RSEMTEST

Execute

History

Unnamed history 0 bytes

1: RNASeqRSEM workflow

RSEM Quantification Table

Galaxy

	control_rep1	control_rep2	control_rep3	exper_rep1	exper_rep2	exper_rep3
gene	transcript					
Zfp536	uc009gkk.1,uc009gkl.1,uc009gkm.1,uc009gko.1,uc012fip.1	0.00	0.00	0.00	0.00	0.00
AK137448	uc009ewo.1	0.00	0.00	0.00	0.00	0.00
4931406P16Rik	uc009giz.1,uc009gja.1,uc009gjb.1	0.00	0.00	0.00	0.00	0.00
Cyp2f2	uc009fuy.1	0.00	0.00	0.00	0.00	0.00
1700049G17Rik	uc009fxh.1,uc009fxi.1,uc009fxj.1	0.00	0.00	0.00	0.00	0.00
Ceacam20	uc012fb1.1	0.00	0.00	0.00	0.00	0.00
Dnaj3	uc007yab.2,uc007yac.2,uc007yad.2,uc012aat.1	2067.00	1340.00	1391.00	1217.00	1606.00
Fam71e2	uc012ewx.2,uc012ewy.2	0.00	0.00	0.00	0.00	0.00
Ppl	uc007ybt.1,uc007ybu.2	0.00	0.00	0.00	0.00	0.00
Nucb1	uc009gvm.2,uc009gvn.2,uc009gvo.2	0.00	0.00	0.00	0.00	0.00
Cd79a	uc009fq.1	0.00	0.00	0.00	0.00	0.00
AI987944	uc009gll.2,uc009glm.2,uc009gln.2	0.00	0.00	0.00	0.00	0.00
Vmn1r135	uc012fdk.1,uc012feo.1	0.00	0.00	0.00	0.00	0.00
Ccdc8	uc012faf.1	0.00	0.00	0.00	0.00	0.00
Ldhc	uc009gzp.1	0.00	0.00	0.00	0.00	0.00
Hspbp1	uc009eyc.1,uc009eyd.1	0.00	0.00	0.00	0.00	0.00
Pth2	uc009gtz.1,uc009gua.1	0.00	0.00	0.00	0.00	0.00
Npas1	uc009fhv.2	0.00	0.00	0.00	0.00	0.00
Vmn1r82	uc009fdx.1	0.00	0.00	0.00	0.00	0.00
Zfp658	uc009gnb.1	0.00	0.00	0.00	0.00	0.00
Siglec5	uc009gml.1,uc009gmm.1,uc009gmn.1,uc012fje.1	0.00	0.00	0.00	0.00	0.00
Vmn2r29	uc009fcn.1,uc012exq.1	0.00	0.00	0.00	0.00	0.00
5730403B10Rik	uc007yaj.1,uc007yak.1,uc012aau.1	1613.31	1202.79	1210.77	1563.76	1742.42
Dbp	uc009gws.2	0.00	0.00	0.00	0.00	0.00
AK131696	uc009exx.1	0.00	0.00	0.00	0.00	0.00
Psg20	uc012faq.1	0.00	0.00	0.00	0.00	0.00
Fcgtr	uc009gte.2,uc009gtf.2,uc009gtg.2	0.00	0.00	0.00	0.00	0.00
Trappc6a	uc009fbm.1,uc009fmc.1,uc009fmd.1	0.00	0.00	0.00	0.00	0.00
Fbxo46	uc009fk.2	0.00	0.00	0.00	0.00	0.00
Cnfn	uc009fsi.1,uc009fsm.1	0.00	0.00	0.00	0.00	0.00
Arhgap33	uc009geg.1,uc009ger.1,uc009ges.1	0.00	0.00	0.00	0.00	0.00
Gnp8	uc009fk.1	0.00	0.00	0.00	0.00	0.00
Gm10046	uc009fbk.2,uc009fxc.2,uc012fgn.1	0.00	0.00	0.00	0.00	0.00
Zfp784	uc009ezn.2	0.00	0.00	0.00	0.00	0.00
Hnrnpull	uc009ftu.1,uc009ftv.1,uc009ftw.1	0.00	0.00	0.00	0.00	0.00
Vmn1r123	uc012fcx.1	0.00	0.00	0.00	0.00	0.00

History
Unnamed history
64.6 KB
2: RNASeqRSEM workflow
1,127 lines
format: csv, database: ?
[RSEMTEST1]
python
/project/umw_biocore/bin/runWorkflow
-/project/umw_biocore/svcgalaxy/bioc
dist/database/files/000/dataset_567.
-d biocoreweb -o
/home/USERNAME/RSEMTEST1 -p
/project/umw_biocore/bin/workflow/:
1 2
1 gene transcript
Zfp536 uc009gkk.1,uc009gkl.1,uc009gkm.1,uc009gko.1,uc012fip.1
AK137448 uc009ewo.1
4931406P16Rik uc009giz.1,uc009gja.1,uc009gjb.1
Cyp2f2 uc009fuy.1
1700049G17Rik uc009fxh.1,uc009fxi.1,uc009fxj.1

Dolphin Job Monitoring
biocore.umassmed.edu for this workshop
bioinfo.umassmed.edu for general usage
Go services=>dolphin after login
double click on a workflow to see the details

Services
double click on a service to see the jobs

Jobs
Double click on a job to see the standard output file of the job. If you get an error consult to these files to track the error.

Output file example for job tracking:
test1234.log

Dolphin Job Monitoring

biocore.umassmed.edu for this workshop

bioinfo.umassmed.edu for general usage

Go services=>dolphin after login

double click on a workflow to see the details

Workflows					
#	Name	Duration	% Complete	Start	Finish
52	RNASeqRSEMWorkflow	183	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 14:04:51	2014-03-10 14:07:54
51	RNASeqRSEMWorkflow	587	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 13:58:13	2014-03-10 14:08:00
49	RNASeqRSEMWorkflow	97	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 11:58:10	2014-03-10 11:59:47
48	FileTransfer	4	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 11:56:51	2014-03-10 11:56:55
39	RNASeqRSEMWorkflow	70	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 11:15:02	2014-03-10 11:16:12
38	RNASeqRSEMWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-10 11:12:54	0000-00-00 00:00:00
37	RNASeqRSEMWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-10 11:07:40	0000-00-00 00:00:00
36	RNASeqRSEMWorkflow	70	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 10:56:39	2014-03-10 10:57:49
35	RNASeqRSEMWorkflow	92	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-10 10:52:42	2014-03-10 10:54:14
34	RNASeqRSEMWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-10 10:46:55	0000-00-00 00:00:00
33	RNASeqRSEMWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-07 20:57:25	0000-00-00 00:00:00
32	TophatDolphinWorkflow	1216	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-06 20:13:47	2014-03-06 20:34:03
31	TophatDolphinWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-06 20:13:32	0000-00-00 00:00:00
30	TophatDolphinWorkflow	1264	<div style="width: 100%; background-color: green; height: 10px;"></div>	2014-03-06 20:13:04	2014-03-06 20:34:08
29	TophatDolphinWorkflow		<div style="width: 100%; background-color: red; height: 10px;"></div>	2014-03-06 20:01:37	0000-00-00 00:00:00

Services

double click on a service to see the jobs

BioCore News Calendar Services ▾ Forum FAQ Team Contact Training ▾ Log Out (kucukura)

Home / Services / Dolphin

[user:ak97w]

Workflows

Showing all 40 rows

Services

#	Name	Duration	% Complete	Start	Finish
189	stepRibo	93	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:04:51	2014-03-10 14:06:24
190	stepRSEM	22	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:06:23	2014-03-10 14:06:45
191	stepIGVTDF	17	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:06:46	2014-03-10 14:07:03
192	stepRSEMCount	4	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:06:47	2014-03-10 14:06:51
195	stepRSEMRawCount	3	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:07:09	2014-03-10 14:07:12
197	stepBam2BW	13	<div style="width: 100%; background-color: #00A000; height: 10px;"></div>	2014-03-10 14:07:31	2014-03-10 14:07:44

Jobs

Double click on a job to see the standard output file of the job. If you get an error consult to these files to track the error.

The screenshot shows a user interface for managing jobs. At the top, there is a navigation bar with links for BioCore, News, Calendar, Services (with a dropdown), Forum, FAQ, Team, Contact, Training (with a dropdown), and Log Out (kucukura). Below the navigation bar, the URL is shown as Home / Services / Dolphin. The main content area is divided into three sections: Workflows, Services, and Jobs. The Jobs section is currently active, indicated by a blue header. A red arrow points to this blue header. The Jobs table has columns for #, Name, Dur..., JobNum, Submission Time, Start, and Finish. The data in the table is as follows:

#	Name	Dur...	JobNum	Submission Time	Start	Finish
1537	stepRSEM	3	78677	2014-03-10 14:06:24	2014-03-10 14:06:25	2014-03-10 14:06:28
1538	stepRSEM_exper_rep3	4	78678	2014-03-10 14:06:25	2014-03-10 14:06:26	2014-03-10 14:06:30
1539	stepRSEM_control_rep3	7	78679	2014-03-10 14:06:26	2014-03-10 14:06:26	2014-03-10 14:06:33
1540	stepRSEM_exper_rep1	7	78680	2014-03-10 14:06:26	2014-03-10 14:06:31	2014-03-10 14:06:38
1541	stepRSEM_control_rep2	5	78681	2014-03-10 14:06:27	2014-03-10 14:06:31	2014-03-10 14:06:36
1542	stepRSEM_control_rep1	7	78682	2014-03-10 14:06:27	2014-03-10 14:06:34	2014-03-10 14:06:41
1543	stepRSEM_exper_rep2	6	78683	2014-03-10 14:06:28	2014-03-10 14:06:38	2014-03-10 14:06:44

A large red arrow points to the bottom right corner of the page.

Output file example for job tracking:

BioCore News Calendar Services ▾ Forum FAQ Team Contact Training ▾ Log Out (kucukura)

Home / Services

stepRSEM_exper_rep3

[user:ak97w]

Workflows

Workflow 1 Workflow 2 Workflow 3 Workflow 4 Show all

Services

Service 1 Service 2 Service 3 Service 4 Show all

Jobs

#	Name
1537	stepRSEM_exper_rep3
1538	stepRSEM_exper_rep3
1539	stepRSEM_exper_rep3
1540	stepRSEM_exper_rep3
1541	stepRSEM_exper_rep3
1542	stepRSEM_exper_rep3
1543	stepRSEM_exper_rep3

Currently Loaded Modulefiles: 1) openssl/1.0.1e 2) python/2.7.5 3) perl/5.18.1
mkdir -p /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/; module load perl/5.18.1; perl /project/umw_biocore/bin/RSEM/rsem-calculate-expression --bowtie-path /project/umw_biocore/bin/bowtie-0.12.9 -p 8 --output-genome-bam --paired-end /home/ak97w/RSEMTEST1/after_ribosome/exper_rep3.1.notR /home/ak97w/RSEMTEST1/after_ribosome/exper_rep3.2.notR /share/data/umw_biocore/genome_data/mousetest/mm10/rsem_ref /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3 /project/umw_biocore/bin/bowtie-0.12.9/bowtie -q --phred33-quals -n 2 -e 99999999 -l 25 -l 1 -X 1000 -p 8 -a -m 200 -S /share/data/umw_biocore/genome_data/mousetest/mm10/rsem_ref -1 /home/ak97w/RSEMTEST1/after_ribosome/exper_rep3.1.notR -2 /home/ak97w/RSEMTEST1/after_ribosome/exper_rep3.2.notR | /project/umw_biocore/bin/RSEM/sam/samtools view -S -b -o /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3.bam - [samopen] SAM header is present: 2058 sequences.
reads processed: 9096
reads with at least one reported alignment: 7486 (82.30%)
reads that failed to align: 1610 (17.70%)
Reported 14675 paired-end alignments to 1 output stream(s)

/project/umw_biocore/bin/RSEM/rsem-parse-alignments /share/data/umw_biocore/genome_data/mousetest/mm10/rsem_ref /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3 /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.stat/rsem.out.exper_rep3.b /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3.bam -t 3 -tag XM
Done!

/project/umw_biocore/bin/RSEM/rsem-build-read-index 32 1 0
/home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3_alignable_1.fq
/home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3_alignable_2.fq
Build Index /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3_alignable_1.fq is Done!
Build Index /home/ak97w/RSEMTEST1/rsem/pipe.rsem.exper_rep3/rsem.out.exper_rep3.temp/rsem.out.exper_rep3_alignable_2.fq is Done!

4:06:28
4:06:30
4:06:33
4:06:38
4:06:36
4:06:41
4:06:44

DESeq Analysis

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools Load Data

DESeq Analysis (version 1.0.0)

Input File: 2: RNASeqRSEM workflow

Library Column Names: control_rep1,control_rep2,control_rep3,exper_rep1,exper_rep2,exper_rep3
control_rep1,control_rep2,control_rep3,exper_rep1,exper_rep2,exper_rep3

Conditions: Control,Control,Control,Treat,Treat,Treat
Control,Control,Control,Treat,Treat,Treat

Column separation: ;

DESeq Method: pooled

Sharing Mode: maximum

Execute

What it does

This function obtains dispersion estimates for a count data set. For each condition (or collectively for all conditions, see 'method' argument below) it first computes for each gene an empirical dispersion value (a.k.a. a raw SCV value), then fits by regression a dispersion-mean relationship and finally chooses

History

Unnamed history 64.6 KB

2: RNASeqRSEM workflow

1,127 lines format: csv, database: ?

[RSEMTEST1]
python
/project/umw_biocore/bin/runWorkflow
-i
/project/umw_biocore/svcgalaxy/bioc
dist/database/files/000/dataset_567.c
-d biocoreweb -o
/home/@USERNAME/RSEMTEST1 -p
/project/umw_biocore/bin/workflow/s

1	2
gene	transcript
Zfp536	uc009gk1,uc009gk1.1,uc
AK17448	uc009ewo.1
4931406P16Rik	uc009giz.1,uc009gja.1,uc
Cyp2f2	uc009fuy.1
1700049G17Rik	uc009fxh.1,uc009fxi.1,uc

DESeq Analysis Results

Fold Change csv file

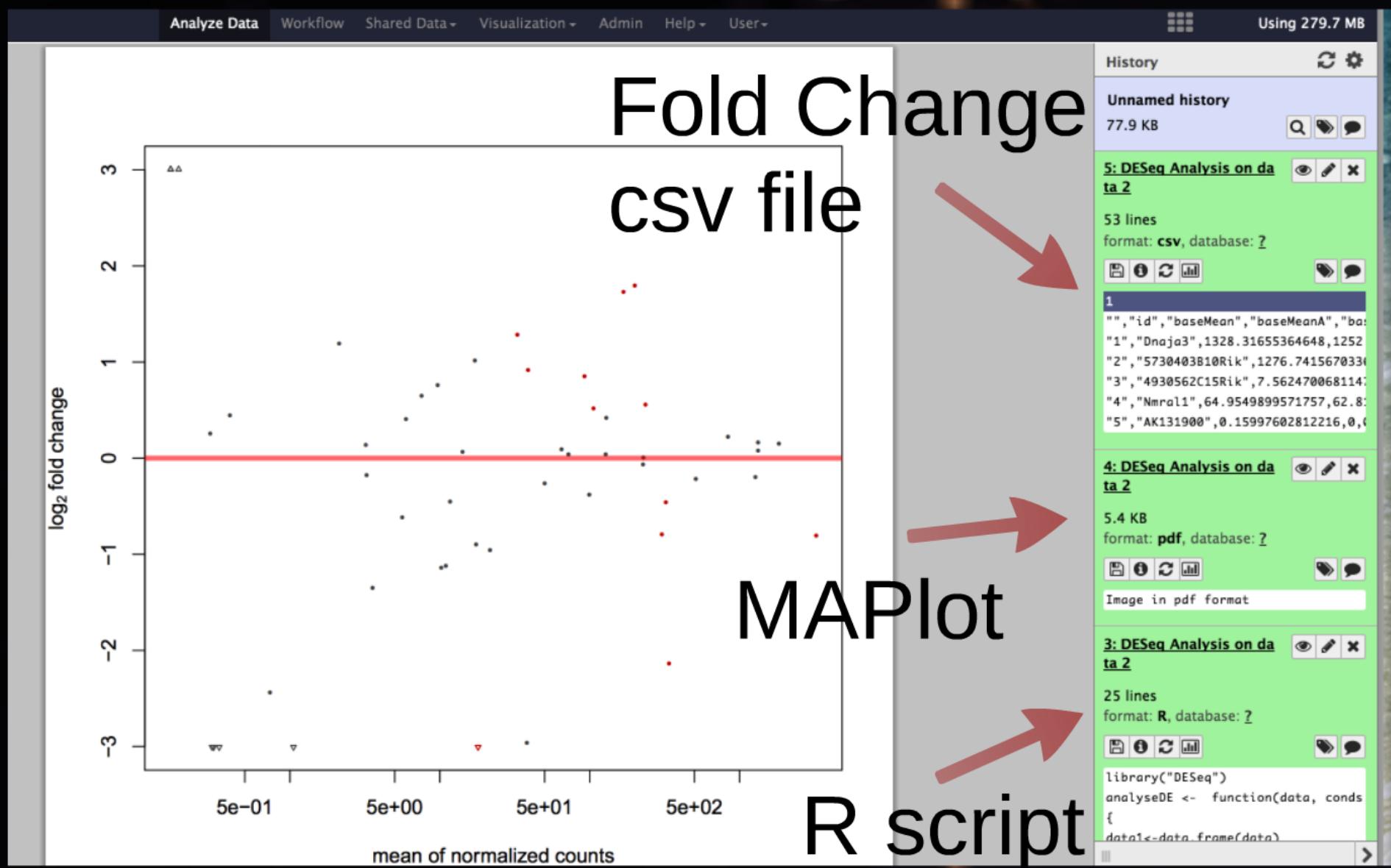
MAPlot

R script

Convert and Filter the Output

padj<0.01 and foldchange > 2 or foldchange<1/2 fold

DESeq Analysis Results



Convert and Filter the Output

The screenshot shows the Galaxy web interface with two tool panels. On the left, under the 'Tools' section, is the 'Convert' tool. It has a dropdown menu set to 'Commas' and is set to 'in Dataset: 5: DESeq Analysis on data 2'. A large blue arrow points from the 'Convert' panel to the right. On the right is the 'Filter' tool, which has a dropdown menu set to '6: Convert on data 5'. The condition input field contains 'c9<0.01 and (c6<0.5 or c6>2)'. Below it, the 'Number of header lines to skip' is set to '1'.

$\text{padj} < 0.01$ and $\text{foldchange} > 2$ or $\text{foldchange} < 1/2$ fold

	"id"	"baseMean"	"baseMeanA"	"baseMeanB"	"foldChange"	"log2FoldChange"	"pval"	"padj"
"19"	"Adcy9"	200.175141012137	89.2449307764603	311.105351247813	3.48597224000394	1.80156108083247	6.84324034375637e-10	8.89621244688328e-09
"30"	"AK208554"	17.9936592931904	35.9873185863809	0	0	-Inf	3.37288854207811e-05	0.000350780408376123
"34"	"Crebbp"	167.892352951942	77.6074506507582	258.177255253125	3.32670707629542	1.73409483941561	5.21761605945395e-19	1.35658017545803e-17
"39"	"Zfp174"	32.9641197742175	19.1618074446482	46.7664321037868	2.4406065157934	1.28723971688363	0.00145516355087421	0.00840761162727324
"51"	"Fgf21"	338.321577346388	551.192747807564	125.450406885212	0.22759807233351	-2.13543975625281	4.55291413039486e-28	2.36751534780533e-26

UCSC Genome Browser

Analyze Data Workflow Shared Data ▾ Visualization ▾ Help ▾ User ▾

Using 395.0 KB

History

17: RNASeqRSEM workflow

3.9 KB

format: **html**, database: ?

(mousetest, mm10, control_rep1
DIR/control_rep1.1
DIR/control_rep1.2_cr__cn__cr__
DIR/control_rep2.1
DIR/control_rep2.2_cr__cn__cr__
DIR/control_rep3.1
DIR/control_rep3.2_cr__cn__cr__
DIR

HTML file

16: RNASeqRSEM workflow

1,127 lines

format: **csv**, database: ?

UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm¹⁰)

move <<< << < > >> zoom in 1.5x 3x 10x base | zoom out 1.5x

chr7:45,613,890-45,615,490 1,601 bp. enter position, gene symbol or search terms

chr7: 45,614,000 45,614,100 45,614,200 45,614,300 45,614,400 45,614,500 45,614,600 45,614,700 45,614,800 45,614,900 45,615,000

500 bases

a control_rep1 track

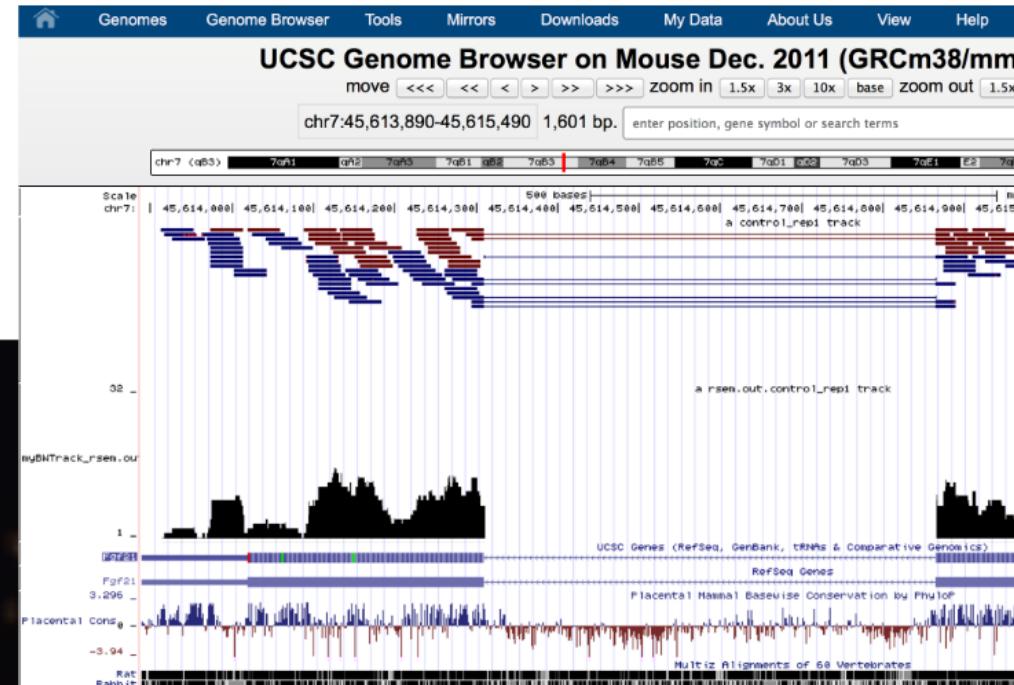
a rsem.out.control_rep1 track

bigWigTrack_rsem.out

Fgf21

Placental conservation

Multiz Alignments of 69 Vertebrates





Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

About Us

View

Help

UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10)

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x

chr7:45,613,890-45,615,490 1,601 bp.

enter position, gene symbol or search terms

chr7 (qB3) 7qA1 qA2 7qA3 7qB1 qB2 7qB3 7qB4 7qB5 7qC 7qD1 qD2 7qD3 7qE1 E2 7qE3

Scale

chr7:

500 bases

mm10

45,614,000| 45,614,100| 45,614,200| 45,614,300| 45,614,400| 45,614,500| 45,614,600| 45,614,700| 45,614,800| 45,614,900| 45,615,

a control_repl track



32

a rsem.out.control_repl track

myBWTrack_rsem.out

1

Fgf21

Fgf21

3.296

Placental Cons

-3.94

Rat

Rabbit

Human

UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)

RefSeq Genes

Placental Mammal Baseewise Conservation by PhyloP

Multiz Alignments of 60 Vertebrates

Thanks!!!



© culture-nomad.com