

Ceto

a Modular Suite of Pipelines for Next Generation Sequence Analysis

Elizabeth Bartom, PhD

Department of Biochemistry and Molecular
Genetics

Northwestern University

ebartom@northwestern.edu

Outline

- What is Ceto?
- Overview of RNAseq Analysis
- RNA-seq Pipeline Details
- Hands-on with RNAseq

Outline

- **What is Ceto?**
- Overview of RNAseq Analysis
- RNA-seq Pipeline Details
- Hands-on with RNAseq

Where does the name come from?

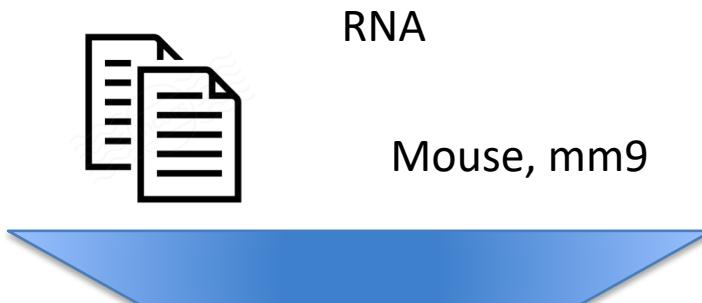
Phorcys — Ceto

Typhon — Echidna



Ceto is the matriarch of a family of hybrid monsters

What is Ceto?

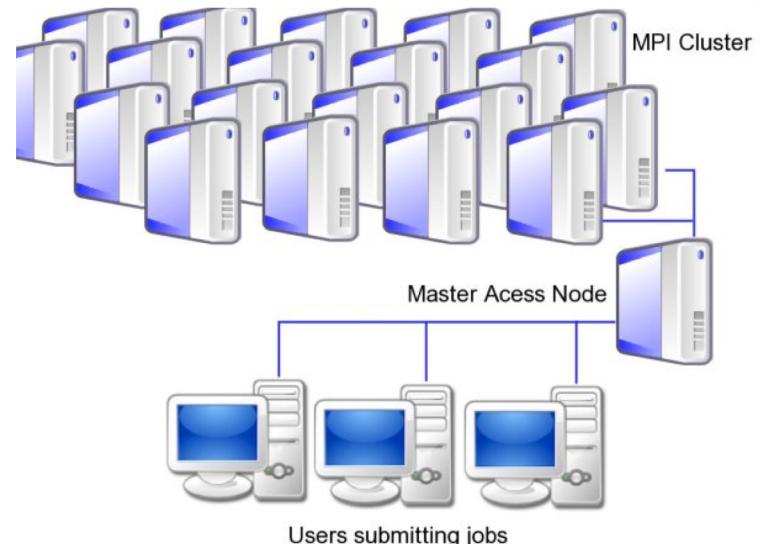
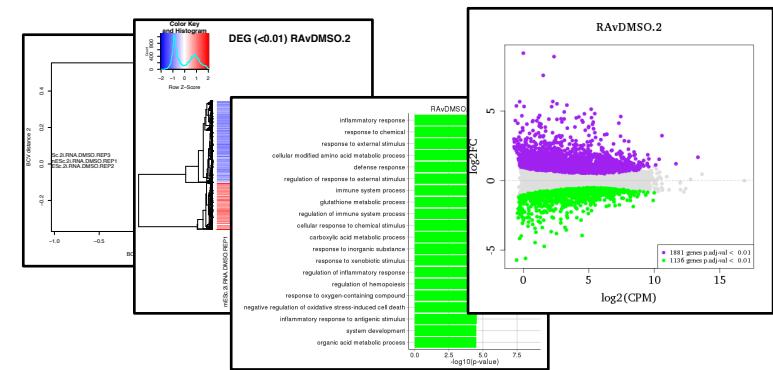


```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=24:00:00
#MSUB -m a
#MSUB -j oe
#MOAB -W umask=0113
#MSUB -N downstreamRNAanalysis
#MSUB -l nodes=1:ppn=4
module load R/3.2.2

# Make Analysis directory for all analysis files.
mkdir ..//results//testRNA/analysis

# Make HTseq counts table.
perl
/projects/p20742//tools/bin/makeHTseqCountsTable.p
l ..//results//testRNA/bam// 
/projects/p20742//anno/Ens/mm9.Ens_67/mm9.Ens_67.c
uff.gtf ..//results//testRNA/bam// 

# Create MDS plot for samples, with count method
htseq
Rscript
/projects/p20742//tools/bin/runEdgeRrnaSeq_2.R --
```



What is Ceto

A decision tree and job handler

Based on input parameters and established pipelines, it will create shell scripts for your samples and submit them to the Quest scheduler

Primary focus: Bulk RNA-seq, ChIP-seq

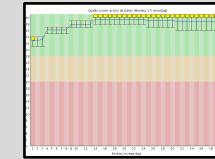
Outline

- What is Ceto?
- **Overview of RNAseq Analysis**
- RNA-seq Pipeline Details
- Hands-on with RNAseq

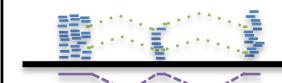


RNA-seq

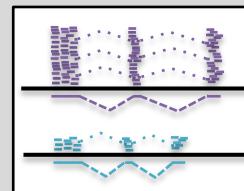
Check Sequence Quality



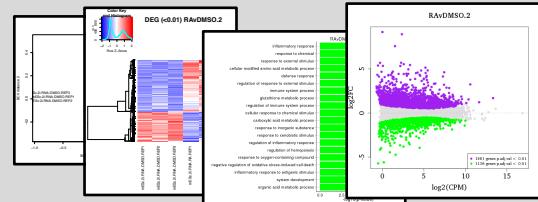
Align to Reference Genome



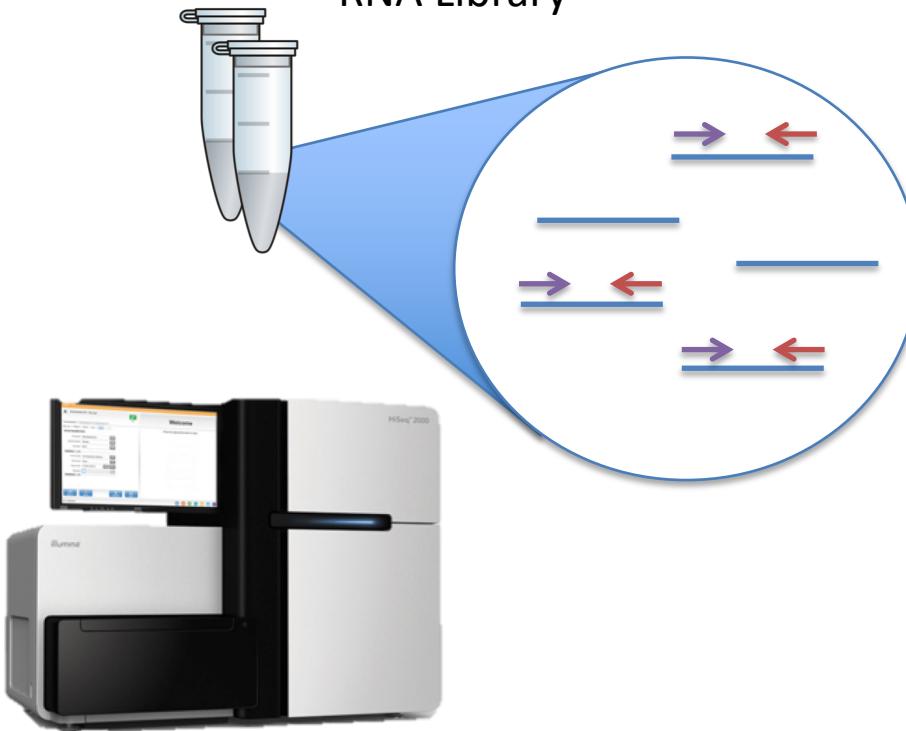
Estimate Gene Expression



Find Differentially
Expressed Genes and
Make Plots



Extracted RNA Library



DNA fragment length
Read length
Paired end or Single

```

@ERR030881.21417929/1
CATGACCCCTGAGTTGTCAAGACAAGTTGCCCTGCAGAGGTAGCTGCC
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.59349201/1
CATGACCCCTGAGTTGTCAAGACAAGTTGCCCTGCAGAGGTAGCTGCC
+
HEGHHHHHFH5' 45544554HHHHHEEEF5*44435555EEGG
@ERR030881.63660328/1
CATGACCCCTGAGTTGTCAAGACAAGTTGCCCTGCAGAGGTAGCTGCC
+
HHHHHHHHHHGG#GFFFDFHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.65131462/1
CGGTTGTCCTCAGNGGCTGCCACACCAGTATGACCCGGAGTCATCTC
+
HHHHHHHHHHFFF#F@=>8FAFC>44445>A@AA9DF:I#####
@ERR030881.40687283/1
TCACAGTCCAAGAACACTGTTTCCAGCGCCAATTCTCAGAGCTGTACCCCTA
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.9914386/1
CTCAGAGCTGTACCCCTCAGCTGAACCCCAAGCAGGAAGCCCCATGACCTT
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.45973010/1
GTTTCCAGCGCCAATTCTCAGAGCTGTACCCCTCAGCTGAACCCCAAGCAG
+

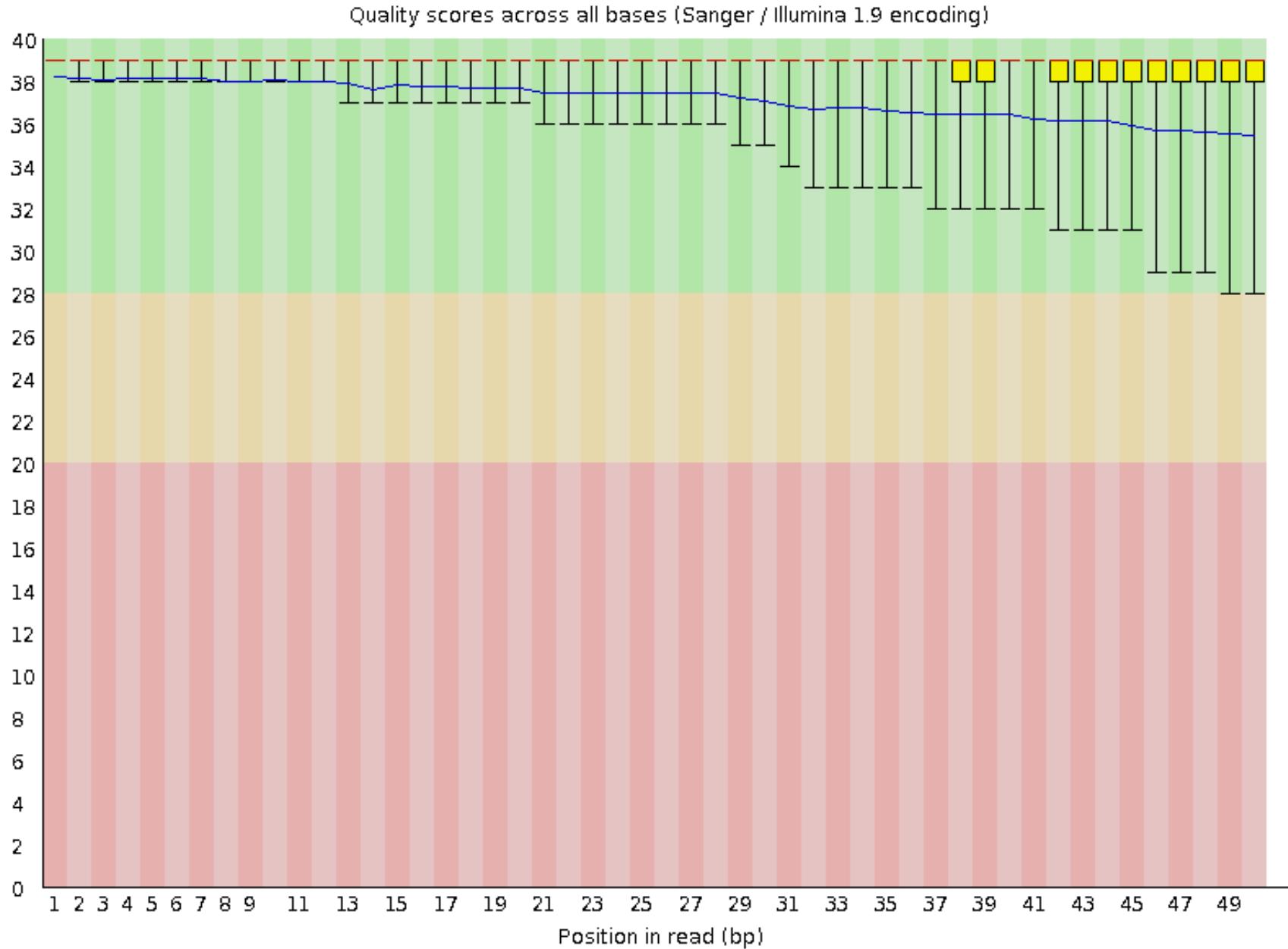
```

```

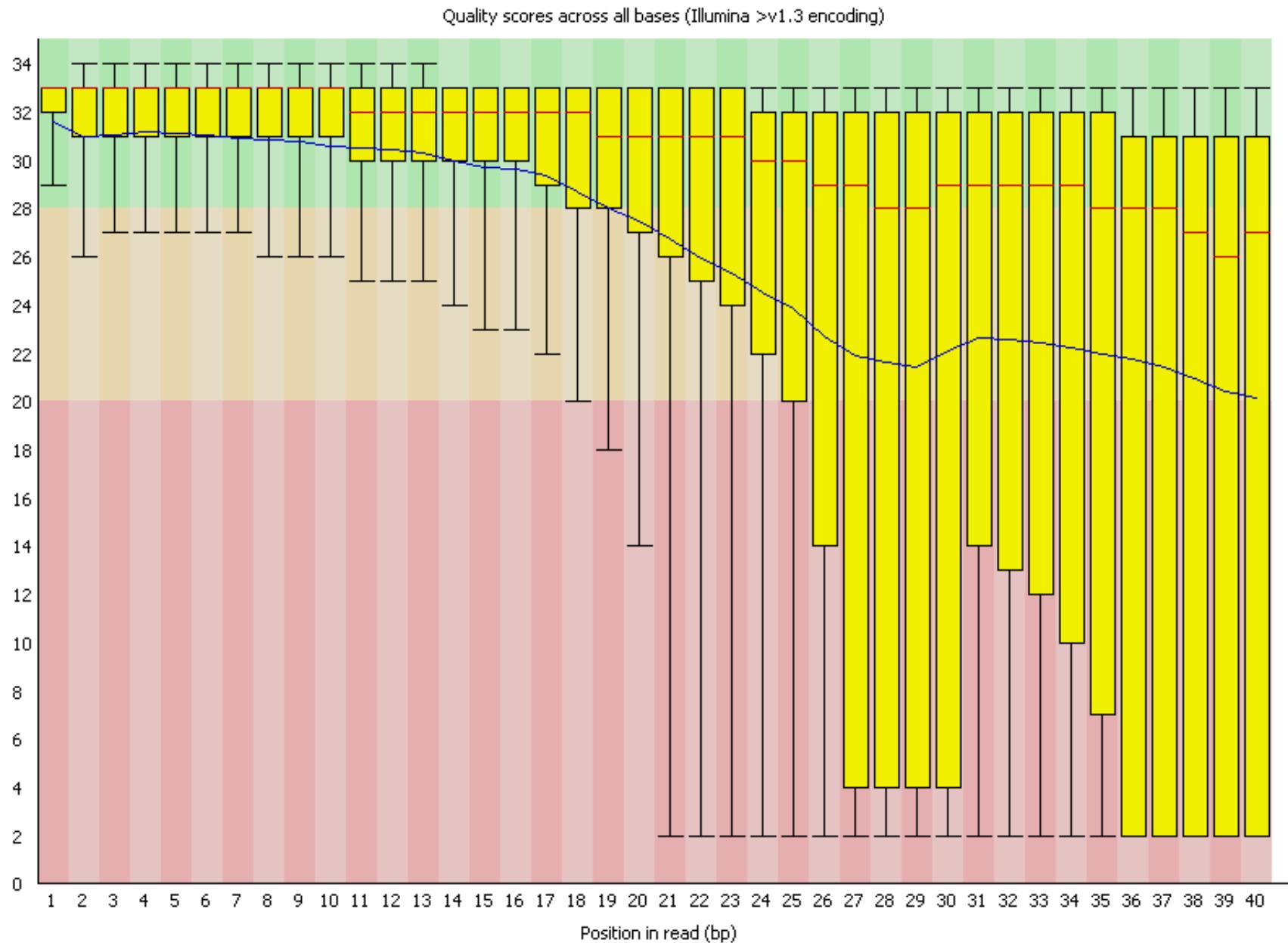
@ERR030881.21417929/2
CCTCACACCACTATGACCCGGAGTGATCTCTGAAGCTGTGGGGATCTGG
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.59349201/2
CCTCACACCACTATGACCCGGAGGGATCTCTGAAGCTGTGGGGATCTGG
+
HHADHGFEEB555445544524+(45544==53?1555,2411435445
@ERR030881.63660328/2
CCTCACACCACTATGACCCGGAGTGATCTCTGAAGCTGTGGGGATCTGG
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.65131462/2
CCAAGAACTGTTGCAGCGCCAATTCTCAGAGCTGTACCCCTCAGCTGAAC
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.40687283/2
CAAACCTTGTGTCCTCAGTGGCTGGCTCACACCAGTATGACCCGGAGTG
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.9914386/2
GGGGGCTCTGTTCCAAACTGGTTGTCCAGTGGCTGGCTCACACCCAG
+
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@ERR030881.45973010/2
GGGGGCTCTGTTCCAAACTGGTTGTCCAGTGGCTGGCTCACACCCAG
+

```

Quality Score Distribution



Quality Score Distribution



Map Raw Reads To Reference

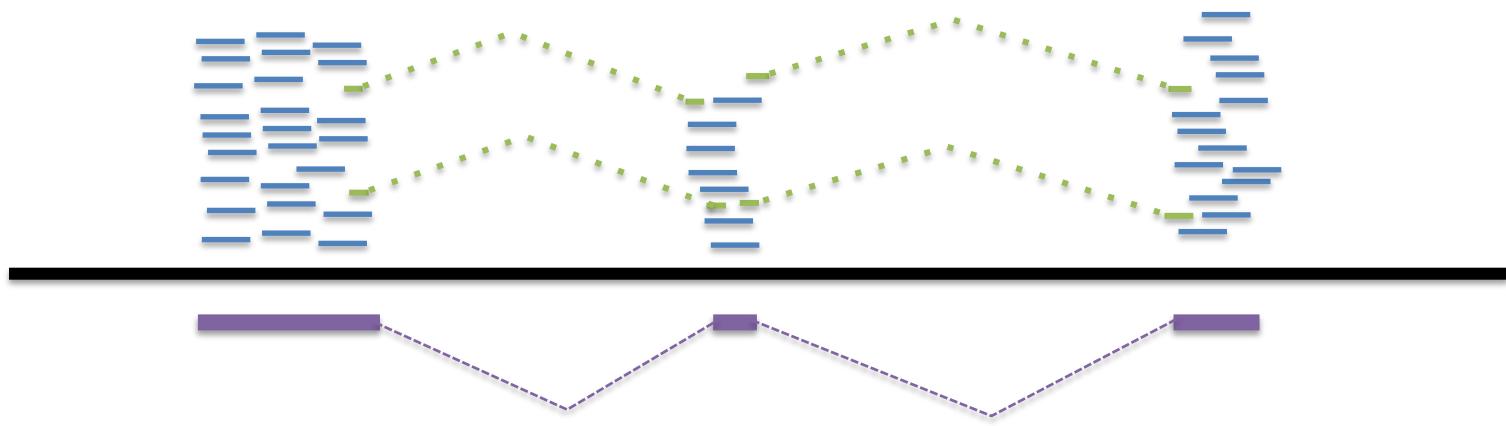


Raw Reads
(x 10,000,000 or so)

Considerations:

Speed? Accuracy? Spliced Reads?

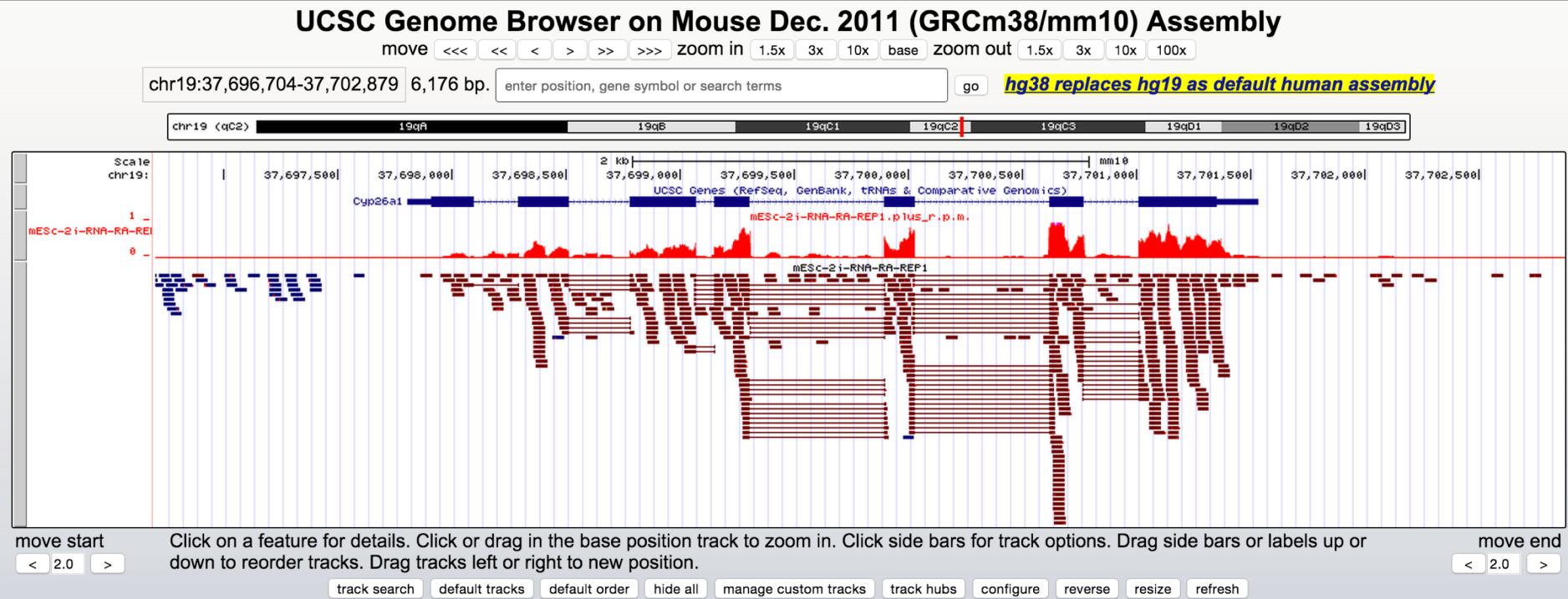
Map Raw Reads To Reference



Bowtie and **Tophat**

<http://tophat.cbcb.umd.edu/>

Visualizing the data



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

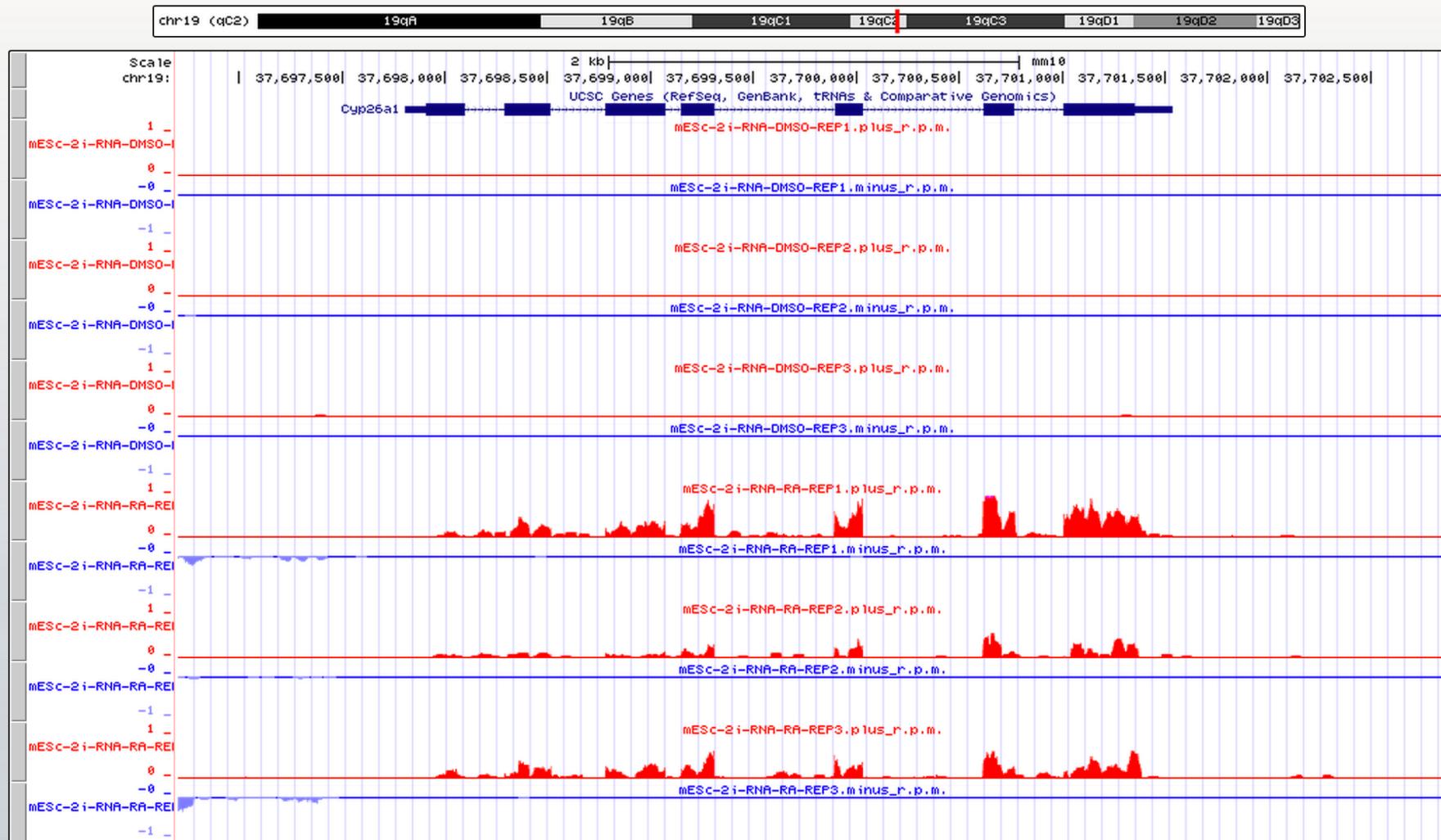
move <<< << < > >>> zoom in 1.5x 3x 10x base ZOOM out 1.5x 3x 10x 100x

chr19:37,696,704-37,702,879 6,176 bp.

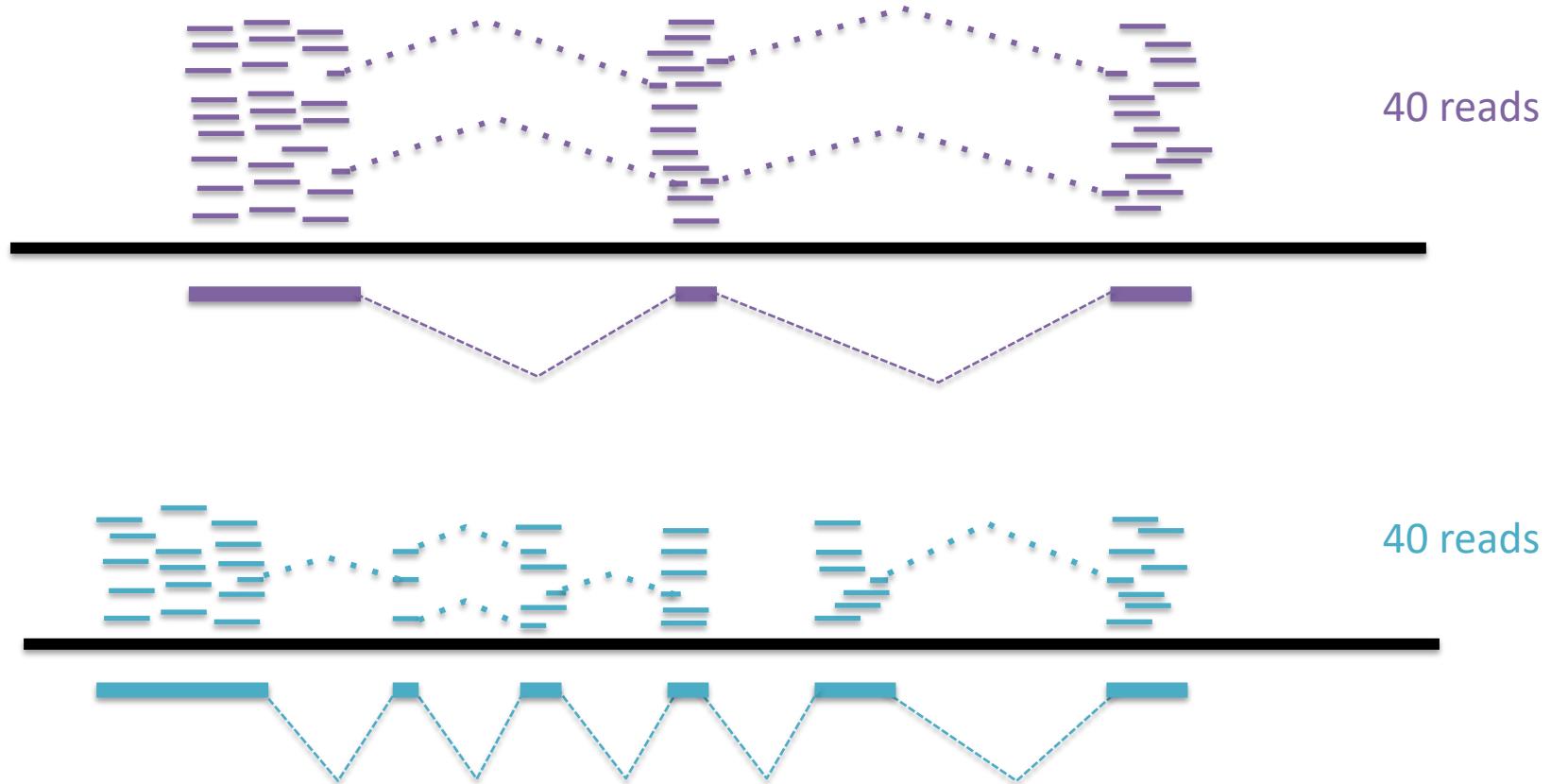
enter position, gene symbol or search terms

go

hg38 replaces hg19 as default human assembly

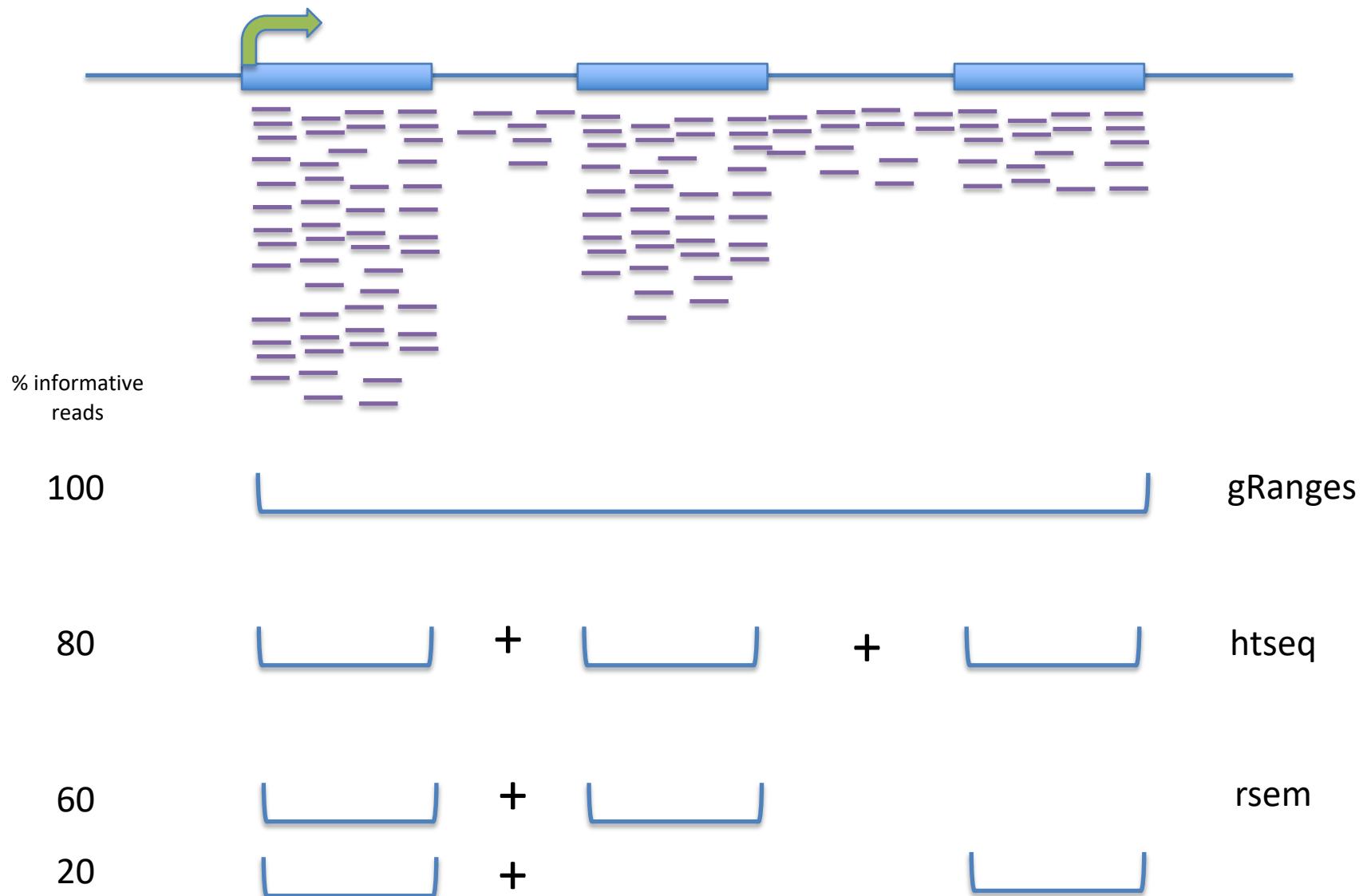


Assign Expression Values to Transcripts



Same RPM / CPM values, different RPKMs

Which reads are informative about a gene?



Counts Table

	seqnames	start	end	width	strand
ENSMUSG000000000001	chr3	108107280	108146146	38867	-
ENSMUSG000000000003	chrX	77837901	77853623	15723	-
ENSMUSG000000000028	chr16	18780447	18811987	31541	-
ENSMUSG000000000031	chr7	142575529	142578143	2615	-
ENSMUSG000000000037	chrX	161117193	161258213	141021	+
ENSMUSG000000000049	chr11	108343354	108414396	71043	+
ENSMUSG000000000056	chr11	121237253	121255856	18604	+
ENSMUSG000000000058	chr6	17281185	17289115	7931	+
ENSMUSG000000000078	chr13	5861489	5870393	8905	+

	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
ENSMUSG000000000001	3313	3020	2834	3270	813	2226
ENSMUSG000000000003	0	0	0	0	0	0
ENSMUSG000000000028	2624	2280	1926	2289	525	1921
ENSMUSG000000000031	3423	3095	2513	4136	917	2215
ENSMUSG000000000037	87	71	55	111	39	125
ENSMUSG000000000049	1	0	0	2	0	0
ENSMUSG000000000056	535	386	386	397	85	268
ENSMUSG000000000058	0	0	0	0	0	0
ENSMUSG000000000078	1240	1138	947	1045	220	701

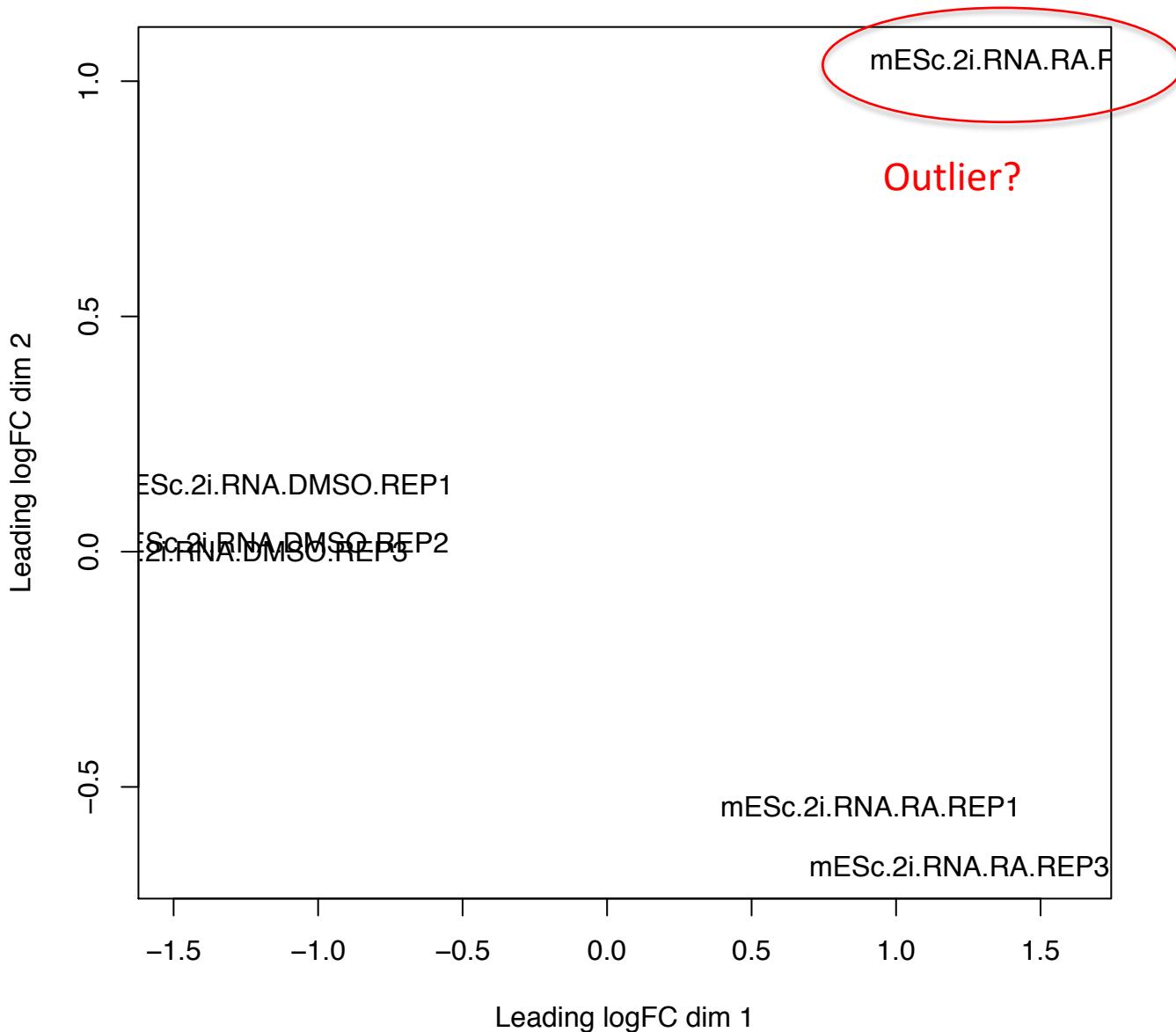
From Counts to RPM

	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
ENSMUSG000000000001	3313	3020	2834	3270	813	2226
ENSMUSG000000000003	0	0	0	0	0	0
ENSMUSG000000000028	2624	2280	1926	2289	525	1921
ENSMUSG000000000031	3423	3095	2513	4136	917	2215
ENSMUSG000000000037	87	71	55	111	39	125
ENSMUSG000000000049	1	0	0	2	0	0
ENSMUSG000000000056	535	386	386	397	85	268
ENSMUSG000000000058	0	0	0	0	0	0
ENSMUSG000000000078	1240	1138	947	1045	220	701

mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
55.2	53.0	45.4	56.2	48.2	49.3

	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
ENSMUSG000000000001	87.1	83.6	94.8	83.8	83.2	68.0
ENSMUSG000000000003	0.0	0.0	0.0	0.0	0.0	0.0
ENSMUSG000000000028	69.0	63.1	64.4	58.6	53.7	58.7
ENSMUSG000000000031	90.0	85.7	84.0	105.9	93.8	67.7
ENSMUSG000000000037	2.3	2.0	1.8	2.8	4.0	3.8
ENSMUSG000000000049	0.0	0.0	0.0	0.1	0.0	0.0
ENSMUSG000000000056	14.1	10.7	12.9	10.2	8.7	8.2
ENSMUSG000000000058	0.0	0.0	0.0	0.0	0.0	0.0
ENSMUSG000000000078	32.6	31.5	31.7	26.8	22.5	21.4

MDS Plot

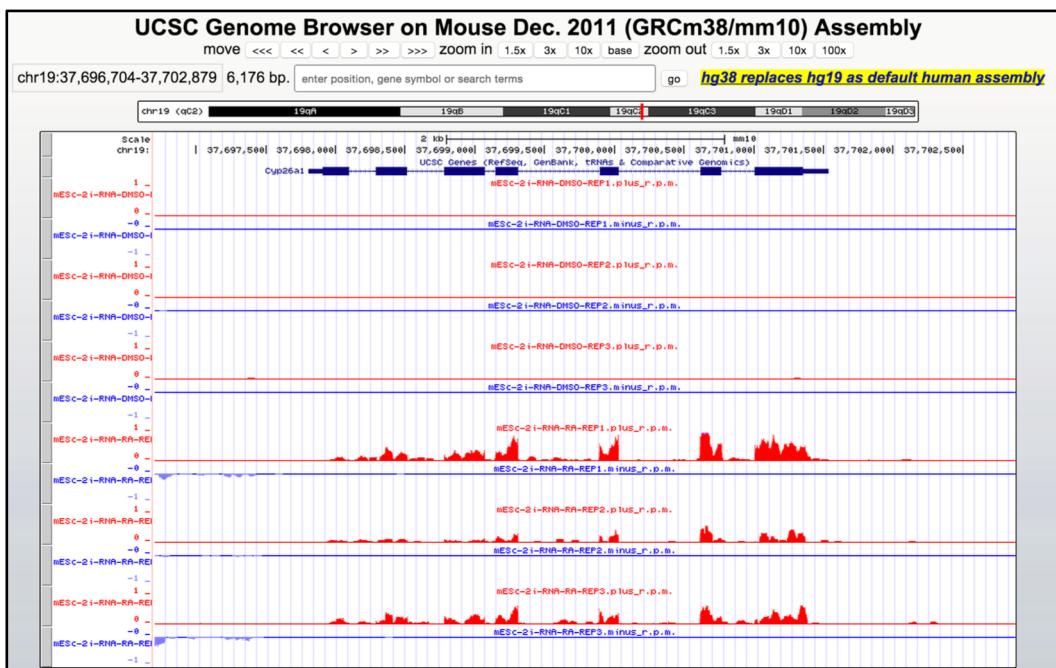


Differentially Expressed?

Comparisons Table

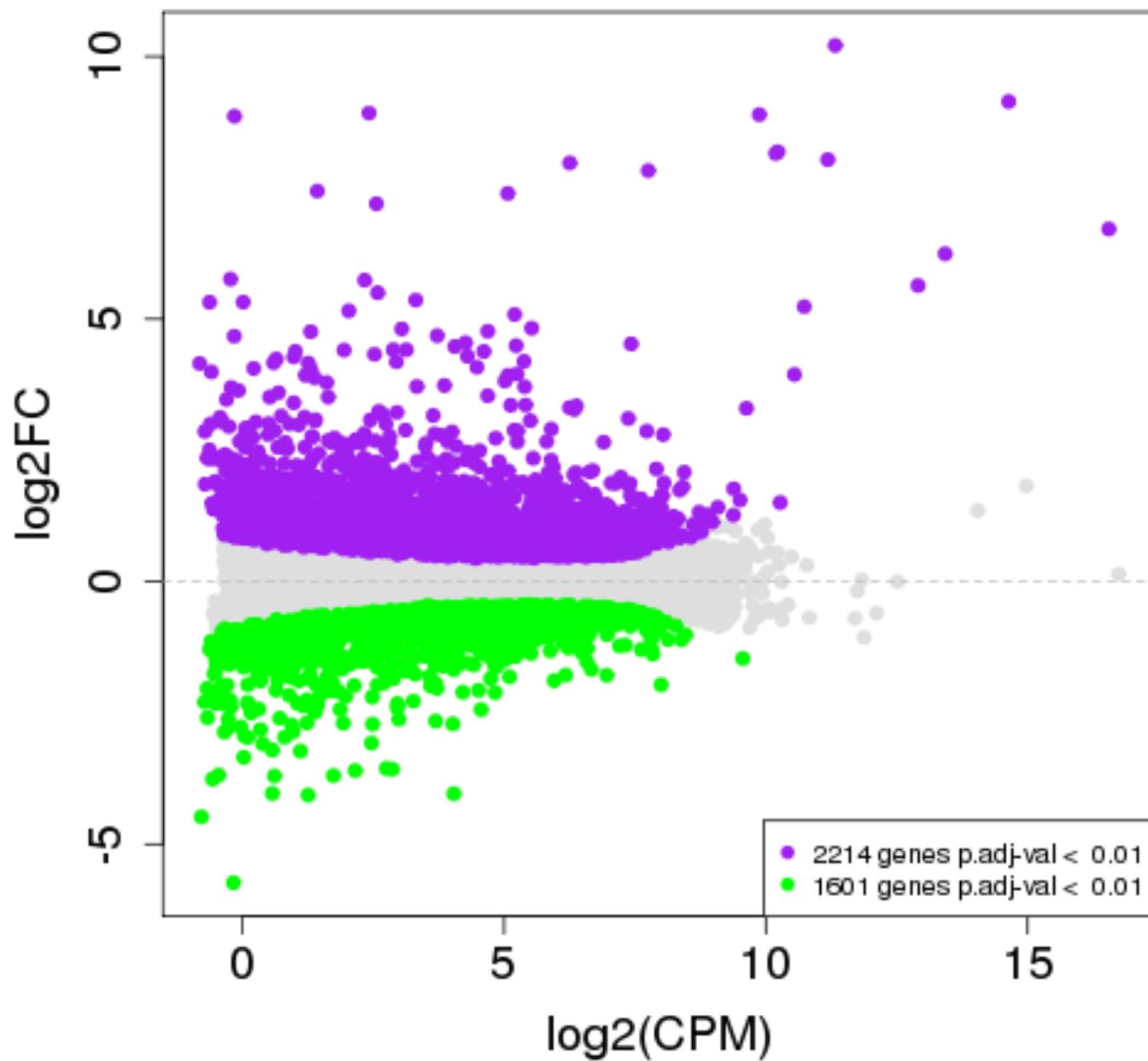
	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
RAvDMSO.1	1	1	1	-1	-1	-1
RAvDMSO.2	1	1	1	-1	0	-1

	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3	logFC	adj.p	gene
ENSMUSG00000029848	0.157184354	0.129567183	0.102811155	4.497875494	1.868627846	4.26999727	5.087518729	1.13E-149	Stra8
ENSMUSG00000024427	0.278930384	0.270539891	0.207311991	5.980218971	2.451647987	5.19150639	4.493456729	5.66E-132	Spry4
ENSMUSG00000091345	0.026851538	0.028999579	0.020826718	0.489887732	0.257296243	0.494874624	4.37275266	3.10E-112	Col6a5
ENSMUSG00000038916	0.084056755	0.095565343	0.09748072	1.441273204	0.748854089	1.798723213	4.191914864	1.38E-107	Soga3
ENSMUSG00000024987	0.008268323	0	0.010192275	3.87109811	1.666948932	2.653880244	8.921873434	9.54E-102	Cyp26a1

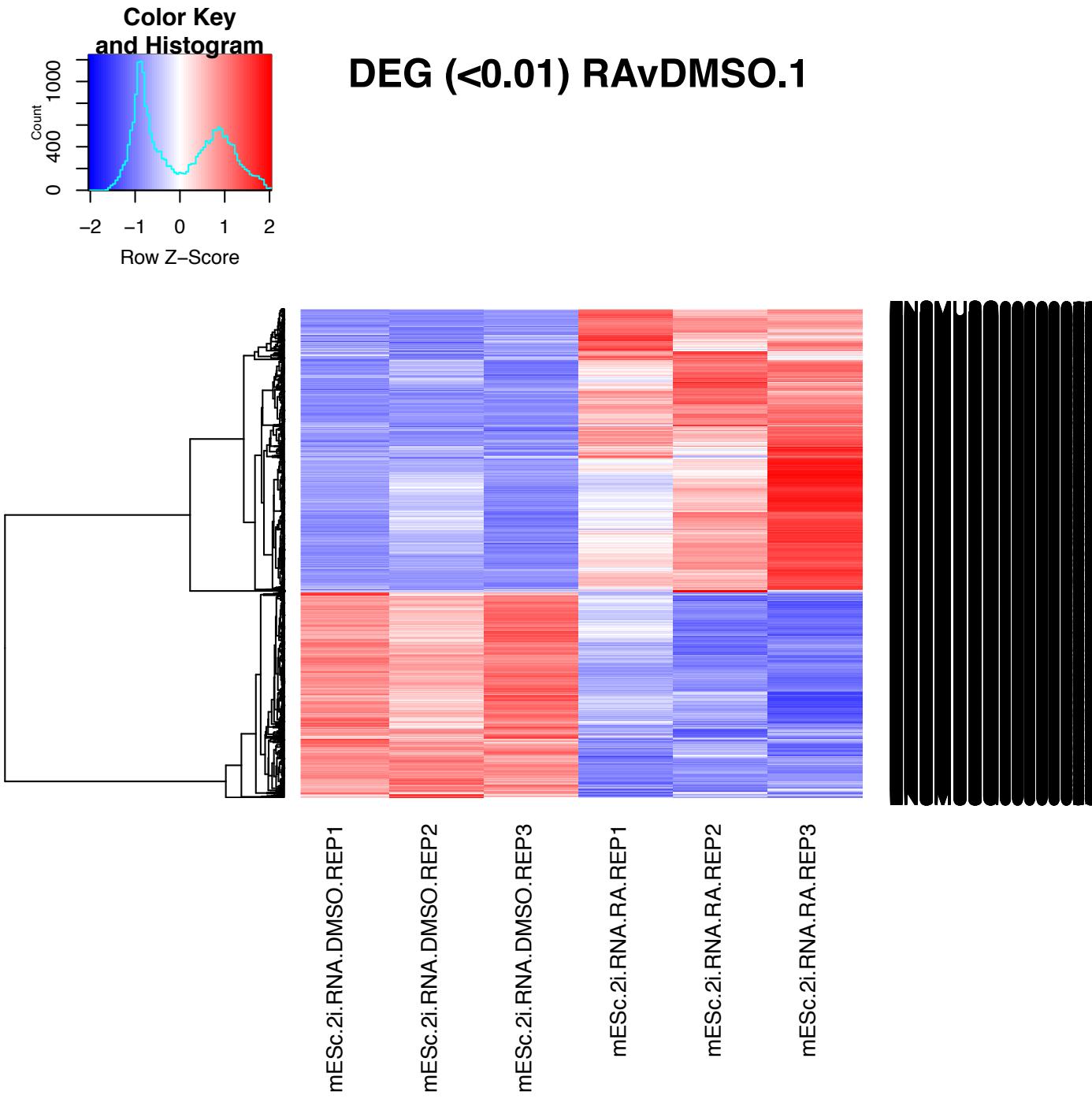


MA plot

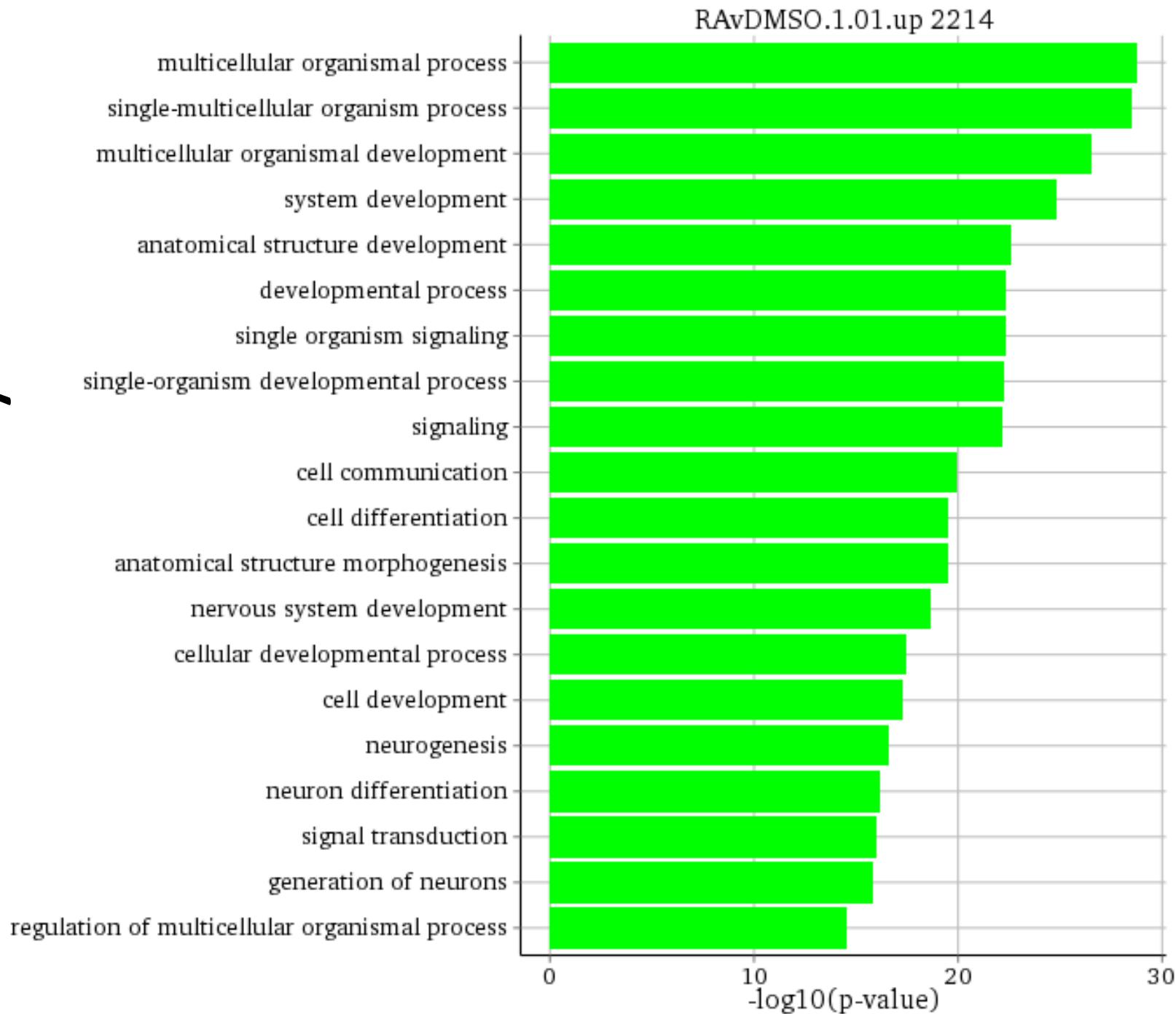
RAvDMSO.1



Expression Heatmap



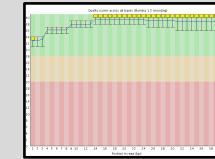
GO Analysis



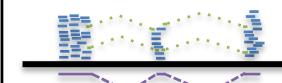


RNA-seq

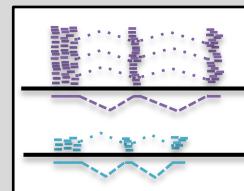
Check Sequence Quality



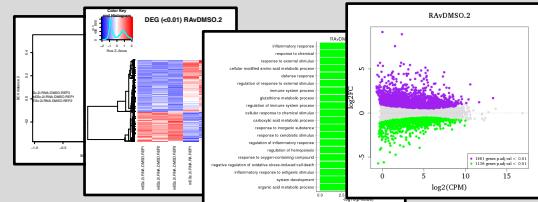
Align to Reference Genome



Estimate Gene Expression



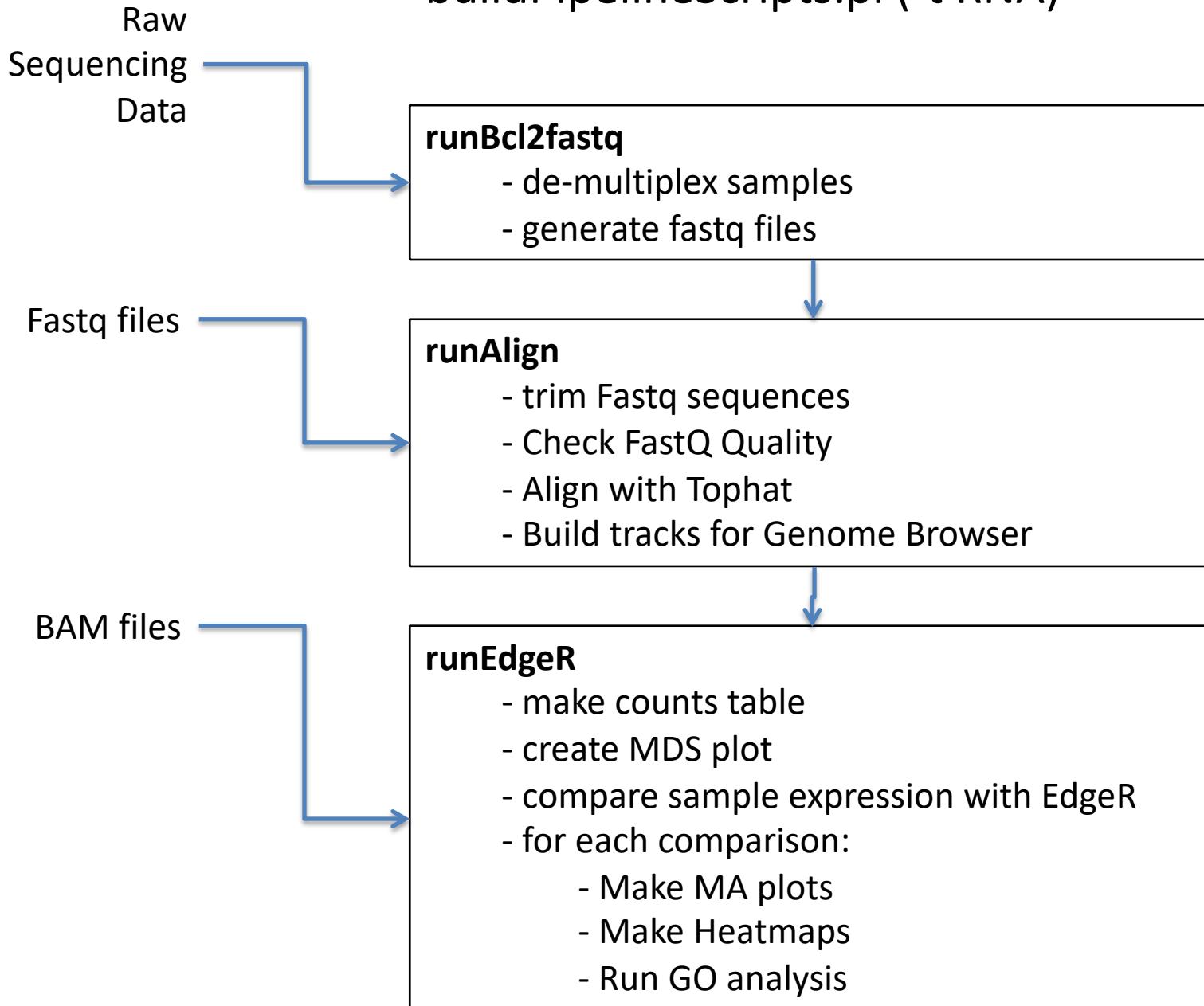
Find Differentially
Expressed Genes and
Make Plots



Outline

- What is Ceto?
- **Overview of RNAseq Analysis**
- RNA-seq Pipeline Details
- Hands-on with RNAseq

buildPipelineScripts.pl (-t RNA)



Default RNAseq Assumptions with Ceto

- Single end data *
- Stranded, reads are reverse complement of transcribed sequence *
- Exonic reads are where it's at; ignore intronic *
- Pairwise comparisons for differentially expr. genes
- No normalization for gene length *
(assume gene length is constant across comparisons)
- Isoforms are pooled to give a single genic expression level *

* non-default alternatives available

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNAAstats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from
Base Space Directory & Sample Sheet

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from
fastq Directory

Must specify assembly!



Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from
bam Directory

Must specify assembly!



Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from
fastq Directory

Must specify assembly!

Choose steps in the analysis

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from fastq Directory

Must specify assembly!

Choose steps in the analysis

Change parameters from defaults if desired

Usage: /projects/p20742/tools/buildPipelineScripts.pl

-o	<outputDirectory>
-bs	<baseSpaceDirectory>
-f	<fastqDirectory>
-bam	<bamDirectory>
-c	<comparisons.csv file>
-chip	<ChIP Description file>
-p	<numProcessors>
-g	<assembly/genome>
-t	<RNA chipseq 4C>
-a	<star tophat bwa bowtie>
-walltime	<hh:mm:ss>
-account	<accountName>
-scientist	<initials>
-buildBcl2fq	<1 0>
-runBcl2fq	<1 0>
-runTrim	<1 0>
-buildAlign	<1 0>
-runAlign	<1 0>
-makeTracks	<1 0>
-uploadASHtracks	<1 0>
-buildEdgeR	<1 0>
-runEdgeR	<1 0>
-runRNASTats	<1 0>
-rsem	<1 0>
-htseq	<1 0>
-granges	<1 0>
-stranded	<1 0>
-runPairedEnd	<1 0>

Essential Arguments

Determine samples from fastq Directory

Must specify assembly!

Choose steps in the analysis

RNA defaults: trimming on, Tophat aligner, htseq exon gene quantification, stranded on, runPairedEnd off.

```

/projects/p20742/tools/buildPipelineScripts.pl \
-t RNA \
-o /projects/w10001/${USER} \
-account w10001 \
-g mm9 \
-f /projects/w10001/testRNA.fastq \
-c /projects/w10001/testRNA.fastq/comparisons.csv \
-uploadASHtracks 0 \
-runAlign 1 \
-runEdgeR 1 \
>& buildPipelineScripts.rna.${USER}.log &

```

Content of Comparisons File:

	mESc.DMSO.REP1	mESc.DMSO.REP2	mESc.DMSO.REP3	mESc.RA.REP1	mESc.RA.REP2	mESc.RA.REP3
RAvDMSO.1	1	1	1	-1	-1	-1
RAvDMSO.2	1	1	1	-1	0	-1

Outline

- What is Ceto?
- Overview of RNAseq Analysis
- RNA-seq Pipeline Details
- **Hands-on with RNAseq**

To set up a new quest account for the pipeline:

```
module load R/3.2.2
R
source("http://bioconductor.org/biocLite.R")
biocLite("topGO")
biocLite("BSgenome.Mmusculus.UCSC.mm10")
biocLite("BSgenome.Mmusculus.UCSC.mm9")
biocLite("BSgenome.Hsapiens.UCSC.hg19")
biocLite("BSgenome.Scerevisiae.UCSC.sacCer3")
biocLite("BSgenome.Rnorvegicus.UCSC.rn6")
q()
```

```
# Define some variables for the locations of the input / output
fastqDirectory=/projects/p20742/testRNA/fastq/
outputDirectory=/projects/w10001/${USER}
comparisonsFile=/projects/p20742/testRNA/comparisons.csv

# Make a directory to write your analysis to
mkdir $outputDirectory

# Change directory (cd) so that you are in the output directory.
cd $outputDirectory

# Launch the RNAseq pipeline
/projects/p20742/tools/buildPipelineScripts.pl \
-t RNA \
-o $outputDirectory \
-g mm9 \
-account w10001 \
-walltime 04:00:00 \
-f $fastqDirectory \
-c $comparisonsFile \
-uploadASHtracks 0 \
-runAlign 1 \
-runEdgeR 1 \
-runRNASTATS 0 \
-a star \
-stranded 1 \
-runPairedEnd 0 \
-rsem 0 \
-htseq 1 \
-granges 0 \
>& $outputDirectory/buildPipelineScripts.rna.${USER}.BDSD.log &

# Return to /projects/w10001/ to see the results.
cd /projects/w10001/
```

Interactive Experimental Design

R Shiny application for
metadata collection

Kirsten Bell Burdett, MB

Department of Preventive Medicine, Division of Biostatistics
Northwestern University
Kirsten.burdett@northwestern.edu

What is R Shiny?

- R package
- Build interactive web applications
- Interactive approach for data:
 - Exploration
 - Collection
 - Visualization
 - Reporting

<https://shiny.rstudio.com/tutorial/>

Northwestern IT Research Computing Services – Christina Maimone

Shiny App Structure

- Experimental design for RNA-seq
 - Extend for various high-throughput data
- Replicate Type
- Sample Quality
- Adjust for Multiple Predictors
 - Allows for expanding beyond pairwise comparisons

Run on Quest Analytics

- Quest Analytics Node
 - <https://rstudio.questanalytics.northwestern.edu/auth-sign-in>
- Enter Credentials (Login with NU NetID)
- In RStudio, verify necessary packages are installed
 - shiny, tidyverse, shinyjs, DT
- Run the following commands

```
> dir <- "/projects/p20742/tools/bin/CETO_RNAseq_DESeq2_app_KB/"  
  
> shiny::runApp(file.path(dir, "app.R"))
```



RNA-seq: Study setup for analysis

This application is used to specify details of the RNA-seq experiment. This will be used in order to determine the input data structure that will be used in CETO.

Input Directory

Output Directory

Experimental Design:

Choose Replicate Type

Technical replicates are defined as libraries derived from the same samples.

Replicate Type

- Technical
- Biological

Samples In Comparison Output

 ▾

Predictors

Only include predictors of interest that will be used for further analysis.

Number of Predictors



Predictor Details

For each predictor, specify the variable name and class.

Predictor 1

Name

Class

 ▾

Design Complete

- **Input Directory**
 - /projects/p20742/testRNA/fastq/
- **Output Directory**
 - /projects/w10001/\${USER}
- **Replicate type**
 - Biological
- **Samples in Comparison Output**
 - All Samples
- **Predictors**
 - Number of Predictors: 1
 - Name: Treatment
 - Class: Categorical
 - Reference Level: DMSO



RNA-seq: Study setup for analysis

This application is used to specify details of the RNA-seq experiment. This will be used in order to determine the input data structure that will be used in CETO.

Input Directory

/projects/p20742/testRNA/fastq/

Output Directory

/projects/w10001/tempuser03

Input and Output Directories

Experimental Design:

Choose Replicate Type

Technical replicates are defined as libraries derived from the same samples.

Replicate Type

- Technical
- Biological

Predictors

Only include predictors of interest that will be used for further analysis.

Number of Predictors



Predictor Details

For each predictor, specify the variable name and class.

Predictor 1

Name	Class	Reference Level
Treatment	Categorical	DMSO

Samples In Comparison Output

High

- High
- Average-and-above
- All Samples

Design Complete

rstudio.questanalytics.northwestern.edu
https://rstudio.questanalytics.northwestern.edu/p/e2d8496a/ | Open in Browser |

Publish

RNA-seq: Study setup for analysis

This application is used to specify details of the RNA-seq experiment. This will be used in order to determine the input data structure that will be used in CETO.

Input Directory

/projects/p20742/testRNA/fastq/

Output Directory

/projects/w10001/tempuser03

Experimental Design:

Choose Replicate Type

Technical replicates are defined as libraries derived from the same samples.

Replicate Type

- Technical
- Biological

Samples In Comparison Output

All Samples

Predictors

Only include predictors of interest that will be used for further analysis.

Number of Predictors



Predictor Details

For each predictor, specify the variable name and class.

Predictor 1

Name

Treatment

Class

Categorical

Reference Level

DMSO

Design Complete

Data Entry

Sample 1

- mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001
 - Quality = High
 - Treatment = DMSO

Data Entry:

Complete the below information for **each observation**. When it is filled, click the **Submit** button.

If you want to delete all the previous observations that have been entered, click the **Delete All Data** button. After all the observations have been entered, click the **Download** button to download the entire dataset.

Sample Name
mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001
Sample Quality
High
Treatment
<input type="text"/>
Submit
Clear Form
Delete All Data
Download Table
Download Comparison

Show 10 entries					Search:
	sample_name	sample_quality	Treatment	submit_time	
1	mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001	High	DMSO	Mon Feb 3 12:10:56 2020	
Showing 1 to 1 of 1 entries					

Samples

- mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001
 - High, DMSO
- mESc-2i-RNA-DMSO-REP2_S2_L001_R1_001
 - High, DMSO
- mESc-2i-RNA-DMSO-REP3_S3_L001_R1_001
 - High, DMSO
- mESc-2i-RNA-RA-REP1_S4_L001_R1_001
 - High, RA
- mESc-2i-RNA-RA-REP2_S5_L001_R1_001
 - Low, RA
- mESc-2i-RNA-RA-REP3_S6_L001_R1_001
 - High, RA

Export Data

Data Entry:

Complete the below information for **each observation**. When it is filled, click the **Submit** button.

If you want to delete all the previous observations that have been entered, click the **Delete All Data** button. After all the observations have been entered, click the **Download** button to download the entire dataset.

Sample Name		
mESc-2i-RNA-RA-REP3_S6_L001_R1_001		
Sample Quality		
High		
Treatment		
Submit	Clear Form	Delete All Data
Download Table		
Download Comparison		

Show 10 entries Search:				
	sample_name	sample_quality	Treatment	submit_time
1	mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001	High	DMSO	Mon Feb 3 14:37:54 2020
2	mESc-2i-RNA-DMSO-REP2_S2_L001_R1_001	High	DMSO	Mon Feb 3 14:38:05 2020
3	mESc-2i-RNA-DMSO-REP3_S3_L001_R1_001	High	DMSO	Mon Feb 3 14:38:12 2020
4	mESc-2i-RNA-RA-REP1_S4_L001_R1_001	High	RA	Mon Feb 3 14:38:22 2020
5	mESc-2i-RNA-RA-REP2_S5_L001_R1_001	Low	RA	Mon Feb 3 14:40:08 2020
6	mESc-2i-RNA-RA-REP3_S6_L001_R1_001	High	RA	Mon Feb 3 14:40:16 2020

Showing 1 to 6 of 6 entries

Previous 1 Next

- Download Table
- Download Comparison

All Samples

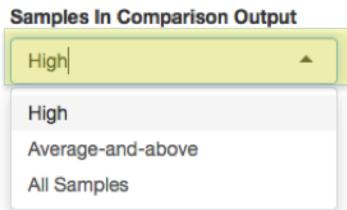
Download Table Output

sample_name	sample_quality	Treatment	submit_time
mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001	High	DMSO	Mon Feb 3 12:10:56 2020
mESc-2i-RNA-DMSO-REP2_S2_L001_R1_001	High	DMSO	Mon Feb 3 12:11:18 2020
mESc-2i-RNA-DMSO-REP3_S3_L001_R1_001	High	DMSO	Mon Feb 3 12:11:26 2020
mESc-2i-RNA-RA-REP1_S4_L001_R1_001	High	RA	Mon Feb 3 12:11:39 2020
mESc-2i-RNA-RA-REP2_S5_L001_R1_001	Low	RA	Mon Feb 3 12:12:10 2020
mESc-2i-RNA-RA-REP3_S6_L001_R1_001	High	RA	Mon Feb 3 12:12:17 2020

Download Comparison Output

	mESc-2i-RNA-DMSO-REP1	mESc-2i-RNA-DMSO-REP2	mESc-2i-RNA-DMSO-REP3	mESc-2i-RNA-RA-REP1	mESc-2i-RNA-RA-REP2	mESc-2i-RNA-RA-REP3
Treatment_RAvDMSO	1	1	1	1	-1	-1

Remove Low Quality Samples



Download Table Output

sample_name	sample_quality	Treatment	submit_time
mESc-2i-RNA-DMSO-REP1_S1_L001_R1_001	High	DMSO	Mon Feb 3 14:37:54 2020
mESc-2i-RNA-DMSO-REP2_S2_L001_R1_001	High	DMSO	Mon Feb 3 14:38:05 2020
mESc-2i-RNA-DMSO-REP3_S3_L001_R1_001	High	DMSO	Mon Feb 3 14:38:12 2020
mESc-2i-RNA-RA-REP1_S4_L001_R1_001	High	RA	Mon Feb 3 14:38:22 2020
mESc-2i-RNA-RA-REP2_S5_L001_R1_001	Low	RA	Mon Feb 3 14:40:08 2020
mESc-2i-RNA-RA-REP3_S6_L001_R1_001	High	RA	Mon Feb 3 14:40:16 2020

Download Comparison Output

	mESc-2i-RNA-DMSO-REP1	mESc-2i-RNA-DMSO-REP2	mESc-2i-RNA-DMSO-REP3	mESc-2i-RNA-RA-REP1	mESc-2i-RNA-RA-REP2	mESc-2i-RNA-RA-REP3
Treatment_RAvDMSO	1	1	1	1	-1	0

Future Work

- Extend to other high-throughput data
 - Currently RNA-seq with EdgeR and DESeq2
- Dynamic Reports
 - Interactively select CETO output
 - User friendly
 - Single report with QC, figures, tables