# RNA-seq Data Analysis on the HPRC Ada Cluster

# Your Login Password

- Both state of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts
- Don’t write down passwords
- Don’t choose easy to guess/crack passwords
- Change passwords frequently

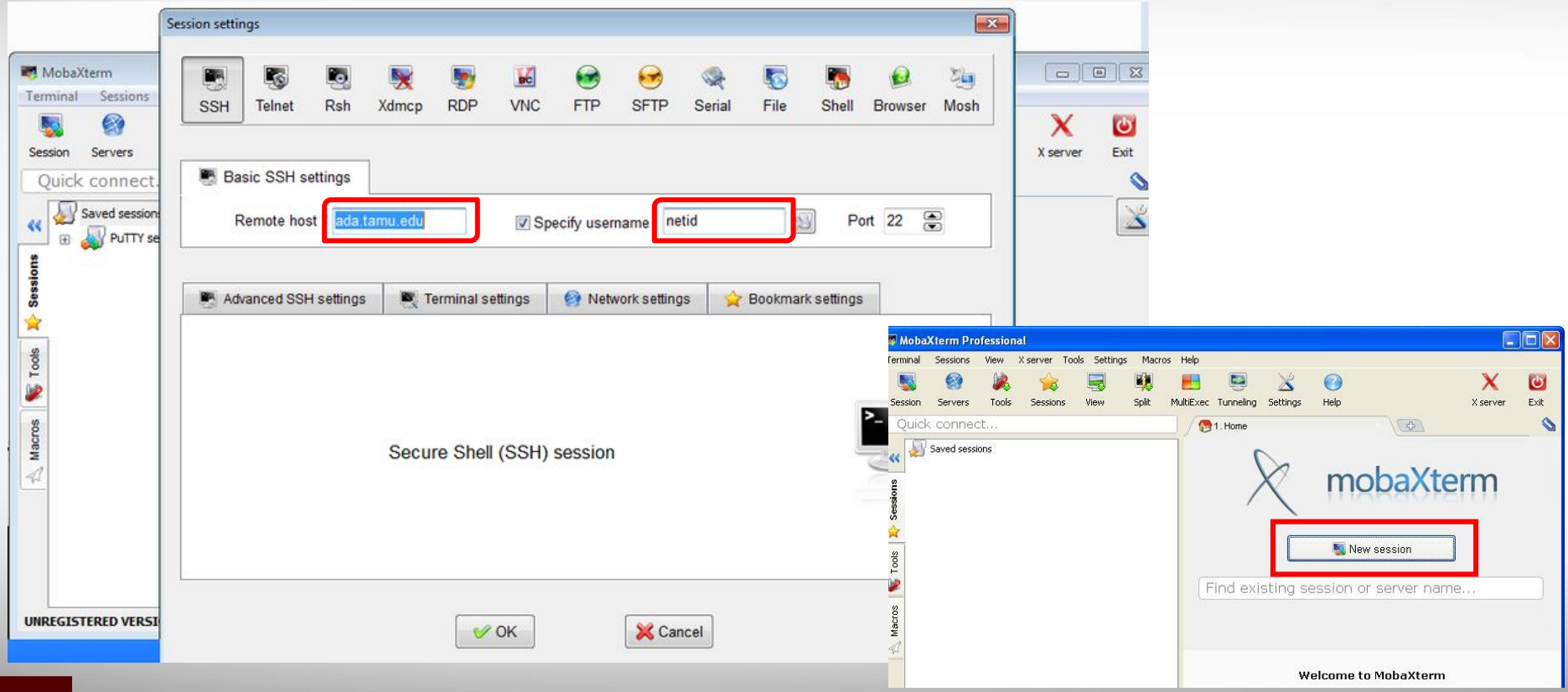
# For More Help...

| | |
|---|---|
| Website: | hprc.tamu.edu |
| Email: | help@hprc.tamu.edu |
| Telephone: | (979) 845-0219 |
| Visit us in person: | Henderson Hall, Room 114A |

**Help us, help you -- we need more info**

- Which Cluster
- UserID/NetID
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem

# Using SSH - MobaXterm (on Windows)

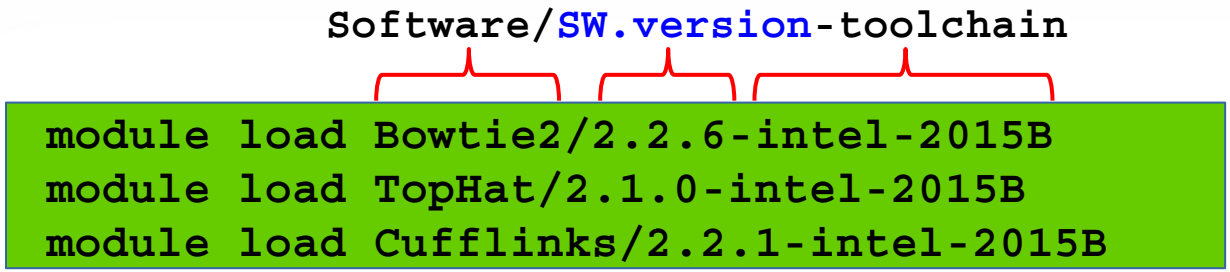


# Where to Find NGS Tools

- TAMU HPRC Documentation
  - https://hprc.tamu.edu/wiki/index.php/Ada:Bioinformatics
- Type the following UNIX **commands** to see which tools are already installed on Ada
  - `module avail`
  - `module spider toolname` (not case sensitive, but read entire output)
  - `module key assembly` (some modules may be missed because this searches tool descriptions)
- If you find a tool that you want installed on Ada, send an email with the URL link to: **help@hprc.tamu.edu**
  - SeqAnswers http://seqanswers.com/wiki/Software/list
  - omictools.com
  - slideshare.net – find shared NGS presentations

# Ada Software Toolchains

- Use the same toolchains in your job scripts




- Avoid loading mixed toolchains:

```
module load Bowtie2/2.2.2-ictce-6.3.5
module load TopHat/2.0.14-goolf-1.7.20
module load Cufflinks/2.2.1-intel-2015B
```

- Avoid loading defaults which may have different toolchains

```
module load Bowtie2 TopHat Cufflinks
```

# Use $TMPDIR whenever possible

- Use the **$TMPDIR** if the application you are running can utilize a temporary directory for writing temporary files which are deleted when the job ends
- A temp directory (**$TMPDIR**) is automatically assigned for each job which uses the disk(s) on the compute node not the **$SCRATCH** shared file system
  - Especially useful when a computational tool writes tens of thousands of temporary files which are deleted when the job is finished and are not needed for the final results
  - This is useful since files on **$TMPDIR** will not count against your file quota
  - Be aware when using **$TMPDIR** if your software uses temporary files for restarting where it left off if it should stop before completion
  - Will significantly speed up an mpiBLAST job

```
run_BUSCO.py --in Trinity.fasta -m transcriptome --tmp_path $TMPDIR /
-l /scratch/datasets/BUSCO/v3.0.2/fungi_odb9 -c 20 --evalue 0.001
```

# Finding NGS job template scripts using GCATemplates on Ada

```
mkdir $SCRATCH/rnaseq_class
cd $SCRATCH/rnaseq_class
module load GCATemplates
gcatemplates
```

For practice, we will copy a template file

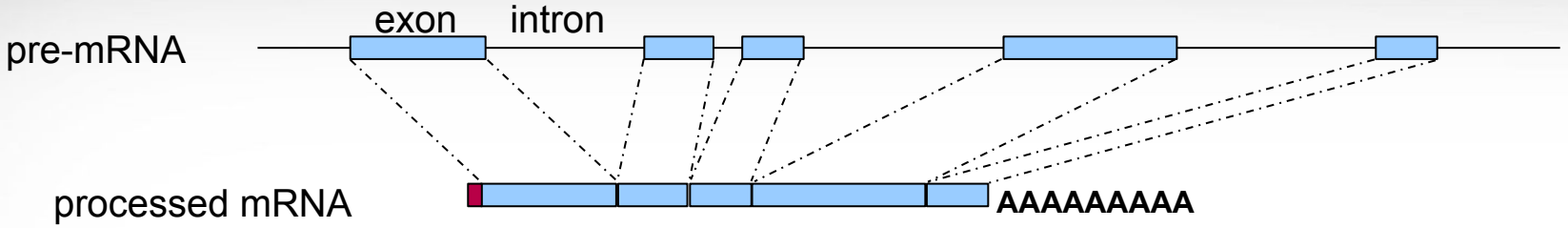
- Select #13 RNA-seq, #1 QC, #1 rnaseqc, #1 two samples
- Final step will save a template job script file to your current working directory
- After you save the template file:

```
module purge
```

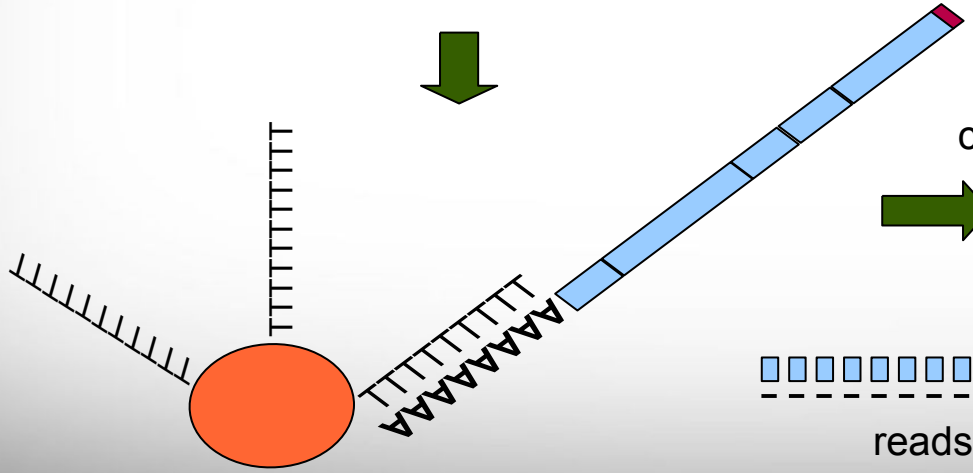
Genomic Computational Analysis Templates

# RNA-seq Overview

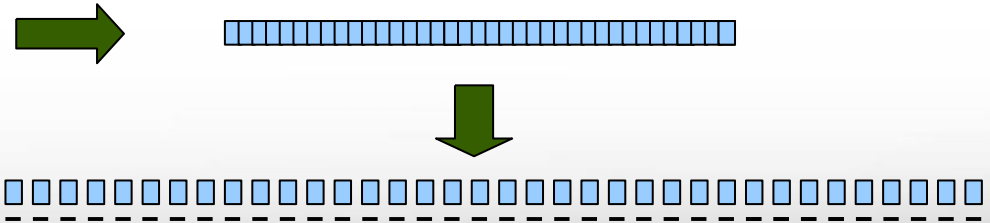
# Example of RNA sequencing




mRNA strands are captured by their Poly(A) tail using Poly(T) coated magnetic beads




cDNA exon region is fragmented and sequenced



reads are then mapped to a reference genome or reference sequences or assembled

# RNA-seq Applications

- Differential Expression (DE) and transcript abundance
  - HISAT2, Bowtie, TopHat, Cufflinks, Cuffmerge, Cuffdiff
  - DESeq and DESeq2 (R package)
  - EdgeR  (R package)
- Transcriptome assembly (find isoforms and rare transcripts)
  - *de novo* (Trinity, Oases, SOAPdenovo-Trans)
  - reference based (Trinity, StringTie)
- Genome Annotation
  - Align to assembly for validation of gene models
- Variant Calling
  - STAR/Picard/GATK (Haplotype Caller (HC) in RNA-seq mode)
- *de novo* genome assembly scaffolding
  - L_RNA_scaffolder
- Identify fusion transcripts
  - tophat-fusion

# Illumina Sequencing Technology

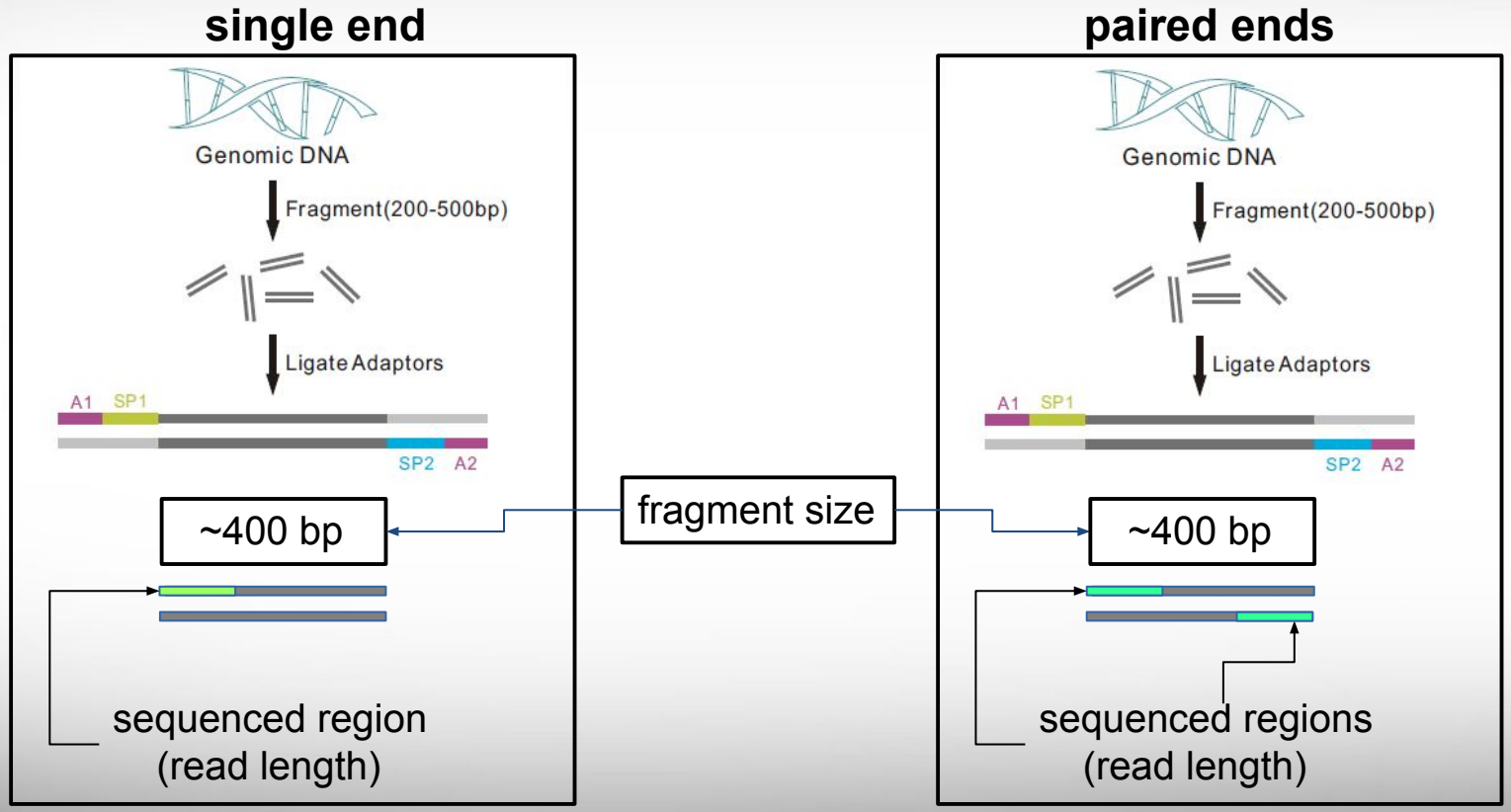
RNA-seq applications

| | MiniSeq System | MiSeq Series | NextSeq Series | HiSeq Series | HiSeq X Series* | NovaSeq 5000 |
|---|---|---|---|---|---|---|
| **Key Methods** | Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing. | Small genome, amplicon, and targeted gene panel sequencing. | Everyday exome, transcriptome, and targeted resequencing. | Production-scale genome, exome, transcriptome sequencing, and more. | Population- and production-scale whole-genome sequencing. | Same as HiSeq |
| **Maximum Output** | 7.5 Gb | 15 Gb | 120 Gb | 1500 Gb | 1800 Gb | 2000 Gb |
| **Maximum Reads per Run** | 25 million | 25 million† | 400 million | 5 billion | 6 billion | 6.6 billion |
| **Maximum Read Length** | 2 × 150 bp | 2 × 300 bp | 2 × 150 bp | 2 × 150 bp | 2 × 150 bp | 2 x 150 bp |
| **Run Time** | 4–24 hours | 4–55 hours | 12–30 hours | <1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500) | <3 days | 19 - 40 hrs |
| **Benchtop Sequencer** | Yes | Yes | Yes | No | No | no |

http://www.illumina.com/systems/sequencing-platforms.html (Oct 2017)

# Illumina Sequencing Libraries

illumina.com



# Biological vs Technical Replication

- Biological replicates include multiple samplings within a population
- Technical replicates include multiple prepping and or resequencing the same individual
- Biological replicates generally increase statistical power more than technical replicates
  - Biological variability is generally greater than technical variability
  - Biological replicates contain both biological and technical variability

# Sequence Depth for RNA-seq Differential Expression

**RNA-seq differential expression studies: more sequence or more replication?**
Liu, Yuwen, Zhou, Jie and White, Kevin P. Bioinformatics. 2014 Feb 1; 30(3): 301–304.
doi: 10.1093/bioinformatics/btt688 PMCID: PMC3904521

- Using more biological replicates instead of increasing sequencing depth resulted in improved accuracy of expression estimation
- Use more biological replicates at lower sequencing depth is more beneficial than fewer samples at a higher sequencing depth
- Increasing sequence depth is beneficial for exon or transcript-specific expression studies

# Quality Control (QC)

# QC Evaluation

```
module spider fastqc
```

- Use FastQC to visualize quality scores
  - Displays quality score distribution of reads
    - Input is a fastq file or files
    - Can disable grouping of sequence regions
  - Will alert you of poor read characteristics
  - Displays a representative sample of the fastq file
  - Can be run as a GUI or a command line interface
- FastQC will process using one CPU core per file
  - If there are 10 fastq files to analyze and 4 cores used
    - 4 files will start processing and 6 will wait in a queue
    - If there is only one fastq file to process then using 10 cores does not speed up the process

# QC Quality Trimming

- Sequence quality trimming tools

```
module spider Trimmomatic
```

recommended tool

- Trimmomatic will maintain paired end read pairing after trimming
- Trim reads based on quality scores
  - Trim the same number of bases from each read or
  - Use a sliding window to calculate average quality at ends of sequences
- Decide if you want to discard reads with Ns
  - some assemblers replace Ns with As or a random base G, C, A or T
- Trim adapter sequences
  - Trimmomatic has a file of Illumina adapter sequences

```
module load Trimmomatic/0.36-Java-1.8.0_92
```

```
ls $EBROOTTRIMMOMATIC/adapters/
```

# RNA-SeQC

```
module spider RNA-SeQC
```

- Provides alignment metrics & graphs all samples together
  - Yield alignment and duplication rates
  - GC bias
  - rRNA content
  - Regions of alignment (exon, intron, intragenic)
  - continuity of coverage
  - 5'/3' bias and much more ...
- Metrics can help identify sample outliers by comparing metrics of all samples

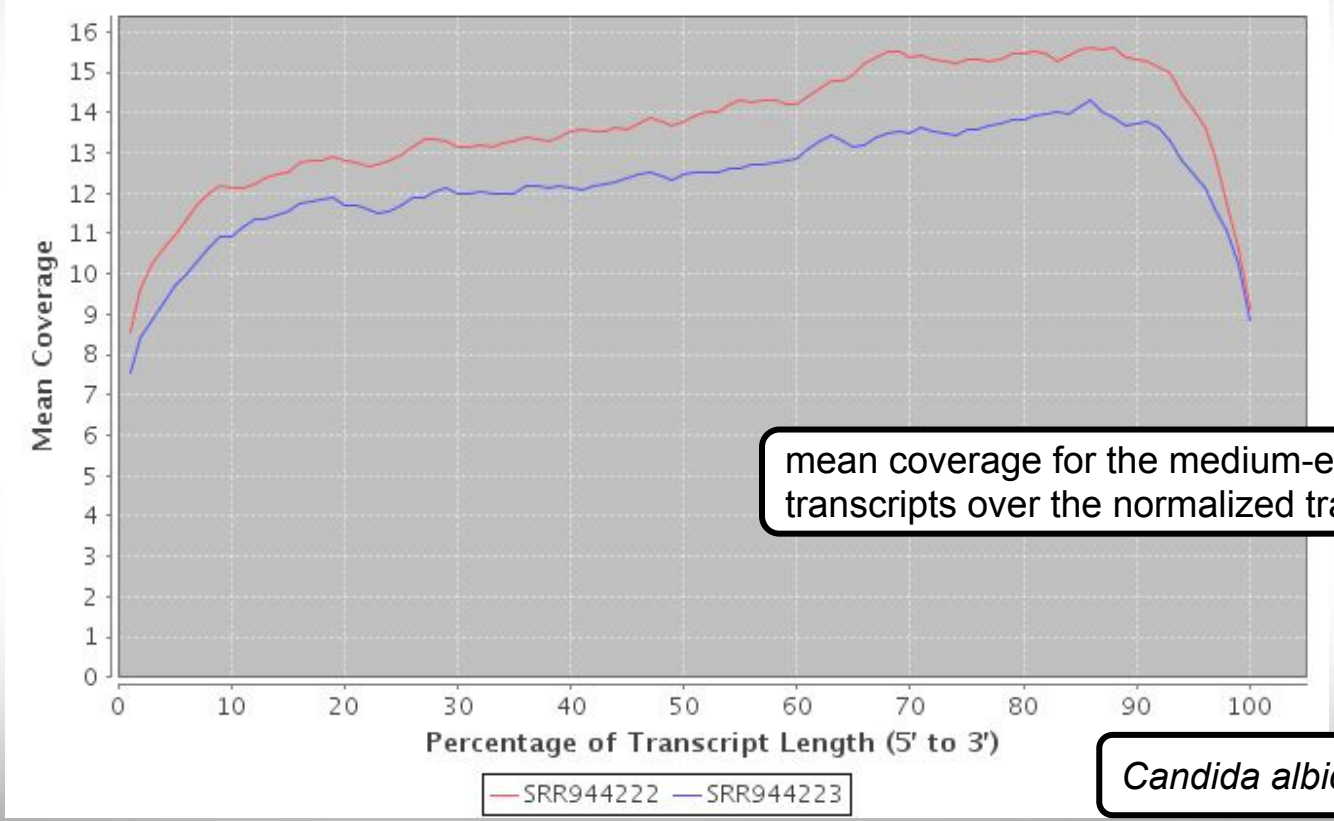
**RNA-SeQC: RNA-seq metrics for quality control and process optimization**

DeLuca, et al. Bioinformatics. 2012 Jun 1; 28(11): 1530–1532. Published online 2012 Apr 25. doi: 10.1093/bioinformatics/bts196PMCID: PMC3356847

# Example of Some of the Metrics Generated by RNA-SeQC

| Sample | SRR944222 | SRR944223 |
|---|---|---|
| Intragenic Rate | 0.94193894 | 0.942244 |
| Num. Gaps | 261 | 267 |
| Exonic Rate | 0.88744026 | 0.88684684 |
| Mapping Rate | 0.9407622 | 0.9328903 |
| Mean Per Base Cov. | 13.762129 | 12.427471 |
| Mapped | 9991016 | 10685055 |
| Intergenic Rate | 0.058061063 | 0.057755996 |
| Mean CV | 0.4337087 | 0.43792304 |
| Transcripts Detected | 6097 | 6028 |
| Cumul. Gap Length | 10387 | 10073 |
| Gap % | 0.0061764293 | 0.0061276583 |
| Unpaired Reads | 10620129 | 11453710 |
| Intronic Rate | 0.05449866 | 0.055397186 |
| Mapped Unique Rate of Total | 0.9407622 | 0.9328903 |
| Expression Profiling Efficiency | 0.8348703 | 0.82733077 |
| Mapped Unique | 9991016 | 10685055 |
| Base Mismatch Rate | 0.0041912985 | 0.0042645987 |

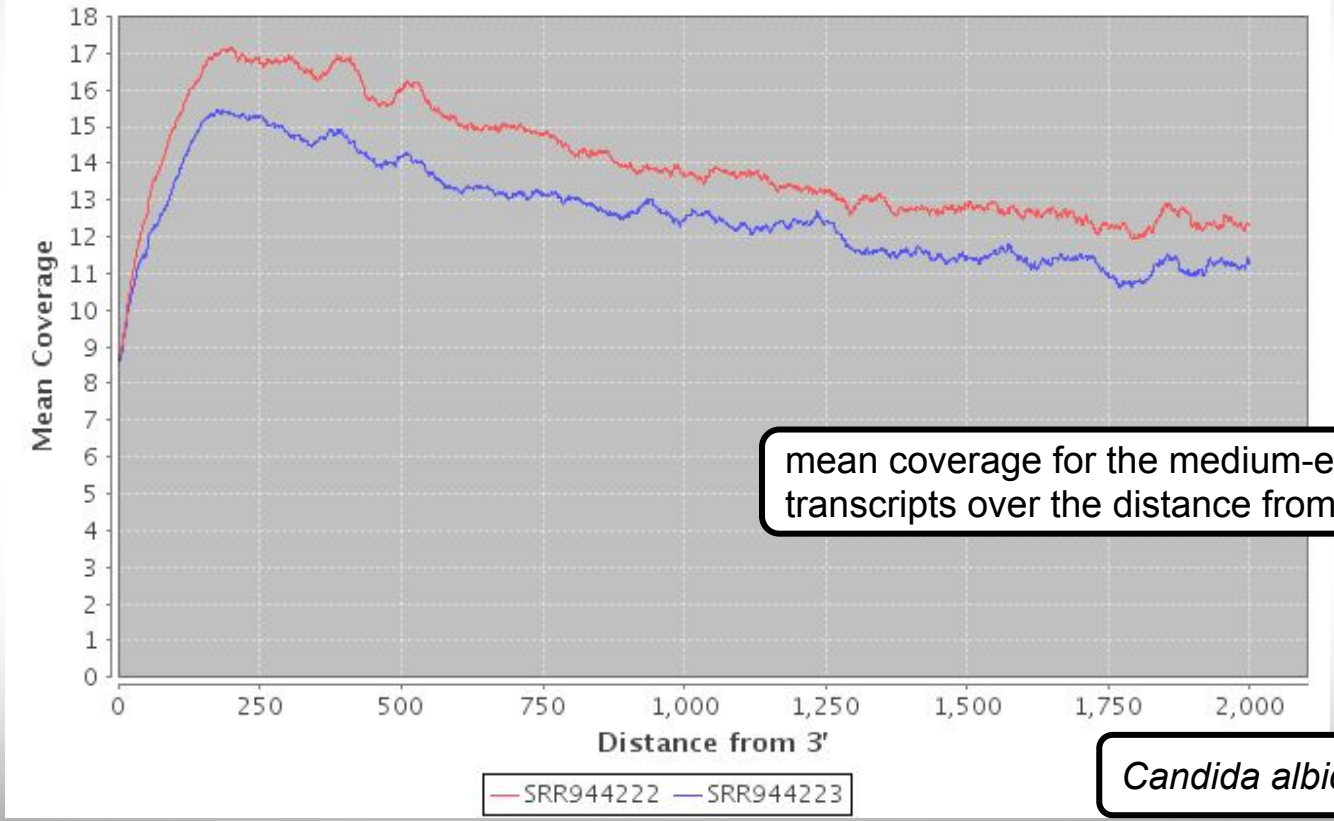
# RNA-SeQC meanCoverageNorm_medium




mean coverage for the medium-expressed transcripts over the normalized transcript length

*Candida albicans* RNA-seq

# RNA-SeQC meanCoverage_medium




mean coverage for the medium-expressed transcripts over the distance from the 3' end

*Candida albicans* RNA-seq

# Mapping RNA-seq Reads to a Reference Assembly

# Splice-Aware Aligners for RNA-seq Short Reads

- HISAT2 which supersedes TopHat2
  - `/scratch/datasets/genome_indexes/ucsc/mm10/hisat2_index/`
- STAR (on Ada as module STAR-STAR)
  - Uses gene annotations in gtf format
    - can use **`gffread`** in Cufflinks module to convert gff3 to gtf
  - supports PacBio but should use non-default settings
    - *Bioinfx study: Optimizing STAR aligner for Iso Seq data*
- BBMap
  - also supports PacBio and Nanopore
- GMap
  - also supports PacBio and Nanopore

# Visualize BAM File Alignments

# Sorting, Viewing sam/bam Files

- Sequence Alignment/Map format (sam)
  - view sam files using the UNIX command:
    ```
    more file.sam
    ```
- Binary Alignment/Map format (bam)
  - Compressed (binary) sam files need samtools to view
    - ```
      module load SAMtools/1.3-intel-2015B
      ```
  - Sort sam/bam file based on coordinate into bam format (10 cores, 2GB mem/core)
  - ```
    samtools sort -@ 10 -m 2G -o file_sorted.bam file.sam
    ```
  - Create an index of the bam file using samtools
    - ```
      samtools index file_sorted.bam
      ```
    - A samtools index is needed prior to viewing alignments in viewers
  - Viewing bam files using samtools (index not required)
    - ```
      samtools view file_sorted.bam | more
      ```
      view only alignments
    - ```
      samtools view -H file_sorted.bam
      ```
      view only header
    - ```
      samtools view -h file_sorted.bam | more
      ```
      view header + alignments

# SAM format

```
samtools view -h file_sorted.bam | more
```

header

alignments

```
@HD	VN:1.0	SO:coordinate
@PG	ID:GMAP PN:gmap VN:2015-09-21	CL:gmap -t 18 -D genome_dir -d ASM678v2 -f
samse --read-group-id=RG1 --read-group-name=ASM678v2 --re
ad-group-library=SRR4289711 --read-group-platform=ILLUMINA
@SQ	SN:AE004092	LN:1852433
@RG	ID:rg1	PL:ILLUMINA	LB:SRR4289711	SM:ASM678v2
1/1	0	AE004092	1	40	4S41M	*	0	0
TAGCTTGTTGATATTCTGTTTTTTCTTTTTTAGTTTTCCACATGA	FEHHHHHHHHIHH
IJJJJJJJJJJJHIIJJJJJJIJJJJJJJJJIJ	RG:Z:rg1	MD:Z:41	NH:i:1	HI:i:1	NM:i:0	SM:i:40
XQ:i:40	X2:i:0	XO:Z:UU	XG:Z:M
2/1	0	AE004092	36	40	45M	*	0	0
ACATGAAAAATAGTTGAAAACAATAGCGGTGTCCCCTTAAAATGG	FFHHHHHJJJJJ
JJJJJJJJJJJJJJIJJHIJJJJJJJJJJJJJI	RG:Z:rg1	MD:Z:45	NH:i:1	HI:i:1	NM:i:0	SM:i:40
XQ:i:40	X2:i:0	XO:Z:UU	XG:Z:M
3/1	0	AE004092	100	40	45M	*	0	0
GAACCCAAATTAACAGTGTTAATTTATTTTCCACAGGTTGTGGAA	DFGFHDFHGIJI
JGHHFHEHGFIIJIIGIGHGIJEGI?DGFDG@H	RG:Z:rg1	MD:Z:45	NH:i:1	HI:i:1
NM:i:0	SM:i:40	XQ:i:40	X2:i:0	XO:Z:UU	XG:Z:M
```

# Integrative Genomics Viewer (IGV) Exercise

- IGV is a genome browser with pre-loaded genomes available in which you can use to view multiple .bed, .sam and .vcf files.
- IGV is launched from a login node not a job script or compute node.

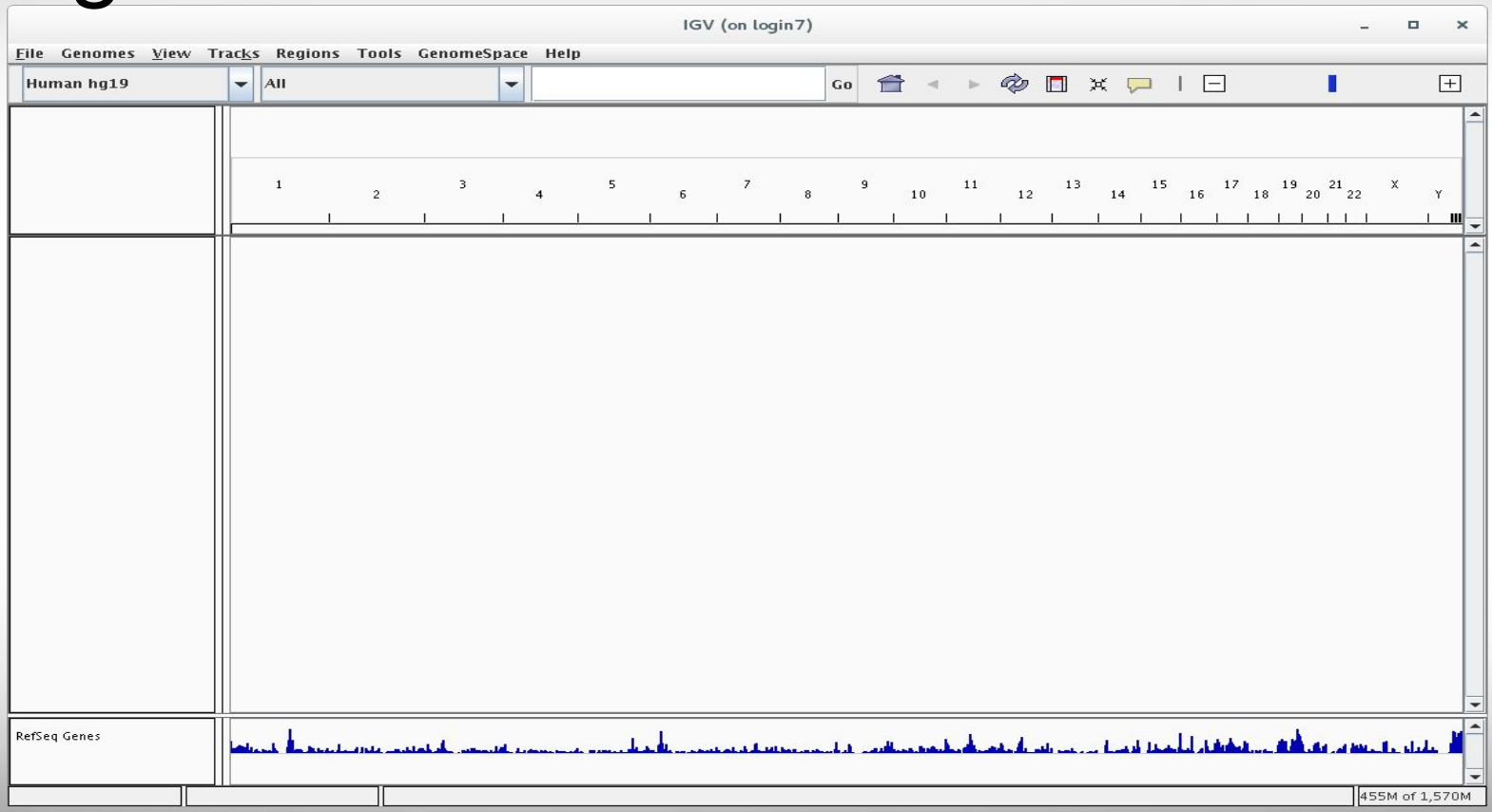
```
module spider IGV
```

```
module load IGV/2.3.68-Java-1.8.0_66
```

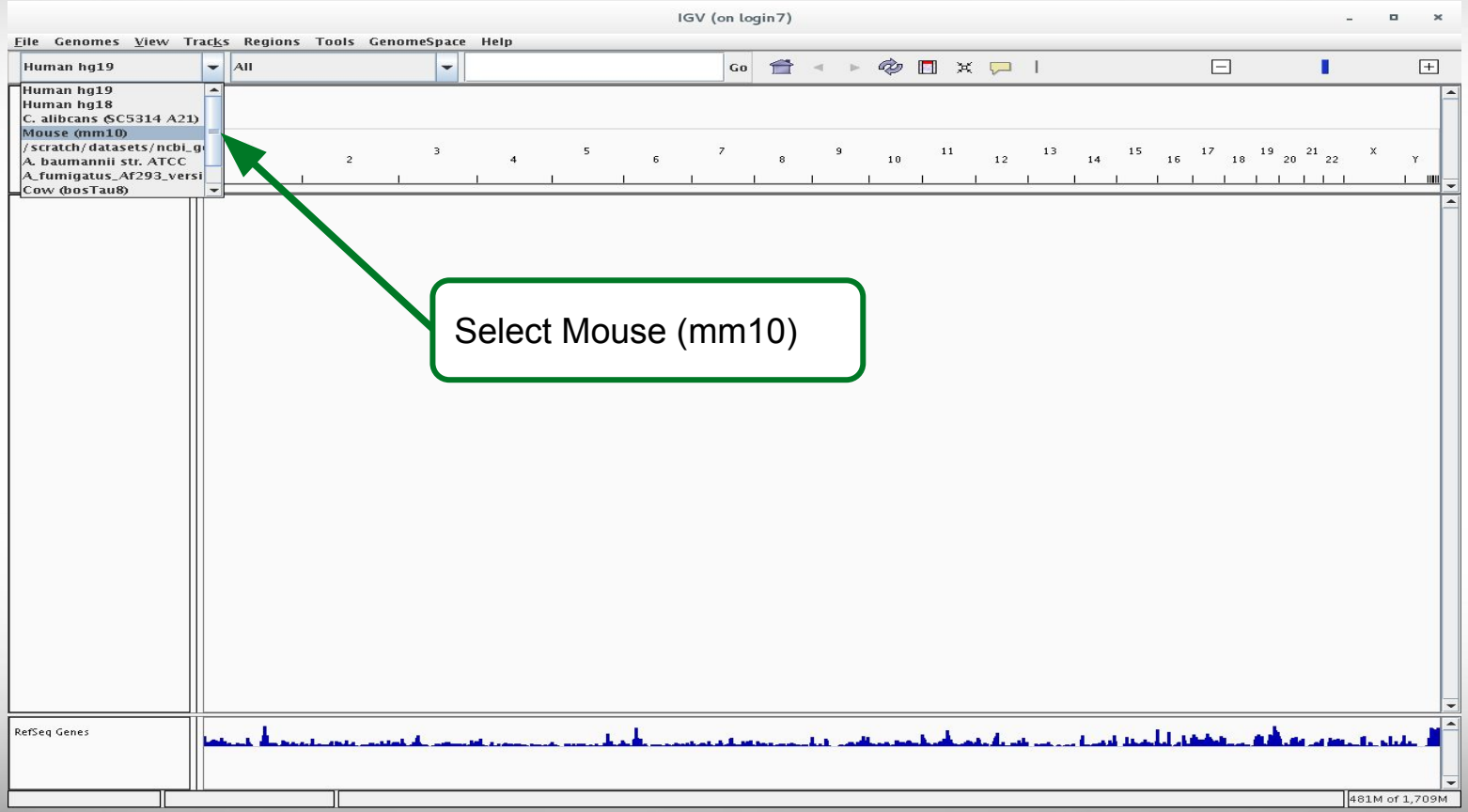
Launch IGV using the igv.sh script (X11 login required)

```
igv.sh
```

# hg19 is default Reference Genome



# Change the Reference Genome

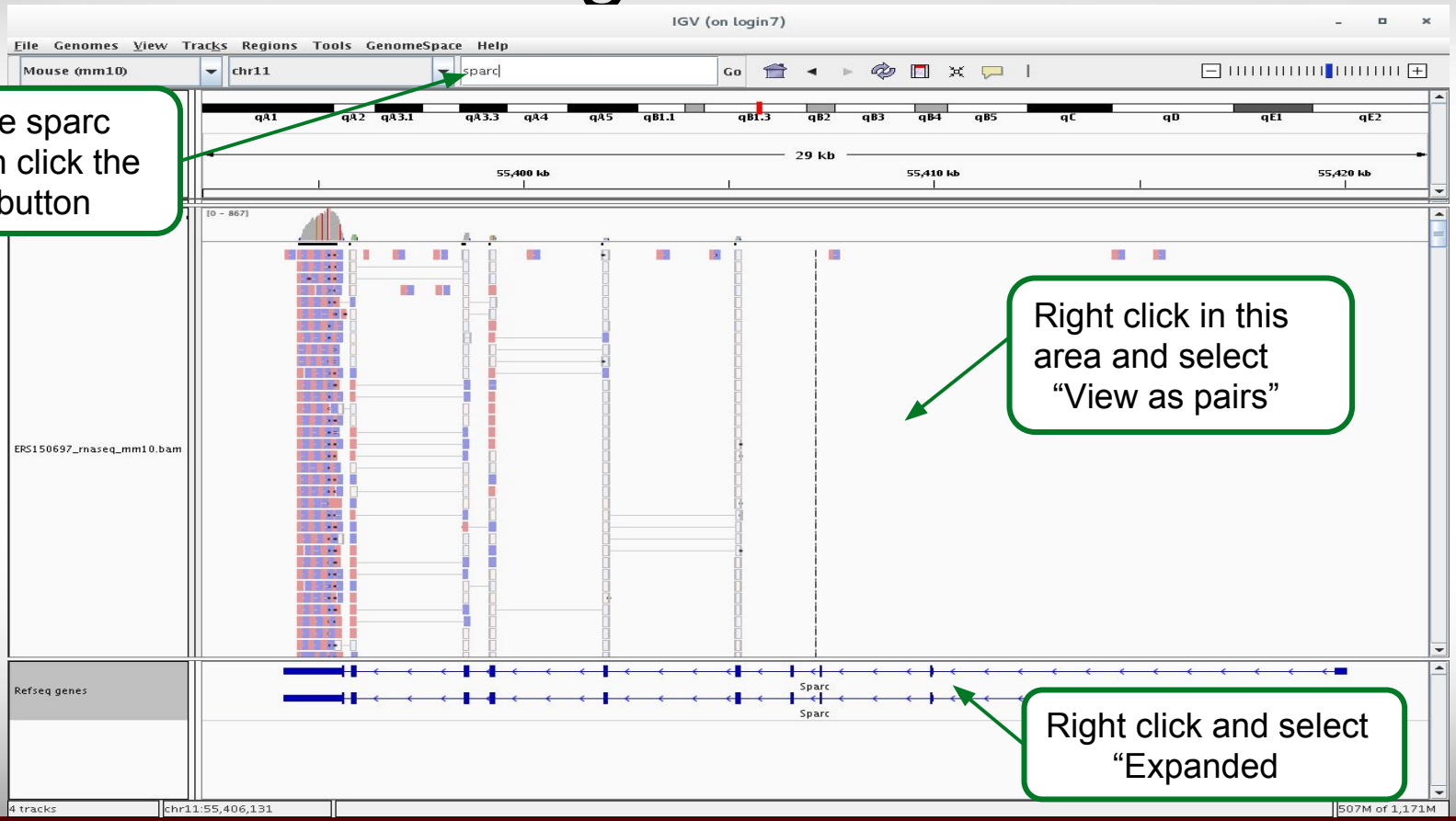


# Load BAM Alignment File

1. Select “File → “Load from file”
2. Search for

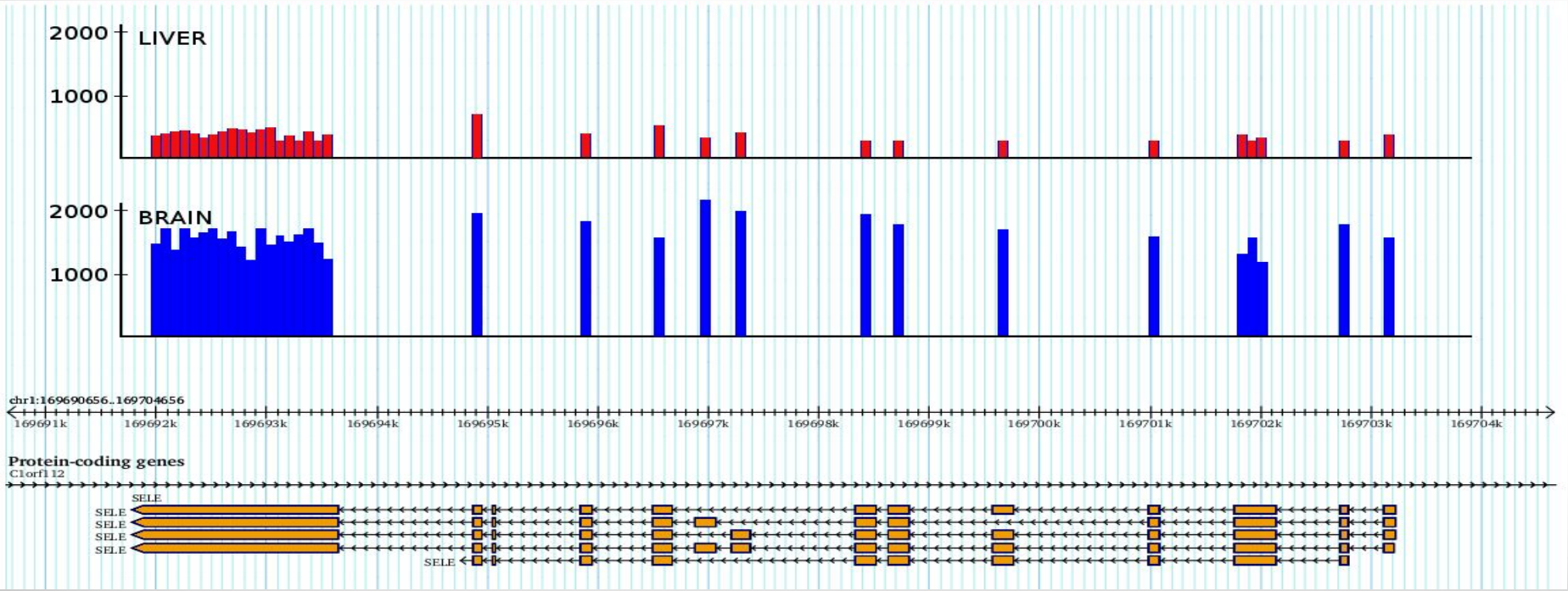
/scratch/helpdesk/ngs/alignments/mm10/ERS150697_rnaseq_mm10.bam

# IGV viewing indexed bam file



# RNA-seq for Differential Expression

# RNA-seq Differential Expression (DE)




http://www.illumina.com/technology/mrna_seq.ilmn

# RNA-seq Sequence Fragment Counting

- Alignment based
  - Non-normalized alignment counts
    - HTSeq-count
  - Normalized (RPKM, FPKM, TPM)
    - eXpress (outputs FPKM)
    - RSEM (isoform/gene level estimates without RPKM or FPKM)
    - Trinity Transcript Quantification
      - A Trinity script can run: Kallisto, RSEM, eXpress, Salmon
- Non-Alignment based
  - Kallisto (pseudoalignment)
  - Salmon (lightweight alignment)
  - Sailfish (k-mer)

# RPKM vs FPKM vs TPM

- The number of **R**eads **P**er **K**ilobase of transcript per **M**illion mapped reads.
  - Intended for single end reads
- The number of **F**ragments **P**er **K**ilobase of transcript per **M**illion mapped reads.
  - Intended for paired-end reads
    - If both paired reads align to a transcript then they are counted as one alignment
- **T**ranscripts **P**er kilobase **M**illion
  - Normalize for gene length first
  - Normalize for sequence depth second

http://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/

# Tuxedo Suite

- HISAT2
  - splice aware mapping of RNA-seq reads
  - TopHat (which uses Bowtie2) and HISAT are superseded by HISAT2
- Cufflinks
  - assembles aligned reads into transcripts and estimates their abundances
- Cuffdiff
  - compares RNA-seq abundance (expression) levels of two samples or groups

| test_id | gene_id | gene | locus | sample_1 | sample_2 | status | value_1 | value_2 | log2(fold_change) | test_stat | p_value | q_value | significant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CAWT_00001 | CAWG_00001 | - | chr_1.1:8373-9093 | q1 | q2 | OK | 111.944 | 163.869 | 0.549763 | 0.768107 | 0.58795 | 0.996768 | no |
| CAWT_00002 | CAWG_00002 | - | chr_1.1:11447-12425 | q1 | q2 | OK | 14.5992 | 30.9037 | 1.08189 | 1.3841 | 0.2921 | 0.98312 | no |
| CAWT_00003 | CAWG_00003 | - | chr_1.1:14130-14451 | q1 | q2 | OK | 248.323 | 259.152 | 0.0615814 | 0.172186 | 0.94685 | 0.996768 | no |
| CAWT_00004 | CAWG_00004 | - | chr_1.1:14890-16045 | q1 | q2 | OK | 60.9546 | 86.0009 | 0.496617 | 0.604904 | 0.6204 | 0.996768 | no |
| … | | | | | | | | | | | | | |
| … | | | | | | | | | | | | | |
| CAWT_01628 | CAWG_01628 | - | chr1.2:664522-665344 | q1 | q2 | OK | 3.56447 | 157.849 | 5.46871 | 6.64693 | 0.00015 | 0.0482417 | yes |

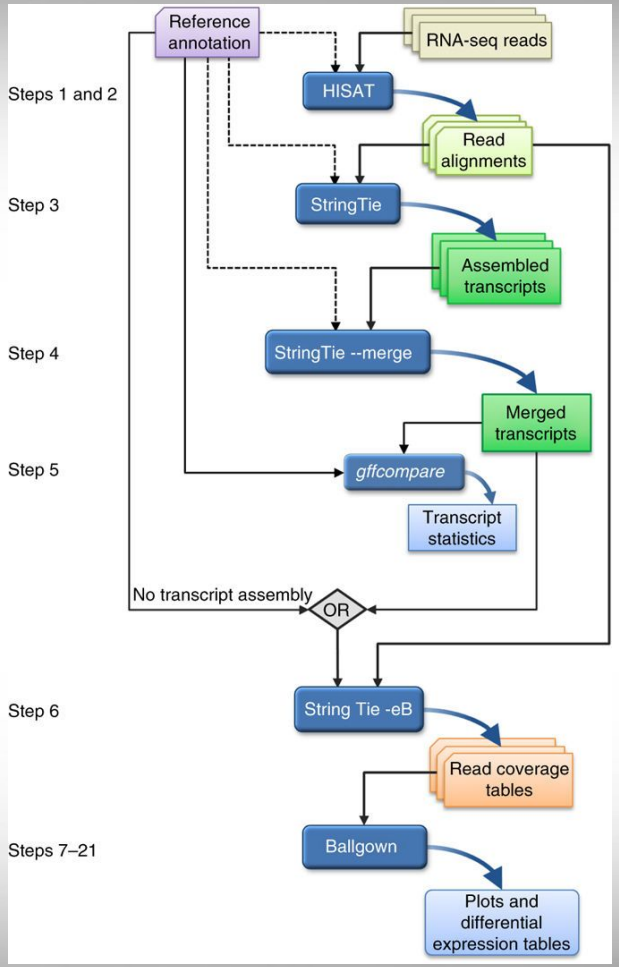
p_value = The uncorrected p-value of the test statistic.
q_value = The FDR-adjusted p-value of the test statistic

# “New Tuxedo” Protocol

**Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown**

Pertea, et al. Nature Protocols 11,1650–1667 (2016)
doi:10.1038/nprot.2016.095

HISAT2 supercedes HISAT



# Sailfish

- Alignment-free isoform quantification from RNA-seq data (uses k-mers)
- Requires a set of target transcripts (fasta)
  - From a reference or a *de novo* assembly
- Requires sequence reads (fasta or fastq)

| Name | Length | EffectiveLength | TPM | NumReads |
|---|---|---|---|---|
| TRINITY_DN30_c0_g1_i1 | 215 | 68.4635 | 236.773 | 233 |
| TRINITY_DN43_c0_g1_i1 | 280 | 102.34 | 5971.5 | 8784 |
| TRINITY_DN88_c0_g1_i1 | 217 | 69.3036 | 191.74 | 191 |
| TRINITY_DN59_c0_g1_i1 | 393 | 194.337 | 4092.64 | 11432 |
| TRINITY_DN98_c0_g1_i1 | 205 | 64.4299 | 1097.09 | 1016 |
| TRINITY_DN17_c0_g1_i1 | 310 | 122.99 | 2634.35 | 4657 |

# R Bioconductor

- Popular R bioconductor packages for RNA-seq
  - CQN – Normalization of RNA-seq data
  - edgeR – Differential gene expression
  - DESeq, DESeq2 – Differential gene expression
  - cummeRbund – analysis/visualization of cufflinks data
- Bioconductor packages can be found in this R version

```
module load R_tamu/3.3.1-intel-2015B-default-mt
```

# Differential Expression (DE) based on alignment counts

Non-normalized abundance counts are used as input for DE analysis in these R Bioconductor Packages

- DESeq2
  - DE for genes not isoforms
- edgeR
  - DE at gene, exon (isoform) or transcript level
- EBSeq
  - DE for isoforms
- DEXSeq
  - DEU differential exon usage

# RNA-seq for Transcriptome Assembly

# RNA-seq Transcriptome Assembly

- Assembly with a reference genome

```
module spider Trinity
```

```
module spider HISAT2 Cufflinks
```

```
module spider Scripture
```

```
module spider StringTie
```

- *de novo a*ssembly without a reference genome

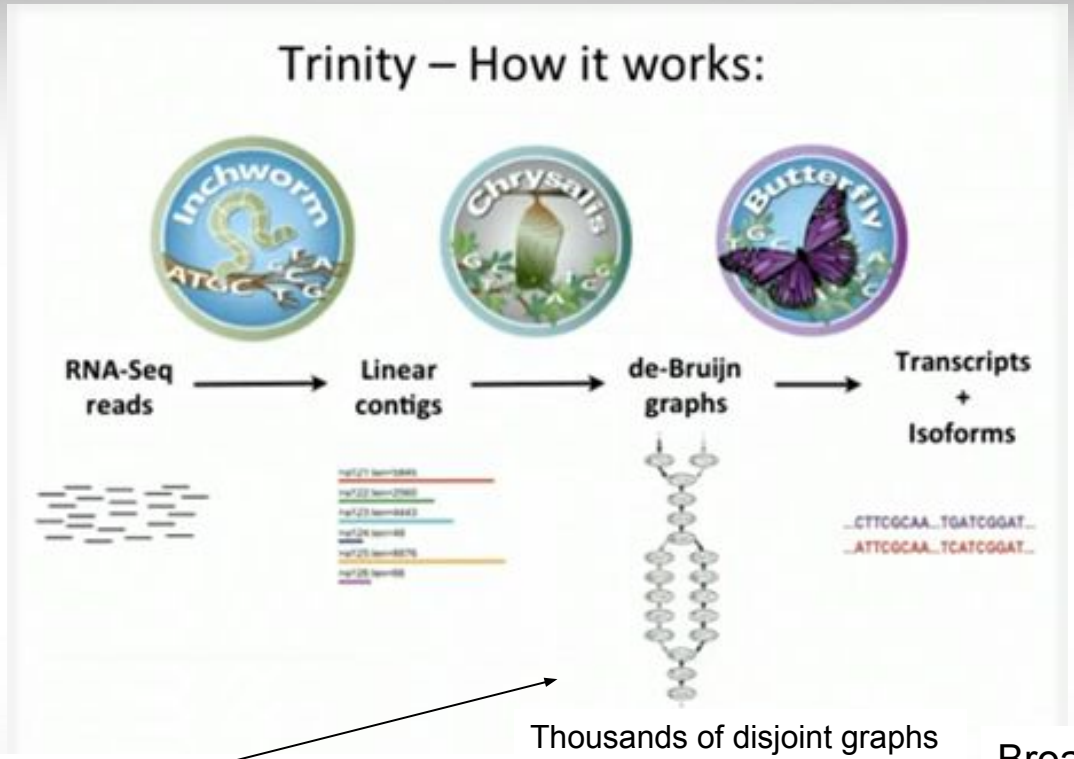
```
module spider Trinity
```

```
module spider Oases
```

# Digital Normalization for Transcriptome Assembly

- Reduce memory requirements by reducing the number of redundant sequence reads if you have a very high sequencing coverage (> 200x)
- Trinity 2.4.0 automatically normalizes reads to a depth of 50
- The **`bbnorm.sh`** script in BBMap can normalize reads

```
module spider BBMap
```




http://www.rna-seqblog.com/a-collection-of-new-rna-seq-videos-from-the-broad-institute/

# Running Trinity on Ada

- Trinity uses 100,000s of intermediate files
  - Contact **help@hprc.tamu.edu** and request a file quota increase before running Trinity
  - Run one Trinity job at a time and check resource usage
    - `showquota`
    - It is recommended not to run multiple Trinity jobs unless you know memory usage and an estimate of the number of temporary files
  - Trinity creates checkpoints and can be restarted if it stops due to file/disk quota met, out of memory or runtime
    - Checkpoints are not available when running Trinity in Galaxy
    - Checkpoints are not available if you use $TMPDIR with Trinity
      - need to rsync results from $TMPDIR at end of job script
      - checkpoints are stored in $TMPDIR which is deleted after job ends
- See GCATemplates for sample Trinity scripts

# Running Trinity on Ada 64GB node

- Use all cores and memory on a node
  - There are 54GB available memory on 64GB nodes

```
#BSUB -n 20
#BSUB -R "span[ptile=20]"
#BSUB -R "rusage[mem=2700]"
#BSUB -M 2700
```

- Recommended Trinity options

```
--max_memory 53G
--CPU 20
--inchworm_cpu 6
--no_version_check
```

# Running Trinity on Ada 256GB node

- Use all cores and memory on a node
  - There are 246GB available memory on 256GB nodes

```
#BSUB -n 20
#BSUB -R "span[ptile=20]"
#BSUB -R "rusage[mem=12300]"
#BSUB -M 12300
#BSUB -R "select[mem256gb]"
```

- Recommended Trinity options

```
--max_memory 245G
--CPU 20
--inchworm_cpu 6
--no_version_check
```

# Running Trinity on Ada 1TB node

- Use all cores and memory on a node
  - There is 1TB avail memory on 1TB nodes

```
#BSUB -n 40
#BSUB -R "span[ptile=40]"
#BSUB -R "rusage[mem=25000]"
#BSUB -M 25000
#BSUB -q xlarge
#BSUB -R "select[mem1tb]"
```

- Recommended Trinity options

```
--max_memory 999G
--CPU 40
--inchworm_cpu 6
--no_version_check
```

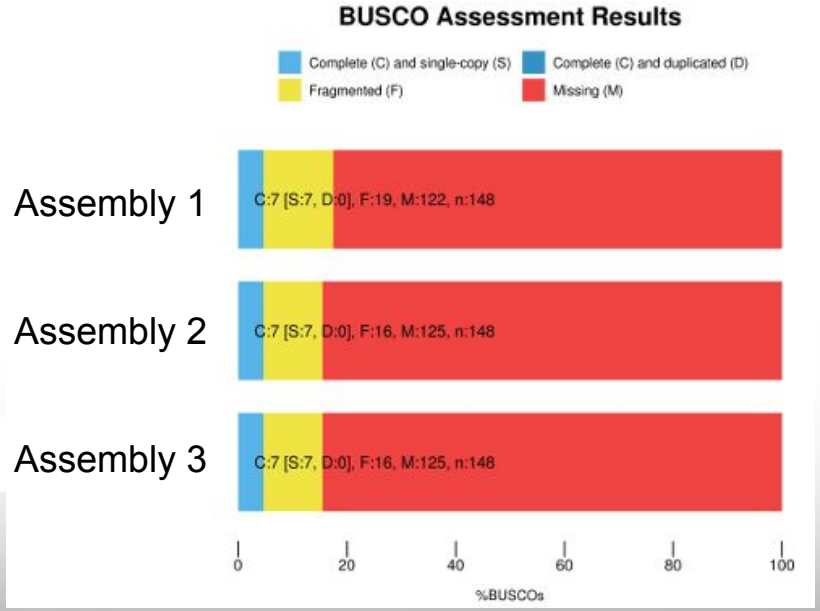
# Transcriptome Assembly Completeness

The completeness of a transcriptome can be estimated by using a set of highly conserved genes that are common to specific taxonomic groups

- 44 taxonomic groups available
  - aves, bacteria, eukaryota, insecta, vertebrata, ...
- BUSCO – uses single-copy genes to access transcriptome assembly and annotation completeness
  - evaluates % complete 'BUSCOs', % fragmented, % missing
  - can run in genome, transcriptome or protein mode
  - `module spider BUSCO`

# Transcriptome Assembly Completeness

BUSCO script (generate_plot.py) can be used to plot multiple BUSCO short summaries to compare different assemblies



# Transcriptome Assembly Evaluation

```
module spider DETONATE
```

- Input transcriptome fasta assembly and sequence reads fastq or fasta files
  - RSEM-EVAL used for reference-free evaluation
  - REF-EVAL used for reference-based evaluation
  - Higher score = better evaluation
- Sample output:

```
Score                        -30198099.46
Number_of_contigs              1976
Number_of_alignable_reads                  1140584
Number_of_alignments_in_total              1434453
```

# Transcriptome Assembly Evaluation

```
module spider Transrate
```

- For *de novo* transcriptome assembly quality analysis
  - Inputs are combinations of the following
    - assembly.fa (one or more assemblies)
    - left.fq (quality trimmed)
    - right.fq (quality trimmed)
    - reference.fa
  - Output
    - Contig metrics (smallest, largest, N50, %GC, …)
    - Supports merging assemblies

# Transcriptome Assembly Annotation

```
module spider Trinotate
```

- You can run each of the following tools individually but Trinotate will run all these tools to annotate an assembly
  - RNAMMER
    - predicts 5s/8s, 16s/18s, 23s/28s ribosomal RNA
  - TransDecoder
    - predicts coding regions
  - BLAST+ (SwissProt db)
  - HMMER (PFAM db)
  - SignalP
    - predicts presence and location of signal peptide cleavage sites in amino acid sequences
  - tmhmm
    - prediction of transmembrane helices in proteins
- Results are saved in SQLite db and as a summary file: Trinotate.xls

DON'T MISS Maverix Biomics Offers Free Trial of Self-Service RNA-Seq Data Analysis

Search...

# RNA-Seq Blog

Transcriptome Sequencing Research & Industry News



HOME | NEWS » | EVENTS » | JOBS » | TECHNOLOGY » | DATA ANALYSIS » | BLOG | CONTACT »

## Tag Archives: TMM

### Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data

October 30, 2015 0 1,501 Views

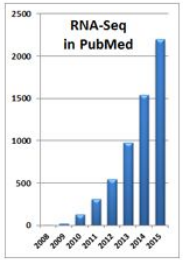
Recently, rapid improvements in technology and decrease in sequencing costs have made RNA-Seq a widely used technique to quantify gene expression levels. Various normalization approaches have been proposed, owing to the importance of normalization in the analysis of RNA-Seq data. …

Read More »

### An iteration normalization and test method for differential expression analysis of RNA-seq data

### STAY CONNECTED

### PUBLICATIONS TREND




### RECENT RNA-SEQ PUBS

RED: A Java-MySQL Software for Identifying

### SUBSCRIBE TO THE RNA-SEQ BLOG

email address

Subscribe

### RNA-SEQ PRODUCTS & SERVICES


# Thank you.

# Any question?