Bioinformatics Short Course: RNA-Seq Data Analysis

Part VI: Expression Analysis (Exercises)

Chuming Chen Ph. D.
University of Delaware
May 22-23, 2012
Summary (Lecture)

- Transcriptome assembly strategies
- Short read aligners
- Alignment format and SAMtools
- Alignment visualization
Software Components of Tuxedo Suite Tools

- **Bowtie** forms the algorithmic core of TopHat, which align reads to the reference genome.
- **TopHat**’s read alignments are assembled by **Cufflinks** and its associated utility program (**Cuffmerge**, **Cuffcompare**) can produce a transcriptome annotation of the genome.
- **Cuffdiff** quantifies this transcriptome across multiple conditions using the TopHat read alignments.
- **CummeRbund** explores and visualizes the differential expression data (Genes and Transcripts) generated by Cuffdiff.

Bowtie

• An ultrafast, memory-efficient short read aligner.
• It uses an extremely economical data structure called the Burrows-Wheeler index to store the reference genome sequence and allows it to be searched rapidly at a rate of tens of millions reads per CPU hour.
• It makes a number of compromises to achieve its high speed:
  – If one or more exact matches exist for a read, it is guaranteed to find one.
  – If the best match is not exact match, then it is not guaranteed in all cases to find the highest quality alignment.
  – It may fail to align reads with multiple mismatches.
• Furthermore, Bowtie does not allow alignments between a read and the genome to contain large gaps; hence, it cannot align reads that span introns. TopHat was created to address this limitation.
• Web Site: http://bowtie-bio.sourceforge.net/index.shtml
TopHat

- Use Bowtie as alignment engine.
- Break up reads Bowtie cannot align into segments then align them independently.
- When several of a read’s segments aligned to the genome far apart, TopHat infers that the read spans a splice junction and estimate the splice site.
- By using the ‘initially unmapped’ reads, TopHat can build an index of splice sites in the transcriptome on the fly without a prior gene or splice site annotations.

(Trapnell et al. Bioinformatics. 2009 May 1;25(9):1105-11)
Cufflinks

- Assembles individual transcripts from RNA-Seq reads that have been aligned to the genome.
- Reports as few full-length transcript fragments or ‘transfrags’ as are needed to ‘explain’ all the splicing events in the input data.
- Quantifies the expression level of each transfrag in the sample using a rigorous statistical model of RNA-Seq to filter out background or artifactual transfrags such as immature primary transcripts.
- Quantifies transcript abundance using a reference annotation.

• In addition to differential expression analysis, people are often interested in discovering new genes and transcripts.

• Gaps in sequencing coverage will cause breaks in transcript reconstruction and make it difficult to distinguish full-length novel transcripts from partial fragments.

• Cuffcompare can compare the Cufflinks assemblies to reference annotation files and help sort out new genes from known ones.

• Web site: http://cufflinks.cbcb.umd.edu/manual.html#cuffcompare
Print Summary Reports

$ ls -tlr cufflinks_out_*/*map
-rw-r--r-- 1 chenc cwu 481209 May 17 14:35 cufflinks_out_FL1/cuffcmp.transcripts.gtf.tmap
-rw-r--r-- 1 chenc cwu 196073 May 17 14:35 cufflinks_out_FL1/cuffcmp.transcripts.gtf.refmap
-rw-r--r-- 1 chenc cwu 479849 May 17 14:35 cufflinks_out_FL2/cuffcmp.transcripts.gtf.tmap
-rw-r--r-- 1 chenc cwu 6144 May 17 14:35 cufflinks_out_FL2/cuffcmp.transcripts.gtf.refmap
-rw-r--r-- 1 chenc cwu 482416 May 17 14:35 cufflinks_out_LL1/cuffcmp.transcripts.gtf.tmap
-rw-r--r-- 1 chenc cwu 4466 May 17 14:35 cufflinks_out_LL1/cuffcmp.transcripts.gtf.refmap
-rw-r--r-- 1 chenc cwu 476989 May 17 14:35 cufflinks_out_LL2/cuffcmp.transcripts.gtf.tmap
-rw-r--r-- 1 chenc cwu 3880 May 17 14:35 cufflinks_out_LL2/cuffcmp.transcripts.gtf.refmap

.tmap
This tab delimited file lists the most closely matching reference transcript for each Cufflinks transcript. There is one row per Cufflinks transcript.

.refmap
This tab delimited file lists, for each reference transcript, which cufflinks transcripts either fully or partially match it. There is one row per reference transcript.

The following bash script prints a simple table for each assembly that lists how many transcripts in each assembly are complete matches to the know transcripts, how many are partial matches etc.

$ find . -name *.tmap | while read file; do echo $file; awk 'NR > 1 { s[$3]++ } END { for (j in s) { print j, s[j] } }' $file; done
Cuffcompare Summary Reports

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>=</td>
<td>Complete match of intron chain</td>
</tr>
<tr>
<td>c</td>
<td>Contained</td>
</tr>
<tr>
<td>j</td>
<td>Potentially novel isoform (fragment): at least one splice junction is shared with a reference transcript</td>
</tr>
<tr>
<td>e</td>
<td>Single exon transfrag overlapping a reference exon and at least 10 bp of a reference intron, indicating a possible pre-mRNA fragment</td>
</tr>
<tr>
<td>i</td>
<td>A transfrag falling entirely within a reference transcript</td>
</tr>
<tr>
<td>o</td>
<td>Generic exonic overlap with a reference transcript</td>
</tr>
<tr>
<td>P</td>
<td>Possible polymerase run-on fragment (within 2Kbases of a reference transcript)</td>
</tr>
<tr>
<td>r</td>
<td>Repeat. Currently determined by looking at the soft-masked reference sequence and applied to transcripts where at least 50% of the bases are lower case</td>
</tr>
<tr>
<td>u</td>
<td>Unknown, intergenic transcript</td>
</tr>
<tr>
<td>x</td>
<td>Exonic overlap with reference on the opposite strand</td>
</tr>
<tr>
<td>s</td>
<td>An intron of the transfrag overlaps a reference intro on the opposite strand (likely due to read mapping errors)</td>
</tr>
<tr>
<td>-</td>
<td>.tracking file only, indicates multiple classification</td>
</tr>
</tbody>
</table>

(http://cufflinks.cbcb.umd.edu/manual.html#cuffcompare)
Cuffmerge

• In multi-sample RNA-Seq experiment, sometime it is necessary to pool the data and assemble them into a comprehensive set of transcripts before differential analysis.
• Pool aligned reads from all samples and run Cufflinks once on them is not recommended:
  – Assembly becomes more computationally expensive as read depth increases.
  – Complex mixture of splice isoforms for many genes may lead to the incorrectly assembled transcripts.
• As a ‘meta-assembler’, Cuffmerge parsimoniously merges the individually assemblies by Cufflinks by treating the assembled transfrags the way Cufflinks treats the reads.
• It can also performs a reference annotation-based transcript (RABT) (Roberts et al. 2011) assembly to merge reference transcripts with assembled sample transfrags to produce a single annotation file for downstream differential analysis.
• Web site: http://cufflinks.cbcb.umd.edu/manual.html#cuffmerge
Cuffdiff

- Calculates expression in two or more samples and tests the statistical significance of each observed change in expression between them.
- The statistical model assumptions:
  - The number of reads produced by each transcript is proportional to its abundance.
  - It fluctuates due to technical variability during library preparation and sequencing, and the biological variability between replicates of the same experiment.
- Allows multiple technical or biological replicate sequencing libraries per condition.
- Reports gene and transcript expression level changes in tabular format, which includes fold change (in log2), P-value (both raw and corrected for multiple hypotheses testing), gene and transcript related information such as name and location in the genome.
Cuffdiff (additional differential analysis)

- Identify genes that are differentially spliced or regulated via promoter switching
- Group isoforms of a gene that have the same TSS (derived from the same pre-mRNA, changes in abundance reflect the differential splicing of common pre-mRNA).
- Total expression levels of a TSS group is the sum of expression levels of the isoforms within it.
- Relative abundance between multiple TSSs reflect the changes in TSS (promoter) preferences between condition.

(Trapnell et al., Nat Protoc. 2012 Mar 1;7(3):562-78)
GTF format

• GTF stands for **Gene Transfer Format**.

• The tab-delimited file includes fields below:
  
  – `<seqname>` <source> <feature> <start> <end> <score> <strand> <frame> [attributes] [comments]

  (http://mblab.wustl.edu/GTF22.html)
GTF Field Definitions

<seqname> - The name of the sequence. Commonly, this is the chromosome ID or contig ID. Note that the coordinates used must be unique within each sequence name in all GTFs for an annotation set.

<source> - The source column should be a unique label indicating where the annotations came from --- typically the name of either a prediction program or a public database.

<feature> - The following feature types are required: "CDS", "start_codon", "stop_codon". The features "5UTR", "3UTR", "inter", "inter_CNS", "intron_CNS" and "exon" are optional. All other features will be ignored. The types must have the correct capitalization shown here.

<start> <end> - Integer start and end coordinates of the feature relative to the beginning of the sequence named in <seqname>. <start> must be less than or equal to <end>. Sequence numbering starts at 1. Values of <start> and <end> that extend outside the reference sequence are technically acceptable, but they are discouraged.

<score> - The score field indicates a degree of confidence in the feature's existence and coordinates. The value of this field has no global scale but may have relative significance when the <source> field indicates the prediction program used to create this annotation. It may be a floating point number or integer, and not necessary and may be replaced with a dot.

)frame> - 0 indicates that the feature begins with a whole codon at the 5' most base. 1 means that there is one extra base (the third base of a codon) before the first whole codon and 2 means that there are two extra bases (the second and third bases of the codon) before the first codon. Note that for reverse strand features, the 5' most base is the <end> coordinate.

(http://mblab.wustl.edu/GTF22.html)
BED format

- BED format provides a flexible way to define the data lines that are displayed in an annotation track. BED lines have 12 fields.
  - Required fields:
    • chrom, chromStart, chromEnd
  - Additional optional fields:
    • name, score, strand, thickStart, thickEnd, itemRgb, blockCount, blockSizes, blockStarts.

(track://genome.ucsc.edu/FAQ/FAQformat#format1)

track name=junctions description="TopHat junctions"
chr1 13983 19738 JUNC00000001 1 + 13983 19738 255,0,0 2 46,34 0,5721
chr1 20459 20649 JUNC00000002 1 + 20459 20649 255,0,0 2 46,55 0,135
chr1 33994 35271 JUNC00000003 1 + 33994 35271 255,0,0 2 28,69 0,1208
chr1 41809 42420 JUNC00000004 2 + 41809 42420 255,0,0 2 97,61 0,550
chr1 42392 45268 JUNC00000005 3 + 42392 45268 255,0,0 2 81,74 0,2802
chr1 46828 48793 JUNC00000006 1 + 46828 48793 255,0,0 2 46,38 0,1927
Exercise 4

View Alignment, Coverage, and Isoforms
(SAMtools, IGV)
Index Alignment Files for IGV

```bash
$ cat ~/.rnaseq-shared/pbs_scripts/samtools_index.qs
#PBS -N  SamtoolsIndex
#PBS -S /bin/bash
#PBS -V
#PBS -l ncpus=1,walltime=16:00:00,cput=10:00:00,mem=2000mb,nodes=1:ppn=4
#PBS -q rnaseq

cd $PBS_O_WORKDIR
samtools faidx index/gallus_chr1.fa
ln -s accepted_hits.bam tophat_out_F1/FL1.bam
ln -s accepted_hits.bam tophat_out_F2/FL2.bam
ln -s accepted_hits.bam tophat_out_L1/LL1.bam
ln -s accepted_hits.bam tophat_out_L2/LL2.bam

samtools index tophat_out_F1/FL1.bam
samtools index tophat_out_F2/FL2.bam
samtools index tophat_out_L1/LL1.bam
samtools index tophat_out_L2/LL2.bam

ln -s transcripts.gtf cufflinks_out_F1/FL1_transcripts.gtf
ln -s transcripts.gtf cufflinks_out_F2/FL2_transcripts.gtf
ln -s transcripts.gtf cufflinks_out_L1/LL1_transcripts.gtf
ln -s transcripts.gtf cufflinks_out_L2/LL2_transcripts.gtf

$ qsub ~/.rnaseq-shared/pbs_scripts/samtools_index.qs
90297.biohen.dbi.local
$ ls -ltr tophat_out_*/
$ ls -tlr index/

total 412968
-rw-r--r-- 1 chenc cwu 48797843 May 17 14:03 gallus_chr1.4.ebwt
-rw-r--r-- 1 chenc cwu 89909 May 17 14:03 gallus_chr1.3.ebwt
-rw-r--r-- 1 chenc cwu 60083414 May 17 14:06 gallus_chr1.1.ebwt
-rw-r--r-- 1 chenc cwu 24398928 May 17 14:06 gallus_chr1.2.ebwt
-rw-r--r-- 1 chenc cwu 60083414 May 17 14:08 gallus_chr1.rev.1.ebwt
-rw-r--r-- 1 chenc cwu 24398928 May 17 14:08 gallus_chr1.rev.2.ebwt
-rw-r--r-- 1 chenc cwu 205013902 May 17 15:07 gallus_chr1.fa
-rw-r--r-- 1 chenc cwu 23 May 17 15:07 gallus_chr1.fa.fai
```
• Your login ends with odd number, i.e. rna7
  ssh –X rna7@glycine.dbi.udel.edu
• Your login ends with even number, i.e. rna8
  ssh –X rna8@biohen.dbi.udel.edu
  Qsub –IXV –q rnaseq
Launch IGV

$ qsub -I -X -V -q rnaseq
$ cd ~/rnaseq-work
$ igv.sh

• Click “File”, then click “Load Genome from File …”, then select gallus_chr1.fa from the directory called reference.
• Click “File”, then click “Load from File …”, then select gallus_chr1.gtf from the directory called reference.
• Click “File”, then click “Load from File …”, then select LL2.bam from the directory called tophat_out_LL2.
• Click “File”, then click “Load from File …”, then select LL2_transcripts.gtf from the directory called cufflinks_out_LL2.
• Type in chr1:79055360-79070712 in the search box and click “Go”.
• Right click LL2_transcripts.gtf track and select “Expanded”.

---

19
View Splice Junctions and Isoforms of MFAP5
View Differentially Expressed Novel Gene

- Type in chr1:171657285-171662936 in the search box and click “Go”.
Exercise 6

Explore Differential Analysis Results
(CummeRbund)
### Tabular view of differentially expressed genes

$ cd ~/rnaseq-work$

$ head -3 cuffdiff_out/gene_exp.diff$

<table>
<thead>
<tr>
<th>test_id</th>
<th>gene_id</th>
<th>gene</th>
<th>locus</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>log2(fold_change)</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_00001</td>
<td>XLOC_00001</td>
<td>LOC425783</td>
<td>chr1:92892486-92925118</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>315.97190,759</td>
<td>-0.728074</td>
<td>3.96029</td>
<td>7.48577e-05</td>
<td>0.00764796</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>XLOC_00002</td>
<td>XLOC_00002</td>
<td>GOLGBl</td>
<td>chr1:33922-67653</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>135.819</td>
<td>155.192</td>
<td>0.192365</td>
<td>-0.820333</td>
<td>0.412026</td>
<td>0.832533</td>
<td>no</td>
</tr>
</tbody>
</table>

$ grep yes cuffdiff_out/gene_exp.diff | cut -f2- $

$ grep yes cuffdiff_out/gene_exp.diff | cut -f2- >> sig_diff_genes.txt$

$ cat sig_diff_genes.txt

<table>
<thead>
<tr>
<th>test_id</th>
<th>gene_id</th>
<th>gene</th>
<th>locus</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>log2(fold_change)</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_00015</td>
<td>XLOC_00015</td>
<td>DOCK4</td>
<td>chr1:28011240-29077867</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>315.97190,759</td>
<td>-0.728074</td>
<td>3.96029</td>
<td>7.48577e-05</td>
<td>0.00764796</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>XLOC_00020</td>
<td>XLOC_00020</td>
<td>LGR5</td>
<td>chr1:38117567-38203984</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>101.9591,290,104</td>
<td>1.51355</td>
<td>-4.5102</td>
<td>6.47663e-06</td>
<td>0.000794035</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>XLOC_00059</td>
<td>XLOC_00059</td>
<td>B3ZB5_CHICK</td>
<td>chr1:71254727-71270462</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>361.21</td>
<td>4752.44</td>
<td>0.392991</td>
<td>-3.55812</td>
<td>0.000373513</td>
<td>0.0190803</td>
<td>yes</td>
</tr>
</tbody>
</table>

$ grep yes cuffdiff_out/gene_exp.diff | cut -f2- $

$ grep yes cuffdiff_out/gene_exp.diff | cut -f2- >> sig_diff_genes.txt$

$ cat sig_diff_genes.txt

<table>
<thead>
<tr>
<th>test_id</th>
<th>gene_id</th>
<th>gene</th>
<th>locus</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>log2(fold_change)</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_00015</td>
<td>XLOC_00015</td>
<td>DOCK4</td>
<td>chr1:28011240-29077867</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>315.97190,759</td>
<td>-0.728074</td>
<td>3.96029</td>
<td>7.48577e-05</td>
<td>0.00764796</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

$ head -1 cuffdiff_out/gene_exp.diff > sig_diff_genes.txt$

$ grep yes cuffdiff_out/gene_exp.diff | cut -f2- >> sig_diff_genes.txt$

$ cat sig_diff_gen...
What is R?

- R is a data analysis software for statistical analysis, data visualization and predictive modeling.
- R is a complete, interactive, object-oriented programming language.
- R is an environment for statistical analysis, providing functions for virtually every data manipulation, statistical modeling.
- R is an open-source software project.
- R is a community of leading statisticians and computer scientists and thousands of contributors.
- [http://www.r-project.org/](http://www.r-project.org/)
CummeRbund

- A user-friendly R package to help manage, visualize and integrate all the data generated by Cuffdiff analysis.
- Simplify the data exploration task such as plotting and cluster analysis of expression data.
- Scripted plotting automates the plot generation and reuse analyses from previous experiments.
- Transform Cuffdiff data into R statistical computing environment enables other advanced statistical analysis and plotting packages.
- Takes the various output files from a Cuffdiff run and creates a SQLite database of the results describing appropriate relationships between genes, transcripts, transcription start sites, and CDS regions to allow efficiently retrieval and exploration.
Start R interactive shell

$ qsub -I -X -V -q rnaseq

$ cd ~/rnaseq-work

$ R

R version 2.15.0 (2012-03-30)
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
Basic Syntax of R language

```r
> x <- c(1,2,3,4,5,6)   # Create ordered collection (vector)
> y <- x^2              # Square the elements of x
> print(y)              # print (vector) y
[1]  1  4  9 16 25 36
> mean(y)               # Calculate average (arithmetic mean) of (vector) y; result is scalar
[1] 15.16667
> var(y)                # Calculate sample variance
[1] 178.9667
> lm_1 <- lm(y ~ x)     # Fit a linear regression model "y = f(x)" or "y = B0 + (B1 * x)" store the results as lm_1
> print(lm_1)           # Print the model from the (linear model object) lm_1

Call:
  lm(formula = y ~ x)

Coefficients:  
(Intercept)        x
-9.333        7.000

> summary(lm_1)     # Compute and print statistics for the fit of the (linear model object) lm_1

Call:  
  lm(formula = y ~ x)

Residuals:  
1         2         3         4         5         6
  3.3333  -0.6667  -2.6667  -2.6667  -0.6667   3.3333

Coefficients:  
              Estimate Std. Error t value  Pr(>|t|)  
(Intercept)   -9.3333     2.8441  -3.282 0.030453 *  
x              7.0000     0.7303   9.585 0.000662 *** 
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.055 on 4 degrees of freedom  
Multiple R-squared: 0.9583, Adjusted R-squared: 0.9478
F-statistic: 91.88 on 1 and 4 DF,  p-value: 0.000662

> library(cairoDevice)  # external package provides Cairo() function.
> Cairo()               # Open an R graphics device based on the Cairo vector graphics
> par(mfrow=c(2,2))     # Request 2x2 plot layout
> plot(lm_1)            # Diagnostic plot of regression model
> pdf("lm_1.pdf")      # starts the graphics device driver for producing PDF file "lm_1.pdf"
> par(mfrow=c(2,2))     # request the same layout
> plot(lm_1)            #Diagnostic plot of regression model
> dev.off()             # shuts down the specified (by default the current) device
> list.files(pattern="pdf")
[1] "lm_1.pdf"
```

Create a CummeRbund Database from Cuffdiff Output

```r
> library(cummeRbund)
Loading required package: RSQLite
Loading required package: DBI
Loading required package: ggplot2
Loading required package: reshape2
>
cuff_data <- readCufflinks('cuffdiff_out')
Creating database cuffdiff_out/cuffData.db
Reading cuffdiff_out/genes.fpkm_tracking
Checking samples table...
Populating samples table...
Writing genes table
Reshaping geneData table
Recasting
Writing geneData table
Reading cuffdiff_out/gene_exp.diff
Writing geneExpDiffData table
Reading cuffdiff_out/promoters.diff
Writing promoterDiffData table
Reading cuffdiff_out/isoforms.fpkm_tracking
Checking samples table...
OK!
Writing isoforms table
Reshaping isoformData table
Recasting
Writing isoformData table
Reading cuffdiff_out/isoform_exp.diff
Writing isoformExpDiffData table
Reading cuffdiff_out/tss_groups.fpkm_tracking
Checking samples table...
OK!
Writing TSS table
Reshaping TSSData table
Recasting
Writing TSSData table
Reading cuffdiff_out/tss_group_exp.diff
Writing TSSExpDiffData table
Reading cuffdiff_out/splicing.diff
Writing splicingDiffData table
Reading cuffdiff_out/cds.fpkm_tracking
Checking samples table...
OK!
Writing CDS table
Reshaping CDSData table
Recasting
Writing CDSData table
Reading cuffdiff_out/cds_exp.diff
Writing CDSExpDiffData table
Reading cuffdiff_out/cds.diff
Writing CDSDiffData table
Indexing Tables...
>
```
Accessing Data

```r
> options(width=200)
> gene.feature <- features(genes(cuff_data))
> head(gene.feature)

<table>
<thead>
<tr>
<th>gene_id</th>
<th>class_code</th>
<th>nearest_ref_id</th>
<th>gene_short_name</th>
<th>locus</th>
<th>length</th>
<th>coverage</th>
<th>gene_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_000001</td>
<td>&lt;NA&gt;</td>
<td>LOC425783</td>
<td>gene_short_name</td>
<td>chr1</td>
<td>6267</td>
<td>21806</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>XLOC_000002</td>
<td>&lt;NA&gt;</td>
<td>GOLGB1</td>
<td>chr1:33922-67653</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>XLOC_000003</td>
<td>&lt;NA&gt;</td>
<td>Q5ZMV0 CHICK</td>
<td>chr1:69103-83497</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>XLOC_000004</td>
<td>&lt;NA&gt;</td>
<td>RABL2B</td>
<td>chr1:99078-102964</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>XLOC_000005</td>
<td>&lt;NA&gt;</td>
<td>ENSGALG00000023985</td>
<td>chr1:108633-114770</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>XLOC_000006</td>
<td>&lt;NA&gt;</td>
<td>SHANK3</td>
<td>chr1:171646-199289</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&lt;NA&gt;</td>
</tr>
</tbody>
</table>

> gene.fpkm <- fpkm(genes(cuff_data))
> head(gene.fpkm)

<table>
<thead>
<tr>
<th>gene_id</th>
<th>sample_name</th>
<th>fpkm</th>
<th>conf_hi</th>
<th>conf_lo</th>
<th>quant_status</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_000001</td>
<td>FL</td>
<td>37.6272</td>
<td>80.6297</td>
<td>0.0000</td>
<td>OK</td>
</tr>
<tr>
<td>XLOC_000001</td>
<td>LL</td>
<td>96.7472</td>
<td>161.2450</td>
<td>32.2491</td>
<td>OK</td>
</tr>
<tr>
<td>XLOC_000002</td>
<td>FL</td>
<td>135.8230</td>
<td>167.5380</td>
<td>104.1070</td>
<td>OK</td>
</tr>
<tr>
<td>XLOC_000002</td>
<td>LL</td>
<td>155.1970</td>
<td>190.3030</td>
<td>120.0920</td>
<td>OK</td>
</tr>
<tr>
<td>XLOC_000003</td>
<td>FL</td>
<td>132.7550</td>
<td>190.9640</td>
<td>74.5450</td>
<td>OK</td>
</tr>
<tr>
<td>XLOC_000003</td>
<td>LL</td>
<td>140.7280</td>
<td>200.0020</td>
<td>81.4548</td>
<td>OK</td>
</tr>
</tbody>
</table>

> isoform.fpkm <- fpkm(isoforms(cuff_data))
> head(isoform.fpkm)

<table>
<thead>
<tr>
<th>isoform_id</th>
<th>sample_name</th>
<th>fpkm</th>
<th>conf_hi</th>
<th>conf_lo</th>
<th>quant_status</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCONS_00000001</td>
<td>FL</td>
<td>37.6272</td>
<td>80.6297</td>
<td>0.0000</td>
<td>OK</td>
</tr>
<tr>
<td>TCONS_00000001</td>
<td>LL</td>
<td>96.7472</td>
<td>161.2450</td>
<td>32.2491</td>
<td>OK</td>
</tr>
<tr>
<td>TCONS_00000002</td>
<td>FL</td>
<td>0.0000355677</td>
<td>3.43438</td>
<td>0.00000</td>
<td>OK</td>
</tr>
<tr>
<td>TCONS_00000002</td>
<td>LL</td>
<td>0.000000000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>OK</td>
</tr>
<tr>
<td>TCONS_00000003</td>
<td>FL</td>
<td>37.313300000</td>
<td>59.59370</td>
<td>15.03290</td>
<td>OK</td>
</tr>
<tr>
<td>TCONS_00000003</td>
<td>LL</td>
<td>22.016000000</td>
<td>40.60530</td>
<td>3.42675</td>
<td>OK</td>
</tr>
</tbody>
</table>

> gene.diff <- diffData(genes(cuff_data))
> head(gene.diff)

| gene_id     | sample_1 sample_2 status  value_1  value_2 log2_fold_change test_stat p_value q_value significant |
|-------------|--------------------------|---------|---------|-------------|-----------------|---------|--------|-------------|-----------|
| XLOC_000001 | FL LL                    NOTEST 37.6272 96.7472 1.36245e+00  -1.42753e+00 0.1534270 1.000000 no       |
| XLOC_000002 | FL LL                    OK 135.8230 155.1970 1.92379e-01  -8.20334e-01 0.4120260 0.829944 no       |
| XLOC_000003 | FL LL                    NOTEST 0.0000 30.2349 1.79769e+308  1.79769e+308 0.0786496 1.000000 no       |
| XLOC_000004 | FL LL                    NOTEST 0.0000 22.0152 1.79769e+308  1.79769e+308 0.2970490 1.000000 no       |
| XLOC_000005 | FL LL                    NOTEST 0.0000 71.7574 44.4741  -6.90162e-01  5.89906e-01 0.5755430 1.000000 no       |
```
Inspect the Differentially Expressed Genes

```r
> cuff_data
CuffSet instance with:
  2 samples
  2445 genes
  4914 isoforms
  3047 TSS
  1578 CDS
  2445 promoters
  3047 splicing
  1351 relCDS
> gene_diff_data <- diffData(genes(cuff_data))
> sig_gene_data <- subset(gene_diff_data, (significant == 'yes'))
> nrow(sig_gene_data)
[1] 15
> sig_gene_data
  gene_id sample_1 sample_2 status  value_1  value_2 log2_fold_change test_stat p_value q_value significant
1   151  XLOC_000151       FL       LL     OK  315.986  190.767        -0.728054   3.96393 7.37261e-05 7.53235e-03         yes
3   509  XLOC_000509       FL       LL     OK 3619.320  4752.600        -0.392999  -3.55611 3.76389e-04 1.92272e-02         yes
4   581  XLOC_000581       FL       LL     OK  291.897  712.350         1.287130  -4.82627 1.39117e-06 2.84262e-04         yes
5   640  XLOC_000640       FL       LL     OK 1500.080 2234.310         0.574785  -4.70454 1.39117e-06 2.84262e-04         yes
6   711  XLOC_000711       FL       LL     OK  209.051  385.690         0.883583  -3.84734 1.19406e-04 9.14950e-03         yes
7   818  XLOC_000818       FL       LL     OK  259.468  592.153         1.190420  -3.65196 2.60251e-04 1.59534e-02         yes
8   875  XLOC_000875       FL       LL     OK  378.188  786.100         1.055610  -8.89649 0.00000e+00 0.00000e+00         yes
9  1059 XLOC_001059       FL       LL     OK 3713.950  5652.920        -0.606044  -3.85972 9.14950e-03         yes
10 1315 XLOC_001315       FL       LL     OK  487.095  772.977         0.666222  -3.56953 3.57626e-04 1.92272e-02         yes
11 1494 XLOC_001494       FL       LL     OK  214.394  428.568         0.999259  -3.24377 1.17959e-03 4.82059e-02         yes
12 1708 XLOC_001708       FL       LL     OK  4509.900  6781.980         1.190420  -3.65196 2.60251e-04 1.59534e-02         yes
13 1826 XLOC_001826       FL       LL     OK 1121.490  496.331        -1.176040   6.86799 6.51124e-12 1.99569e-09         yes
14 1875 XLOC_001875       FL       LL     OK  369.099  197.538        -0.901878   3.33496 8.53125e-04 3.75475e-02         yes
15 2164 XLOC_002164       FL       LL     OK  366.506  156.929        -1.223730  3.80781 9.54943e-04         yes
```
Inspect the Differentially Expressed Transcripts

```r
> isoform_diff_data <- diffData(isoforms(cuff_data), 'FL', 'LL')
> sig_isoform_data <- subset(isoform_diff_data, (significant == 'yes'))
> nrow(sig_isoform_data)
[1] 72
> head(sig_isoform_data, 20)

<table>
<thead>
<tr>
<th>isoform_id</th>
<th>isoform_id.1</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>log2_fold_change</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCONS_00000036</td>
<td>TCONS_00000036</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>184.9200</td>
<td>0.0000</td>
<td>-1.79769e+308</td>
<td>-1.79769e+308</td>
<td>1.13350e-05</td>
<td>3.50250e-04</td>
<td>yes</td>
</tr>
<tr>
<td>TCONS_00000058</td>
<td>TCONS_00000058</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>156.3640</td>
<td>0.0000</td>
<td>-1.79769e+308</td>
<td>-1.79769e+308</td>
<td>2.00928e-03</td>
<td>3.24290e-02</td>
<td>yes</td>
</tr>
<tr>
<td>TCONS_000000186</td>
<td>TCONS_000000186</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>55.8696</td>
<td>239.2600</td>
<td>2.09844e+00</td>
<td>-4.11086e+00</td>
<td>3.91484e-05</td>
<td>9.36945e-04</td>
<td>yes</td>
</tr>
<tr>
<td>TCONS_000000328</td>
<td>TCONS_000000328</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>190.1450</td>
<td>46.9916</td>
<td>-2.01663e+00</td>
<td>4.14806e+00</td>
<td>3.35304e-05</td>
<td>8.28871e-04</td>
<td>yes</td>
</tr>
<tr>
<td>TCONS_000000458</td>
<td>TCONS_000000458</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>90.6414</td>
<td>191.2480</td>
<td>1.07720e+00</td>
<td>-2.98451e+00</td>
<td>2.84032e-03</td>
<td>3.02641e-02</td>
<td>yes</td>
</tr>
</tbody>
</table>

# remaining rows...
```
Inspect the Differentially Expressed TSS Groups (optional)

```r
> tss_diff_data <- diffData(TSS(cuff_data), 'FL', 'LL')
> sig_tss_data <- subset(tss_diff_data, (significant == 'yes'))
> nrow(sig_tss_data)
[1] 26
> sig_tss_data

<table>
<thead>
<tr>
<th>TSS_group_id</th>
<th>TSS_group_id.1</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>log2_fold_change</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>TSS1028</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>149.2180</td>
<td>0.0000</td>
<td>-1.79769e+308</td>
<td>7.16239e-05</td>
<td>4.24930e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>TSS1103</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>378.1880</td>
<td>786.1000</td>
<td>1.05561e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>TSS1139</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>518.8100</td>
<td>1450.4400</td>
<td>1.48321e+00</td>
<td>-3.40323e+00</td>
<td>6.65948e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>368</td>
<td>TSS1329</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>1692.8200</td>
<td>3086.8200</td>
<td>8.66697e+00</td>
<td>-5.53756e+00</td>
<td>3.06716e-08</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>653</td>
<td>TSS1586</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>2214.9800</td>
<td>1726.3500</td>
<td>-3.59568e+00</td>
<td>3.10359e+00</td>
<td>1.91186e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>685</td>
<td>TSS1614</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>320.7450</td>
<td>121.4850</td>
<td>-1.40065e+00</td>
<td>3.82435e+00</td>
<td>3.1117e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>697</td>
<td>TSS1625</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>232.5470</td>
<td>64.4287</td>
<td>-1.85175e+00</td>
<td>3.28215e+00</td>
<td>1.03019e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>882</td>
<td>TSS1792</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>4292.0800</td>
<td>5177.5000</td>
<td>8.66697e+00</td>
<td>-3.17760e+00</td>
<td>1.48498e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>956</td>
<td>TSS1859</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>214.0030</td>
<td>415.5230</td>
<td>9.57295e-01</td>
<td>-3.60364e+00</td>
<td>3.17025e-02</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1035</td>
<td>TSS193</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>315.9860</td>
<td>190.7670</td>
<td>-7.28054e+00</td>
<td>7.37261e-05</td>
<td>4.24930e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1172</td>
<td>TSS2052</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>316.8100</td>
<td>154.9110</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1242</td>
<td>TSS2115</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>2850.3400</td>
<td>4538.4600</td>
<td>6.71070e+00</td>
<td>-4.90994e+00</td>
<td>4.29626e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1243</td>
<td>TSS2116</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>429.4060</td>
<td>104.6850</td>
<td>-2.03629e+00</td>
<td>3.34006e+00</td>
<td>8.37602e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1331</td>
<td>TSS2196</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>2261.0030</td>
<td>1349.8600</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1381</td>
<td>TSS2240</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>2261.0030</td>
<td>1349.8600</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1417</td>
<td>TSS2273</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>316.8100</td>
<td>154.9110</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1489</td>
<td>TSS2338</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>316.8100</td>
<td>154.9110</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1518</td>
<td>TSS2364</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>316.8100</td>
<td>154.9110</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1780</td>
<td>TSS260</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>90.6414</td>
<td>220.4320</td>
<td>1.28209e+00</td>
<td>-3.79027e+00</td>
<td>1.50481e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1894</td>
<td>TSS2702</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>366.5060</td>
<td>156.9290</td>
<td>1.28209e+00</td>
<td>-3.79027e+00</td>
<td>1.50481e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2413</td>
<td>TSS427</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>2261.0030</td>
<td>1349.8600</td>
<td>1.79769e+00</td>
<td>-8.88649e+00</td>
<td>0.00000e+00</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2503</td>
<td>TSS508</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>817.5910</td>
<td>300.9690</td>
<td>-1.44690e+00</td>
<td>5.33737e+00</td>
<td>9.43041e-08</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2576</td>
<td>TSS574</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>173.9980</td>
<td>313.3070</td>
<td>-1.79769e+00</td>
<td>1.79769e+00</td>
<td>2.50821e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2590</td>
<td>TSS587</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>1364.3900</td>
<td>313.3070</td>
<td>-2.12261e+00</td>
<td>3.87503e+00</td>
<td>1.06612e-04</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2747</td>
<td>TSS728</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>291.8970</td>
<td>659.7650</td>
<td>1.17649e-01</td>
<td>-4.19242e+00</td>
<td>2.75992e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2830</td>
<td>TSS802</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>1500.0800</td>
<td>2234.3100</td>
<td>5.74785e-01</td>
<td>-4.70454e+00</td>
<td>2.54439e-06</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
>
```
Inspect the Differentially Expressed Coding Sequences (optional)

```r
> cds_diff_data <- diffData(CDS(cuff_data), 'FL', 'LL')

> sig_cds_data <- subset(cds_diff_data, (significant=='yes'))

> nrow(sig_cds_data)
[1] 0

> options(width=300)
> sig_cds_data

   CDS_id     CDS_id.1    sample_1     sample_2   status   value_1   value_2 log2_fold_change test_stat    p_value   q_value significant
[1] <0 rows> (or 0-length row.names)
```
## Inspect the Differentially Spliced TSS Groups

```r
> sig_splicing_data <- subset(sig_splicing_data, (significant == 'yes'))
> nrow(sig_splicing_data)
```

### sig_splicing_data

<table>
<thead>
<tr>
<th>TSS_group_id</th>
<th>gene_id</th>
<th>sample_1 sample_2 status</th>
<th>value_1</th>
<th>value_2</th>
<th>JS_dist</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1484</td>
<td>TSS2332</td>
<td>XLOC_001873</td>
<td>FL</td>
<td>LL</td>
<td>0.213655</td>
<td>0.009315</td>
<td>4.93062e-03</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1571</td>
<td>TSS2411</td>
<td>XLOC_001937</td>
<td>FL</td>
<td>LL</td>
<td>0.353909</td>
<td>0.006755</td>
<td>1.63424e-02</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1609</td>
<td>TSS2446</td>
<td>XLOC_001962</td>
<td>FL</td>
<td>LL</td>
<td>0.351265</td>
<td>0.015730</td>
<td>3.14600e-02</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1669</td>
<td>TSS250</td>
<td>XLOC_000199</td>
<td>FL</td>
<td>LL</td>
<td>0.265453</td>
<td>0.007515</td>
<td>1.78357e-02</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1819</td>
<td>TSS2635</td>
<td>XLOC_002109</td>
<td>FL</td>
<td>LL</td>
<td>0.388909</td>
<td>0.000045</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>TSS2708</td>
<td>XLOC_002170</td>
<td>FL</td>
<td>LL</td>
<td>0.213887</td>
<td>0.000030</td>
<td>1.13429e-02</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>TSS2755</td>
<td>XLOC_002204</td>
<td>FL</td>
<td>LL</td>
<td>0.313655</td>
<td>0.000045</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td>TSS2756</td>
<td>XLOC_002206</td>
<td>FL</td>
<td>LL</td>
<td>0.832410</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2039</td>
<td>TSS2833</td>
<td>XLOC_002266</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2294</td>
<td>TSS32</td>
<td>XLOC_000026</td>
<td>FL</td>
<td>LL</td>
<td>0.234199</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2296</td>
<td>TSS32</td>
<td>XLOC_000028</td>
<td>FL</td>
<td>LL</td>
<td>0.831733</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2376</td>
<td>TSS34</td>
<td>XLOC_000030</td>
<td>FL</td>
<td>LL</td>
<td>0.351265</td>
<td>0.000045</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2458</td>
<td>TSS44</td>
<td>XLOC_000032</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2468</td>
<td>TSS47</td>
<td>XLOC_000039</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2503</td>
<td>TSS50</td>
<td>XLOC_000413</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2552</td>
<td>TSS52</td>
<td>XLOC_000451</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2580</td>
<td>TSS58</td>
<td>XLOC_000465</td>
<td>FL</td>
<td>LL</td>
<td>0.571820</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2736</td>
<td>TSS70</td>
<td>XLOC_000057</td>
<td>FL</td>
<td>LL</td>
<td>0.831733</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2830</td>
<td>TSS80</td>
<td>XLOC_000060</td>
<td>FL</td>
<td>LL</td>
<td>0.831733</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2966</td>
<td>TSS92</td>
<td>XLOC_000071</td>
<td>FL</td>
<td>LL</td>
<td>0.831733</td>
<td>0.000010</td>
<td>4.84211e-05</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
```

---

**Note:** The above data represents a subset of the Differentially Spliced TSS Groups where the significant (p-value) is less than 0.05.
Inspect the Genes with Differential Promoter Usage (optional)

```r
> promoter_diff_data <- distValues(promoters(cuff_data))

> sig_promoter_data <- subset(promoter_diff_data, (significant == 'yes'))

> nrow(sig_promoter_data)
[1] 22

> sig_promoter_data

<table>
<thead>
<tr>
<th>gene_id</th>
<th>sample_1</th>
<th>sample_2</th>
<th>status</th>
<th>value_1</th>
<th>value_2</th>
<th>JS_dist</th>
<th>test_stat</th>
<th>p_value</th>
<th>q_value</th>
<th>significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLOC_000010</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.350489</td>
<td>2.05866e-06</td>
<td>0.000100</td>
<td>0.000750000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000026</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.163359</td>
<td>1.88976e-03</td>
<td>0.003890</td>
<td>0.017426500</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000076</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.341414</td>
<td>2.98360e-03</td>
<td>0.014295</td>
<td>0.048733000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000349</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.101063</td>
<td>6.15428e-03</td>
<td>0.008725</td>
<td>0.034440800</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000370</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.314046</td>
<td>6.62848e-04</td>
<td>0.009325</td>
<td>0.034968700</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000413</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.495559</td>
<td>0.000000e+00</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000462</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.528164</td>
<td>7.99361e-15</td>
<td>0.000015</td>
<td>0.000140625</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000465</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.181609</td>
<td>1.91158e-04</td>
<td>0.003285</td>
<td>0.017075000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000470</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.565331</td>
<td>1.89777e-11</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000620</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.347070</td>
<td>6.35768e-04</td>
<td>0.005325</td>
<td>0.022187500</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000813</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.426070</td>
<td>1.55431e-15</td>
<td>0.000065</td>
<td>0.000541667</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000818</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.595225</td>
<td>1.28807e-05</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_000903</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.494402</td>
<td>2.22045e-16</td>
<td>0.000015</td>
<td>0.000140625</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001055</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.438843</td>
<td>1.09297e-03</td>
<td>0.002580</td>
<td>0.014884600</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001304</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.431044</td>
<td>3.05717e-07</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001335</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.275093</td>
<td>9.07483e-03</td>
<td>0.013430</td>
<td>0.047964300</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001450</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.301625</td>
<td>5.66491e-05</td>
<td>0.003950</td>
<td>0.017426500</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001656</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.169095</td>
<td>1.31302e-10</td>
<td>0.000535</td>
<td>0.003647730</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001658</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.829814</td>
<td>0.000000e+00</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_001767</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.503301</td>
<td>0.000000e+00</td>
<td>0.000010</td>
<td>0.000125000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_002177</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.304342</td>
<td>1.76425e-05</td>
<td>0.001160</td>
<td>0.007250000</td>
<td>yes</td>
</tr>
<tr>
<td>XLOC_002265</td>
<td>FL</td>
<td>LL</td>
<td>OK</td>
<td>0</td>
<td>0</td>
<td>0.512821</td>
<td>1.04072e-05</td>
<td>0.003415</td>
<td>0.017075000</td>
<td>yes</td>
</tr>
</tbody>
</table>
```
Inspect the Genes with Differential CDS Output (optional)

```r
> relCDS_diff_data <- distValues(relCDS(cuff_data))

> sig_relCDS_data <- subset(relCDS_diff_data, (significant == 'yes'))

> nrow(sig_relCDS_data)
[1] 2

> sig_relCDS_data
gene_id sample_1 sample_2 status value_1 value_2 JS_dist test_stat p_value q_value significant
1036 XLOC_001767       FL       LL     OK       0       0 0.503301 0.00000e+00   1e-05   4e-05         yes
1269 XLOC_002206       FL       LL     OK       0       0 0.728721 1.33311e-06   1e-05   4e-05         yes
```
Exercise 7

Visualizing the Differential Analysis Results (CummeRbund)
Plot the Distribution of Expression Levels

```r
> library(cairoDevice)
> Cairo()
> csDensity(genes(cuff_data))
```
Boxplot View of Expression Levels

```r
> csBoxplot(genes(cuff_data))
```
Compare the Expression Levels of Genes

> csScatter(genes(cuff_data), 'FL', 'LL')
Create a Volcano Plot to Inspect DE Genes

> csVolcano(genes(cuff_data), 'FL', 'LL')
> pdf("volcano.pdf")
> csVolcano(genes(cuff_data), 'FL', 'LL')
> dev.off()
Plot the Expression Levels for Genes of Interest

```r
> mygene <- getGene(cuff_data, 'MXRA5')
> expressionPlot(mygene)
> mygene <- getGene(cuff_data, 'PROS1')
> expressionPlot(mygene)
```
Barplot of the Expression Levels for Genes of Interest

```r
> mygene <- getGene(cuff_data, 'MXRA5')
> expressionBarplot(mygene)
> mygene <- getGene(cuff_data, 'PROS1')
> expressionBarplot(mygene)
```
Barplot of Isoform Expression Levels for Genes of Interest (optional)

> mygene <- getGene(cuff_data, 'MXRA5')
> expressionBarplot(isoforms(mygene))
> mygene <- getGene(cuff_data, 'PROS1')
> expressionBarplot(isoforms(mygene))
Create Gene Set from Significantly Regulated Genes

```r
> mySigGeneIds <- getSig(cuff_data, alpha=0.05, level = "genes")
> head(mySigGeneIds)

  [1] "XLOC_000151" "XLOC_000206" "XLOC_000509" "XLOC_000581" "XLOC_000640" "XLOC_000711"

> length(mySigGeneIds)

  [1] 15

> mySigGenes <- getGenes(cuff_data, mySigGeneIds)

Getting gene information:
  FPKM
  Differential Expression Data
  Annotation Data
Getting isoforms information:
  FPKM
  Differential Expression Data
  Annotation Data
Getting CDS information:
  FPKM
  Differential Expression Data
  Annotation Data
Getting TSS information:
  FPKM
  Differential Expression Data
  Annotation Data
Getting promoter information:
  distData
Getting splicing information:
  distData
Getting relCDS information:
  distData

> mySigGenes

CuffGeneSet instance for 15 genes

Slots:
  annotation
  fpkm
diff
  isoforms                                      CuffFeatureSet instance of size 68
  TSS                                           CuffFeatureSet instance of size 27
  CDS                                           CuffFeatureSet instance of size 0
  promoters                                     CuffFeatureSet instance of size 15
  splicing                                      CuffFeatureSet instance of size 27
  relCDS                                        CuffFeatureSet instance of size 15
```
Heatmap of Significantly Regulated Genes

By default, the Jensen-Shannon distance is used as the clustering metric.
Barplot of Significantly Regulated Genes

> expressionBarplot(mySigGenes)
 ymax not defined: adjusting position using y instead
>

```r
> expressionBarplot(mySigGenes)
 ymax not defined: adjusting position using y instead
```
Scatter Plot of Significantly Regulated Genes

> csScatter(mySigGenes, 'FL', 'LL', smooth=T)
Using tracking_id, sample_name as id variables
>
Volcano Plot of Significantly Regulated Genes

> csVolcano(mySigGenes, 'FL', 'LL')
Heatmap of Significantly Regulated Isoforms (optional)

By default, the Jensen-Shannon distance is used as the clustering metric.

```r
> csHeatmap(isoforms(mySigGenes), cluster="both")
Using tracking_id, sample_name as id variables
Using  as id variables
> 
```
Barplot of Significantly Regulated Isoforms (optional)

> expressionBarplot(isoforms(mySigGenes))
ymax not defined: adjusting position using y instead
>
Scatter Plot of Significantly Regulated Isoforms (optional)

```r
> csScatter(isoforms(mySigGenes), 'FL', 'LL', smooth=T)
Using tracking_id, sample_name as id variables
```
Volcano Plot of Significantly Regulated Isoforms (optional)

```r
> csVolcano(isoforms(mySigGenes), 'FL', 'LL')
Warning message:
Removed 17 rows containing missing values (geom_point).
> 
```
Map Significantly Regulated Genes for iProXpress

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Effect</th>
<th>Log2 Fold Change (FL:LL)</th>
<th>Log2 p-value (pval)</th>
<th>q-value (qval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1NKB3</td>
<td>decrease</td>
<td>-0.728074</td>
<td>7.48577e-05</td>
<td>0.00764796</td>
</tr>
<tr>
<td>E1C2F8</td>
<td>increase</td>
<td>1.51355</td>
<td>6.47663e-06</td>
<td>0.000794035</td>
</tr>
<tr>
<td>F1NXW6</td>
<td>increase</td>
<td>0.392991</td>
<td>0.000373513</td>
<td>0.0190803</td>
</tr>
<tr>
<td>Q5P3N3</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>0.883563</td>
<td>2.50853e-06</td>
<td>0.000384432</td>
</tr>
<tr>
<td>F1NUC1</td>
<td>increase</td>
<td>1.066211</td>
<td>2.50853e-06</td>
<td>0.000384432</td>
</tr>
<tr>
<td>F1NXW6</td>
<td>increase</td>
<td>0.392991</td>
<td>0.000373513</td>
<td>0.0190803</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
<tr>
<td>F1NUC1</td>
<td>increase</td>
<td>1.066211</td>
<td>2.50853e-06</td>
<td>0.000384432</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
<tr>
<td>F1NTD7</td>
<td>increase</td>
<td>2.1.28712</td>
<td>1.39072e-06</td>
<td>0.000284171</td>
</tr>
</tbody>
</table>

- Copy the output as shown in blue above, and paste it into the text box on the iProXpress web site at the URL below
  [http://pir18.georgetown.edu/iproxpress2/](http://pir18.georgetown.edu/iproxpress2/)
- Click submit button, you can now do GO Slim analysis and other analyses from there.
- Next section of this short course will cover more on this topic using full set of significantly regulated genes.
Summary

• Use Cuffdiff to identify differentially expressed genes/transcripts.
• Use CummeRbund to explore the Cufflinks RNA-Seq output.
References


