RNA-Seq Tutorial 1

John Garbe
Research Informatics Support Systems, MSI
September 23, 2014

Slides available at www.msi.umn.edu/tutorial-materials
RNA-Seq Tutorials

• Tutorial 1
  – RNA-Seq experiment design and analysis
  – Instruction on individual software will be provided in other tutorials

• Tutorial 2
  – Advanced RNA-Seq Analysis topics

• Hands-on tutorials
  – Analyzing human and potato RNA-Seq data using Tophat and Cufflinks in Galaxy
Galaxy.msi.umn.edu

Web-based platform for bioinformatic analysis
Introduction

- Gene expression
- RNA-Seq
- Platform characteristics
Central dogma of molecular biology

92–94% of human genes undergo alternative splicing, 86% with a minor isoform frequency of 15% or more

Introduction

• RNA-Seq
  – High-throughput sequencing of RNA
  – Transcriptome assembly (optional)
    • Qualitative identification of expressed sequence
  – Differential expression analysis
    • Quantitative measurement of transcript expression
Sample 1

mRNA isolation

Fragmentation RNA -> cDNA

Sequence fragment end(s)

Map reads

Calculate transcript abundance

<table>
<thead>
<tr>
<th># of Reads</th>
<th>Gene A</th>
<th>Gene B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reads per kilobase of exon</th>
<th>Gene A</th>
<th>Gene B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Sample 2</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reads per kilobase of exon per million mapped reads (RPKM)</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>.7</td>
<td>.3</td>
<td>6</td>
</tr>
<tr>
<td>Sample 2</td>
<td>.6</td>
<td>.3</td>
<td>12</td>
</tr>
</tbody>
</table>
**a**

Technical replicates

\[ R^2 = 0.96 \]

**b**

- **Intergenic** (3%)
- **Introns** (4%)
- **Exons** (93%)

**c**

Sensitivity and dynamic range

\[ R^2 = 0.99 \]
Experimental Design

- Biological comparison(s)
- Paired-end vs single end reads
- Read length
- Read depth
- Replicates
- Pooling
Experimental design

• Simple designs (Pairwise comparisons)

Two group
Drug effect

Control
Experimental (drug applied)

• Complex designs

Consult a statistician

Two factor
Cancer type X drug

Cancer sub-type 1
With drug

Cancer sub-type 2
With drug

Matched-pair

Normal
Cancer
Experimental design

• What are my goals?
  – Transcriptome assembly?
  – Differential expression analysis?
  – Identify rare transcripts?

• What are the characteristics of my system?
  – Large, complex genome?
  – Introns and high degree of alternative splicing?
  – No reference genome or transcriptome?
Experimental design

Cost = library prep ($141/sample) + sequencing

<table>
<thead>
<tr>
<th>HiSeq 2000 Rates</th>
<th>10 million reads (1/16 lane)</th>
<th>20 million reads (1/8 lane)</th>
<th>40 million reads (1/4 lane)</th>
<th>80 million reads (1/2 lane)</th>
<th>160 million reads (1 lane)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-read (1x50 cycles)</td>
<td>$208</td>
<td>$275</td>
<td>$410</td>
<td>$680</td>
<td>$1,219</td>
</tr>
<tr>
<td>$1,078/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-read (1x100 cycles)</td>
<td>$225</td>
<td>$310</td>
<td>$480</td>
<td>$820</td>
<td>$1,499</td>
</tr>
<tr>
<td>$1,358/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired-end read (2x50 cycles)</td>
<td>$250</td>
<td>$360</td>
<td>$579</td>
<td>$1,018</td>
<td>$1,1895</td>
</tr>
<tr>
<td>$1,754/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired-end (2x100 cycles)</td>
<td>$268</td>
<td>$395</td>
<td>$649</td>
<td>$1,157</td>
<td>$2,147</td>
</tr>
<tr>
<td>$2,033/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UMGC RNA-Seq Price list

- Valid September, 2014
- Volume discounts available
# Experimental design

<table>
<thead>
<tr>
<th>HiSeq 2000 Rates</th>
<th>10 million reads (1/16 lane)</th>
<th>20 million reads (1/8 lane)</th>
<th>40 million reads (1/4 lane)</th>
<th>80 million reads (1/2 lane)</th>
<th>160 million reads (1 lane)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-read (1x50 cycles)</td>
<td>$208</td>
<td>$275</td>
<td>$410</td>
<td>$680</td>
<td>$1,219</td>
</tr>
<tr>
<td>$1,078/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-read (1x100 cycles)</td>
<td>$225</td>
<td>$310</td>
<td>$480</td>
<td>$820</td>
<td>$1,499</td>
</tr>
<tr>
<td>$1,358/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired-end read (2x50 cycles)</td>
<td>$250</td>
<td>$360</td>
<td>$579</td>
<td>$1,018</td>
<td>$1,1895</td>
</tr>
<tr>
<td>$1,754/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired-end (2x100 cycles)</td>
<td>$268</td>
<td>$395</td>
<td>$649</td>
<td>$1,157</td>
<td>$2,147</td>
</tr>
<tr>
<td>$2,033/lane</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10 million reads per sample, 50bp single-end reads

- Small genomes with no alternative splicing
## Experimental design

<table>
<thead>
<tr>
<th>HiSeq 2000 Rates</th>
<th>10 million reads (1/16 lane)</th>
<th>20 million reads (1/8 lane)</th>
<th>40 million reads (1/4 lane)</th>
<th>80 million reads (1/2 lane)</th>
<th>160 million reads (1 lane)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-read (1x50 cycles) $1,078/lane</td>
<td>$208</td>
<td>$275</td>
<td>$410</td>
<td>$680</td>
<td>$1,219</td>
</tr>
<tr>
<td>Single-read (1x100 cycles) $1,358/lane</td>
<td>$225</td>
<td>$310</td>
<td>$480</td>
<td>$820</td>
<td>$1,499</td>
</tr>
<tr>
<td>Paired-end read (2x50 cycles) $1,754/lane</td>
<td>$250</td>
<td>$360</td>
<td>$579</td>
<td>$1,018</td>
<td>$1,1895</td>
</tr>
<tr>
<td>Paired-end (2x100 cycles) $2,033/lane</td>
<td>$268</td>
<td>$395</td>
<td>$649</td>
<td>$1,157</td>
<td>$2,147</td>
</tr>
</tbody>
</table>

20 million reads per sample, 50bp paired-end reads

- Mammalian genomes (large transcriptome, alternative splicing, gene duplication)
Experimental design

<table>
<thead>
<tr>
<th>HiSeq 2000 Rates</th>
<th>10 million reads (1/16 lane)</th>
<th>20 million reads (1/8 lane)</th>
<th>40 million reads (1/4 lane)</th>
<th>80 million reads (1/2 lane)</th>
<th>160 million reads (1 lane)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-read (1x50 cycles) $1,078/lane</td>
<td>$208</td>
<td>$275</td>
<td>$410</td>
<td>$680</td>
<td>$1,219</td>
</tr>
<tr>
<td>Single-read (1x100 cycles) $1,358/lane</td>
<td>$225</td>
<td>$310</td>
<td>$480</td>
<td>$820</td>
<td>$1,499</td>
</tr>
<tr>
<td>Paired-end read (2x50 cycles) $1,754/lane</td>
<td>$250</td>
<td>$360</td>
<td>$579</td>
<td>$1,018</td>
<td>$1,1895</td>
</tr>
<tr>
<td>Paired-end (2x100 cycles) $2,033/lane</td>
<td>$268</td>
<td>$395</td>
<td>$649</td>
<td>$1,157</td>
<td>$2,147</td>
</tr>
</tbody>
</table>

40-160 million reads per sample, 100bp paired-end reads
• Transcriptome Assembly (100X coverage of transcriptome)

50bp Paired-end >> 100bp Single-end
Experimental design

• Technical replicates
  – Not needed: low technical variation
    • Minimize batch effects
    • Randomize sample order

• Biological replicates
  – Not needed for transcriptome assembly
  – Essential for differential expression analysis
  – Difficult to estimate
    • 3+ for cell lines
    • 5+ for inbred lines
    • 20+ for human samples
Experimental design

• Pooling samples
  – Limited RNA obtainable
    • Multiple pools per group required
  – Transcriptome assembly
Experimental design

RNA-seq: technical variability and sampling
Lauren M McIntyre, Kenneth K Lopiano, Alison M Morse, Victor Amin, Ann L Oberg, Linda J Young and Sergey V Nuzhdin
BMC Genomics 2011, 12:293

Statistical Design and Analysis of RNA Sequencing Data
Paul L. Auer and R. W. Doerge

Calculating Sample Size Estimates for RNA Sequencing Data

subSeq: Determining Appropriate Sequencing Depth Through Efficient Read Subsampling.
David G. Robinson and John D. Storey, Bioinformatics first published online September 3, 2014

Focus on RNA sequencing quality control (SEQC)
Nature Biotechnology 32, vii (2014)
Sequencing

- Platforms
- Library preparation
- Multiplexing
- Sequence reads
Sequencing

- **Illumina sequencing by synthesis**
  - HiSeq2000 (dominant RNA-Seq platform)
  - HiSeq2500
  - MiSeq
    - low throughput, longer reads (2x300), fast turnaround

- **PacBio SMRT** (Mayo Clinic)
  - Very long reads, low throughput, high cost

Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study

Sequencing

• **Library preparation** (Illumina TruSeq protocol for HiSeq)
  – RNA isolation
  – Ploy-A purification
  – Fragmentation
  – cDNA synthesis using random primers
  – Adapter ligation
  – Size selection
  – PCR amplification (~16 cycles)
Sequencing

• Flowcell
  – 8 lanes
  – 160+ Million reads per lane
  – Multiplex 16 or more samples on one lane using barcodes
Sequencing

Adapter  cDNA insert  Adaptor

Barcode

Read 1
“left” read

Index read

Read 2
“right” read (optional)

Adaptor contamination
Sequencing

• Library types
  – Polyadenylated RNA > 200bp (standard method)
  – Small RNA
  – Strand-specific (+$40/sample)
    • Gene-dense genomes (bacteria, archaea, lower eukaryotes)
    • Antisense transcription (higher eukaryotes)
    • De novo transcriptome assembly
  – Low input
  – Total RNA
  – Library capture
Sequence Data Format

- **Data delivery**
  - `/home/PI-groupname/data_release/umgc/hiseq/120318_SN261_0348_A81JUMABXX`
  - Upload to Galaxy

- **File names**
  - `cancer1_GCGCATTA_L002_R1.fastq`
  - `cancer1_GCGCATTA_L002_R2.fastq`

- **Fastq format** (Illumina Casava 1.8.0)
  - 4 lines per read

  - **Read ID**
  - **Sequence**
  - **Quality score**
  - **Phred+33**
  - **Machine ID**
  - **QC Filter flag**
    - Y=bad
    - N=good
  - **barcode**
  - **Read pair #**

  ```
  @HWI-M00262:4:000000000-A0ABC:1:1:18376:2027 1:N:0:AGATC
  TTCAGAGAGAATGAATTGTACGTGCTTTTTTTGT
  +
  +
  =1:?7A7+?77+<<@AC<3,<33@A;<A?A=4=
  ```

http://en.wikipedia.org/wiki/FASTQ_format
Data Quality Control

- Quality assessment
- Trimming and filtering
Data Quality Assessment

• Evaluate read library quality
  – Identify poor/bad samples

• Software
  – FastQC (recommended)
    • Command-line, Java GUI, or Galaxy
Data Quality Assessment

• Trimming: remove bad bases from (end of) read
  – Adaptor sequence
  – Low quality bases

• Filtering: remove bad reads from library
  – Low quality reads
  – Contaminating sequence
  – Low complexity reads (repeats)
  – Short reads
    • Short (< 20bp) reads slow down mapping software
    • Only needed if trimming was performed

• Software
  – Trimmomatic
  – Cutadapt
  – Galaxy, many options including cutadapt (NGS: QC and manipulation)
  – Many others: http://seqanswers.com/wiki/Software/list
Data Quality Assessment - FastQC

Phred 30 = 1 error / 1000 bases
Phred 20 = 1 error / 100 bases
Data Quality Assessment - FastQC

Quality scores across reads

Good

Bad
Filtering needed
Data Quality Assessment - FastQC

**Good**

**Bad**
Data Quality Assessment - FastQC

High level of sequencing adapter contamination, trimming needed

### Overrepresented sequences

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Count</th>
<th>Percentage</th>
<th>Possible Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTATTACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>820428</td>
<td>2.8366639370528275</td>
<td>Illumina Paired End PCR Primer 2 (100% over 43bp)</td>
</tr>
<tr>
<td>GTATACAGATCGGAAGAGCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>749728</td>
<td>2.5922157461699773</td>
<td>Illumina Paired End PCR Primer 2 (100% over 44bp)</td>
</tr>
<tr>
<td>CGGTACAGAGGGAATGCCGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>648852</td>
<td>2.243432780066747</td>
<td>Illumina Paired End Adapter 2 (100% over 31bp)</td>
</tr>
<tr>
<td>GATCGGAAGAGGCTGTACAGAGGGAATGCCGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>176765</td>
<td>0.6111723403310748</td>
<td>Illumina Paired End PCR Primer 2 (97% over 36bp)</td>
</tr>
<tr>
<td>ACGTGCTAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>143840</td>
<td>0.4973327832615156</td>
<td>Illumina Paired End PCR Primer 2 (100% over 43bp)</td>
</tr>
<tr>
<td>GTATACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>124281</td>
<td>0.4297067271727257</td>
<td>Illumina Paired End PCR Primer 2 (100% over 44bp)</td>
</tr>
<tr>
<td>GTATACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>99207</td>
<td>0.3430123917842867</td>
<td>Illumina Paired End PCR Primer 2 (100% over 45bp)</td>
</tr>
<tr>
<td>GATCGGAAGAGGCTGTACAGAGGGAATGCCGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>96289</td>
<td>0.33292322279941655</td>
<td>Illumina Paired End PCR Primer 2 (100% over 50bp)</td>
</tr>
<tr>
<td>CGGAAGAGGCTGTCAGAGGGAATGCCGAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>93842</td>
<td>0.3244626185124245</td>
<td>Illumina Paired End PCR Primer 2 (96% over 33bp)</td>
</tr>
<tr>
<td>CGGTACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>75370</td>
<td>0.2605949101391854</td>
<td>Illumina Paired End PCR Primer 2 (100% over 43bp)</td>
</tr>
<tr>
<td>CGGTACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>63691</td>
<td>0.22021428183196043</td>
<td>Illumina Paired End PCR Primer 2 (100% over 44bp)</td>
</tr>
<tr>
<td>AGTACAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>56765</td>
<td>0.19626734873359242</td>
<td>Illumina Paired End PCR Primer 2 (100% over 46bp)</td>
</tr>
<tr>
<td>TACTGTAAGATCGGAAGAGCCTGTCAGAGAATGCCGAGAGACCTCTCGG</td>
<td>42991</td>
<td>0.14864317078139472</td>
<td>Illumina Paired End PCR Primer 2 (100% over 43bp)</td>
</tr>
</tbody>
</table>
Data Quality Assessment - FastQC

Normal level of sequence duplication in 20 million read mammalian sample

Sequence Duplication Level ≥ 61.84%

%Duplicate relative to unique
Data Quality Assessment - FastQC

Normal sequence bias at beginning of reads due to non-random hybridization of random primers
Data Quality Assessment

• Recommendations
  – Generate quality plots for all read libraries
  – Trim and/or filter data if needed
    • Always trim and filter for de novo transcriptome assembly
  – Regenerate quality plots after trimming and filtering to determine effectiveness
Read Mapping

- Pipeline
- Software
- Input
- Output
Mapping – with reference genome

Reference Genome

Millions of short reads

Spliced aligner

Reads aligned to genome

Abundance estimation

Differential expression analysis
Mapping – with reference genome

- Reference Genome
- Millions of short reads
  - Fastq
- Spliced aligner
  - Unmapped reads
  - Reference splice Junction library
    - Fasta/GTF
  - De novo splice junction library
    - Fasta/GTF
- Reads aligned to genome
  - SAM/BAM
- Abundance estimation
- Differential expression analysis

3 exon gene junction library
Mapping

• Alignment algorithm must be
  – Fast
  – Able to handle SNPs, indels, and sequencing errors
  – Allow for introns for reference genome alignment (spliced alignment)

• Burrows Wheeler Transform (BWT) mappers
  – Faster
  – Few mismatches allowed (< 3)
  – Limited indel detection
  – Spliced: Tophat, MapSplice, STAR (super-fast)
  – Unspliced: BWA, Bowtie

• Hash table mappers
  – Slower
  – More mismatches allowed
  – Indel detection
  – Spliced: GSNAP, MapSplice
  – Unspliced: SHRiMP, Stampy

Systematic evaluation of spliced alignment programs for RNA-seq data
Mapping

• **Input**
  - Fastq read libraries
  - **Reference genome index** *(software-specific: /panfs/roc/riisdb/genomes)*
  - **Insert size mean and stddev** *(for paired-end libraries)*
    • Map library (or a subset) using estimated mean and stddev
    • Calculate empirical mean and stddev
      – Galaxy: NGS Picard: insertion size metrics
      – Cufflinks standard error
    • Re-map library using empirical mean and stddev
Mapping

- **Output**
  - SAM (text) / BAM (binary) alignment files
    - SAMtools – SAM/BAM file manipulation
    - Picard-tools – SAM/BAM file manipulation
  - **Summary statistics** (per read library)
    - % reads with unique alignment
    - % reads with multiple alignments
    - % reads with no alignment
    - % reads properly paired (for paired-end libraries)
Differential Expression

- Cuffdiff
- EdgeR
- Ballgown
**Differential Expression**

- **Cuffdiff (Cufflinks package)**
  - Pairwise comparisons
  - Differential gene, transcript, and primary transcript expression; differential splicing and promoter use
  - Easy to use, well documented
  - Input: transcriptome, SAM/BAM read alignments (abundance estimation built-in)

- **EdgeR**
  - Complex experimental designs using generalized linear model
  - Information sharing among genes (Bayesian gene-wise dispersion estimation)
  - Difficult to use R package
  - Input: raw gene/transcript read counts (calculate abundance with separate software not cufflinks)

- **Ballgown**
  - Differential gene, transcript, and exon expression
  - Complex experimental designs: multi-group, timecourse, continuous covariates; adjust for confounding factors
  - R package, works with Cufflinks
Differential Expression

• Others
  – DESeq - R package
  – DEXseq – R package

• Abundance estimation
  – HTseq
  – RSEM
Transcriptome Assembly

- Pipeline
- Software
- Input
- Output
**RNA-Seq**
- Reference genome
- Reference transcriptome

**Experimental Design**
- RNA

**Sequencing**
- RNA

**Data Quality Control**
- fastq

**Read mapping**
- Reference Genome
  - fasta
  - SAM/BAM

**Reference Transcriptome**
- GFF/GTF

**Differential Expression Analysis**
- GFF/GTF

---

**RNA-Seq**
- Reference genome
- No reference transcriptome

**Experimental Design**
- RNA

**Sequencing**
- RNA

**Data Quality Control**
- fastq

**Read mapping**
- Reference Genome
  - fasta
  - SAM/BAM

**Transcriptome assembly**
- fasta

**Differential Expression Analysis**
- GFF/GTF

---

**RNA-Seq**
- No reference genome
- No reference transcriptome

**Experimental Design**
- RNA

**Sequencing**
- RNA

**Data Quality Control**
- fastq

**Read mapping**
- Reference Transcriptome
  - SAM/BAM

**Differential Expression Analysis**
- GFF/GTF
Transcriptome Assembly - with reference genome

Reference transcriptome available

- Reads aligned to genome
  - SAM/BAM
  - UCSC
  - Ensembl
  - Reference transcriptome
  - GFF/GTF

- Abundance estimation

No/poor reference transcriptome available

- Reads aligned to genome
  - With de novo junction library

- Assembler
  - De novo transcriptome
  - GFF/GTF

- Annotate

- Abundance estimation
Transcriptome Assembly -with reference genome

- Reference genome based assembly
  - Cufflinks, Scripture
- Reference annotation based assembly
  - Cufflinks
- Transcriptome comparison
  - Cuffcompare
- Transcriptome Annotation
  - Generate cDNA fasta from annotation (Cufflinks’ gffread program)
  - Align to library of known cDNA (RefSeq, GenBank)

Assessment of transcript reconstruction methods for RNA-seq
Transcriptome Assembly – no reference genome

- Millions of short reads
- Unspliced Aligner
- Assembler
- Abundance estimation
- Differential expression analysis

- Fastq
- SAM/BAM
- Trinity
- Trans-ABySS
- Velvet/oasis

- Computationally intensive

De novo Transcriptome
Further Reading

Bioinformatics for High Throughput Sequencing
Rodríguez-Ezpeleta, Naiara.; Hackenberg, Michael.; Aransay, Ana M.;
SpringerLink New York, NY : Springer c2012

RNA sequencing: advances, challenges and opportunities
Fatih Ozsolak1 & Patrice M. Milos1
Nature Reviews Genetics 12, 87-98 (February 2011)

Computational methods for transcriptome annotation and quantification using RNA-seq
Manuel Garber, Manfred G Grabherr, Mitchell Guttman & Cole Trapnell
Nature Methods 8, 469–477 (2011)

Next-generation transcriptome assembly
Jeffrey A. Martin & Zhong Wang
Nature Reviews Genetics 12, 671-682 (October 2011)

Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks
Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David Kelley, Harold Pimentel, Steven Salzberg, John L Rinn & Lior Pachter

SEQanswers.com
biostar.stackexchange.com

Online access through U library
Table of RNA-Seq software
Popular bioinformatics forums
Questions / Discussion