Protocol

# An RNA-Seq Protocol for Differential Expression Analysis

Nick D.L. Owens,[1,2] Elena De Domenico,[1,2] and Michael J. Gilchrist[1,3]

[1]*The Francis Crick Institute, NW1 1ST London, United Kingdom*

Here we consider RNA-Seq, used to measure global gene expression through RNA fragmentation, capture, sequencing, and subsequent computational analysis. *Xenopus*, with its large number of RNA-rich, synchronously developing, and accessible embryos, is an excellent model organism for exploiting the power of high-throughput sequencing to understand gene expression during development. Here we present a standard RNA-Seq protocol for performing two-state differential gene expression analysis (between groups of replicates of control and treated embryos) using Illumina sequencing. Samples contain multiple whole embryos, and polyadenylated mRNA is measured under relative normalization. The protocol is divided into two parts: wet-lab processes to prepare samples for sequencing and downstream computational analysis including quality control, quantification of gene expression, and differential expression.

## MATERIALS

**It is essential that you consult the appropriate Material Safety Data Sheets and your Institution's Environmental Health and Safety Office for proper handling of equipment and hazardous materials used in this protocol.**

### Reagents

Chloroform

Developmentally synchronized *Xenopus* embryos at stage(s) appropriate for the research question being addressed

  *This protocol is appropriate for both* X. laevis *and* X. tropicalis.

Ethanol (75%–80%, freshly prepared)

Isopropanol

Liquid nitrogen

Lithium chloride (10 M)

Qubit RNA HS Assay kit (Thermo Fisher Scientific)

RNase*Zap* Decontamination Solution (Invitrogen)

TRIzol (Thermo Fisher Scientific)

TruSeq Stranded mRNA Kit (Illumina)

Water, nuclease-free

### Equipment

Benchtop fluorometer (Qubit fluorometer, Thermo Fisher Scientific)

Capillary electrophoresis machine to analyze RNA, DNA, and protein (Agilent 2100 Bioanalyzer)

Glass Pasteur pipette

Magnetic stand

Microcentrifuge tubes (RNAse-free, 1.5-mL)

Microvolume spectrophotometer (NanoDrop, Thermo Fisher Scientific)

> *This item is optional (see Step 23). It can also replace the Qubit, although this is not recommended.*

PCR tubes (RNase-free)

Pestle for 1.5-mL microcentrifuge tube (optional; see Step 2)

Refrigerated microcentrifuge

Thermocycler

Vortex mixer

*Xenopus* genome and/or *Xenopus* transcriptome reference data

### Software

Bowtie2

FastQC

Integrative Genomics Viewer (for visualization if required)

Kallisto (optional mapping software; see Step 36)

R

R packages DESeq2 and tximport

RSEM

STAR

## METHOD

### Wet-Lab Processes

*Before starting this protocol, it is necessary to decide on the experimental design, particularly how many biological replicates in each state (experimental condition). There is a necessary tradeoff between sequencing depth and numbers of biological replicates, with additional replicates preferred over deeper sequencing (Liu et al. 2014), but a minimum of ~10 million reads per sample. We recommend 15–20 million reads per sample, with at least four biological replicates per state (see Schurch et al. 2016).*

*Wet lab processes include sample collection, RNA extraction, and sequencing library preparation. Care should be taken to make sure the methods and kits used match your samples' RNA content. Quality controls are available at various stages, and poorly performing samples should be excluded as necessary.*

*While working with RNA, it is critical to avoid RNase contamination by using sterile, RNase-free solutions and plastic ware. RNase/DNase decontamination solution must be used on all work surfaces and equipment before starting.*

#### Sample Collection

1. Prepare embryos via in vitro fertilization as described (see Protocol: **Egg Collection and In Vitro Fertilization of the Western Clawed Frog *Xenopus tropicalis*** [Showell and Conlon 2009]).

2. When the embryos have reached the desired stage(s), collect groups of 10 embryos per sample in RNAse-free tubes using a glass Pasteur pipette.

3. Remove as much liquid as possible and snap freeze the tube in liquid nitrogen. Store the sample at −80°C.

   > *If samples are collected over a period of time, it may help to use an RNA stabilizing agent.*

#### RNA Extraction

*High-quality RNA is necessary for accurate expression quantification. Avoid overamplification of low-RNA samples during library preparation as this can lead to artifacts and spurious differential expression. RNA extraction should be performed consistently, preferably in a single batch with randomized processing order. Several RNA extraction protocols are available; here we use Trizol extraction followed by precipitation with LiCl.*

4. Thaw the tubes of collected embryo samples on ice and add 1 mL TRIzol to each tube. If RNA stabilizing agent has been used during the collection process, follow the manufacturer's instructions before starting.

5. Homogenize the embryos with a pipette, a pestle, or using a vortexer. Leave homogenate at room temperature for 5 min.

6. Add 200 µL chloroform.

7. Mix by tipping the tube until the solution is homogenous.

8. Centrifuge the tube at top speed in a microcentrifuge for 30 min at 4°C.

9. Transfer the top (aqueous) layer to a new RNase-free tube; avoid touching the white layer (protein) in the middle.

10. Add an equal volume of isopropanol.

11. Mix, tipping the tube and place for at least 30 min at −20°C.

12. Centrifuge the tube at top speed in a microcentrifuge for 20 min at 4°C.

13. Pour off the liquid, taking care not to lose the pellet. Add 500 µL of 75% ethanol and mix.

14. Centrifuge the tube at top speed in a microcentrifuge for 20 min at 4°C.

15. Remove all ethanol with a pipette and allow the pellet to dry with the tube lid open until it become translucent.

16. Resuspend the pellet in 40 µL nuclease free water.

17. Add 40 µL 10 M LiCl and place overnight at −20°C (or at least 1 h).

18. Centrifuge the tube at top speed in a microcentrifuge for 30 min at 4°C.

19. Pour off the liquid, taking care not to lose the pellet. Add 200 µL of 75% ethanol and mix.

20. Centrifuge the tube at top speed in a microcentrifuge for 20 min at 4°C.

21. Air-dry the pellet as in Step 15 and resuspend, initially, in 20 µL nuclease-free water. Once the RNA concentration has been assessed the sample can be further diluted if required.

22. Measure the concentration of extracted RNA in the samples using the Qubit according to the manufacturer's instructions. Consult the library preparation guide (see below) for the acceptable range of RNA input concentrations and dilute the samples accordingly.

    *If the samples cannot be brought within this range, consult the manufacturers' guidelines for a more appropriate kit.*

23. Assess the quality of the extracted RNA samples using the Bioanalyzer to measure RNA Integrity Numbers (RIN) for each sample.

    *Any sample with a RIN <8 should be discarded, or will run a serious risk of generating sequencing artifacts.*

    *In addition, the Nanodrop may be used to measure the $A_{260}/A_{280}$ ratio, which estimates the purity of the RNA. Samples should lie in the range 1.8–2.0.*

    *Samples falling outside acceptable thresholds may be taken through sequencing with the risk of generating unacceptable data and should therefore be examined carefully at postquantification quality control (QC).*

### Sequencing Library Construction

24. Prepare the sequencing libraries.

    *The number of samples to multiplex at this stage will depend on the yield per lane of the sequencer being used. The sequencing facility should be able to provide guidance based on the number of samples and the desired number of reads per sample.*

    *For most applications, enrichment of polyadenylated mRNA from total RNA using poly-T oligo-attached magnetic beads is suitable. The resulting data will comprise coding and noncoding polyadenylated RNAs. We recommend following the Illumina Stranded mRNA Sequencing Sample Preparation guide*

*(TruSeq Stranded mRNA Sample Prep Guide), which will generate libraries from samples with 100–1000 ng total RNA. Let the beads from the TruSeq kit stand at room temperature for at least 30 min before use.*

### Library Quality Check

25. Before sequencing, check the fragment size distribution of the library using the Bioanalyzer. Follow the manufacturer's instructions and study the graphical output for the fragment size distribution.

    *There should be a well-separated broad central peak with a maximum around 250 bp. Additional significant smaller peaks around 120–130 bp may represent adapter dimers and should be removed by repeating the "Clean Up PCR" step in the TruSeq protocol.*

    *Library concentration can also be assessed using qRT-PCR (Hawkins and Guest 2018).*

### Sequencing the Library

26. Send the libraries to a sequencing facility.

    *Sequencing read length will probably be determined by this facility. Sequencing depth will be determined by the number of samples multiplexed together. Helpful considerations about these choices can be found on the Illumina website: https://support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html.*

    *Sequencing data will likely be returned to you as a set of fastq files annotated by sample name.*

## Downstream Computational Analysis

*Computational analysis is usually performed at the command line. Reads are mapped to a reference transcriptome or genome, and then quantified to gene counts per gene or transcript. These are then used for differential expression analysis. For mapping and quantification we recommend the RSEM + STAR (Li and Dewey 2011) pipeline. This approach gives accurate quantification but is computationally intensive (Dobin et al. 2013). An alternative, less computationally intensive approach, also presented here, uses Kallisto (Bray et al. 2016) and can be run on a laptop. New methods continue to be developed, so it is advisable to read recent reviews (e.g., Conesa et al. 2016). We recommend the Linux or Mac operating systems to run the tools described here, although some tools can be run on Windows.*

### Postsequencing Quality Control

27. Use FastQC (Andrews and Others 2010) to detect and provide evidence of problematic sequencing data. Run this postsequencing quality assessment as follows:

    ```
    $> fastqc sample_01.fastq.gz
    ```

    *Postsequencing quality assessment can identify problematic libraries: such as those with low quality base-call scores, a shift from the expected GC-content (∼40% in Xenopus), or overrepresented adapter sequences.*

28. Because failure of quality tests does not guarantee a low-quality sample and vice-versa, proceed with analysis and discard samples based on postquantification QC.

### Read Alignment to Reference Sequence Data

*Read alignment (mapping) assigns reads to their possible reference sequence locations. Mapping to a reference transcriptome is computationally efficient but restricted to defined transcript models and may perform poorly with repetitive sequences. Mapping to a reference genome can be performed with or without aid from reference gene models. Genomic alignment can be visualized with the UCSC genome browser (Kent et al. 2002) or with the integrative genomics viewer (IGV) (Thorvaldsdóttir et al. 2013).*

29. Perform read alignment using the splice-aware (capable of spanning exon–exon junctions) genomic aligner, STAR (Dobin et al. 2013). If gene expression is to be quantified by RSEM, skip this section and proceed with the RSEM alignment/quantification described below. For accurate quantification it is important to retain ambiguously mapping reads from the initial mapping stage. This is particularly important for the homeologous paired genes in *X. laevis*.

N.D.L. Owens et al.

> *For transcriptome mapping we recommend Bowtie 2 (Langmead and Salzberg 2012). The quantification tool RSEM (Li and Dewey 2011) (described below) is able to perform transcriptomic (via Bowtie 2) or genomic (via STAR) alignments and quantification in a single command. We describe Bowtie 2 only as part of the RSEM pipeline.*

30. Proceed with the two stages of alignment: one-off construction of an *index* from reference sequences, followed by mapping reads from each sample using this index.

31. Download *Xenopus* reference sequence files from Xenbase (Karimi et al. 2018). Create an index from the genome reference file:

```
$> STAR --runThreadN [N] --runMode genomeGenerate
        --genomeDir <directory> --genomeFastaFiles <xen_genome.fa>
```

> *In this and the following text, basic commands are shown with a placeholder for command line [options], which are expanded underneath, and <file> denotes a required input file.*

32. For genomes with large numbers of small scaffolds, such as current *X. tropicalis* and *X. laevis* genome builds, STAR recommends setting "`--genomeChrBinNbits=min(18, log2(GenomeLength/NumberOfReferences))`".

> *STAR index creation requires memory proportional to the number of sequences in the genome assembly. For the* X. laevis *genome v9.2 of ~2.7 × 10⁹ bp over ~108,000 sequences this implies* `--genomeChrBinNbits 15`.

33. Perform alignment as follows:

```
$> STAR [options] --outFileNamePrefix <directory_output> --genomeDir <directory_genome> \
                          --readFilesIn sample_01_r1.fastq.gz sample_01_r2.fastq.gz
```

| [options] | Explanation |
|---|---|
| --runThreadN <N> | Number of threads (processors) |
| --readFilesCommand zcat | Compressed FASTQ files (*.gz format) |
| --sjdbGTFfile <gtf_file.gtf> | Annotation for guiding genomic alignment |
| --outFilterMultimapNMax 200 | Report max 200 ambiguous alignments |
| --outSAMtype BAM Unsorted | Output alignments in BAM format |
| --quantMode GeneCounts | Count reads mapping uniquely to genes, see below. |

### Quantification of Gene Expression Data

*Here we describe quantification of gene expression directly from reads and read alignments. Transcript isoform usage can also be quantified with the tools here, but the specifics are beyond the scope of this protocol.*

*We recommend that quantification should include statistical assignment of ambiguous reads, as performed by RSEM/Kallisto. Alternatively, one can count total unique aligning reads for each gene, and although this will underestimate total reads mapping to certain genes (especially* X. laevis *homeologs) it can be robust in quantifying genes with pathogenic repeats. This is done by STAR with the option "--quantMode GeneCounts".*

#### Recommended Quantification: RSEM + STAR

*RSEM quantifies transcript/gene expression from genomic or transcriptomic alignments; the associated pipeline generates the required alignments as necessary. We recommend RSEM+STAR alignment, as it is the current gold standard for RNA-Seq quantification.*

34. Build an RSEM reference as follows (required for all uses of RSEM):

```
$> rsem-prepare-reference [options] <genome/transcriptome>.fa <genome/transcriptome name>
```

| [options] | Explanation |
|---|---|
| -p <N> | Number of threads |
| --gtf <file> | GTF file describing gene models **for genomic alignments** |
| --transcript-to-gene-map <file> | File with rows "gene_id<TAB>transcript_id" **for transcriptomic alignments**. |
| --bowtie2 | Generate bowtie2 index from transcriptome (if desired) |
| --star | Generate STAR index from genome (if desired) |

*For genome builds with a large number of scaffolds such as the* X. laevis *v9.2 it may be necessary to set STAR options as previously described, and so the STAR genome index should be built directly with STAR, not using* rsem-prepare-reference.

35. To quantify expression, run RSEM in one of three modes depending on single/paired end reads (we recommend paired end) and the requirement for alignments:

```
$> rsem-calculate-expression [options]   <reads>.fastq.gz \
                                         <genome/transcriptome name> <sample_name>

$> rsem-calculate-expression [options]   --paired-end <reads_r1>.fastq.gz
                                         <reads_r2>.fastq.gz \ <genome/transcriptome
                                          name> <sample_name>

$> rsem-calculate-expression [options]   --alignments [--paired-end] <alignments>.
                                         bam \
                                         <genome/transcriptome name> <sample_name>
```

| [options] | Explanation |
|---|---|
| -p <N> | Number of threads |
| --paired-end | Paired end reads |
| --bowtie | Align with bowtie2 |
| --star | Align with STAR |
| --star-gzipped-read-file | If STAR is used and reads are compressed (.gz format) |
| --star-output-genome-bam | Output genome bam file generated by STAR, for visualization |
| **Advanced options:** | |
| --calc-pme | Calculate posterior mean estimates, for assessing the impact of ambiguous reads |
| --calc-ci | Calculate 95% confidence intervals for the pme. |
| --estimate-rspd | Estimate the "read start position distribution", account for biases in read coverage over transcripts, e.g. the 3′ bias of polyA+ RNA-Seq. |

### Alternative Quantification: Kallisto

*Kallisto is an alignment free quantification method (Bray et al. 2016), which employs pseudoalignment to identify the set of transcripts consistent with each read. This approach significantly reduces the computational burden, reducing processing times by two orders of magnitude, whilst retaining quantification accuracy.*

36. Construct a kallisto index as follows:

```
$> kallisto index -i <xen_transcriptome> <xen_transcriptome>.fa
```

37. Quantify fastq files using kallisto:

```
$> kallisto quant [options] -i <xen_transcriptome> -o <sample_output_dir> \
        <sample_01_r1.fastq.gz> <sample_01_r2.fastq.gz>
```

| [options] | Explanation |
|---|---|
| -t <N> | Number of threads |
| --single | Input data is single end |
| -l | Estimated fragment length mean, **for single end reads.** |
| -s | Estimated fragment length standard deviation, **for single end reads.** |

### Normalization

38. Normalize RNA-Seq data for sequencing depth (total reads per sample) to compare a gene's expression *between* samples.

39. To compare the expression of *different* genes *within* a sample, further normalize data by transcript/gene length to units of reads (or fragments for paired reads) per kilobase per million: R(F)PKM, or as transcripts per million (TPM).

40. For differential expression analysis do not perform length normalization (Love et al. 2014). Appropriate depth normalization is provided by the package DESeq2 (Love et al. 2014), which we describe below.

### Postquantification QC

41. Identify samples with quality problems by one of the following methods:
    i. Evaluate pairwise sample correlations or total genes expressed above a given threshold.
    ii. Perform principal components analysis (PCA), which can indicate the existence of outlying samples suggesting quality issues. A PCA appropriate for RNA-Seq is offered by DESeq2.

### Differential Expression Analysis

*Differentially expressed genes are characterized by smaller variation in gene expression* within *each condition than* between *multiple conditions.*

42. Options for statistical testing appropriate to count data (counting discrete items, i.e., sequence reads) are as follows:
    i. Use a test designed for count data, such as DESeq2. We recommend DESeq2, which is also more suited to small numbers of replicates (Soneson and Delorenzi 2013; Liu et al. 2014; Zhang et al. 2014; Seyednasrollah et al. 2015).
    ii. Or transform the data to abrogate the technical effects of count data and then apply a noncount data test (Ritchie et al. 2015).

43. Install R (R Core Team 2018) and R packages DESeq2 (Love et al. 2014) and tximport (Soneson et al. 2016).

    *We describe a basic DESeq2 analysis below. More information and examples of usage are provided by DESeq2. The tximport package is used to import RSEM quantifications into DESeq2; it can also import expression data from other sources.*

44. Perform DESeq2 as shown in the following example. In this example, we compare two conditions A and B, with three replicates per condition and sample names Sample_N for $N = 1,2,…,6$.

45. Prepare an output folder for each sample from RSEM in a path which we will refer to as dir, such that dir/Sample_N/Sample_N.genes.results is the output quantification of Sample_N.

46. Create a tab or comma delimited file dir/samples.txt that describes the identity of the samples as follows:

Cold Spring Harbor Protocols

www.cshprotocols.org

| Sample | Condition |
|--------|-----------|
| Sample_1 | A |
| Sample_2 | A |
| Sample_3 | A |
| Sample_4 | B |
| Sample_5 | B |
| Sample_6 | B |

47. Use the following R code to load the data and run DESeq2 with default settings comparing condition A vs condition B. Save the results.

```
1    # Load packages
2    library("tximport")                              # For loading RSEM data
3    library("DESeq2")                                # Load DESeq2
4
5
6    # Load Sample tables and quantification files
7    samples <- read.table(file.path(dir, "samples.txt"),    # Load Sample table
     header=TRUE)
8    files <- file.path(dir, samples$Sample,          # Find RSEM
     paste0(samples$Sample, ".genes.results"))        results files
9    names(files) <- samples$Sample
10   txi <- tximport(files, type = "rsem", txIn = FALSE,    # Import RSEM data
     txOut = FALSE)
11   txi$length[txi$length == 0] <- 1                 # RSEM sets some transcript#
                                                       lengths to 0, set these to 1
12
13   # DESeq2
14   dds <- DESeqDataSetFromTximport(txi, colData     # Load Data into DESeq2
     = samples, design = ~ Condition)
15   dds <- DESeq(dds)                                # Run differential
                                                       expression
16   res <- results(dds, name="Condition_B_vs_A")     # Collect results
17   summary(res)                                     # Show summary of results
18
19
20   # Write results
21   resOrdered <- res[order(res$padj),]              # Order results for padj
22   write.csv(as.data.frame(resOrdered),             # Write results
     file=file.path(dir, "deseq_results.csv"))
```

*The result of this analysis will be a comma-separated file describing the results of the differential expression tests for all genes ordered by the adjusted p-value. Columns in the file include: mean expression, fold change between conditions, relevant statistics on the statistical test, p-value and adjusted p-value. The adjusted p-values are derived from the p-values accounting for the number of tests performed and should be used in preference over the p-values.*

## RELATED INFORMATION

A two-state experiment is described here, but more complex designs are possible. For example, *Xenopus* is ideal for time-resolved experimental design to understand the developmental timing of gene expression (Collart et al. 2014; Owens et al. 2016).

Protocols are available for measuring differing types and levels of RNA. Standard polyA+ protocols can be applied with as few as five *Xenopus* embryos per sample. Low-RNA kits (SMART-Seq v4 Ultra

N.D.L. Owens et al.

Low Input RNA Kit [Takara] or Ovation SoLo RNA-Seq Library Preparation Kit [NuGEN]) and protocols are available for both single embryos and individual cells.

## REFERENCES

Andrews S, et al. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34:** 525–527. doi:10.1038/nbt.3519

Collart C, Owens NDL, Bhaw-Rosun L, Cooper B, De Domenico E, Patrushev I, Sesay AK, Smith JN, Smith JC, Gilchrist MJ. 2014. High-resolution analysis of gene activity during the *Xenopus* mid-blastula transition. *Development* **141:** 1927–1939. doi:10.1242/dev.102012

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Wojciech Szcześniak M, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* **17:** 13. doi:10.1186/s13059-016-0881-8

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Hawkins SFC, Guest PC. 2018. Rapid and easy protocol for quantification of next-generation sequencing libraries. *Methods Mol Biol* **1735:** 343–350. doi:10.1007/978-1-4939-7614-0_23

Karimi K, Fortriede JD, Lotay VS, Burns KA, Wang DZ, Fisher ME, Pells TJ, et al. 2018. Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Res* **46:** D861–D868. doi:10.1093/nar/gkx936

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006. doi:10.1101/gr.229102

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323. doi:10.1186/1471-2105-12-323

Liu Y, Zhou J, White KP. 2014. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30:** 301–304. doi:10.1093/bioinformatics/btt688

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Owens NDL, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KWY, Khokha MK. 2016. Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep* **14:** 632–647. doi:10.1016/j.celrep.2015.12.050

R Core Team. 2018. *R: a language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria.

Ritchie ME, Belinda Phipson DW, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43:** e47. doi:10.1093/nar/gkv007

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, et al. 2016. Erratum: how many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **22:** 1641. doi:10.1261/rna.058339.116

Seyednasrollah F, Laiho A, Elo LL. 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* **16:** 59–70. doi:10.1093/bib/bbt086

Showell C, Conlon FL. 2009. Egg collection and in vitro fertilization of the western clawed frog *Xenopus tropicalis*. *Cold Spring Harb Protoc* **2009:** pdb.prot5293. doi:10.1101/pdb.prot5293

Soneson C, Delorenzi M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14:** 91. doi:10.1186/1471-2105-14-91

Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4:** 1521. doi:10.12688/f1000research.7563.2

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:** 178–192. doi:10.1093/bib/bbs017

TruSeq Stranded mRNA Sample Prep Guide (accessed 18th December 2018) https://support.illumina.com/downloads/truseq_stranded_mrna_sample_preparation_guide_15031047.html.

Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, et al. 2014. A comparative study of techniques for differential expression analysis on RNA-seq data. *PLoS One* **9:** e103207. doi:10.1371/journal.pone.0103207

Cite this protocol as *Cold Spring Harb Protoc;* doi:10.1101/pdb.prot098368

## An RNA-Seq Protocol for Differential Expression Analysis

Nick D.L. Owens, Elena De Domenico and Michael J. Gilchrist

| | |
|---|---|
| **Email Alerting Service** | Receive free email alerts when new articles cite this article - **click here.** |
| **Subject Categories** | Browse articles on similar topics from *Cold Spring Harbor Protocols.*<br><br>Analysis of Gene Expression (189 articles)<br>Analysis of Gene Expression, general (129 articles)<br>Bioinformatics/Genomics, general (188 articles)<br>Developmental Biology (709 articles)<br>High-Throughput Analysis, general (150 articles)<br>mRNA (129 articles)<br>RNA (310 articles)<br>Xenopus (196 articles) |