***List of useful databases***

| Category/database name | URL | Comment |
| --- | --- | --- |
| **Plant (*Arabidopsis*) transcription factors** | | |
| RARTF | <http://rarge.gsc.riken.jp/rartf/> |  |
| AGRIS | <http://arabidopsis.med.ohio-state.edu/AtTFDB/> |  |
| DATF | <http://datf.cbi.pku.edu.cn/>  A part of a plant transcription factor database |  |
| PlnTFDB | <http://plntfdb.bio.uni-potsdam.de/v2.0/index.php?sp_id=ATH>  Data of other plants are also stored |  |
| Molbiol Tools | <https://molbiol-tools.ca/Transcriptional_factors.htm> |  |
| RARTF | <http://rarge.gsc.riken.jp/rartf/>  transcription factors on the Arabidopsis thaliana genome based on PSI-BLAST search |  |
| **Conserved domain search** | | |
| InterProScan | <http://www.ebi.ac.uk/Tools/InterProScan/>  Perform against various Conserved Domain databases and provides sophisticated graphical output | For known motifs |
| MEME | <http://meme.sdsc.edu/meme/intro.html>  Finding known or unknown CDs among a set of proteins | For discovering unknown motifs |
| SALAD database | <http://salad.dna.affrc.go.jp/salad/en/> | For known and unknown motifs |
| **Homology search** | | |
| TAIR BLAST | <http://www.arabidopsis.org/Blast/index.jsp> | For *Arabidopsis* |
| NCBI BLAST | <http://blast.ncbi.nlm.nih.gov/Blast.cgi> | For multispecies search |
| **Prediction of subcellular localization**  (Because TFs cannot function outside the nucleus) | | |
| SUBAII | <http://www.plantenergy.uwa.edu.au/suba2/>  provides hydropathy plots of all Arabidopsis proteins | Experimental data are also stored |
| SubLoc, TargetP , WoLF PSORT |  |  |
| LocDB | <https://www.rostlab.org/services/locDB/index.php>  Protein Localization Database for Human and Arabidopsis |  |
| **Protein–protein interaction**  Many TFs are known to form functional complexes. A number of TFs are known to interact with kinases, resulting in TF phosphorylation (He et al. [2002](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2709548/#B35), Furihata et al. [2006](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2709548/#B25), Robertson et al. [2008](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2709548/#B93)) | | |
| *Arabidopsis* predicted interactome | <http://www.arabidopsis.org/portals/proteome/proteinInteract.jsp> |  |
| EBI IntAct | <http://www.ebi.ac.uk/intact/site/index.jsf>  stores continuously updated PPI information of all organisms based on literature curation | For all organisms |
| AtPID | <http://atpid.biosino.org/index.php>   a search facility with graphical output against a predicted and literature-curated Arabidopsis PPI data set |  |
| **Small RNAs**  200 genes are predicted to be targets of known miRNAs, 69 of these genes (35%) encode putative TFs. | | |
| ASRP | <http://asrp.cgrb.oregonstate.edu/db/> | Includes data of miRNA, siRNA and ta-siRNA |
| **Repository of microarray data** | | |
| NCBI GEO | <http://www.ncbi.nlm.nih.gov/geo/> |  |
| EBI ArrayExpress | <http://www.ebi.ac.uk/microarray-as/ae/> |  |
| NASCArrays | <http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl> |  |
| **Browsing microarray data and co-expression analysis** | | |
| ATTED-II | <http://atted.jp/> |  |
| Genevestigator | <https://www.genevestigator.com/gv/index.jsp> |  |
| BAR eFP browser | <http://bbc.botany.utoronto.ca/efp/cgi-bin/efpWeb.cgi> |  |
| **Finding novel *cis*-elements** | | |
| TAIR motif analysis | <http://www.arabidopsis.org/tools/bulk/motiffinder/index.jsp> |  |
| **Database of known *cis*-elements** | | |
| PLACE | <http://www.dna.affrc.go.jp/PLACE/> | No longer updated after 2007 |
| AGRIS ATCISDB | <http://arabidopsis.med.ohio-state.edu/AtcisDB/> |  |
| **GO categorization** | | |
| TAIR GO | <http://www.arabidopsis.org/tools/bulk/go/index.jsp> | annotation search |
| agriGO  (for Agricultural Community) | http://bioinfo.cau.edu.cn/agriGO/index.php | GO Analysis Toolkit and Database |

**Reference**:

[Plant Cell Physiol](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2709548/). 2009 Jul; 50(7): 1232–1248.

Published online 2009 May 28. doi:  [10.1093/pcp/pcp075](https://dx.doi.org/10.1093%2Fpcp%2Fpcp075)

PMCID: PMC2709548

|  |  |
| --- | --- |
| Functional Analysis of Transcription Factors in Arabidopsis Methods and tools for RNA-seq-based co-expression network analysis | |
| **Tool/method** | **Description** |
| Quality control | |
| FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ | • A tool that uses .fastq, .bam or .sam files to identify and highlight potential issues in the data, such as low base quality scores, low sequence quality and GC content biases.  + Can be used either with or without user interface.  − Uses only the first 200 000 sequences in the file. |
| RSeQC  http://rseqc.sourceforge.net/ | + A tool with a wider range of quality control measures than FastQC.  + Can also be used on mapped data to obtain information on metrics such as the prevalence of splicing events. |
| QoRTs http://hartleys.github.io/QoRTs/ | + This is a similar tool to RSeQC but incorporates more quality control metrics. |
| Read Mappers | |
| Bowtie/Tophat/Tophat2 https://ccb.jhu.edu/software/tophat/index.shtml | • The first widely used mapping tool.  + Detects splice variants.  − Currently much slower than most other mappers and requires a relatively large amount of memory. |
| STAR https://code.google.com/p/rna-star/ | • A widely used tool to align reads to a genome.  + Maps ∼50 times faster than Tophat and Tophat2.  + Commonly used tool to detect novel splice variants.  − Uses a large amount of memory (>20 GB for mapping to the human genome). |
| HISAT http://www.ccb.jhu.edu/software/hisat/index.shtml | • A widely used tool to align reads to a genome at a faster rate than STAR with comparable accuracy.  + HISAT2 is expected to be the core of the next version of Tophat (Tophat3).  + Detects novel splice variants.  + The newer HISAT2 version aligns to genotype variants, likely achieving higher accuracy.  + Uses less memory than STAR (<8 GB for mapping to the human genome using default settings). |
| BWA | • A commonly used aligner for species in which splicing does not occur.  − Does not detect splice variants. |
| Kallisto https://pachterlab.github.io/kallisto/about.html | • A tool that uses a pseudoalignment strategy to assign expression values to transcripts/genes to achieve optimal speed.  • Comparable accuracy to other tools using real alignment strategies.  • Reports reads/expression per gene instead of read alignment coordinates (which are commonly used to acquire the expression per gene).  + Uses little memory and can be run on a regular desktop computer.  − Does not identify novel splice variants |
| Salmon  http://combine-lab.github.io/salmon/ | • Another pseudoalignment tool. Performance comparable with Kallisto.  • Reports reads/expression per gene instead of read alignment coordinates (which are commonly used to acquire the expression per gene).  − Does not identify novel splice variants. |
| Read counting tools | |
| HTseq  http://www-huber.embl.de/HTSeq/doc/overview.html | • A tool that assigns expression values to genes based on reads that have been aligned with, e.g. STAR or HISAT. |
| + Well documented and supported. |
| FeatureCounts http://bioinf.wehi.edu.au/featureCounts/ | + A tool that is similar to HTseq but much faster. Results are slightly different owing to slightly different expression assignment strategies. |
| SpliceNet http://jjwanglab.org/SpliceNet/ | • A tool that divides the reads mapping to an exon shared with two isoforms proportionally to the total expression of each of the two whole isoforms.  + Estimates expression more accurately when multiple genes/transcripts partly share the same genome regions. |
| Normalization | |
| FPKM/RPKM | • Widely used normalization methods that correct for the total number of reads in a sample while accounting for gene length.  − TMM has been suggested as a better alternative. |
| TPM | • A method similar to FPKM, but normalizes the total expression to 1 million, i.e. the summed expression of TPM-normalized samples is always 1 million. |
| TMM | • Similar to FPKM/RPKM but puts expression measures on a common scale across different samples. |
| RAIDA | • A method that uses ratios between counts of genes in each sample for normalizations.  + Avoids problems caused by differential transcript abundance between samples (resulting from differential expression of highly abundant gene transcripts). |
| DEseq2 | • A normalization method that adjusts the expression values of each gene in a sample by a set factor. This factor is determined by taking the median gene expression in a sample after dividing the expression of each gene by the geometric mean of the given gene across all samples. This differs from the normalization implemented in the DEseq2 differential expression analysis.  • Implemented into the DEseq2 R package. |
| Correction for batch effects | |
| Limma-remove BatchEffect | • A method which uses linear models to correct for batch effects. |
| Svaseq https://github.com/jtleek/svaseq | • This method estimates biases based on genes that have no phenotypic expression effects, which are then used for correction of the data.  • Specifically designed for RNA-seq data. |
| Combat http://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html | • A method that is robust to outliers and also effective at batch effect correction in small sample sizes (<25). |
| Co-expression module detection | |
| WGCNA https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/ | • A tool that constructs a co-expression network using Pearson correlation (default) or a custom distance measure.  • Uses hierarchical clustering and has various ‘tree cutting’ options to identify modules.  + Most widely used tool, well supported and documented. |
| DiffCoEx | • A method that uses a similar approach to WGCNA to identify and group differentially co-expressed genes instead of identifying co-expressed modules.  • Identifies modules of genes that have the same different partners between different samples. |
| DICER | • A method that identifies modules that correlate differently between sample groups, e.g. modules that form one large interconnected module in one group compared with several smaller modules in another group. |
| CoXpress http://coxpress.sourceforge.net/ | • A tool that identifies co-expression modules in each sample group and tests whether the genes within these modules are also co-expressed in other groups. |
| DINGO | • DINGO is a more recent tool that groups genes based on how differently they behave in a particular subset of samples (representing e.g. a particular condition) from the baseline co-expression determined from all samples |
| GSCNA | • A tool that tests whether a predefined defined gene set is differentially expressed between two sample groups. |
| GSVD | • A method that identifies ‘genelets’, which can be interpreted as modules representing partial co-expression signals from multiple genes. These signals are then compared between two groups to identify genelets unique to samples and genelets that are shared between the two groups. |
| HO-GSVD https://github.com/aanchan/hogsvd-python/blob/master/README.md | • A tool similar to GSVD, but that can be used across multiple sample groups rather than only two. |
| Biclustering | • A group of methods that identify modules that are unique to a subpopulation of samples without the need for prior grouping of samples. |
| Functional enrichment | |
| DAVID https://david.ncifcrf.gov/ | • A widely used tool with an online web interface. Users supply a list of genes and select the annotation categories from various sources to identify enrichment. |
| PANTHER http://pantherdb.org/ | • A tool that uses a comprehensive protein library combined with human curated pathways and evolutionary ontology.  • If a gene is not in the library, it is classified based on its protein sequence conservation and by finding a related gene. |
| g:Profiler http://biit.cs.ut.ee/gprofiler/ | • A tool that performs enrichment analyses for gene ontologies, KEGG pathways, protein–protein interactions, TF and miRNA binding sites.  + Also available as an R package. |
| ClusterProfiler https://github.com/GuangchuangYu/clusterProfiler/blob/master/vignettes/clusterProfiler.Rmd | • An R package for overrepresentation and gene set enrichment analyses for several curated gene sets.  + Allows users to compare the results of analyses performed on several gene sets. |
| Enrichr http://amp.pharm.mssm.edu/Enrichr/ | • An intuitive web tool for performing gene overrepresentation analyses using a comprehensive set of functional annotations. |
| ToppGene https://toppgene.cchmc.org/ | • An intuitive tool that determines enrichment of different categories such as GO terms, chromosomal locations and disease associations.  + Also has other functions, such as candidate gene prioritization, based on network structures. |
| Regulatory network inference | |
| ARACNE  <http://califano.c2b2.columbia.edu/aracne> | • A tool that removes indirect connections between genes (i.e. partners of a gene that have a stronger correlation with each other than with the gene itself), leaving only those connections that are expected to be regulatory.  + Creates directional networks. |
| Genie3 <https://bioconductor.org/packages/release/bioc/html/GENIE3.html> | • A tool that incorporates TF information to construct a regulatory network by determining the TF expression pattern that best explains the expression of each of their target genes.  + Creates directional networks.  − Requires TF information. |
| CoRegNet | • A tool that identifies co-operative regulators of genes from different data types. |
| cMonkey | • Calculates joint bicluster membership probability from different data types by identifying groups of genes that group together in multiple data types. |
| Visualization | |
| Cystoscape <http://www.cytoscape.org/> | • A widely used tool for the visualization of networks. |
| + Has many plug-ins available for specific analyses. |
| BioLayout <http://www.biolayout.org/> | • Similar to Cytoscape but less widely used. |
| + Can load and visualize much larger networks than Cytoscape. |
| Co-expression databases | |
| COXPRESdb <http://coxpresdb.jp/> | • A web resource incorporating 12 co-expression networks for different species created from ∼157 000 microarrays and 10 000 RNA-seq samples. Has a focus on protein-coding RNAs. |
| GeneFriends <http://www.genefriends.org/> | • Human and mouse gene and transcript co-expression networks.  • Networks constructed from ∼4000 RNA-seq samples each.  + Includes a number of non-coding RNAs (∼10 000 for mouse and ∼25 000 for human). |
| GeneMANIA <http://www.genemania.org/> | • Also includes physical and genetic interaction, co-localization, pathway and shared protein domain information data sets.  + Networks for nine species. |
| GENEVESTIGATOR <https://genevestigator.com/gv/> | • A database constructed using ∼145 000 samples.  + Curated database.  + Networks for 18 species.  + Multiple data types. |
| GIANT <http://giant.princeton.edu> | • Tissue-specific interaction network database.  • Includes 987 Datasets encompassing 38 000 conditions describing 144 tissues types.  + Integrates physical interaction, co-expression, miRNA binding motif and TF binding site data. |

**Reference**:

Briefings in Bioinformatics, bbw139

Published online 10 January 2017. doi:  <https://doi.org/10.1093/bib/bbw139>

Gene co-expression analysis for functional classification and gene–disease predictions

|  |  |
| --- | --- |
| Protein Dynamics | |
| Server Name | Description |
| MakeMultimer  <http://watcut.uwaterloo.ca/tools/makemultimer/> | Calculates the coordinates of the missing subunits of pdb structures that represent multimeric molecules |
| Protein Interaction Server  <http://pic.mbu.iisc.ernet.in> | It recognizes various kinds of interactions within a protein or between proteins in a complex. |
| The ConSurf Server  <http://consurf.tau.ac.il/2016/> | Identification of Function regions of a protein |
| meta-PPISP  <http://pipe.scs.fsu.edu/meta-ppisp.html> | Binding site Prediction - Protein |
| Protein Plus  <https://proteins.plus/> | Binding site Prediction – Small Ligands |
| Hotspot Wizard  <https://loschmidt.chemi.muni.cz/hotspotwizard/> | Design of Mutations and smart libraries in protein engineering |
| CAVER  <http://www.caver.cz/index.php> | Analysis and visualization of tunnels and channels in protein structures |
| Nucleic Acid Database  <http://ndbserver.rutgers.edu> | Experimentally-determined nucleic acids and complex assemblies |
| EM Databank  <http://www.emdatabank.org> | Unified Data Resource for 3-Dimensional Electron Microscopy |
| DAVID  <https://david.ncifcrf.gov/home.jsp> | DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. |
| PDB - Europe  <https://www.ebi.ac.uk/pdbe/> | European resource for the collection, organisation and dissemination of data on biological macromolecular structures |

Other Databases and Websites

|  |  |
| --- | --- |
| MSigDB  <http://software.broadinstitute.org/gsea/msigdb/index.jsp> | Gene set Enrichment Analysis: The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. |
| Melina II  <http://melina2.hgc.jp/public/index.html> | motif search tool : Melina is composed of several published programs such as CONSENSUS, GIBBS DNA, MEME and Coresearch, which are considered to be the most progressive motif search algorithms |
| Prosite  <https://prosite.expasy.org/prosite.html> | Database of protein families and domains. It is based on the observation that, while there is a huge number of different proteins, most of them can be grouped, on the basis of similarities in their sequences, into a limited number of families. |
| Interpro  <http://www.ebi.ac.uk/interpro/> | Protein sequence analysis & classification. InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. |
| Pscan  <http://159.149.160.88/pscan/> | Finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes |
| psRNATarget  <http://plantgrn.noble.org/psRNATarget/> | The psRNATarget was specifically developed to identify target transcripts of these sRNAs through i) analyzing complementary matching between sRNA and target using a predefined scoring schema |
| P3DB  <http://p3db.org/index.php> | Plant Protein Phosphorylation DataBase |
| Java Heatchart  <http://www.javaheatmap.com> |  |
| miARma-Seq  <https://www.nature.com/articles/srep25749> | a comprehensive tool for miRNA, mRNA and circRNA analysis |
| miRGen  <http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=mirgenv3%2Findex> | Discover the connection between miRNAS and IncRNAs |
| TarBase  <http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=tarbasev8%2Findex/> | A Database of experimentally supported miRNA gene interactions |
| ViennaRNA Web Services  <http://rna.tbi.univie.ac.at> | server provides programs, web services, and databases, related to our work on RNA secondary structures |
| SNPin  <http://korkinlab.org/snpintool/> | Predictin effects of non-synonymous SNPs in Interactions |
| CyTrargetLinker  <https://projects.bigcat.unimaas.nl/cytargetlinker/tutorial-1/> | Cytoscape with our CyTargetLinker |
| CoExpress  <http://sablab.net/coexpress.html> | A Tool for an Effective Co-Expression Analysis of Large Microarray Data Sets |
| PMRD : Plant microRNA Database  <http://bioinformatics.cau.edu.cn/PMRD/> | Plant microRNA database |
| Arabidopsis miRNA candidates  <http://sundarlab.ucdavis.edu/mirna/search_candidates.html> | Computational Prediction of miRNAs in Arabidopsis thaliana |
| MetaboAnalyst  <http://www.metaboanalyst.ca> | Statistical, Functional and Integrative analysis of metabolomics data |
| Trava  <http://travadb.org> | Transcriptome Variation Analysis: the database of gene expression profiles in Arabidopsis thaliana based on RNA-seq analysis |
| MicroRNA  <http://sundarlab.ucdavis.edu/mirna/> | Precursor Candidates for Arabidopsis thaliana |
| Cyverse  <https://wiki.cyverse.org/wiki/dashboard.action> | NGS analysis pipeline |
| Ensembl Plants  <https://plants.ensembl.org/info/website/ftp/index.html> |  |
| OmicsDB  <https://www.omicsdi.org/home> |  |
| Arabidopsis eFP Browser  <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi> | Gene Expression and Protein Tools |
| Expression Angler  <http://bar.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi> |  |
| RNA-seqlopedia  <https://rnaseq.uoregon.edu/#figure3.4> | Overview of RNA-seq and of the choices necessary to carry out a successful RNA-seq experiment |
| RNA Seq Tutorial  <https://github.com/griffithlab/rnaseq_tutorial> |  |
| Arabidopsis Hormone Database  <http://ahd.cbi.pku.edu.cn> |  |
| pssRNAMiner  <http://bioinfo3.noble.org/pssRNAMiner/> | A plant short small RNA regulatory cascade analysis server |
| AtPIN  <https://atpin.bioinfoguy.net/cgi-bin/atpin.pl> | Arabidopsis thaliana protein interaction network |
| UbPred  <http://www.ubpred.org> | Predictor of protein ubiquitination sites |
| miRNEST 2.0  <http://rhesus.amu.edu.pl/mirnest/copy/home.php> | integrative collection of animal, plant and virus microRNA data |
| RNA Tools  <https://web.njit.edu/~wangj/rna/sequence.htm>  <http://rna.informatik.uni-freiburg.de/INFORNA/Input.jsp> | Collection of RNA Tools |
| GSA  <http://bigd.big.ac.cn/gsa/> | Genome Sequence Archive |
| PMirKB  <http://bis.zju.edu.cn/pmirkb/index.php> | Plant microRNA Knowledge Base |
| iPTMNet  <https://research.bioinformatics.udel.edu/iptmnet/> | integrated understanding of protein post-translational modifications (PTMs) in systems biology |
| PaxDB  <https://pax-db.org> | Protein Abundance Database |
| MicroRPM  <http://microrpm.itps.ncku.edu.tw> | Predicting mature miRNAs from NGS Reads |
| Interactome 3D  <https://interactome3d.irbbarcelona.org/index.php> | Structural annotation of protein-protein interaction networks |
| RNA Server  <http://rna.tbi.univie.ac.at> | ViennaRNA Web Services |
| GREIN : GEO RNA-seq Experiments Interactive Navigator  <https://shiny.ilincs.org/grein/?gse>= | Explore and analyze GEO RNA-seq data |
| KAAS - KEGG Automatic Annotation Server  <https://www.genome.jp/tools/kaas/> | Functional annotation of genes by BLAST |

**RNA-Seq library protocols** (KR Kukurba et al ‎2015)

| **Library design** | **Usage** | **Description** |
| --- | --- | --- |
| Poly-A selection | Sequencing mRNA | Select for RNA species with poly-A tail and enriches for mRNA |
| Ribo-depletion | Sequencing mRNA, pre-mRNA, ncRNA | Removes ribosomal RNA and enriches for mRNA, pre-mRNA, and ncRNA |
| Size selection | Sequencing miRNA | Selects RNA species using size fractionation by gel electrophoresis |
| Duplex-specific nuclease | Reduce highly abundant transcripts | Cleaves highly abundant transcripts, including rRNA and other highly expressed genes |
| Strand-specific | De novo transcriptome assembly | Preserves strand information of the transcript |
| Multiplexed | Sequencing multiple samples together | Genetic barcoding method that enables sequencing multiple samples together |
| Short-read | Higher coverage | Produces 50–100 bp reads; generally higher read coverage and reduced error rate compared to long-read sequencing |
| Long-read | De novo transcriptome assembly | Produces >1000 bp reads; advantageous for resolving splice junctions and repetitive regions |

**Widely used RNA-Seq software packages**

| **Primary category** | **Tool name** | **Notes** |
| --- | --- | --- |
| Splice-aware read alignment | GEM | Filtration-based approach to approximate string matching for alignment |
| GSNAP | Based on seed and extend alignment algorithm aware of complex variants |
| MapSplice | Based on Burrows-Wheeler Transform (BWT) algorithm |
| RUM | Integrates alignment tools Blat and Bowtie to increase accuracy |
| STAR | Based on seed searching in an uncompressed suffix arrays followed by seed clustering and stitching procedure; fast but memory-intensive |
| TopHat | Uses Bowtie, based on BWT, to align reads; resolves spliced reads using exons by split read mapping |
| Transcript assembly and quantification | Cufflinks | Assembles transcripts to reference annotations or de novo and quantifies abundance |
| FluxCapacitor | Quantifies transcripts using reference annotations |
| iReckon | Models novel isoforms and estimates their abundance |
| Differential expression (DE) | BaySeq | Count-based approach using empirical Bayesian method to estimate posterior likelihoods |
| Cuffdiff2 | Isoform-based approach based on beta negative binomial distribution |
| DESeq | Exon-based approach using the negative binomial model |
| DEGSeq | Isoform-based approach using the Poisson model |
| EdgeR | Count-based approach using empirical Bayes method based on the negative binomial model |
| MISO | Isoform-based model using Bayes factors to estimate posterior probabilities |
| Other tools | HCP | Normalizes expression data by inferring known and hidden factors with prior knowledge |
| PEER | Normalizes expression data by inferring known and hidden factors using a probabilistic estimation based on the Bayesian framework |
| Matrix eQTL | Fast eQTL detection tool that uses linear models (linear regression or ANOVA) |

Ref: <https://github.com/griffithlab/rnaseq_tutorial>

| **Category** | **Representative tools** |
| --- | --- |
| **Raw data QC**[[25577376](http://www.ncbi.nlm.nih.gov/pubmed/25577376), [25150838](http://www.ncbi.nlm.nih.gov/pubmed/25150838)] | FastQC, HTQC [[23363224](http://www.ncbi.nlm.nih.gov/pubmed/23363224)], QC3 [[24703969](http://www.ncbi.nlm.nih.gov/pubmed/24703969)], kPAL [[25514851](http://www.ncbi.nlm.nih.gov/pubmed/25514851)]. |
| **Read trimming**[[24376861](http://www.ncbi.nlm.nih.gov/pubmed/24376861)] | Trimmomatic [[24695404](http://www.ncbi.nlm.nih.gov/pubmed/24695404)], Skewer [[24925680](http://www.ncbi.nlm.nih.gov/pubmed/24925680)], Flexbar [[24832523](http://www.ncbi.nlm.nih.gov/pubmed/24832523)], FASTX. |
| **Alignment (splice aware, for alignment to a reference genome)** [[24185836](http://www.ncbi.nlm.nih.gov/pubmed/24185836)] | TopHat [[19289445](http://www.ncbi.nlm.nih.gov/pubmed/19289445), [23618408](http://www.ncbi.nlm.nih.gov/pubmed/23618408)], STAR [[23104886](http://www.ncbi.nlm.nih.gov/pubmed/23104886)], HISAT [[25751142](http://www.ncbi.nlm.nih.gov/pubmed/25751142)], [HISAT2](http://ccb.jhu.edu/software/hisat2/index.shtml), segemehl [[24512684](http://www.ncbi.nlm.nih.gov/pubmed/24512684)], [SubRead](http://subread.sourceforge.net/), GSNAP, MapSplice [[20802226](http://www.ncbi.nlm.nih.gov/pubmed/20802226)], JAGuaR [[25062255](http://www.ncbi.nlm.nih.gov/pubmed/25062255)], SpliceMap [[20371516](http://www.ncbi.nlm.nih.gov/pubmed/20371516)], HMMSplicer [[21079731](http://www.ncbi.nlm.nih.gov/pubmed/21079731)], TrueSight/UnSplicer [[24259430](http://www.ncbi.nlm.nih.gov/pubmed/24259430)]. |
| **Alignment (non splice aware for alignment to a reference transcriptome)**[[23060614](http://www.ncbi.nlm.nih.gov/pubmed/23060614), [23758764](http://www.ncbi.nlm.nih.gov/pubmed/23758764)] | BowTie [[19261174](http://www.ncbi.nlm.nih.gov/pubmed/19261174)], Bwa [[19451168](http://www.ncbi.nlm.nih.gov/pubmed/19451168)]. |
| **Post-alignment QC**[[24185836](http://www.ncbi.nlm.nih.gov/pubmed/24185836)] | FastQC, samtools [[19505943](http://www.ncbi.nlm.nih.gov/pubmed/19505943)], QuaCRS [[25368506](http://www.ncbi.nlm.nih.gov/pubmed/25368506)], RSeQC [[22743226](http://www.ncbi.nlm.nih.gov/pubmed/22743226)], RNA-SeQC [[22539670](http://www.ncbi.nlm.nih.gov/pubmed/22539670)], Picard CollectRnaSeqMetrics, BAMstats, SAMstat [[21088025](http://www.ncbi.nlm.nih.gov/pubmed/21088025)], BlackOPs [[23935067](http://www.ncbi.nlm.nih.gov/pubmed/23935067)], seqbias [[22285831](http://www.ncbi.nlm.nih.gov/pubmed/22285831)]. |
| **Gene/transcriptome annotation** [[24722185](http://www.ncbi.nlm.nih.gov/pubmed/24722185), [25319663](http://www.ncbi.nlm.nih.gov/pubmed/25319663)] | Annocript [[25701574](http://www.ncbi.nlm.nih.gov/pubmed/25701574)], XSAnno [[24884593](http://www.ncbi.nlm.nih.gov/pubmed/24884593)], GeneMark-ET [[24990371](http://www.ncbi.nlm.nih.gov/pubmed/24990371)], WImpiBLAST [[24979410](http://www.ncbi.nlm.nih.gov/pubmed/24979410)], RNASEG [[24780064](http://www.ncbi.nlm.nih.gov/pubmed/24780064)], TSSAR [[24674136](http://www.ncbi.nlm.nih.gov/pubmed/24674136)], Vicinal [[24623808](http://www.ncbi.nlm.nih.gov/pubmed/24623808)], OMIGA [[24609470](http://www.ncbi.nlm.nih.gov/pubmed/24609470)], CoRAL [[24145223](http://www.ncbi.nlm.nih.gov/pubmed/24145223)], AfterParty [[24093729](http://www.ncbi.nlm.nih.gov/pubmed/24093729)], ShortStack [[23610128](http://www.ncbi.nlm.nih.gov/pubmed/23610128)], CIRI [[25583365](http://www.ncbi.nlm.nih.gov/pubmed/25583365)]. |
| **Small RNA identification and characterization (e.g., miRNAs)** [[25319663](http://www.ncbi.nlm.nih.gov/pubmed/25319663), [23720668](http://www.ncbi.nlm.nih.gov/pubmed/23720668)] | ShortStack [[23610128](http://www.ncbi.nlm.nih.gov/pubmed/23610128)], CoRAL [[24145223](http://www.ncbi.nlm.nih.gov/pubmed/24145223)], MTide [[25256573](http://www.ncbi.nlm.nih.gov/pubmed/25256573)], FlaiMapper [[25338717](http://www.ncbi.nlm.nih.gov/pubmed/25338717)], miRPlant [[25117656](http://www.ncbi.nlm.nih.gov/pubmed/25117656)], PROmiRNA [[23958307](http://www.ncbi.nlm.nih.gov/pubmed/23958307)], omiRas [[23946503](http://www.ncbi.nlm.nih.gov/pubmed/23946503)], DREAM [[25840043](http://www.ncbi.nlm.nih.gov/pubmed/25840043)]. |
| **Transcript assembly (reference genome guided)** [[24185837](http://www.ncbi.nlm.nih.gov/pubmed/24185837), [21897427](http://www.ncbi.nlm.nih.gov/pubmed/21897427), [23393030](http://www.ncbi.nlm.nih.gov/pubmed/23393030)] | Cufflinks [[20436464](http://www.ncbi.nlm.nih.gov/pubmed/20436464)], Scripture [[20436462](http://www.ncbi.nlm.nih.gov/pubmed/20436462)], StringTie [[25690850](http://www.ncbi.nlm.nih.gov/pubmed/25690850)], bayesembler [[25367074](http://www.ncbi.nlm.nih.gov/pubmed/25367074)], IsoLasso [[21951053](http://www.ncbi.nlm.nih.gov/pubmed/21951053)]. |
| **Transcript assembly (de novo, reference genome free)**[[21897427](http://www.ncbi.nlm.nih.gov/pubmed/21897427), [23393030](http://www.ncbi.nlm.nih.gov/pubmed/23393030), [23056003](http://www.ncbi.nlm.nih.gov/pubmed/23056003), [23666209](http://www.ncbi.nlm.nih.gov/pubmed/23666209), [25084827](http://www.ncbi.nlm.nih.gov/pubmed/25084827), [25279728](http://www.ncbi.nlm.nih.gov/pubmed/25279728), [25788326](http://www.ncbi.nlm.nih.gov/pubmed/25788326)] | Trinity [[23845962](http://www.ncbi.nlm.nih.gov/pubmed/23845962)], Trans-ABySS [[20935650](http://www.ncbi.nlm.nih.gov/pubmed/20935650)], Oases [[22368243](http://www.ncbi.nlm.nih.gov/pubmed/22368243)], RSEM [[21816040](http://www.ncbi.nlm.nih.gov/pubmed/21816040)], DETONATE [[25608678](http://www.ncbi.nlm.nih.gov/pubmed/25608678)], SEECER (sequencing error correction for assembly) [[23558750](http://www.ncbi.nlm.nih.gov/pubmed/23558750)], BRANCH [[23493323](http://www.ncbi.nlm.nih.gov/pubmed/23493323)] uses partial or related genomics sequences as a guide, EBARDenovo [[23457040](http://www.ncbi.nlm.nih.gov/pubmed/23457040)], Bridger [[25723335](http://www.ncbi.nlm.nih.gov/pubmed/25723335)]. |
| **Transcript abundance or expression estimation (FPKM/RPKM)**[[24185837](http://www.ncbi.nlm.nih.gov/pubmed/24185837), [24885830](http://www.ncbi.nlm.nih.gov/pubmed/24885830), [24109770](http://www.ncbi.nlm.nih.gov/pubmed/24109770), [24685233](http://www.ncbi.nlm.nih.gov/pubmed/24685233)] | Cufflinks [[20436464](http://www.ncbi.nlm.nih.gov/pubmed/20436464)], eXpress [[23160280](http://www.ncbi.nlm.nih.gov/pubmed/23160280)], RSEM [[21816040](http://www.ncbi.nlm.nih.gov/pubmed/21816040)], Sailfish (alignment free) [[24752080](http://www.ncbi.nlm.nih.gov/pubmed/24752080)], RNA-Skim (alignment free) [[24931995](http://www.ncbi.nlm.nih.gov/pubmed/24931995)], MITIE [[23980025](http://www.ncbi.nlm.nih.gov/pubmed/23980025)], ireckon [[23204306](http://www.ncbi.nlm.nih.gov/pubmed/23204306)], DRUT [[23202426](http://www.ncbi.nlm.nih.gov/pubmed/23202426)], Kallisto (alignment free) [[arXiv](http://arxiv.org/abs/1505.02710)]. |
| **Obtaining raw transcript/gene read counts (FPM/RPM)**[[21176179](http://www.ncbi.nlm.nih.gov/pubmed/21176179)] | HTSeq [[25260700](http://www.ncbi.nlm.nih.gov/pubmed/25260700)], FeatureCounts [[24227677](http://www.ncbi.nlm.nih.gov/pubmed/24227677)], Rcount [[25322836](http://www.ncbi.nlm.nih.gov/pubmed/25322836)], maxcounts [[24564404](http://www.ncbi.nlm.nih.gov/pubmed/24564404)], FIXSEQ (adjusts counts to compensate for overdispersion) [[24603409](http://www.ncbi.nlm.nih.gov/pubmed/24603409)], Cuffquant. |
| **Differential expression**[[25119138](http://www.ncbi.nlm.nih.gov/pubmed/25119138), [24300110](http://www.ncbi.nlm.nih.gov/pubmed/24300110), [24020486](http://www.ncbi.nlm.nih.gov/pubmed/24020486), [25024085](http://www.ncbi.nlm.nih.gov/pubmed/25024085)] | Cuffdiff [[23222703](http://www.ncbi.nlm.nih.gov/pubmed/23222703)], limma [[25605792](http://www.ncbi.nlm.nih.gov/pubmed/25605792)], DESeq2 [[25516281](http://www.ncbi.nlm.nih.gov/pubmed/25516281)], EdgeR [[19910308](http://www.ncbi.nlm.nih.gov/pubmed/19910308)], Corset (for de novo assembled transcriptomes) [[25063469](http://www.ncbi.nlm.nih.gov/pubmed/25063469)], sSeq [[23589650](http://www.ncbi.nlm.nih.gov/pubmed/23589650)], BADGE [[25252852](http://www.ncbi.nlm.nih.gov/pubmed/25252852)], compcodeR [[24813215](http://www.ncbi.nlm.nih.gov/pubmed/24813215)], metaRNASeq [[24678608](http://www.ncbi.nlm.nih.gov/pubmed/24678608)], Characteristic Direction [[24650281](http://www.ncbi.nlm.nih.gov/pubmed/24650281)], NPEBseq [[23981227](http://www.ncbi.nlm.nih.gov/pubmed/23981227)]. |
| **Alternative splicing, alternative expression**[[24447644](http://www.ncbi.nlm.nih.gov/pubmed/24447644), [24885830](http://www.ncbi.nlm.nih.gov/pubmed/24885830), [24058384](http://www.ncbi.nlm.nih.gov/pubmed/24058384), [24549677](http://www.ncbi.nlm.nih.gov/pubmed/24549677), [24951248](http://www.ncbi.nlm.nih.gov/pubmed/24951248), [25511303](http://www.ncbi.nlm.nih.gov/pubmed/25511303)] | Cuffdiff [[23222703](http://www.ncbi.nlm.nih.gov/pubmed/23222703)], DEXSeq [[22722343](http://www.ncbi.nlm.nih.gov/pubmed/22722343)], ALEXA-seq [[20835245](http://www.ncbi.nlm.nih.gov/pubmed/20835245)], IUTA [[25283306](http://www.ncbi.nlm.nih.gov/pubmed/25283306)], FineSplice [[24574529](http://www.ncbi.nlm.nih.gov/pubmed/24574529)], PennSeq [[24362841](http://www.ncbi.nlm.nih.gov/pubmed/24362841)], FlipFlop [[24813214](http://www.ncbi.nlm.nih.gov/pubmed/24813214)], SNPlice [[25481010](http://www.ncbi.nlm.nih.gov/pubmed/25481010)], spliceR [[24655717](http://www.ncbi.nlm.nih.gov/pubmed/24655717)], GESS [[24447644](http://www.ncbi.nlm.nih.gov/pubmed/24447644)], RNASeq-MATS [[23872975](http://www.ncbi.nlm.nih.gov/pubmed/23872975)], SplicingCompass [[23449093](http://www.ncbi.nlm.nih.gov/pubmed/23449093)], DiffSplice [[23155066](http://www.ncbi.nlm.nih.gov/pubmed/23155066)], SigFuge [[25030904](http://www.ncbi.nlm.nih.gov/pubmed/25030904)], SUPPA [bioRXiv], CLASS [bioRXiv], SplAdder [bioRXiv], SplicePie [[25800735](http://www.ncbi.nlm.nih.gov/pubmed/25800735)]. |
| **Variant (e.g., SNP) and mutation detection**[[23555596](http://www.ncbi.nlm.nih.gov/pubmed/23555596), [24075185](http://www.ncbi.nlm.nih.gov/pubmed/24075185), [22468815](http://www.ncbi.nlm.nih.gov/pubmed/22468815)], germline or somatic, and eQTL/sQTL characterization [[25733796](http://www.ncbi.nlm.nih.gov/pubmed/25733796)] | GATK (Best Practices Guide) [[20644199](http://www.ncbi.nlm.nih.gov/pubmed/20644199)], samtools [[19505943](http://www.ncbi.nlm.nih.gov/pubmed/19505943)], SNVMix [[20130035](http://www.ncbi.nlm.nih.gov/pubmed/20130035)], SNPlice [[25481010](http://www.ncbi.nlm.nih.gov/pubmed/25481010)], eSNV-detect [[25352556](http://www.ncbi.nlm.nih.gov/pubmed/25352556)], RVboost [[25170027](http://www.ncbi.nlm.nih.gov/pubmed/25170027)], sQTLseekeR [[25140736](http://www.ncbi.nlm.nih.gov/pubmed/25140736)], eQTL/ASE, [BioRXiv], SNiPloid [[24163691](http://www.ncbi.nlm.nih.gov/pubmed/24163691)], SNPiR [[24075185](http://www.ncbi.nlm.nih.gov/pubmed/24075185)], QualitySNPng [[23632165](http://www.ncbi.nlm.nih.gov/pubmed/23632165)], RNAmapper [[23299976](http://www.ncbi.nlm.nih.gov/pubmed/23299976)], CRAC [[23537109](http://www.ncbi.nlm.nih.gov/pubmed/23537109)], RADIA [[25405470](http://www.ncbi.nlm.nih.gov/pubmed/25405470)]. |
| **RNA editing**[[22327324](http://www.ncbi.nlm.nih.gov/pubmed/22327324), [23291724](http://www.ncbi.nlm.nih.gov/pubmed/23291724), [23598527](http://www.ncbi.nlm.nih.gov/pubmed/23598527), [25859542](http://www.ncbi.nlm.nih.gov/pubmed/25859542)] | REDItools [[23742983](http://www.ncbi.nlm.nih.gov/pubmed/23742983)], GIREMI [[25730491](http://www.ncbi.nlm.nih.gov/pubmed/25730491)], ICEBreaker [[25855956](http://www.ncbi.nlm.nih.gov/pubmed/25855956)]. |
| **Allele specific expression** [[23919664](http://www.ncbi.nlm.nih.gov/pubmed/23919664), [25183311](http://www.ncbi.nlm.nih.gov/pubmed/25183311), [25339465](http://www.ncbi.nlm.nih.gov/pubmed/25339465)] | AlleleSeq [[21811232](http://www.ncbi.nlm.nih.gov/pubmed/21811232)], Allim [[23615333](http://www.ncbi.nlm.nih.gov/pubmed/23615333)], mamba [[25819081](http://www.ncbi.nlm.nih.gov/pubmed/25819081)], EMASE, MBASED [[25315065](http://www.ncbi.nlm.nih.gov/pubmed/25315065)], limma [[25605792](http://www.ncbi.nlm.nih.gov/pubmed/25605792)]. |
| **Viral detection**[[23740984](http://www.ncbi.nlm.nih.gov/pubmed/23740984), [23279287](http://www.ncbi.nlm.nih.gov/pubmed/23279287), [22647373](http://www.ncbi.nlm.nih.gov/pubmed/22647373)] | VirusSeq [[23162058](http://www.ncbi.nlm.nih.gov/pubmed/23162058)], VirusFinder [[23717618](http://www.ncbi.nlm.nih.gov/pubmed/23717618)], RNA CoMPASS [[24586784](http://www.ncbi.nlm.nih.gov/pubmed/24586784)]. |
| **Fusion detection**[[25500544](http://www.ncbi.nlm.nih.gov/pubmed/25500544), [25266161](http://www.ncbi.nlm.nih.gov/pubmed/25266161), [23815381](http://www.ncbi.nlm.nih.gov/pubmed/23815381), [23555082](http://www.ncbi.nlm.nih.gov/pubmed/23555082), [25286921](http://www.ncbi.nlm.nih.gov/pubmed/25286921)] | FusionQ [[23815381](http://www.ncbi.nlm.nih.gov/pubmed/23815381)], TRUP [[25650807](http://www.ncbi.nlm.nih.gov/pubmed/25650807)], Dissect [[22689759](http://www.ncbi.nlm.nih.gov/pubmed/22689759)], Trans-ABySS [[20935650](http://www.ncbi.nlm.nih.gov/pubmed/20935650)], PRADA (RNA-seq pipeline with a fusion module) [[24695405](http://www.ncbi.nlm.nih.gov/pubmed/24695405)], Pegasus (used for fusion annotation) [[25183062](http://www.ncbi.nlm.nih.gov/pubmed/25183062)], FusionCatcher, ChimeraScan [[21840877](http://www.ncbi.nlm.nih.gov/pubmed/21840877)], TopHat-fusion [[21835007](http://www.ncbi.nlm.nih.gov/pubmed/21835007)], BreakFusion [[22563071](http://www.ncbi.nlm.nih.gov/pubmed/22563071)], deFuse [[21625565](http://www.ncbi.nlm.nih.gov/pubmed/21625565)], FusionHunter [[21546395](http://www.ncbi.nlm.nih.gov/pubmed/21546395)], EricScript [[23093608](http://www.ncbi.nlm.nih.gov/pubmed/23093608)], Barnacle [[23941359](http://www.ncbi.nlm.nih.gov/pubmed/23941359)], bellerophontes [[22711792](http://www.ncbi.nlm.nih.gov/pubmed/22711792)], Chimera (merge results from multiple fusion algorithms) [[25286921](http://www.ncbi.nlm.nih.gov/pubmed/25286921)], GFML (format for representing fusion data) [[23072312](http://www.ncbi.nlm.nih.gov/pubmed/23072312)]. |
| **Visualization**[[24792048](http://www.ncbi.nlm.nih.gov/pubmed/24792048), [25757788](http://www.ncbi.nlm.nih.gov/pubmed/25757788)] | SplicingViewer [[22226708](http://www.ncbi.nlm.nih.gov/pubmed/22226708)], IGV [[22517427](http://www.ncbi.nlm.nih.gov/pubmed/22517427)], Sashimi plots [[25617416](http://www.ncbi.nlm.nih.gov/pubmed/25617416)], IGB (splicing visualization protocol) [[24792048](http://www.ncbi.nlm.nih.gov/pubmed/24792048)], PrimerSeq (Visualize RNA-seq data for primer design) [[24747190](http://www.ncbi.nlm.nih.gov/pubmed/24747190)], ASTALAVISTA [[25577392](http://www.ncbi.nlm.nih.gov/pubmed/25577392)], Circos [[19541911](http://www.ncbi.nlm.nih.gov/pubmed/19541911)], Epiviz [[25086505](http://www.ncbi.nlm.nih.gov/pubmed/25086505)], RNAbrowse [[24823498](http://www.ncbi.nlm.nih.gov/pubmed/24823498)], ZENBU [[24727769](http://www.ncbi.nlm.nih.gov/pubmed/24727769)], RNAseqViewer [[24215023](http://www.ncbi.nlm.nih.gov/pubmed/24215023)], viRome [[23709497](http://www.ncbi.nlm.nih.gov/pubmed/23709497)], miRseqViewer [[25322835](http://www.ncbi.nlm.nih.gov/pubmed/25322835)], Circleator [[25075113](http://www.ncbi.nlm.nih.gov/pubmed/25075113)], RNASeqBrowser [[25766521](http://www.ncbi.nlm.nih.gov/pubmed/25766521)]. |
| **Integration of DNA-seq and RNA-seq data**[[23499923](http://www.ncbi.nlm.nih.gov/pubmed/23499923)] | Veridical [[24741438](http://www.ncbi.nlm.nih.gov/pubmed/24741438)], SpliceFinder [[24498620](http://www.ncbi.nlm.nih.gov/pubmed/24498620)], nFuse [[22745232](http://www.ncbi.nlm.nih.gov/pubmed/22745232)], RADIA [[25405470](http://www.ncbi.nlm.nih.gov/pubmed/25405470)]. |