**Data Analysis, Modeling, Visualization (Bioconductor)**

|  |
| --- |
| **Bioconductor Learning Resources** |
| |  | | --- | | **BioC 2020 Conference** | | **Conference Presentations Videos**  <http://bioc2020.bioconductor.org/schedule>  100: **Annotating inter-sample DNA methylation and ATAC-seq variation with COCOA**  100: **Human RNA-seq data from recount2 and related packages**  100: **Introduction to Bioconductor annotation resources**  100: **A tidy transcriptomics introduction to RNA-Seq analyses**  200: **Functional enrichment analysis of high-throughput omics data**  200: **Best practices for ATAC-seq QC and data analysis**  200: **Copy number variation analysis with Bioconductor**  200: **Interactive visualization of SummarizedExperiment objects with iSEE**  200: **Integrated ChIP-seq data analysis workshop**  200: **An introduction to matrix factorization and principal component analysis in R**  500: **Bioconductor toolchain for usage and development of reproducible bioinformatics pipelines in CWL**  500: **Effectively Using the DelayedArray Framework to Support the Analysis of Large Datasets**  100: **Cloud-based genomics using Terra/AnVIL  Workshop Notes**  <https://bioc2020.bioconductor.org/workshops> |  |  | | --- | | **Bioconductor Events**  [**https://bioconductor.org/help/events/**](https://bioconductor.org/help/events/) | | **BioC 2019 Workshops**  <http://biocworkshops2019.bioconductor.org.s3-website-us-east-1.amazonaws.com/>  [**https://bioconductor.github.io/BiocWorkshops/BioC2018.pdf**](https://bioconductor.github.io/BiocWorkshops/BioC2018.pdf)  **BioC 2018 Workshops**  <https://bioconductor.github.io/BiocWorkshops/>  **BioC 2017 Workshops**  <http://bioconductor.org/help/course-materials/2017/BioC2017/> |  |  | | --- | | **Bioconductor.Org** | | **Bioconductor**  <http://bioconductor.org/>  **Bioconductor Courses & Conferences**  <http://bioconductor.org/help/course-materials/>  <http://bioconductor.org/help/course-materials/2015/useR/ad-hoc-exercises.html> **Community Contributed Help Resources** <http://bioconductor.org/help/community/> **Bioconductor Support**  **Bioconductor Q&A Forum**  <https://support.bioconductor.org/> <https://support.bioconductor.org/accounts/login/?next=/new/post/>  <https://support.bioconductor.org/t/Tutorials/>  **F1000 Research Bioconductor**  [**https://f1000research.com/gateways/bioconductor**](https://f1000research.com/gateways/bioconductor)  <https://f1000research.com/articles/5-2122/v1> |  |  | | --- | | **Bioconductor References** | | **R & Bioconductor Manual**  <http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual>  Thomas Girke, UC Riverside  **High-Throughput (HT) Sequence Analysis with R and Bioconductor**  <http://manuals.bioinformatics.ucr.edu/home/ht-seq>  **Bioc-refcard**  <https://github.com/mikelove/bioc-refcard>  Mike Love |  |  | | --- | | **Bioconductor YouTube Channel** | | **Videos:**  <https://www.youtube.com/user/bioconductor> |      |  | | --- | | **Bioconductor Docker** | | <https://www.bioconductor.org/help/docker/>  <https://hub.docker.com/r/bioconductor/bioconductor_docker> |     <https://broadinstitute.github.io/2019_scWorkshop/introduction-rbioconductor.html>  <https://www.bioconductor.org/help/course-materials/2013/SeattleMay2013/IntermediateSequenceAnalysis2013.pdf>  <https://bioconductor.github.io/BiocWorkshops/analysis-of-single-cell-rna-seq-data-dimensionality-reduction-clustering-and-lineage-inference.html>  **GitHub Bioconductor**  <https://github.com/Bioconductor/>  **Linux Basics - Finding Things**  <http://hpcc.ucr.edu/manuals_linux-basics_finding-things.html> |

|  |
| --- |
| **Bioconductor R Packages** |
| |  | | --- | | **GenomicRanges**  <http://bioconductor.org/packages/release/bioc/html/GenomicRanges.html> | | **Software for Computing and Annotating Genomic Ranges**  <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003118> | | **GenomicRanges HOWTOs**  April 27, 2020  <http://bioconductor.org/packages/release/bioc/vignettes/GenomicRanges/inst/doc/GenomicRangesHOWTOs.pdf> | | **A quick introduction to GRanges and GRangesList objects**  July 2015  <http://bioconductor.org/packages/release/bioc/vignettes/GenomicRanges/inst/doc/GRanges_and_GRangesList_slides.pdf> |  |  | | --- | | **GenomicAlignments**  <https://bioconductor.org/packages/release/bioc/html/GenomicAlignments.html> | | **Package GenomicAlignments** October 17, 2020  <https://bioconductor.org/packages/release/bioc/manuals/GenomicAlignments/man/GenomicAlignments.pdf>  Title: Representation and manipulation of short genomic alignments  **An Introduction to the GenomicAlignments Package** April 27, 2020  <https://bioconductor.org/packages/release/bioc/vignettes/GenomicAlignments/inst/doc/GenomicAlignmentsIntroduction.pdf> |  |  | | --- | | **IRanges**  <http://bioconductor.org/packages/release/bioc/html/IRanges.html> | | **An Overview of the IRanges package** May 21 2020  <http://bioconductor.org/packages/release/bioc/vignettes/IRanges/inst/doc/IRangesOverview.pdf>  **Package IRanges** October 17, 2020  <http://bioconductor.org/packages/release/bioc/manuals/IRanges/man/IRanges.pdf>  Foundation of integer range manipulation in Bioconductor |  |  | | --- | | **Tidyverse** | | **Introduction to dplyr**  <https://dplyr.tidyverse.org/articles/dplyr.html>  **Data Transformation with dplyr :: Cheat Sheet**  <https://github.com/rstudio/cheatsheets/blob/master/data-transformation.pdf>  **Bioinformatics in the Tidyverse**  <https://chapmandu2.github.io/post/2017/02/21/bioinformatics-in-the-tidyverse/>  2017/02/21  **The tidyverse: dplyr, ggplot2, and friends**  <https://monashbioinformaticsplatform.github.io/r-more/topics/tidyverse.html#an-rna-seq-example>  <https://www.tidyverse.org/>  <https://dplyr.tidyverse.org/>  <https://tidyr.tidyverse.org/>  <https://tibble.tidyverse.org/> | | <https://chapmandu2.github.io/post/2017/02/21/bioinformatics-in-the-tidyverse/> | |

|  |
| --- |
| **Bioconductor: Bioinformatics** |
| **Most Common Tools Used for the Analysis of WGS Data**  <https://www.researchgate.net/figure/NGS-and-analysis-pipelines-Most-common-tools-used-for-the-analysis-of-WGS-data-QC_fig2_317413533> **Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools** **Available in Clinical Oncology** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6895801/>  **Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data** 2019  <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2599-6>  **Most Common Tools Used for the Analysis of WGS Data**  <https://www.researchgate.net/figure/NGS-and-analysis-pipelines-Most-common-tools-used-for-the-analysis-of-WGS-data-QC_fig2_317413533> **Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6895801/>  **Comparitive Analysis of Differential Gene Expression Analysis Tools for Single-Cell Sequencing Data**  <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2599-6>  **FASTQC A quality control tool for high throughput sequence data**. 2014 September 29  <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>  **Free RNA-seq Analysis Tools – Software and Resources**  <https://bioinformaticshome.com/tools/rna-seq/rna-seq.html> |

**RNA-seq workflow: gene-level exploratory analysis and differential expression**

<http://master.bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html>

|  |
| --- |
| **Bioconductor**  **Analysis of single cell RNA-seq data (**University of Cambridge Bioinformatics)  <https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/public/index.html>  quality control, visualisation, data normalisation, exploratory data analysis, clustering, trajectory (pseudotime) inference, differential expression, batch correction, combining datasets, data integration, confounders, latent spaces, cell annotation, case studies |
| **2-day Course: (16 hours of video) RStudio**  Day 1: <https://www.youtube.com/watch?v=thHgPqQpkE4&feature=emb_err_woyt>  Processing Raw scRNA-Seq Data  Construction of Expression Matrix  Intro to R/Bioconductor  Seurat  Day 2: <https://www.youtube.com/watch?v=7dQ_pleDO2Y&feature=emb_err_woyt>  Clustering example  Feature Selection  Pseudotime Analysis  Differental Expression Analysis  DE Real Dataset  Comparing/Combining scRNA  Search scRNA-Seq Data  Seurat  scRNA-Seq Pipeline |

|  |
| --- |
| **Bioconductor**  **SEURAT: R toolkit for single cell genomics**  <https://satijalab.org/seurat/v3.1/pbmc3k_tutorial.html> |
| Installation Instructions for Seurat  <https://satijalab.org/seurat/install.html>  Vignettes: Guided Analyses  <https://satijalab.org/seurat/vignettes.html>  **Seurat - Guided Clustering Tutorial** 2020 April  Setup the Seurat Object  Standard pre-processing workflow  Normalizing the data  Identification of highly variable features (feature selection)  Scaling the data  Perform linear dimensional reduction  Determine the ‘dimensionality’ of the dataset  Cluster the cells  Run non-linear dimensional reduction (UMAP/tSNE)  Finding differentially expressed features (cluster biomarkers)  Assigning cell type identity to clusters |

|  |
| --- |
| **Bioconductor  Statistics for Genomics**  <https://www.youtube.com/playlist?list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr> |
| **17 Videos:** 2017 May  **Statistics for Genomcs: Distances and Clustering**  <https://www.youtube.com/watch?v=wQhVWUcXM0A&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=2>  **Statistics for Genomics Lab: Quick Introduction to R and Bioconductor**  <https://www.youtube.com/watch?v=J5h5WxOn3Gw&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=11>  **Statistics for Genomics Lab: Distances and Clustering RStudio**  <https://www.youtube.com/watch?v=PArRvqLUP6o&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=7>  **Statistics for Genomics: Introduction to RNAseq**  <https://www.youtube.com/watch?v=C8RNvWu7pAw&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=12>  **Statistics for Genomics: Advanced Differential Expression**  <https://www.youtube.com/watch?v=QINX3cI7qgk&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=15>  **Statistics for Genomics: Useful plots and bad plots**  <https://www.youtube.com/watch?v=46-t2jOYsyY&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=17> |

|  |
| --- |
| **Principal Components Analysis / Feature Selection** |
| **bioRxiv Search Results**  **mixOmics: an R package for omics feature selection and multiple data integration** 2017 Aug  <https://www.biorxiv.org/content/10.1101/108597v4.full>  **Differential Principal Components Reveal Patterns of Differentiation in Case/Control Studies** 2019 Feb  <https://www.biorxiv.org/content/10.1101/545798v1.full>  **pathwayPCA: an R package for integrative pathway analysis with modern PCA methodology and gene selection** 2019 April  <https://www.biorxiv.org/content/10.1101/615435v1.full>  **Accurate and Fast feature selection workflow for high-dimensional omics data** 2017 June  <https://www.biorxiv.org/content/10.1101/144162v1.full> |

|  |
| --- |
| **Bioconductor: DESeq2** Differential gene expression analysis based on the negative binomial distribution |
| **Analyzing RNA-seq data with DESeq2** 2020 Oct  <https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html>  **RNA-seq workflow: gene-level exploratory analysis and differential expression** 2019 Oct  <http://master.bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html>  **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2** 2014 Dec  <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>  **The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads**  <https://academic.oup.com/nar/article/47/8/e47/5345150>  **Data preprocessing and creation of the data objects pasillaGenes and pasillaExons** 2020 May  <http://bioconductor.org/packages/release/data/experiment/vignettes/pasilla/inst/doc/create_objects.html> |