

News from the Bioconductor Project

by the Bioconductor Team

The Bioconductor project provides tools for the analysis and comprehension of high-throughput genomic data. The 934 software packages available in Bioconductor can be viewed at <http://bioconductor.org/packages/release/>. Navigate packages using 'biocViews' terms and title search. Each package has an html page with a description, links to vignettes, reference manuals, and usage statistics. Start using Bioconductor and R version 3.1 with

```
source("http://bioconductor.org/biocLite.R")
biocLite()
```

Install additional packages and dependencies, e.g., **GenomicAlignments**, with

```
source("http://bioconductor.org/biocLite.R")
biocLite("GenomicAlignments")
```

Upgrade installed packages with

```
source("http://bioconductor.org/biocLite.R")
biocLite()
```

Bioconductor 3.0 Release Highlights

Bioconductor 3.0 was released on 14 October 2014. It is compatible with R 3.1 and consists of 934 software packages, 219 experiment data packages, and more than 870 current annotation packages. In total, the release includes 114 new software packages and many updates and improvements to existing packages. The [release announcement](#) includes descriptions of new packages and updated NEWS files provided by package maintainers.

The variety of research areas represented by Bioconductor packages are organized (and searched) via the **biocViews** interface. Here we highlight a few topics covered by the new packages. Methods for differential expression analyses are offered in **ballgown** (assembled transcriptomes), **derfinder** (RNA-seq data at base-pair resolution) and **csaw** (differentially bound regions in ChIP-seq data). Quantitative trait loci (QTL) analysis for 1H NMR data are provided in **mQTL.NMR**; **DOQTL** analyzes QTLs in multi-parent outbred populations. Copy number analysis in tumoral phenotypes and genomic focal aberrations are available in **facopy** and **focalCall**, respectively. Additions to the flow cytometry family include **flowcatchR** with tools for analyzing in vivo microscopy imaging data of flowing blood cells, **flowCHIC** for analyzing flow data of microbial communities based on histogram images and **flowDensity** which provides tools for automated sequential gating (analogous to manual gating) based on data density. Several new packages take the pipeline approach and facilitate steps from raw data to final analysis: **groHMM** (GRO-seq data), **FourCSeq** (multiplexed 4C sequencing data), and **systemPipeR** (NGS applications such as RNA-Seq, ChIP-Seq, VAR-Seq).

Bioconductor is built on the mature and flexible 'Ranges' infrastructure defined in packages such as **IRanges**, **GenomicRanges**, **GenomicAlignments**, and **GenomicFeatures**. Many packages rely on the Ranges framework for interoperable, re-usable analysis; [Lawrence et al. \(2013\)](#) provide an introduction and [Lawrence and Morgan \(2014\)](#) review strategies for processing, summarizing and visualizing large genomic data.

Our collection of microarray, transcriptome and organism-specific *annotation packages* use the 'select' interface (keys, columns, keytypes) which enable programmatic access to the databases they contain. The **AnnotationHub** complements our traditional offerings with diverse whole genome annotations from Ensembl, ENCODE, dbSNP, UCSC, and elsewhere.

Other activities

Bioconductor offers an [Amazon Machine Image](#) optimized for running Bioconductor in the Amazon Elastic Compute Cloud (EC2). The AMI comes pre-loaded with the latest release version of R, and a subset of Bioconductor packages. The AMI can be customized by installing R packages with 'biocLite()' or system-level packages with the Ubuntu package manager 'apt-get'. Files can be transferred to the EC2 instance via scp or the Rstudio interface.

A recent addition is our collection of Bioconductor [Docker Images](#). These self-contained environments run on Linux, Windows and Mac as well as virtual machines. The containers provide convenient access to a 'fresh R session' or specific version of Bioconductor without the overhead of installing packages and dependencies. Analysis-specific images come pre-loaded with packages of a common topic such as flow, proteomics, microarray and sequencing.

New Bioconductor package contributors are encouraged to consult the [package guidelines](#) and [Package Submission](#) sections of the Bioconductor web site, and use the new [BiocCheck](#) package, in addition to R CMD check, for guidance on conforming to Bioconductor package standards.

The Bioconductor web site advertises [training and community events](#); [mailing lists](#) connect users with each other, to domain experts, and to maintainers eager to ensure that their packages satisfy the needs of leading edge approaches. Keep abreast of packages added to the 'devel' branch and other activities by following @Bioconductor on Twitter.

Bibliography

- M. Lawrence and M. Morgan. Scalable genomics with R and Bioconductor. *Statistical Science*, 29:214–226, 2014. doi: 10.1214/14-STS476. URL <http://arxiv.org/abs/1409.2864>. [p]
- M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. Software for computing and annotating genomic ranges. *PLoS Comput Biol*, 9(8):e1003118, 08 2013. doi: 10.1371/journal.pcbi.1003118. URL <http://dx.doi.org/10.1371/journal.pcbi.1003118>. [p]

The Bioconductor Team
Program in Computational Biology
Fred Hutchinson Cancer Research Center