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 1024 software packages available in Bioconductor can be viewed at http://bioconductor.org/packages/. 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 version 3.1 by installing R 3.2.1 and evaluating the commands

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

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

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

Bioconductor 3.1 release highlights

Bioconductor 3.1 was released on 17 April 2015. It is compatible with R 3.2 and consists of 1024 software packages, 214 experiment data packages, and more 917 up-to-date annotation packages. There are 95 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. There are a great diversity of packages represented. Highlights include: **sincell** for assessment of cell-state hierarchies from single-cell data; gene set and network analysis packages (e.g., **RCyjs mogsa**, **nethet**, **pandaR**, **pwOmics**, **seq2pathway**); packages for methylation (e.g., **BEclear**, **conumee**, **DMRcaller**, **ENmix**, **RnBeads**, **skewr**), flow cytometry (e.g., **FlowRepositoryR**, **FlowSOM**, **flowVS**, **immunoClus**), and other domain-specific analysis (e.g., **LEA**, for landscape genomics); access to multiple-sequence alignment algorithms (**muscle**, **msa**); visualization packages such as **gtrellis** (genome level Trellis graph visualizes), **ggtree** (phylogenetic tree and associated annotation data), **ComplexHeatmap**, **seqPattern** (oligonucleotide patterns and sequence motifs centred at a common reference point), and **soGGi** (genomic interval aggregate and summary plots of signal or motif occurrence); and infrastructure packages such as **Rhtslib** (wrapping a recent version of the htslib C library for processing CRAM, BAM, and other high-throughput sequence files) and **GoogleGenomics** (to interact with the Google Genomics interface).

Our collection of microarray, transcriptome and organism-specific *annotation packages* use the 'select' interface (keys, columns, keytypes) to access static information on gene annotations (**org.*** packages) and gene models (**TxDb.*** packages); these augment packages such as **biomaRt** for interactive querying of web-based resources and **VariantAnnotation**, **VariantFiltering**, and **ensemblVEP** for annotation of DNA sequences. Use of these resources are documented in updated annotation workflows. The **AnnotationHub** complements our traditional offerings with diverse whole genome annotations from Ensembl, ENCODE, dbSNP, UCSC, and elsewhere; a recent addition includes Roadmap Epigenomics resources, with use described in the AnnotationHub How-To vignette.

Other activities

Several enhancements to the Bioconductor web and support sites aim to help users to identify appropriate packages while encouraging developers to provide high-quality software. In addition to biocViews terms to classify packages, each 'landing' page (e.g., http://bioconductor.org/packages/DESeq2) contains graphical shields that highlight cross-platform availability, download percentile, support site queries and responses,

years in Bioconductor, and maintenance activity during the last 6 months. Landing pages also contain shields more useful to package maintainers, including current build status and test coverage. Users posting questions on the support site are encouraged to add tags identifying the package the question is about; maintainers are subscribed to the tag and receive email notification of the question. This enables maintainers to stay informed of problems with their package, without requiring daily monitoring of the support site.

Continued availability of Bioconductor Docker and Amazon images provides a very effective on-ramp for power users to rapidly obtain access to standardized and scalable computing environments. Docker images are available for release and development versions of Bioconductor, with analysis-specific images pre-loaded with packages relevant to common analysis scenarios, e.g., of sequencing, microarray, flow cell, or proteomic data. Both Amazon and Docker images include RStudio Server for easy web-browser based access.

New Bioconductor package contributors are encouraged to consult the package guidelines and submission sections of the Bioconductor web site, and use the **BiocCheck** package, in addition to R CMD check, for guidance on conforming to Bioconductor package standards. New package submissions are automatically built across Linux, Mac, and Windows platforms, providing an opportunity to address cross-platform issues; many new package contributors take advantage of this facility to refine their package before it is subject to technical preview. Keep abreast of packages added to the 'devel' branch and other activities by following @Bioconductor on Twitter.

The Bioconductor web site advertises training and community events, including the BioC 2015, the Bioconductor annual conference, to be held in Seattle, 20–22 July.

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