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 824 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 3.1.0 with

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

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

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

Upgrade installed packages with

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

Bioconductor 2.14 Release Highlights

Bioconductor 2.14 was released on 14 April 2014. It is compatible with R 3.1.0 and consists of 824 software packages, 200 experiment data packages, and more than 860 current annotation packages. The release includes 77 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 current package maintainers.

New packages continue to represent a wide variety of research areas. **Rariant** identifies single nucleotide variants (SNVs) based on the difference of binomially distributed mismatch rates between matched samples. **SomaticSignatures** identifies the mutational signatures of SNVs. Filtering of SNVs based on inheritance models, amino acid change consequence, and minor allele frequency is offered through **VariantFiltering**. The **monocle** package, currently in the 'devel' branch, preforms differential expression and time series analysis for single-cell expression experiments. CRISPR/Cas is a compelling new molecular biology technique used for gene editing; **CRIPSRseek** helps find potential guide RNAs for input target sequences. **CCREPE** assesses the significance of similarity measures in 'compositional' data sets, as found in microbial abundance studies. **MIMOSA** models count data using Dirichlet-multinomial and beta-binomial mixtures with applications to single-cell assays. Machine learning methods such as SVM, Random Forest and CART are applied to RNASeq data in **MLSeq**; clustering and classification methods are used to summarize active paths in genome-scale metabolic and reaction networks in **NetPathMiner**. Bioconductor hosts a number of packages relevant to chemical compound discovery. The latest additions integrate bioinformatics and chemoinformatics into a molecular informatics platform in **Rcpi**, and performs alternating least squares analysis on chemical data in **alsace**.

In addition to these contributed packages, 'Ranges' infrastructure packages such as **GenomicRanges**, **GenomicAlignments**, and **GenomicFeatures** provide an extensive, mature and extensible framework for interacting with high throughput sequence data. One recent addition is the **GenomeInfoDb** package, which contains functions that allow translation between different chromosome sequence naming conventions (e.g., UCSC versus NCBI). Many packages rely on the Ranges infrastructure for interoperable, re-usable analysis; Lawrence et al. (2013) provides an introduction.

Our large collection of microarray, transcriptome and organism-specific *annotation packages* have been updated to include current information. Most of these packages now provide

access to the 'select' interface (keys, columns, keytypes and select) which enable programmatic access to the databases they contain. The **AnnotationHub**, with 10,780 entries, complements our traditional offerings with diverse whole genome annotations from Ensembl, ENCODE, dbSNP, UCSC, and elsewhere.

Other activities

The Bioconductor Git-SVN Bridge allows developers to synchronize a GitHub repository with the canonical Bioconductor SVN package repository. Commits made in SVN are propagated to GitHub and vice versa. This was driven by developer requests for access to social coding features, such as issue tracking and pull requests. The service has been well received, with 73 bridges established as of June 2014.

The Bioconductor Amazon Machine Instance is now compatible with the StarCluster toolkit. This enhancement makes it straightforward to configure a cluster with nodes that communicate via MPI, SSH or Sun Grid Engine, and to control jobs via the **BiocParallel** and **BatchJobs** packages. Details are available at the AMI page `http://www.bioconductor.org/help/bioconductor-cloud-ami/#using_cluster`.

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, 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. [p184]

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