News from the Bioconductor Project

by the Bioconductor Team
Program in Computational Biology
Fred Hutchinson Cancer Research Center

We are pleased to announce Bioconductor 2.2, released on May 1, 2008. Bioconductor 2.2 is compatible with R 2.7.0, and consists of 260 package. The release includes 35 new packages, and many improvements to existing packages.

New packages

New packages address a diversity of topics in high-throughput genomic analysis. Some highlights include:

Advanced statistical methods for analysis ranging from probe-level modeling (e.g., plw) through gene set and other functional profiling (e.g., GSEAlm, goProfiles).

New problem domains addressed by packages such as snpMatrix, offering classes and methods to compactly summarize large single nucleotide polymorphism data sets.

Integration with third-party software including the GEOmetadb package for accessing GEO metadata and AffyCompatible and additions to affxparser for accessing microarray vendor resources.

Graphical tools in packages such as GenomeGraphs and rtracklayer effectively visualize complex data in an appropriate genomic context.

New technical approaches in packages such as affyPara and xps explore the significant computational burden of large-scale analysis.

The release also includes packages to support two forthcoming books: by Gentleman (2008), about using R for bioinformatics; and by Hahne et al. (2008), presenting Bioconductor case studies.

Annotations

The 'annotation' packages in Bioconductor have experienced significant change. An annotation package contains very useful biological information about microarray probes and the genes they are meant to interrogate. Previously, these packages used an R environment to provide a simple key-value association between the probes and their annotations. This release of Bioconductor sees widened use of SQLite-based annotation packages, and SQLite-based annotations can now be used instead of most environment-based packages.

SQLite-based packages offer several attractive features, including more efficient use of memory, representation of more complicated data structures, and flexible queries across annotations (SQL tables). Most users access these new annotations using familiar functions such as `mget`. One useful new function is the `revmap` function, which has the effect (but not the overhead!) of reversing the direction of the map (e.g., mapping from gene symbol to probe identifier, instead of the other way around). Advanced users can write SQL queries directly.

The scope of annotation packages continues to expand, with a more extensive 'organism'-centric (e.g., org.Hs.eg.db, representing *Homo sapiens*) annotations. New 'homology' packages summarize the InParanoid data base, allowing between-species identification of homologous genes.

Other developments and directions

Bioconductor package authors continue to have access to a very effective package repository and build system. All packages are maintained under subversion version control, with the latest version of the package built each day on a diversity of computer architectures. Developers can access detailed information on the success of their package builds on both release and development platforms (e.g., `http://bioconductor.org/checkResults/`). Users access successfully built packages using the `biocLite` function, which identifies the appropriate package for their version of R.

New Bioconductor packages contributed from our active user / developer base now receive both technical and scientific reviews. This helps package authors produce quality packages, and benefits users by providing a more robust software experience.

The 2.3 release of Bioconductor is scheduled for October 2008. We expect this to be a vibrant release cycle. High-throughput genomic research is a dynamic and exciting field. It is hard to predict what surprising packages are in store for future Bioconductor releases. We anticipate continued integration with diverse data sources, use of R's advanced graphics abilities, and implementation of cutting edge research algorithms for the benefit of all Bioconductor users. Short-read DNA resequencing technologies are one area where growth seems almost certain.

Bibliography

R. Gentleman. *Bioinformatics with R*. Chapman & Hall/CRC, Boca Raton, FL, 2008. ISBN 1-420-06367-7.

F. Hahne, W. Huber, R. Gentleman, and S. Falcon. *Bioconductor Case Studies*. Springer, 2008.