News from the Bioconductor Project

by Bioconductor Core Team

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. The project has entered its twentieth year, with funding for core development and infrastructure maintenance secured through 2025 (NIH NHGRI 2U24HG004059). Additional support is provided by NIH NCI, Chan-Zuckerberg Initiative, National Science Foundation, Microsoft, and Amazon. In this news report, we give some details about the software and data resource collection, infrastructure for building, checking, and distributing resources, core team activities, and some new initiatives.

Software ecosystem

Bioconductor 3.15 was released on 27 April, 2022. It is compatible with R 4.2.0 and consists of 2140 software packages, 410 experiment data packages, 990 up-to-date annotation packages, 29 workflows, and 3 books. Books are built regularly from source and therefore fully reproducible; an example is the community-developed Orchestrating Single-Cell Analysis with Bioconductor. The Bioconductor 3.15 release announcement includes descriptions of 78 new software packages, and updates to NEWS files for many additional packages.

Infrastructure updates

- Thanks to a generous allocation (BIR190004, "Engineering and disseminating a software and analysis ecosystem for genomic data science") provided through the National Science Foundation ACCESS (formerly XSEDE) program, academic cloud resources including GPUs and highly accessible object storage systems are being integrated into project operations.

- Transition of primary funding administration from Roswell Park Comprehensive Cancer Center to Dana-Farber Cancer Institute has led to a number of changes to platforms in use for the checking and production of binary package images.

    - Linux builds occur at Dana-Farber Cancer Institute.
    - Windows builds occur in machinery provided by Microsoft Genomics in the Azure cloud environment.
    - MacOS builds occur at Dana-Farber Cancer Institute. Work on the support of ARM Mac systems occurs at MacStadium.
    - Details on the configurations of builders (e.g., the Linux builder for the devel branch) are available at the Build reports link at bioconductor.org.

- An interactive app for surveying adverse conditions arising for package install, build, and check processes has been introduced for release and devel branches.

- Cloud-based workshop delivery systems have been an integral part of Bioconductor conferences and teaching activities.

    - Workshops from Bioconductor 2022 are continuously available for inspection and hands-on exercises at `http://app.orchestra.cancerdatasci.org`, thanks to cloud computing support provided by Dr. Sean Davis of University of Colorado.
    - `http://workshop.bioconductor.org` is a Galaxy-based workshop collection deployed on Jetstream2 in NSF ACCESS.

Core team updates

- After six years of highly effective work in the core, Nitesh Turaga has left for a position in industry. We will miss him!

- New core developers Jen Wokaty and Alexandru Mahmoud have joined. Jen is a member of the Waldron Lab at CUNY. Alex works at Channing Division of Network Medicine.

- Jen and Alex are joined by long-term core members Lori Kern of Roswell Park Comprehensive Cancer Center, Marcel Ramos of CUNY and Roswell, and Hervé Pages of Fred Hutchinson Cancer Research Center.

New initiatives

- Thanks to efforts of members of the Technical and Community Advisory Boards and community members, a collection of working groups has been defined to achieve new project aims. An overview of currently active working groups is available, along with guidelines for proposing new working groups.

- The objectives of the bioconductor-teaching working group are stated at the associated repository:

  > The Bioconductor teaching committee is a collaborative effort to consolidate Bioconductor-focused training material and establish a community of Bioconductor trainers. We define a curriculum and implement online lessons for beginner and more advanced R users who want to learn to analyse their data with Bioconductor packages.

- A mentoring program for new developers has taken flight.

- Thanks to an Essential Open Source Software grant from the Chan-Zuckerberg Initiative, we have partnered with the Dana-Farber Cancer Institute YES for CURE (Young Empowered Scientists for Continued Research Engagement) program to offer instruction in cancer data science to interested undergraduates. A pkgdown site includes current curricular materials.

- With the NSF-based academic cloud resources previously mentioned, we have begun gestation of G-DADS, a program for Genomic Data and Analysis Development Services, with the objectives of providing publicly accessible storage and compute on exemplars of the latest high-volume experimental modalities, and of promoting GPUs to first-class citizenship in our build and check systems.

Using Bioconductor

Start using Bioconductor by installing the most recent version of R and evaluating the commands

```
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install()
```

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

```
BiocManager::install("SingleCellExperiment")
```

Docker images provides a very effective on-ramp for power users to rapidly obtain access to standardized and scalable computing environments. Key resources include:

- bioconductor.org to install, learn, use, and develop Bioconductor packages.

- A list of available software, linking to pages describing each package.

- A question-and-answer style user support site and developer-oriented mailing list.

- A community slack (sign up) for extended technical discussion.

- The F1000Research Bioconductor channel for peer-reviewed Bioconductor work flows.

- The Bioconductor YouTube channel includes recordings of keynote and talks from recent conferences including Bioc2022, EuroBioC2022, and BiocAsia2021, in addition to video recordings of training courses.

- Our package submission repository for open technical review of new packages.

Recent Bioconductor conferences include BioC 2022 (July 27-29), and European Bioconductor Meeting (September 14-16). Each had invited and contributed talks, as well as workshops and other sessions to enable community participation. Slides, videos, and workshop material for each conference are, or will soon be, available on each conference web site as well as from the Courses and Conferences section of the Bioconductor web site.

The Bioconductor project continues to mature as a community. The Technical and Community Advisory Boards provide guidance to ensure that the project addresses leading-edge biological problems with advanced technical approaches, and adopts practices (such as a project-wide Code of Conduct) that encourages all to participate. We look forward to welcoming you!

Bioconductor Core Team
Channing Division of Network Medicine
Mass General Brigham
Harvard Medical School, Boston, MA

Department of Data Science
Dana-Farber Cancer Institute
Harvard Medical School, Boston, MA

Biostatistics and Bioinformatics
Roswell Park Comprehensive Cancer Center, Buffalo, NY

Fred Hutchinson Cancer Research Center, Seatlle, WA

CUNY Graduate School of Public Health, New York, NY
   maintainer@bioconductor.org