Bioc Technical Advisory Board Minutes

4 January 2024

Present: Vince Carey, Charlotte Soneson, Laurent Gatto, Lori Kern, Robert Shear, Michael Love, Levi Waldron, Marcel Ramos, Hervé Pagès, Henrik Bengtsson, Alexandru Mahmoud, Brian Schilder, Sean Davis, Jen Wokaty, Kasper Hansen, Stephanie Hicks (from :25), Rafael Irizarry (from :25)
Apologies: Wolfgang Huber, Helena Crowell, Ludwig Geistlinger, Davide Risso

:02 - :03 Previous [minutes](#) approved

:03 - :14 Welcome and notifications
- bibliometrics - we don't formally keep track of publications related to packages, but it would be useful to have an easy way to generate publication/citation statistics. CITATION files in packages are not always updated after package submission. Tracking down publications would need manual work - bibliometrics working group.
  - Sean will map packages to DOIs to publications from CITATION files and present results.
  - Could have formal checks e.g. for presence of a DOI.
  - Provide examples of CITATION files.
  - usethis::use_citation()
  - CRAN asks maintainers to include references in the Description field in the DESCRIPTION file. Helps promote citing work you rely on.
- Recent issues with VariantAnnotation:
  https://github.com/Bioconductor/VariantAnnotation/issues/79,
  https://github.com/Bioconductor/VariantAnnotation/issues/78
  - Maintainability concerns for core packages as original developers leave.
  - The ability of Bioconductor to handle genomic data on the variant level is key.
  - How deep is the 'contract'? What is generally useful, and what are edge cases?

:15 - :16 Discussion/voting process on TAB governance changes.

:20 - :32 Question from the BioC planning committee about overall goals and positioning of the BioC conference.
- Comments
  - Recent movement towards EuroBioC and BioC Asia becoming more 'independent'/similar to the US conference. The naming ('BioC') is historic.
  - Historically, one of the main goals of the meeting was for the core developers to meet in person.
  - Historically, many keynote speakers have not had strong direct relation to Bioconductor (but could rather contribute something new to learn, which would be useful for the developments within the project).
  - Is 'keynote' speakers the right notation? Or is 'invited speaker' better?

- ○ Inviting speakers is a significant part of the conference budget, especially if they are expected to be presenting in person.
  - ○ Propose more symmetry among the three major annual conferences in branding/advertisement.
  - ○ BioC makes efforts to be accessible to global attendees also through its attention to the online component of the hybrid conference, and through provision of travel scholarships.

:32 - :42 Rafa comments on extensions for SummarizedExperiment.
- Historically, several classes existed that were basically 'renamings' of ExpressionSet. Renaming was still useful, as it indicated something about what the values inside the object represent.
- More recently, efforts aimed to keep the raw data and the summaries in the same object (e.g. dimension reductions in SCE) e.g. to plot them jointly. A bit different from the 'classical' R API with a separate object for results of a fit etc.
- ExpressionSet/SE are very flexible and can encompass almost any type of data. As a consequence, some slots won't make sense for some types of data - having more, dedicated classes would address this (but then there would be more classes, which may also be confusing).
- Hard to predict what workflows will look like in the future before developing new classes.
- Dependency info: 45 packages DEPEND on SingleCellExperiment and 25 SUGGEST it.
  - ○ Might be worth looking at SingleCellExperiment "methods" to see how often the extensions (above SummarizedExperiment) are used 'in the wild'.
- Other similar objects outside Bioconductor (AnnData, SeuratObject) have similar slots as SingleCellExperiment.

:42 - :48 https://pubmed.ncbi.nlm.nih.gov/37704947/ - "Bioconductor-friendly GPU-accelerated pipeline for bisulfite sequencing"
- Should Bioc build/test also this type of applications?
- Nextflow is testable, but requires a cluster environment, and this workflow relies on third-party tools.
- We should be aware of such applications, and can publicize them, but testing would be a big commitment.
- Even thinking about how it *could* be done within the build system might be helpful for general ideas about build system design.
- Farm out to/build on top of Dockstore? There are currently 36 Dockstore workflows matching the search 'Bioconductor': https://dockstore.org/search?entryType=workflows&search=bioconductor

:48 - :58 Think about a new web resource "science.bioconductor.org" which is targeted at investigators and funders to promote understanding of the impact of the project with highlighted new publications, new packages and new technical solutions under development.
- Who would be the users - funders are unlikely to randomly hit such a webpage.

- Structure the yearly report to have sections that can easily be extracted and put into grants, easier ways to generate citation counts, …
- How would such a webpage be kept up to date? Community-driven? How to choose what goes on the page?
- Could also make blog posts with a specific tag. Support site handles news and tutorial posts.
- R consortium infrastructure steering committee - hands out funding to projects. Operating under the Linux foundation. Has a marketing group, which has made a big effort to reach out to previous projects -> blog posts, outreach on social media. Helps bring awareness to industry.
- Lots of packages (that depend on Bioc packages) live only on GitHub and are not contributed back to the project.
- Encourage blog posts about science being accomplished with Bioconductor.

:58 - :60 Vince and Alex presentation on workshop.bioconductor.org.
- ~510 registered users total since launch in May 2023.
- 491 launches since Sep 15th 2023 (job db restarted (separate from auth db)).
- Bot added for workshop authors to be able to get stats for their workshops, example.
- BiocWorkshopSubmit Shiny app available.

---

almahmoud commented 2 weeks ago   (Member) ···

/stats id="scdneyspatialabacbs2023"

🙂

---

almahmoud commented 2 weeks ago   (Member) ···

AUTO-GENERATED RESPONSE
The workshop with id 'scdneyspatialabacbs2023' has been launched a total of 126 times, by 36 unique users.

🙂