Bioc Technical Advisory Board Minutes

6 June 2019

Attending: Martin Morgan, Wolfgang Huber, Robert Gentleman, Vincent Carey, Rafael Irizarry, Matt Ritchie, Kasper Hansen
Regrets: Levi Waldron, Aedin Culhane, Michael Lawrence, Sean Davs

Schedule

:00 - :05 Check-in

:05 - :10 Minutes from previous meeting
- Review of minutes & action items
- Vote to approve

:10 - :15 Material review
- Funding
    - [CZI seed network](#) -- $2M total, spread over 3 years, 8 groups. Funded (no public announcements until Friday, June 21).
    - U41: establishing build node with Vince
    - U24 new proposal due June 11
    - U24 / U01 ITCR face-to-face presentations from Martin, Vince; R / Bioc mentioned implicitly or explicitly in at least 10 presentations
- Events
    - CSAMA
        - Updates on computing or lab material dissemination?
        - Planning document; [Announcement](#); [Webpage](#)
    - BiocEurope: Brussels Dec 9-10; Laurent Gatto local organizer
    - BiocAsia: Sydney Dec 5-6; Peter Hickey (WEHI), Stephen Pederson (University of Adelaide), Saskia Freytag (Harry Perkins Institute) and Dario Strbenac (University of Sydney) are organizing.
    - Meetups
- Committees
    - Conference (BioC2019; below)
    - Emerging Topics
- Working groups
    - Developer forum, coordinated by Mike Smith. Monthly presentations. Michael Lawrence & Wolfgang Huber
        - Should be ready to launch.
    - Code of Conduct
        - Aim: adopt Project-wide Code of Conduct at July TAB

- Other
    - Mike Love's [Literate Publishing](#) talk
    - [Notes from HCA meeting from DCP breakout session](#) (Matt Ritchie)
        - Nils Gehlenborg interested in collaboration in connection with UpSetR
        - The zarr format as a fundamental component of HCA DCP was noted
        - Marcus Kinsella interested in MatrixBrowser, Marcel Ramos will update

:15 - :20 Conferences & workshops
- Scope for realistic testing of workshop activities
- Other logistics
    - Registration closed at 201 for main conference; 30+ on waitlist. Growing user base; earlier content & program development; location.

:20 - :30: [Governance](#)
- TAB Call for Nominees
- Open nominations June 10, close nominations July 3
- Proposal: closed yes/no vote on each nominee, any nominee with >⅔ support is on the board up to maximum (15, or leave some open?), if too many reach ⅔, then keep those with most yeses.
- Up to 4 positions available

:30 - :40 Collaborative project opportunities, inspired by n[ew CZI Funding](#) opportunity -- 1 year, 50 - 250k; due Aug 1.
- Ideas: new build system (Martin)
- annotation server
- [FacileData](#) (that's work by Steve Lianoglou et al; close connection to MultiAssayExpt)
- Ideas (Aedin)
    - 1) core infrastructure development of core bioc classes and deprecation of redundant classes, better documentation on core classes , core annotation, support for spatial data, single cell proteomics, image files
    - 2) web site/user support site, focus groups, training and hackathons to increase community engagement and cross-community (eg to python, Stan, etc) engagement.
    - 3) documentation/credit to package developers for good documentation 1) develop template to provide the ability to publish vignettes on F1000 channel 2) bookdown is clunky but could be developed to generate custom books of documentation from vignettes, 3) searchable vignettes to auto generate "task views" summary of packages.
    - 4) bioc-bloggers or bioc-gists.
    - 5) Hire someone for technical documentation (cheat-sheets, improve some key vignettes)
- Coordinated hackathons; data competitions; sand pits

- Survey the community (perhaps as part of BioC2019) to see what others might want? Feedback-driven project improvement (which of the options above are most pressing)

:40 - 60 Review of Current Technical Priorities

- CRAM (Martin)
    - Requires updates to Rhtslib, as well as moderate-sized updates to Rsamtools
- kubernetes (k8s) / cloud (Martin)
    - Fun: deploy 'manager' and *n* BiocParallel workers using RedisParam to communicate in, e.g., the Google computational cloud
    - k8s-redis-bioc-example ('basic' k8s deployment) and helm chart (parameterized deployment, e.g., edit values.yaml to launch 30 workers on customized containers)
- Annotation packages (Martin)
    - Annotation builds will be moved up more than a month before the release, to avoid confusion & overlapping responsibilities at release time
    - For the next annotation builds, we'll dedicate some time to improving (rather than just 'turning the crank') our current annotation build scripts
    - Revisions will include consideration of MANE (Ensenbl / NCBI) transcript annotation
- Scalability
    - hail (RStudio) could be interfaced with Bioconductor. Vincent will follow up on hail repo
    - altrep: important to be aware of, but early days of development see Jeifie's sharedObject.
    - There are some significant unresolved performance issues in DelayedArray. Using within a for loop is problematic. Chunking must be optimized for specific backend and application, or code can become 100x-1000x slower.
    - tileDB (Martin / Qian) -- Pretty performant. Fast updates. Native sparse representation. tiledb R package (github) is quite immature.
    - scalability working group will summarize current issues.
- Other priorities
    - Brainstorming
    - Do any brainstorming activities translate into action items?

Action Items -- new and ongoing
- Announced TAB publicly; nomination process underway.
- Review annotation process. Martin has discussed w/ core team; summarized earlier in meeting
- Scalability task force. Vince, coordinating with Michael & Kasper; scalability slack