

Bioc Technical Advisory Board Minutes

3 March 2022

Attending: Vince Carey, Lori Shepherd, Shila Ghazanfar, Charlotte Soneson, Alexandru Mahmoud, Levi Waldron, Stephanie Hicks, Marcel Ramos, Rafael Irizarry, Robert Gentleman, Laurent Gatto, Jennifer Wokaty, Kasper Hansen, Mike Love, Wolfgang Huber, Aedin Culhane (from :40)

Regrets: Aaron Lun, Hector Corrada Bravo, Nitesh Turaga

:03 - :04 [Prior minutes](#) approved

:05 - :06 Disclosure of tentative plan for releasing 3.15 (Lori). This is all tentative based on an email exchange with Tomas that the proposed R 4.2 release date is April 22nd.

- March 7 announce key dates publicly
- March 22 have .db0, TxDb, and OrgDb packages in devel
- As closely after March 22 as possible, have TxDb, OrgDbs in hub
- April 1 is new package submission deadline (to have at least one review before release - still subject to passing full review)
- April 11-12 stop current release 3.14 builds to prep new (3.16) builders
- April 12 candidate 3.15 release feature freeze
- April 22 recommended date for all changes to be committed to devel
- April 25 absolute last day to commit changes before freeze to make release branch
- April 26 freeze devel branch and make release branch - builds run overnight
- April 27 Announce release

:06 - :30 Ad hoc [presentation](#), Stephanie [The random short talk series with [assignments for future meeings](#)]

Comments/questions:

- Will the SpatialExperiment class work for all (foreseeable) types of spatial data? Attempt to make it general enough at least for barcode data. Also for MERFISH, seqFISH.
- Will there ever be 3d (rather than 2d) coordinates? To some extent the contiguous technical replicate slice is already in the z-plane. For imaging-based spatial genomics the distance between z-slices can be smaller and could be continuous if segmentation is done in 3D.
- Different platforms/technologies generate very different data types/formats/resolution - do we need multiple classes? Need to think carefully about what the class should achieve - how does it help users. Classes provide standardized infrastructure, convenience functions, etc.
- Is there scope for a SpatialMolecules class? (Rather than a BumpyMatrix)
- Make sure to name classes appropriately (there will be spatial versions of other assay types).

- Some current workflows keep transcripts in a DataFrame format - assign them to cells according to some segmentation mask, then import into SingleCellExperiment/SpatialExperiment object. Would like to be able to apply multiple segmentation masks to a raw data frame. Should this be within the SpatialExperiment class, or just stick to DataFrame?
- Are the 10x Visium H&E images just a single channel image? i.e. no z-plane or colour channels?
- Are there enough example data sets? STExampleData (one slice of a Visium data), TENxVisiumData, one sample from mouse embryogenesis SeqFISH data, MouseGastrulationData package, MERFISH example being put together.
- Be aware of HuBMAP efforts in this area.
- Should we organize a BioC2022 session on the topic? And/or a session at SDSS?

:30 - :35 Directional info – reminders; these were all discussed last meeting

- "Mirrors" – new approach being developed by cloning mirror to XSEDE Jetstream, "smart" selection of nearby resource to enhance BiocManager::install, reduce usage of AWS.
- 10 TB at OpenStorageNetwork – If you have a project with large public (or maybe temporarily private) data volume let's talk about how to use this.
- HCA-oriented matrix API planning with CZI/TileDb – [issue stream](#).
- La Piana recommendations on governance received. Was discussed 21 Feb 730am ET.
- Git and GitHub strategy considerations ongoing: change master to main/devel, improve self-service in credentials management, increase effectiveness of GitHub Actions in general testing/maintenance/artifact creation.
 - TAB members who make significant use of GitHub Actions are invited to add comments here.
- Binary repository production and management for container usage.
- EBI Biostudies used for deep archiving of legacy Bioconductor.
- AWS Open Data proposal by Enis Afgan accepted, he invited us to add to the proposal.
 - Move ExperimentData and AnnotationData packages.
 - Build cloud-native interfaces to make AWS happy.

:35 - :40 Working Groups / Committee Page for community involvement.

- <http://workinggroups.bioconductor.org/>
- Stephanie initialized. Kevin set up bookdown. Lori populated template starter info. What do we think?
 - VJC → too much room devoted to logo and general comments on project.
- List all working groups, with lead, information about how to join or start a new working group.

:40 - :45 CAB liaison – (Aedin and Lori). For the future, move this reporting to be mostly in writing, and provide a shorter update in meeting. Important to maintain two-way communication.

- [EuroBioC2022](#) postponed to 14-16 September 2022.
- Election time.

- Nominations for new member spots and officers due Feb 25th.
 - 11 applied
 - 2-3 running for office.
- Due to governance, 2022 will have expanded board. 2 people voluntarily stepped down (Katerina, Saskia).
- Presentation by CSCCE
 - Helping hire community manager and will offer training and support.
 - Community manager currently advertised as a 2 year full time position.
 - Resources at: <https://www.cscce.org/resources>
 - Training opportunities: <https://www.cscce.org/trainings/>
 - Community managers: <https://www.cscce.org/get-involved/>
 - Research: <https://www.cscce.org/research/>
- Reviewed CAB Working Groups for inclusion on workinggroups.bioconductor.org.
 - Current status (active, non active).
 - Current members.
 - Does it need/accepting more members.
 - How often does it meet.
 - Communication: Slack channel, contact info.
- Scanpy developers considering a community manager - meeting with Wolfgang Huber (and Mike) to discuss how Bioconductor are doing this.

:45 - :50 (optional) Comments from core.

- Release schedule is tentatively planned. See above.
- AnnotationHub default resources follow up:
 - In the last meeting there was some discussion about what resources we provide by default in the hubs. Currently at release time we provide in the hubs the equivalent databases for package versions of TxDBs and OrgDBs and a limited selection of what we have termed “non standard” OrgDBs (meaning the non package OrgDb). We provide the top 1000 taxonomy ids (and aliases) as determined by having the most coverage for genes using data from <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/>.
- GENCODE - would be good to have in Bioconductor. Is there a way to do automated processing? Or find someone to do it manually?

:50 - :60 open

- Discussion about potential publication venues for workflows. Interactions with EMBO Press, F1000 Research, eLife. Would like to have a setup with more continuous releases, rather than one static version, and to keep the document in (e.g.) markdown format across revisions/updates. Could we host the computational infrastructure on our side (publishers may be reluctant)? XSEDE might play a role in this.