

Bioc Technical Advisory Board Minutes

18 July 2024

Present: Wolfgang Huber, Laurent Gatto, Vince Carey, Charlotte Soneson, Marcel Ramos, Hervé Pagès, Helena Crowell, Lori Kern, Robert Shear, Rafael Irizarry, Davide Risso, Alexandru Mahmoud, Jen Wokaty, Erdal Cosgun, Sean Davis

Apologies: Michael Love, Stephanie Hicks, Levi Waldron, Kasper Hansen, Ludwig Geistlinger, Henrik Bengtsson

:03 - :04 What to do about packages with non-compliant licenses:

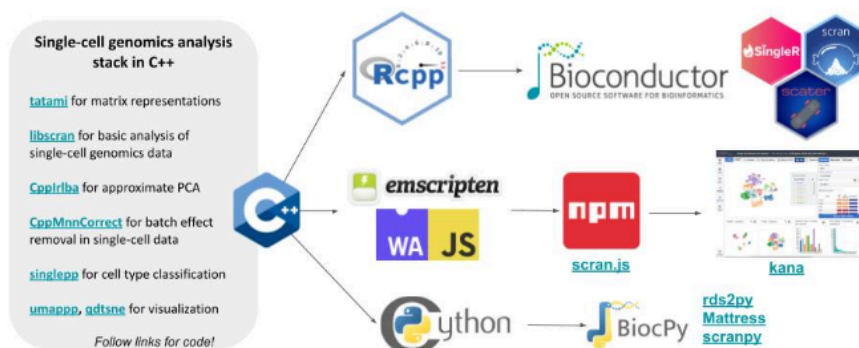
- [Bioc-devel] Package with restrictive licenses, not suitable for Bioconductor [[1](#), [2](#)], and [Bioconductor package licenses](#).

:04 - :05 Previous [minutes](#) approved

:05 - :30 Technical topics.

- C++ foundations and BiocPy. <https://github.com/LTLA/scrان.chan> -> perform scrان computations without the entire Bioc stack. How to relate BiocPy to AnnData?

Case study with single-cell genomics



- "Workflow packages" and workflow language (e.g., nextflow, snakemake, cwl) interfaces.
 - Looking for an alternative to F1000Research.
 - Meeting planned August 5, 10am EDT to discuss renaming and publishing.
 - "Tutorial", "How-To".
 - Tutorial is often used for material for novices (as opposed to manuals, for more experienced users).
 - Need for a citable, "proper" peer-reviewed publication, as well as continuous integration as teaching material used in courses.
- Spatial transcriptomics coordination process.
 - OSTA book.

- SpatialData by Helena; relation to Marconato/python. Missing Zarr representation.
 - Zarr v3 spec, official python bindings not yet released.
 - Image-based spatial transcriptomics - how streamlined is the generation of the count matrix from the images?
 - Molecule data is large - e.g. 2 x 30Gb files per sample. Cell-level count data is much smaller.
 - What to share e.g. via ExperimentHub? Try to be as language-agnostic as possible. 10x is sharing Parquet files - output from other vendors can be converted. DuckDB for querying.
 - Probably we don't want to go to the raw images (with molecule detection data). Nanostring have additional images (DAPI, cell surface markers) that are useful to keep. Transcript-level Parquet files. Polygon files with cell segmentation (can also be Parquet files).
 - Put some example data in the Open Storage Network.
- GitHub Enterprise and Actions. Relation to build/check/distribute processes.
- With more bioinformaticians willing and able to "scramble" to get working environments with python, conda, docker, is Bioconductor's "coherence and convenience" losing recognition and relevance?

:30 - :45 Governance and upcoming elections for TAB.

- Interfacing to CAB (joint meeting?)
- Governance changes approved with 11 yes/1 no votes.
- Suggested timeline for elections (coordinated with CAB):
 - Announce at BioC2024
 - Discuss nominees at September meeting
 - Vote after September meeting
 - New members joining in October
- Nomination form discussed
 - Add questions about pronouns and current geographical location.
- Meet the CAB/TAB/core team session of BioC2024 scheduled for July 24, 4:30pm Eastern time.
- Scheduling of upcoming meetings
 - Regular August meeting (August 1) too close to the July meeting? -> Keep
 - September meeting (September 5) overlaps with EuroBioC -> Keep

:45 - :50 Publications in progress.

- Teaching paper.
- BiocManager and BiocCheck, gatekeeper concepts.
- Matrix representations.
- Very speculative: reproducibility in LLM applications, case of Geneformer. [MSoft attack](#) is challenging to reproduce. Benchmarking of LLMs is difficult.
- DelayedArray.

:50 - :60 Discussion.

- C++ libraries from Aaron Lun, tatami, singlepp, libscrn – Vince has submitted an abstract to the scverse conference aiming to discuss "language-agnostic" development and prospects for symbiosis between scverse and Bioc.
- Performant algorithms in low-level languages with bindings in R/python/... -> practical?
- Vince Buffalo is working on a Rust package that combines bedtools (piping in BED data) and plyranges functionality (e.g. group by and summarize).
- How to increase awareness of the Bioc ecosystem and what it can offer? Focus should be on robust and performant methods. Presence at conferences, online courses.