Bioc Technical Advisory Board Minutes

2 October 2025

Present: Vince Carey, Gabriele Sales, Wolfgang Huber, Henrik Bengtsson, Robert Castelo, Laurent Gatto, Jacques Serizay, Hugo Gruson, Lori Kern, Michael Lawrence, Sehyun Oh, Ludwig Geistlinger, Levi Waldron, Marcel Ramos, Andres Wokaty, Helena Crowell
Absent: Kasper Hansen, Rafael Irizarry

:00 - :05 Previous minutes approved.

:06 - :15 Welcomed new/re-elected TAB members.

- Gabriele Sales, Laurent Gatto, Hugo Gruson, Robert Castelo

:16 - :25 Recap of EuroBioc2025 (September 17-19, Barcelona, Spain)

- 170 participants from 24 different countries (good model!)
- Proposed dates for 2026: June 1-5 in Turku, Finland (to consider: effects on university schedules, October weather is challenging)
- useR will take place in July 6-9, 2026 in Warsaw, Poland
- nf-core discussion as BoF
  > Trying to follow up with Seqera as a collaborator on a European proposal for funding a PhD network involving a number of European Bioconductor developers, seqera/nfc has expressed interest in integration (bioconda and biocontainers)

:25 - :30 CAB updates.

- Johannes Reiner lead a discussion around interoperability of software for mass spectrometry data
- Licensing of logo discussed (CC-BY-SA 4.0)
  > Johannes Reiner will make PR to add to logo page on website
- Presented Governance Draft and asked to provide feedback/comments
- CAB elections took place (new members invited to October meeting)

:30 - :35 Working group updates.

- Discussion/comments on BLAS (see also Appendix 1 on "pent up" discoveries about GRanges, performance, etc.)
- "refreshing infrastructure concept" (e.g., GRanges) as a new working group?
- Two new package reviewers have been added
- BiocClasses working group cancelled for EuroBioC
- MS proteomics group inactive at the moment

:35 - :45 Invited speaker spreadsheet: Vince Carey on U24 proposal and related topics (slides).

- Eight components to the Bioc ecosystem,
  core has to contribute to most in one way or another
- How to prioritize/plan is challenging (maintenance)
- Should Bioc try to do something for Python (the way it did/does for R)?
  > making Bioc scripting language-agnostic/an ecosystem shared across languages
- U24 renewal is submitted (five-year funding)
- No explicit allocation of time/money for maintenance
  of core infrastructure/packages (e.g., IRanges)
- Ideas from EuroBioc2025:
  connecting linters with BiocCheck;
  meta-benchmarking for optimization of functions
- Work to improve/complement BiocViews tagging
  of packages with EDAM ontology is underway
- Separate build system to hand out BiocPy?
  User benefits of current bi-annual releases are substantial

:45 - :50 Discussion of talk.

- Problem with conda and consistency, CXX ...
- Stack adds folders in HPC on shared filesystems,
  loading software is slowed down
  > use of conda might be discouraged
- mamba and pixi are alternatives
- Many are moving to uv for python package management these days
- lag with respect to current Bioc version is problematic (bugs remain, etc.)
- ronda downloads R (CRAN & Bioc) sources & builds packages automatically;
  typically build might only last < 24h (dependencies change, etc.)

:50 - :60 Open floor.

- Bioc v3.22 release scheduled October 29th
- Move to runiverse? Current focus on release. Both systems
  aimed to run in parallel for the next release. Goal: 1-1.5 years.
- Working group on Julia has been proposed
- Previous meeting on Rust, no activity since.
  Rust support at base R level is lacking
- Sign-up sheet aimed at giving TAB members a chance
  to control the agenda, give a presentation, etc.

Appendices

Appendix 1: Håkon Tjeldnes concerning GRanges.

- Dozens of issues in mind, not sure how to proceed
- E.g., Views issue in IRanges here

(quote)

My main 4 topics would be:

- Speed improvements
- New classes to existing packages:
    - experiment-class: A master config object (like SummarizedExperiment but 1 level up) for all files related to a study: bam files, converted file formats, count tables (summarized experiments), linking to organism and annotation, directory linking.
    - covRle-class, a list of RleList coverage per strand with seqinfo, this bypasses the whole coverage() call in coverageByTranscripts)
- New methods and functions (e.g., as.character(GRangesList), makeGRangesListFromCharacter)
- New packages: Top priority would be TranscriptRanges and TranscriptCoverage.
    - TranscriptRanges: A layer on top of GRangesList which requires equal strand for all GRanges elements per group and strand specific sorting, to sort a - strand transcript should give high end coordinate first etc: it also have unique functions like startSites() which is the start of first lowest start for + strands and highest end coordinate for - strand transcripts etc. startCodons and stopCodons are two other examples.
    - TranscriptCoverage: The single nucleotide and region count methods are now a mess in many packages: coverageByTranscripts is hidden in GenomicFeatures, while countOverlaps have a lot of inefficient implementation for many inputs. This should be all standardized, and support a myriad of input classes.
      Example in ORFik I support giving 2 bigwig files (one per strand) or 1 file (unstranded), to get random access compute of coverage over a transcript etc (a lot of packages would love this function to be in the core packages).

As I see it, speed improvements and new methods are very easy to merge in, new classes are harder as implementations for all generics etc takes time. As for new packages, it would be somewhere in between I guess.

Let me know what you guys think, so I don't overflow you with stuff that won't be merged in anyways or you don't have time to follow up.

(end quote)

Appendix 2: Giotto publication.

Quote below from new Nature Methods paper on Giotto. The group was in touch a few years ago about joining Bioconductor but package size standards were too stringent. The paper generously references Bioconductor and DelayedArray, and notes the affordance of various bidirectional converters.

"The programming language R is widely used for analyzing biological data. It has seen a strong increase in users across the fields of biomedical sciences because it offers an array of tools for statistical and genomics analysis. For instance, the Bioconductor project has a central role in more advanced omics analysis and has created a collaborative community for method development and interoperability [17]. In parallel, the geospatial field has a long history of working with various spatial data, including raster-based and vector-based data types [18]. However, implementations of the many associated and established spatial data analysis or simulation methods are less frequent within the field of biomedical sciences. In the present study, we leveraged our previous expertise from Giotto [19,20] to create a rich and inclusive ecosystem for spatial data analysis and engineering."

Appendix 3: Loading a GRanges.

```
Loading required package: GenomicRanges
Loading required package: stats4
Loading required package: BiocGenerics

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:stats':

    IQR, mad, sd, var, xtabs

The following objects are masked from 'package:base':

    anyDuplicated, aperm, append, as.data.frame, basename, cbind,
    colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
    get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
    match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
    Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
    table, tapply, union, unique, unsplit, which.max, which.min

Loading required package: S4Vectors

Attaching package: 'S4Vectors'

The following object is masked from 'package:utils':

    findMatches

The following objects are masked from 'package:base':

    expand.grid, I, unname

Loading required package: IRanges
Loading required package: GenomeInfoDb
```