Bioc Technical Advisory Board Minutes

6 November 2025

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

:04 - :05 Previous minutes approved.

:06 - :14 Introduction of new TAB members.

- **Gabriele Sales** (GPU acceleration in R,
  bridging with/reusing infrastructure available in Python,
  distribution to users not trivial > how to facilitate installation for end-users?)
- **Robert Castelo** (scaling GSVA to single-cell & spatial data,
  genomic DNA contamination in clinical samples – gDNAx,
  quantification of expression in transposable elements – atena,
  training working group, EuroBioC organization)
- **Hugo Gruson** (software engineering background, e.g., hdf5, Rarr;
  involvement in other communities, e.g., rOpenSci & carpentries;
  interested in meta-tooling, tools for developers, open software development)
- **Laurent Gatto** (mass spec proteomics, metabolomics,
  involvement in teaching working group, EuroBioC organization)

:15 - :17 Bioconductor 3.22 has been released.

- Infrastructure problems were encountered for builders in DFCI
  (internet outages > delay in delivery)
- Windows binary production/testing not functional (government shutdown)
- Annotation build currently missing > will be up shortly

:17 - :18 CAB updates.

- Licensing (CC-BY-SA 4.0) added to logo page
- Welcome and Introduction to 7 new members
- Aedin CAB lead discussion: Federated Genomics Data across countries
- Reminder to look at Levi's suggested Bioc charter (no discussion yet)

:18 - :36 momics overview from Jacques Serizay.

- Aim: Organization from (multi-modal) experiments to data files
- Sequencing and processing > Consolidation of data files > Multi-omics analysis
- momics = managing collections of genome-wide datasets, represents format & query framework, using TileDB arrays as backend, default one-hot encoding
- momics repos disk space less vs. bigwig, fasta(.gz) files
- Past data consolidation: Backup, sharing, cloud access (with pre-configured ways to query)
- momics API implements scalable, efficient programmatic queries
- MomicsDataset = class for tensor-flow based analyses & development for momics datasets/repos, support for direct queries of input & output data, useful for ML-based applications
- Python API already available > implementation for R/Bioc underway (using CLI in Python)
- Currently fully tied to TileDB > alternatives, (e.g., .parquet, as backend not supported)
- Metadata-based optimization (e.g., 1,000s of samples) is a concern/worth considering for future developments

:36 - :45 Working groups updates (see https://workinggroups.bioconductor.org/).

- VOTE to deprecate Cloud methods & pangenomes working groups: unanimously approved.
- **Workflows**: (Michael Love)
    - Agreement with JOSE (and the backend updates required on their side) still in the setup phase, blog post in preparation.
- **Training**: (CAB)
    - Discussing how to move forward and solicit more contributions to BiocHowTos.
    - Several Bioconductor Carpentry Workshops in the past few months, first 'co-branded' workshops related to the Physalia collaboration In Nov (https://training.bioconductor.org/workshops/index.html).
- **Multi-omic Integration of Histopathology Image Analysis**: (Sehyun Oh)
    - Working on data sharing through the imageTCGA package.
    - Working on image feature converter/importer.
    - Use case development around OV and CRC datasets.
    - Plan to resume a quarterly working group meeting in early Dec.

:44 - :50 Governance (charter revisions).

- Current charter drafted >10y ago, not an accurate reflection today
  > this presents risks, but also opportunities
- Relationships b/w boards not well/"officially" defined
- Vote on executive coordinating council proposal/
  revised governance document (> postponed)
- Election processes: consider https://opavote.com/pricing/,
  (e.g., membership elections)

:50 - :54 Core efforts for 3.23.

- NumFOCUS subcontract to Alex Mahmoud for
  production of containers and binaries for 3.22
- R-universe transition (target: max. 1y from now)
- "annotation pipeline" needs significant revision
- blend new developments with 'preprint' authoring
  (see Appendix for some ideas) > Bioc underpublished?
- Lori R50 Work: 'pkgmaintainer' app summary (yearly validation email to opt in on CoC, maintainer expectations, licensing, etc.; automate mechanisms against disappearance of maintainers going unnoticed, e.g., email validation)

:55 - :56: Spatial omics topics

- OSTA book processing seems satisfactory.
- Helmholtz grant awarded for Bioconductor/scverse interoperability
  (venue & attendees TBD, target date around spring 2026)
- Pure Bioconductor-centric hackathon 19-23 Apr 2026 in Venice, Italy,
  (incl. remote follow-up events) on spatial omics, imaging & serialization;
  planned with EMBL (Davide Risso, Helena Crowell, Wolfgang Huber)

:57 - :60: 'pkgmaintainer' app demo from Lori.

Appendix: Some topics for preprints.

Papers that are needed

- **Array processing methods for genomics in Bioconductor**
  - DelayedArray
  - SparseArray
  - NaArray
- **The future of functional object-oriented programming strategies in Bioconductor**
  - S4Vectors
  - GenomicRanges ecosystem
  - SummarizedExperiment ecosystem
  - Adoption of S7 for visualization and AI
- **Curating data and annotation for genomics with Bioconductor**
  - Data scope: Human Cell Atlas, GEO, EBI ArrayExpress, ...
  - Annotation scope: NCBI, UCSC, Ensembl, Biomart, ENCODE, ..., GO, KEGG, reactome, ...
  - Hub concepts
  - pipeline and forge
- **Developer experience enhancements for creating genomic data analysis methods in Bioconductor**
  - unified and universal version control: git.bioconductor.org
  - contributions guidelines: interoperable methods linked by common data structures
  - static analysis with BiocCheck
  - BBS: build reports
  - platform family, including GPU
  - package dashboard
- **Bioconductor's involvement with AI for genomic data science**
  - deep learning methods support
  - LLM utilization for package metadata
  - BiocHF: harmonized interfaces for models and weights