

2025



Modern benchmarking

**Advancing computational methods
in molecular biology**

Monte Verità, Ascona, Switzerland

March 23 – 27

Organised by:

Mark D Robinson (U of Zurich)

Jean Yang (U of Sydney)

Serghei Mangul (USC)

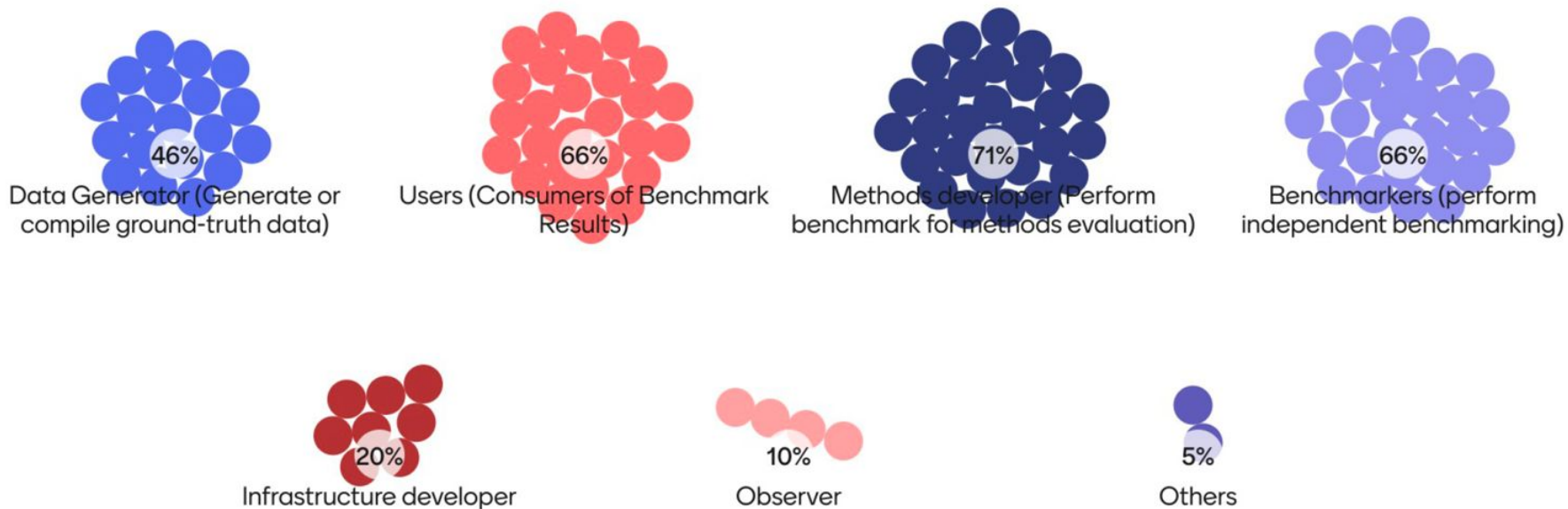


<https://sites.google.com/view/ascona2025>

The background is a solid teal color. In the top-left and top-right corners, there are decorative borders made of blueberries. The blueberries are arranged in a slightly curved line, with some showing their stems and leaves. The lighting on the blueberries is soft, giving them a realistic appearance.

Summary of panel discussion

What role do you see yourself as part of the benchmarking community ?



Which professional society do you participate in ?

| | | |
|------------------|----------------------|--|
| ISMB/ECCB x 8 | ISCB x 2 | SIB x 6 |
| ECR/RECOMB x 3 | SIB x2 | ABACBS / Oz Single Cell x3 |
| Bioconductor x 4 | Human Cell Atlas x 2 | BSI |
| DZL x 2 | Genome Informatics | ASA / IMS (Statistics conference) x 2 |
| ECCB x 3 | EUPA | Computer Science / Machine Learning / Single Cell Conference x 2 / Epigenomics / CZI / nf-core / NeurIPS |

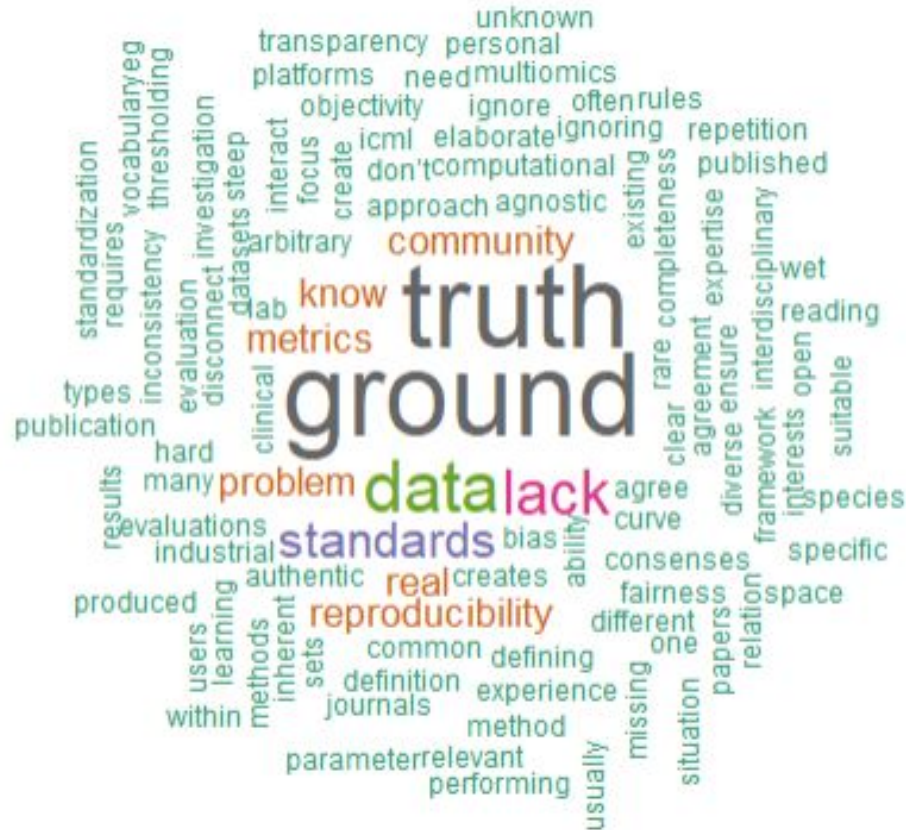
The background is a solid teal color. In the top-left and top-right corners, there are decorative elements consisting of a series of blue, textured spheres arranged in a diagonal line, resembling a molecular chain or a string of beads.

PART A – Current challenges

What do you think is the biggest challenge in benchmarking?

| | | | |
|--|--|---|--|
| suitable data sets | ICML | Disconnect between computational and wet lab | agreement within the community performing the benchmark |
| Clinically-relevant benchmarking datasets | Many platforms with steep learning curve | The unknown ground truth | Vocabulary; e.g. defining multi-omics |
| Lack of ground truth data | Ground truth data | Evaluation metrics inconsistency | Repetition |
| Missing species-agnostic evaluations | How to ensure that the ground truth is authentic? | Having the ability to interact with the methods and results of an existing/published benchmark | thresholding |
| Lack of ground truth and clear problem definition | Reproducibility | Publication focus and journals create an arbitrary situation where we as users usually ignore completeness or objectivity when reading benchmarking papers. | Ground truth data are often produced by one specific method or approach, this creates inherent bias. |
| There is no 'ground truth'. | agree on standards | What is real? | Personal interests |
| Benchmarking requires a rare interdisciplinary expertise and industrial experience. It's hard to know what you don't know. | Transparency | Need more open standards | ground truth |
| reproducibility | Ignoring parameter space | Fairness | Lack of standardization |
| Ground Truth Data | Elaborate rules for a common framework for different types of benchmarking | Community | |
| Standards | Lack of investigation of diverse metrics | | |
| benchmark relation to real problem | Consensus | | |
| | | | |

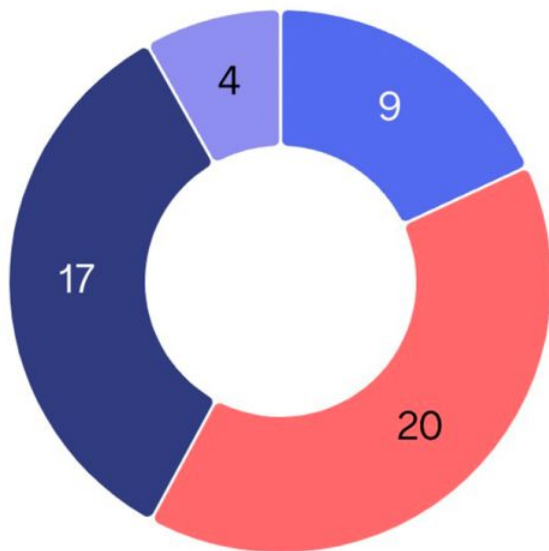
What do you think is the biggest challenge in benchmarking?



The background is a solid teal color. In the top-left and top-right corners, there are decorative elements consisting of a series of blue, textured spheres arranged in a diagonal line, resembling a molecular chain or a string of beads.

PART B – Benchmark system

Which one would you prefer to engage with ?

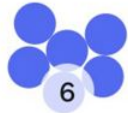


- 9 Centralized (one unified platform for data, tools, and results)
- 20 Semi-centralized
- 17 Decentralized (independent groups managing their methods but using shared benchmarked datasets) benchmarks)
- 4 No strong preference
- 0 Others

What would encourage you to share your data with a database of BMdata with ground truth ?



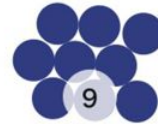
What would encourage you to contribute with an existing living benchmark platform?



Access to high-quality datasets



Improved visibility of your work



Proper credit (e.g. publication)



Easy submission process



Opportunities to contribute to
community standards

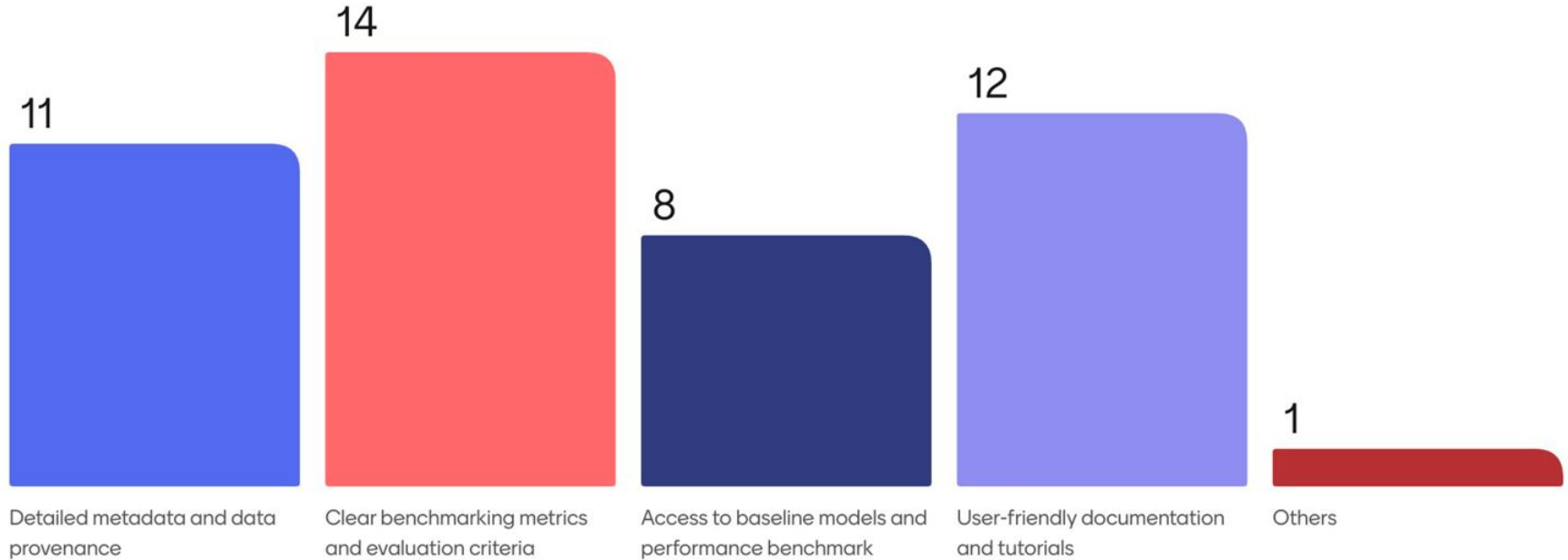
0

Not my role



Others

If you were to use a centralized benchmarking system, what information would you find most valuable?



Any other comments

There was not much emphasis on closed benchmarks (e.g. CASP competitions) where method developers do not have access to data. What do you think about it?

How far are we from having living benchmarks for tasks, and journals requiring it first step to submit to benchmark before publishing. What would be cons/pros?

We need to create a committee for open standards for computational benchmarking, including defining standards for metadata, quality controls, data formats, study design.

Benchmarks exist in their own bubble and not being propagated to a broader community - mostly executed by purely computational scientists. There is a big disconnect in benchmarking/application.

Dataset Bias

When performing collaborative benchmarks, it's important decide whether to benchmark the code, the models, the predictions

What is semi-centralized? The goal for a centralized system can also be to give method developers easy access to evaluation for their models

There should be more emphasis on problem definition for different benchmarks

We need a committee for defining open standards for computational Omics benchmarking -- robrecht

It would be cool to have a type of publication where you define the living benchmark, metrics and methods and encourage people to contribute. Then, once a year publish results with new methods

what is the goal of benchmarking? seems that there are different goals around. this needs to be clearly defined and communicated.

The background is a solid teal color. In the top-left and top-right corners, there are decorative elements consisting of a series of blue, textured spheres arranged in a diagonal line, resembling a string of beads or a molecular chain. The spheres have a slightly irregular, organic texture.

PART C – Going forward

What would you like to see in future efforts related to benchmarking?

Open standards for computational Omics benchmarking -- robrecht

Use our clout to pressure editors

Definitely open a specific benchmarking journal

Reduce work burden for benchmarking, i.e. easy and accessible

Datasets generated specifically for benchmarks

A benchmarking journal

Better collaborations on ground truths

MNIST/ImageNet for biology

Replace benchmarking papers with community driven living benchmarks.

Funded consortia

Centralized efforts

Hackathons every 6 months -- robrecht

More engagement with users/biologists/clinicians

Community efforts to test with tools-methods developers, and create standards for input output

Sharing of same metrics evaluation, eg by opening a benchmarking journal

What would you like to see in future efforts related to benchmarking?

Perhaps a short term goal could be a guideline similar to Greg's template but more accessible/easier

Consolidate infrastructure efforts so that everyone works together to solve problems and not reinvent their own wheel. Something like sklearn, but for benchmarking – one package/platform to go to

Community guidelines for benchmarker

Thinking about community governance structures

Design biology experiment to collect specific mechanisms

Dedicated journals and splitting responsibilities in engineering (packaging), tools/methodology development and benchmarking.

Regular conference. Maybe a ISCB Community of Special Interest?

Benchmark journal.

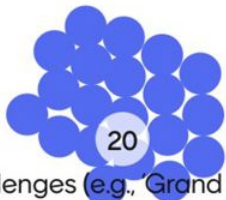
Transparency in how benchmarks are done.

Standards for generating new benchmarks (metric evaluation, dataset metadata) - malte

I expect data to be shared much systematically and openly than it currently is. I wish that the publication acceptance is not bound to data access.

Get a way to recognize contributions of benchmark data contributors to promote their career development

If you were to write an opinion piece based on this meeting, what focus would best represent our discussions?



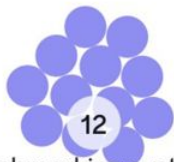
Current challenges (e.g., 'Grand Challenges' of benchmarking)



Recommended best practices



Current landscape and future directions of benchmarking



Review of benchmarking systems/platforms/suites (technical details)



Others

Other suggestions

Common pitfalls section in the paper

If we were to start a consortium, what do you think should be the top priority?

have fun

Creating standard guidelines and evaluation criterias.

Leverage enforcement if best practices

Ambitious long term goals, reaching which would revolutionize the field

Leverage enforcement if best practices

Interdisciplinarity

Gold standard datasets and meaningful metrics

Setting consensus standards and incentivize their use

Accessibility

Cross-discipline communication, evaluation metrics and dedicated journal

Work on community driven living benchmarks

journals to also become aware of the existing benchmarking ecosystems

establish one consortium per topic (e.g., transcriptomics; microbiome)

Open standards on computational omics benchmarks, and how do they contribute to Sustainable Development goals

Community building and engagement

If we were to start a consortium, what do you think should be the top priority?

Work on community driven living benchmarks

journals to also become aware of the existing benchmarking ecosystems

Community engagement through incentives

I like the guidelines on good statistics for evaluation and their limitations.

establish one consortium per topic (e.g. transcriptomics; microbiome)

Open standards on computational omics benchmarks, and how do they contribute to Sustainable Development goals

Get key journals to adopt consensus benchmarking best practice for bioinformatics software.

I'm skeptical. Several such papers exist. What would make this one different?

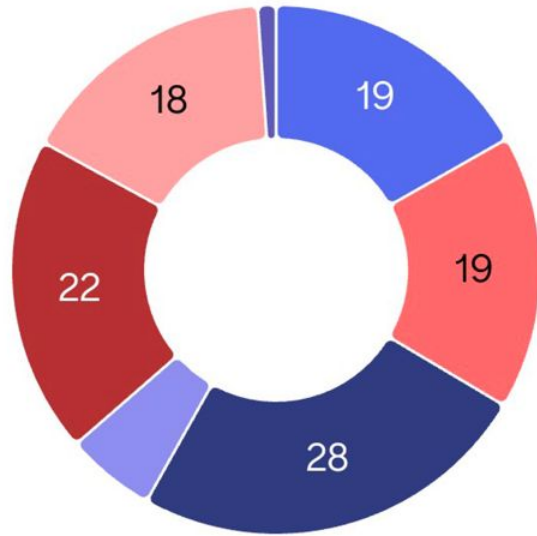
Make benchmarking a well recognized need (by journals for example)

Provide a solid foundation for advancing science and methods

Common pitfalls.

A journals and evaluation metrics

What is next (after this week)?



- 19 Organizing a conference
- 19 Forming a dedicated working group
- 28 Publishing a consensus statement or white paper
- 6 Developing new benchmarking tools/resources
- 22 Start a new journal
- 18 Conduct a benchmarking survey to comprehensively engage with the global community
- 1 Others

What is next (others) ?

establish funded consortia

make benchmarking more visible, communicate its importance. Get funding for performing benchmarks.

EU COST action funding?

Prize for PhD thesis that focus on benchmarks

Funding this initiative is critical

write a joint ontology

communicate with existing funding committees about the importance of benchmarking funds

Share this menti results

Create a compilation of posters to show how benchmark is integrated in interdisciplinary fields, also makes all the work more visible

mailing list & wiki

SHare the power point presentations

Forming a community across different domains

Hackathons every 3-6 months