

iCOBRA: open, reproducible, standardized and live method benchmarking

To the Editor: Modern life science research tasks often involve ranking or classifying items. For example, in studies of differential expression, genes can be ranked by the estimated P value or, using a cutoff on the P value, classified as either ‘significantly different’ or ‘not significantly different’ between conditions of interest. A wide range of computational methods dedicated to these tasks exist^{1–3}, many of which rely on accurate quantification of underlying entities such as abundance levels. As methods are developed and refined, static benchmarking studies quickly become outdated. Moreover, a standard way to present results from method comparisons is lacking, and raw reference data are not always made available. This often makes it difficult for method researchers to reproduce published evaluations or explore them from different angles. Here we present iCOBRA (for “interactive comparative evaluation of binary classification and ranking methods”), a benchmarking platform for both users and developers of methods that promotes open, standardized and reproducible evaluations. iCOBRA consists of an R package and a flexible, interactive web application that can rapidly evaluate methods for binary classification, ranking and continuous target estimation against a ground truth. In addition, we have collected a set of benchmarking data sets in standard formats (a link is provided at <https://github.com/markrobinsonuzh/iCOBRA>) to lower barriers for new method developers as well as to facilitate standardized method evaluations in the future. We envision that this resource

will be extended over time, and we encourage the community to contribute data (for example, simulations) and method assessments. In **Supplementary Note 1**, we show how iCOBRA can be used to exactly reproduce and visualize results from recent benchmarking studies.

iCOBRA’s web application (**Fig. 1**) is based on the Shiny framework and can be run via our public server (accessible from <https://github.com/markrobinsonuzh/iCOBRA>), which makes it platform agnostic and eliminates the need for knowledge about installing or running R. Extensive documentation is included in the app (**Supplementary Note 2**). Underlying the application is an R package (available via Bioconductor) that can be used both to run the interactive application locally and to generate result visualizations directly from the R console, facilitating both interactive exploration and integration in programming pipelines. In contrast to R packages dedicated to evaluating classifiers (for example, ROCR⁴), which generate static performance plots, the Shiny framework is interactive and lets the user include or exclude methods from a comparison, change the appearance of the plots or stratify the results by a provided annotation with minimal effort. The input format is simple and generic (tab-delimited text files), leading to increased ease and range of use compared to other performance evaluators (for example, compcodeR⁵) for which the data representation format and/or choice of evaluation metrics are specifically tailored to certain types of data. The application accepts several input types (nominal P values, adjusted P values and a general ‘score’), allowing for greater flexibility than afforded by existing applications such as BDTCOMPARATOR⁶, which compares two categorizations and is thus strictly limited to classification evaluation.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.3805).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Charlotte Soneson^{1,2} & Mark D Robinson^{1,2}

¹Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. ²Swiss Institute of Bioinformatics (SIB), Zurich, Switzerland. e-mail: mark.robinson@imls.uzh.ch

1. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. *Bioinformatics* **26**, 139–140 (2010).
2. Paulson, J.N., Stine, O.C., Bravo, H.C. & Pop, M. *Nat. Methods* **10**, 1200–1202 (2013).
3. Anders, S., Reyes, A. & Huber, W. *Genome Res.* **22**, 2008–2017 (2012).
4. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. *Bioinformatics* **21**, 3940–3941 (2005).
5. Soneson, C. *Bioinformatics* **30**, 2517–2518 (2014).
6. Fijorek, K., Fijorek, D., Wisniewska, B. & Polak, S. *Bioinformatics* **27**, 3439–3440 (2011).



Figure 1 | Screenshot of the iCOBRA interactive application interface. The sidebar along the left-hand side contains the input controls, and the tabs in the main panel display different aspects of the performance evaluation.