

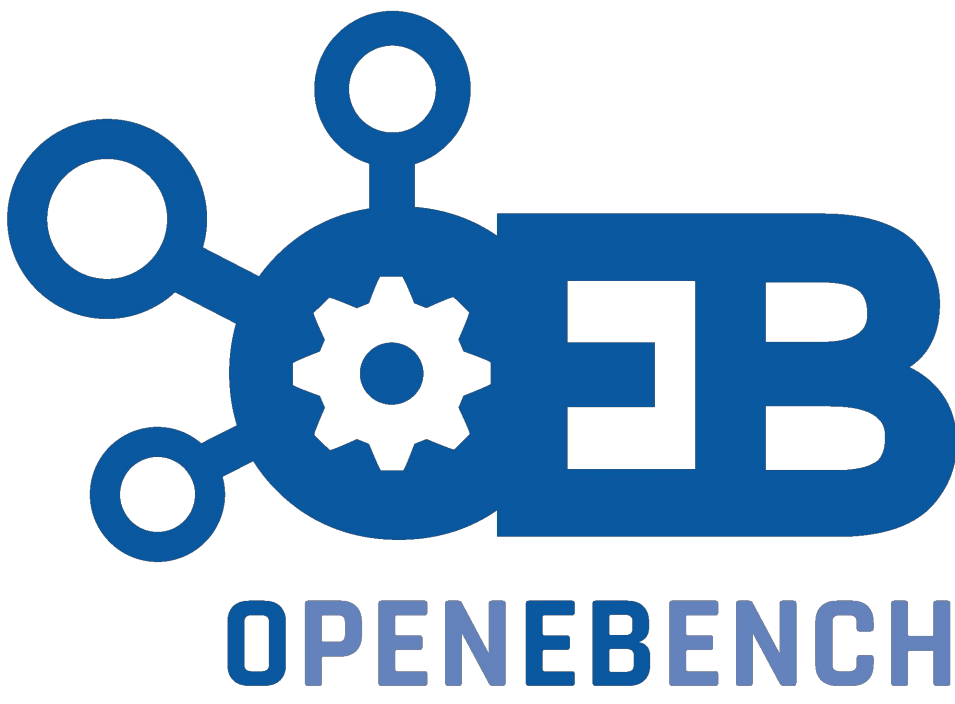
Scientific Communities in OpenEBench

Vicky Sundesha ^{1,2}, Javier Garrayo ^{1,2}, José M^a Fernández ^{1,2}, Laia Codó ^{1,2}, Dmitry Repchevsky ^{1,2}, Juergen Haas ⁵, Adrian Altenhoff ⁵, Christophe Dessimoz ⁶, Isabel Cuesta ⁷, Sara Monzon ⁷, Daniel López ⁸, Javier Perez Florido ⁸, Joaquín Dopazo ⁸, Alfonso Valencia ^{1,2,4}, Josep Ll Gelpi ^{1,2,3} and Salvador Capella-Gutiérrez ^{1,2}

¹ Barcelona Supercomputing Center (BSC), Barcelona Spain.
² Spanish National Bioinformatics Institute (INB), Barcelona Spain.
³ Universidad de Barcelona (UB), Barcelona Spain.
⁴ Institució Catalana de Reserca i Estudis Avançats (ICREA), Barcelona Spain.
⁵ Swiss Institute of Bioinformatics (SIB), Lausanne/Basel Switzerland.
⁶ University of Lausanne (UNIL), Lausanne Switzerland.
⁷ Instituto de Salud Carlos III (ISCIII), Madrid Spain.
⁸ Fundacion Progreso y Salud, Sevilla, Spain



OpenEBench helps determining the precision, recall and other metrics of bioinformatics resources in unbiased scenarios, which have been set up through reference databases, ad-hoc input and test data sets reflecting specifying scientific challenges. Chosen metrics allow to objectively evaluate the relative scientific performance of the different participating resources. It is even possible to understand what are the software potential biases, strengths and weaknesses and/or under which conditions do they perform better or worse.



Unbiased and objective evaluations of bioinformatics resources are challenging to set-up and can only be effective when built and implemented around community driven efforts. Several communities from different scientific domains collaborate with OpenEBench in order to set-up, host and further develop their scientific efforts. Communities can focus on specific problems, e.g. Quest for Orthologs (**QFO**) and The Cancer Genome Atlas (**TCGA**); or having a broader spectrum e.g. Spanish Network of Biomedical Research Centers on Rare Diseases (**CIBERER**); or covering different challenges on each of their editions, e.g. **DREAM** Challenges. Benchmarking efforts led by scientific communities might have a national scope e.g. CIBERER; or a global one e.g., Global Microbial Identifier Initiative (**GMI**) (**coming soon**).

Benchmarking events from scientific communities currently on OpenEBench

Centro de investigación biomédica de enfermedades raras (**CIBERER**)

- ☐ CIBERER -2019 NGS Benchmarking *



Dialogue on Reverse Engineering Assessment and Methods (**DREAM**)

- ☐ DREAM Challenges 7th Edition - Year 2012



Quest for Orthologs (**QFO**) [1]

- ☐ Quest for Orthologs - 2011 Reference Proteomes
- ☐ Quest for Orthologs - 2018 Reference Proteomes *



The Cancer Genome Atlas (**TCGA**)

- ☐ 2018 PanCancer data - Driver Genes prediction benchmark



QFO 2018 - Reference proteomes *

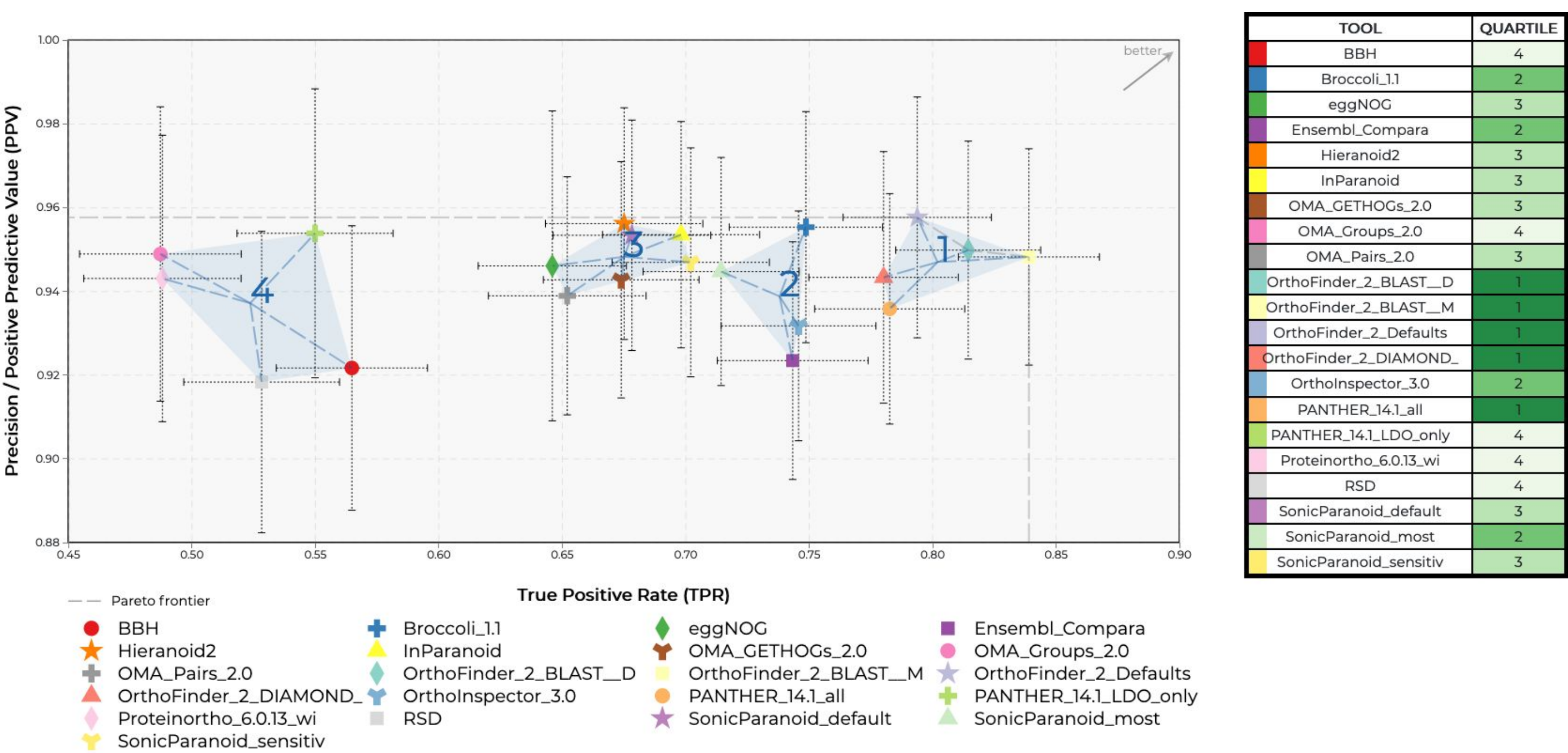


Figure 1. Results for the Agreement with Reference Gene Phylogenies: SwissTree benchmark challenge from the QFO 2018 - Reference proteomes benchmarking event using "K-MEANS Clustering" and "optimal view" displayed on OpenEbench.

CIBERER 2019 - NGS Bechmarking *

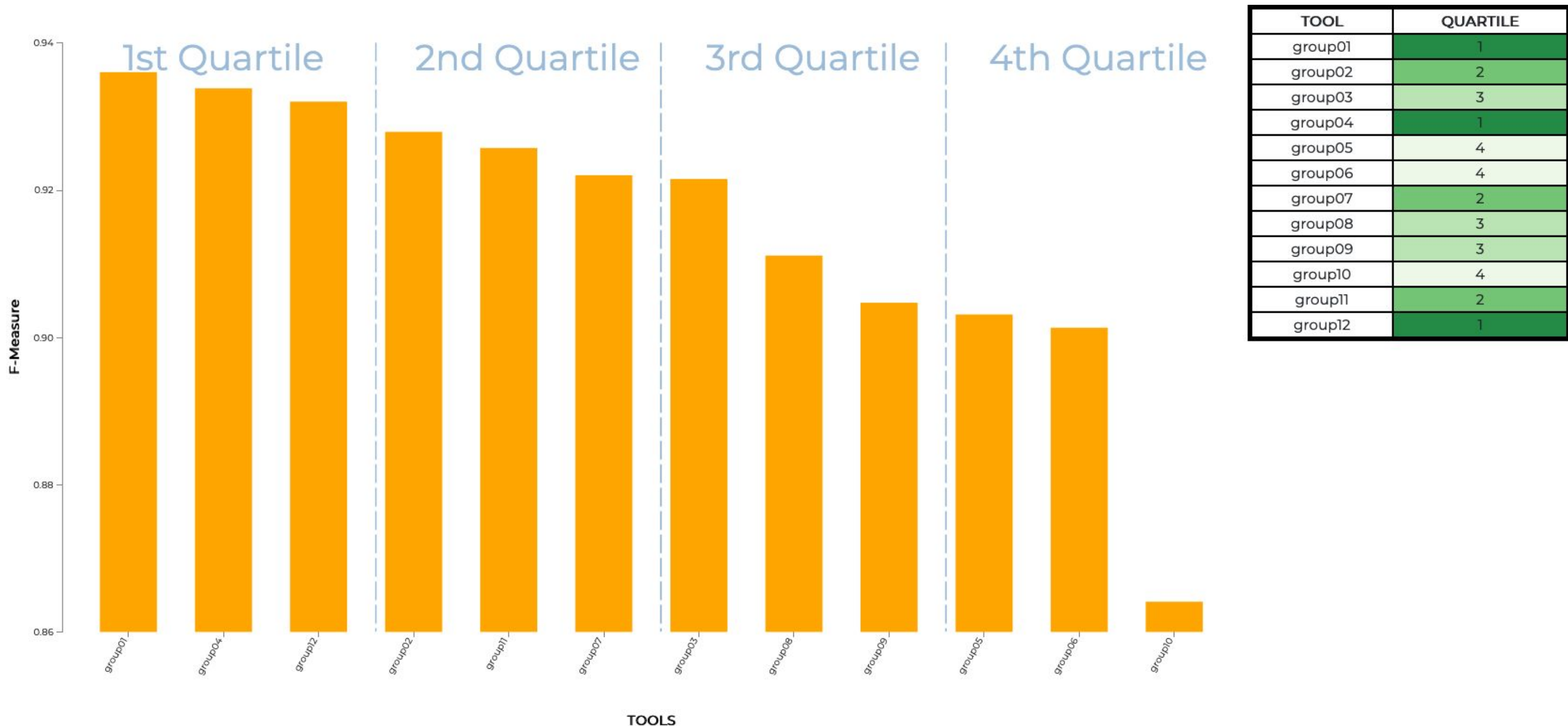


Figure 2. Results for the Germinal SNPs and INDELs Prediction Benchmark in GIAB high confidence dataset challenge from CIBERER 2019 - NGS benchmarking event classified by quartiles displayed on OpenEbench.

[1] Altenhoff, A.M., Garrayo-Ventas J., Cosentino S. *et al.* The Quest for Orthologs benchmark service and consensus calls in 2020. Nucleic Acids Research, gkaa308, <https://doi.org/10.1093/nar/gkaa308>

Contact:

Vicky Sundesha
vicky.sundesha@bsc.es

INB/ELIXIR-ES Coordination Hub
Barcelona Supercomputing Center (BSC).

ELIXIR is partly funded by the European Commission within the Research Infrastructures programme of Horizon 2020.

