October 16, 2019
Updated: February 6, 2020

CAFA4 Rules

Preamble: Introducing the Critical Assessment of Functional Annotations (CAFA)
The Critical Assessment of protein Function Annotation algorithms (CAFA) is a challenge designed to provide a large-scale assessment of computational methods dedicated to predicting protein function, using a time challenge mechanism. The CAFA organizers provide a large number of protein sequences. The predictors then predict the function of these proteins by associating them with Gene Ontology (GO) terms, Human Phenotype Ontology (HPO) terms, and/or (new this year) Disorder Ontology (DO) terms. Following the prediction deadline, we wait for several months. During that time, some proteins whose functions were unknown experimentally have received experimental verification. Those proteins constitute the benchmark, against which the methods are tested.

CAFA is a community-wide effort whose goal is to help understand the state of affairs in computational protein function prediction and drive the field forward. The challenge started in 2010 and is held every three years. See more at our web site:

https://www.biofunctionprediction.org/cafa/

Introductions to CAFA are available in:

1. Friedberg I, Radivojac P. Community-wide evaluation of computational function prediction. *Methods in Molecular Biology* (2017) 1446: 133-146.
2. A (not so) Quick Introduction to Protein Function Prediction (intended primarily for computer scientists with little background in biology):
   https://www.ccs.neu.edu/home/radivojac/papers/radivojac_cafa_2013.pdf

More thorough reading:

1. Radivojac *et al.* A large-scale evaluation of computational protein function prediction. *Nat. Methods* (2013) 10(3): 221-227.
2. Jiang *et al.* An expanded evaluation of protein function prediction methods shows an improvement in accuracy. *Genome Biol.* (2016) 17(1): 184.
3. Zhou N, *et al.* The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome Biol.* (2019), accepted and *bioRxiv:* 653105.

Timeline for the CAFA4 challenge

Target release: mid-October, 2019

Submission Deadline: February 12, 2020 at 11:59pm, anywhere on Earth

Initial assessment: will be provided ad ISMB 2020, mid-July 2020.

Final assessment: on or about October 2020

Important: all CAFA4 participants must register to the email list: afpcafa@iastate.edu

This is the sole venue of general communication between the organizers and the predictors. This email list will serve to communicate any announcements and changes regarding the CAFA4 challenge. You can register at

https://mailman.iastate.edu/mailman/listinfo/afpcafa

Challenges:

1. Prediction of Gene Ontology terms (GO release 2019-10-07).
2. Prediction of Human Phenotype Ontology terms (HPO release 2019-09-06).
3. Prediction of Disorder Ontology terms (DO release 0.1.0)

The ontologies and training sets are publicly available at the following web sites and also on the CAFA4 web site. The new and old versions of these ontologies are available from their web sites. For example,

- Gene Ontology: http://current.geneontology.org/ontology/go.obo
- Human Phenotype Ontology: https://raw.githubusercontent.com/obophenotype/human-phenotype-ontology/master/hp.obo
- Disorder Ontology: https://www.disprot.org

Rules

1. The evaluation will be performed for the target sequences that accumulate experimental annotations between the submission deadline (February 10, 2020) for predictions and the time of evaluation. The initial evaluation will take place in July 2020, however, CAFA4 assessors will carry out evaluations following this date, as more experimentally annotated results accumulate. We expect the final evaluation to be done by the end of 2020. The evaluation will be carried out following the previous publications from CAFA1, CAFA2, and CAFA3.
2. One team may test up to 3 different prediction models (named MODEL 1, MODEL 2, and MODEL 3) during submission. MODEL 1 will be officially evaluated by the organizers, but other models will also be considered. A group should use its best algorithm as MODEL 1.
3. A team may choose to predict using any of the following ontologies: Molecular Function Ontology (MFO), Biological Process Ontology (BPO), Cellular Component Ontology (CCO) from GO, Disorder Ontology for the function of intrinsically disordered proteins and Human Phenotype Ontology (HPO) for human proteins only. The evaluation will be performed separately for each ontology. A team may choose to predict function using one or more of the above ontologies, and does not have to predict using all of them.

File Format for Submissions

The prediction output file format is shown in Figure 1. The file a group submits should be in text format (.txt) *or compressed* (.zip) text file. Predictions can be uploaded any number of times. The ones with the most recent timestamp in the system at the submission deadline will be used for evaluation.

- The AUTHOR line lists the team name that the team leader used during registration.
- The MODEL line contains numbers 1, 2, or 3 and corresponds to the prediction model used as described in bullet (2)
- The KEYWORDS line contains a list of keywords that describe the methodology used. Keywords line uses a comma-separated list, ending with a full stop, of one or more of the following pre-specified keywords: sequence alignment, sequence-profile alignment, profile-profile alignment, phylogeny, sequence properties, physicochemical properties, predicted properties, protein interactions, gene expression, mass spectrometry, genetic interactions, protein structure, literature, genomic context, synteny, structure alignment, comparative model, predicted protein structure, de

novo prediction, machine learning, genome environment, operon, ortholog, paralog, homolog, hidden Markov model, clinical data, genetic data, natural language processing, other functional information.

- The ACCURACY lines are optional. If present, they must contain the group's estimate of the accuracy of their method for each of the three modes of evaluation (see (1) above). Each line contains estimated precision (PR) and recall (RC) for the F_{max} point exactly as evaluated in the CAFA 2010-2011 manuscript [1] and after. The number indicates one of the three evaluation modes as described under point (1) above. Both numbers must be in the interval [0.00, 1.00]. Two significant figures are required (e.g., 0.70 is valid but 0.7 or .70 are not). The ACCURACY lines may be different in each submitted file (all targets are broken up based on species). If so, a weighted average will be used to estimate the final accuracy of the model. Weights will be determined by the number of proteins from each target file that accumulate experimental terms between the submission deadline and the time of evaluation.
- The list of predictions contains a list of pairs between protein targets and GO terms, followed by the probabilistic estimate of the relationship (one association per line). The target name must correspond to the target ID listed in the target files (in the FASTA header for each sequence). The Gene Ontology ID must correspond to valid terms in GO's version listed above. MFO, BPO, and CCO are to be combined in the prediction files, but they will be evaluated independently by the assessment team. The score must be in the interval (0.00, 1.00] and contain two significant figures. A score of 0.00 is not allowed; that is, the team should simply not list such pairs. In case the predictions are not propagated to the root of ontology, the assessors will recursively propagate them by assigning each parent term a score that is the maximum score among its children's scores. Finally, to limit prediction file sizes, one target cannot be associated with more than 1500 terms for MFO, BPO, and CCO combined. The assessors will provide software so the groups will be able to check the format of their prediction files. Please submit only files that are verified for correctness. The assessors will not analyze submissions that are in incorrect format. If your method does not output a score associated with predicted terms, but rather just a set of terms, the team can set scores for all such predictions to the same value (e.g,. 1.00). Such methods will be characterized by a single precision/recall point, instead of a precision/recall curve.
- The Human Phenotype Ontology annotations for human targets shall be submitted separately using the HPO annotation (HPO version stated above). Figure 2 shows the format for HPO submissions. Only two accuracy estimations are necessary: one for the proteins/genes that have not been associated with any HPO terms before the submission deadline; and the other for the remaining targets.
- Disorder Ontology (DO). These assessments will be carried out in the same way as for HPO.
- The prediction file must end with the keyword END in a line of its own.
- Allowed delimiters are tab and whitespace only.

Prediction File Name Format

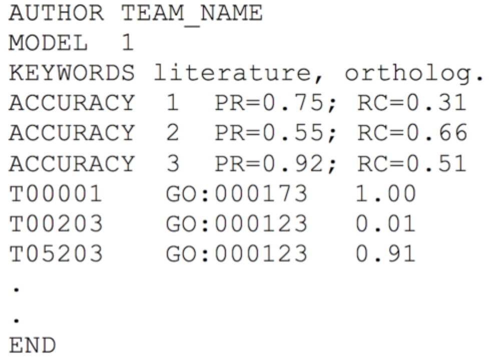
Format for GO predictions

Use team ID, model number, and taxon IDs as follows: teamID_modelNo_taxonID_go.txt

Example for team JohnDoe group, model 1, human sequences GO prediction file name format: (johndoegroup_1_9606_go.txt)

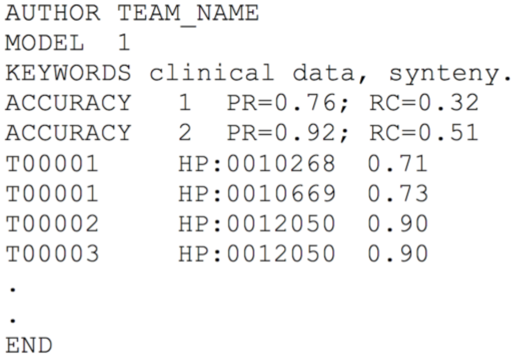
Example for team JohnDoe group, model 3, mouse sequences GO prediction file name format: (johndoegroup_3_10090_go.txt)

All files for a single team should be uploaded as a zipped directory.



```
AUTHOR TEAM_NAME
MODEL  1
KEYWORDS literature, ortholog.
ACCURACY  1  PR=0.75; RC=0.31
ACCURACY  2  PR=0.55; RC=0.66
ACCURACY  3  PR=0.92; RC=0.51
T00001    GO:000173   1.00
T00203    GO:000123   0.01
T05203    GO:000123   0.91
.
.
END
```

Figure 1. File format for GO predictions.



```
AUTHOR TEAM_NAME
MODEL  1
KEYWORDS clinical data, synteny.
ACCURACY   1  PR=0.76; RC=0.32
ACCURACY   2  PR=0.92; RC=0.51
T00001     HP:0010268  0.71
T00001     HP:0010669  0.73
T00002     HP:0012050  0.90
T00003     HP:0012050  0.90
.
.
END
```

Figure 2. File format for HPO predictions.

Format for HPO predictions
Use team ID, model number, and acronym HPO as follows: teamID_modelnum_hpo.txt
Example for team JohnDoe , model 2, HPO prediction: (johndoegroup_2_hpo.txt)

Format for DO predictions
Same as for GO, except use “do” instead of “go” in the file name. One file per species may be submitted.

Submission folder format (NEW to CAFA4)
Submit your predictions in one folder, zipped. No subfolders, no double zipping (zo zipped files in a zipped folder). No tarballs, zip only. **Your zipped archive must have your team name in the filename**, and that team name must be the same as the team name in the individual filenames within the zipped folder.

Example:
johndoegroup.zip
| *(zip archive contains)*
|-- johndoegroup_1_9606_go.txt
|-- johndoegroup_2_9606_go.txt
|-- johndoegroup_1_hpo.txt

Prediction and Evaluation Types
Protein-centric predictions. A protein-centric prediction addresses the following question: “given a protein, what are all the ontology terms associated with it” (Radivojac et al, 2013; Jiang et al, 2016). This is the main mode of evaluation in CAFA. **Term-centric predictions.** A term-centric prediction means: “given a specific ontological term, which genes in an organism fit that term?” All protein centric predictions will also be evaluated in term-centric mode. However, predictors can submit files with term-centric predictions and ask not to be evaluated in the protein-centric mode. ***The filenames of predictions designated for term-centric only evaluations *must* be prefixed with “TC_”, or they will be evaluated both term-centric and protein-centric.*** (the file name would be TC_johndoegroup_3_10090_go.txt) For example, if a team has an algorithm that predicts beta-amylases, they can submit a file that *only* predicts what genes in a given organism are beta amylases. For more about term-centric evaluations, see the CAFA1, CAFA2, and CAFA3 papers.

Submission of predictions:

All submissions must be made by February 10, 2020, 11:59pm in the time zone of your choice. The submissions will be made via the biofunctionprediction.org website. The submission mechanism will be announced by January 2020, it is important therefore to register to the mailing list to receive timely updates.

Important Policies

For the policies on data sharing, anonymity, de-identification, or withdrawal from the entire experiment (after the submission deadline) see the Data Sharing, Anonymity, and Withdrawal Policy supplement below.

A group can appeal their prediction evaluation to the CAFA assessment team, and to the CAFA organizers. Appeals will be discussed and ruled upon by the assessment team and the organizers. All such rulings are final.

Data Sharing, Anonymity, and Withdrawal Policy

All prediction data on the set of benchmark proteins, those that accumulate experimentally validated annotation, will be made public. However, groups and methods will only be identified by their numbers to preserve anonymity. Note that the following exceptions apply:

A group may withdraw one or all of its methods from experiment at any point before or after the submission deadline. Should a group withdraw prior to the 1st of July 2020, its predictions will not be made public, not even as anonymized predictions.

2) Groups that perform in the top 10 according to the major metrics will not retain anonymity. The rationale is that well-performing methods should be made public.

3) If a method is published independently by a group, using CAFA data or citing CAFA participation, its CAFA results will be made public and identifiable by the CAFA organizers.

4) A group may choose at any time to de-anonymize any of its methods by written request to the CAFA organizers.

5) Teams can allow CAFA organizers to release and publish the team's prediction data while preserving anonymity. In previous years, only the methods were released anonymously. To choose this option, share the project that stores your CAFA data with the team CAFA Admin by using the share button in the top right of your project dashboard. Search for "CAFA Admin" and give them "can view" permission. The CAFA Admin team will not attempt to open or view any files within the project. By sharing with CAFA Admin, you are giving the CAFA organizers permission to anonymously publish your predictions data along with your anonymous methods and results. The only data the CAFA organizers will publish will be the CAFA submission data that your team submits during the open submission period.

6) If the option in Step 5 is not chosen, the organizers will direct all inquiries for complete prediction data to the Principal Investigator of each team.

The CAFA organizers will make best efforts to maintain anonymity when requested subject to the limitations above, but cannot guarantee it in any instance due to resource limitations.