CAFA π Submission Rules

1. Team Name

Each team should have a unique identifying team name. This team name will be used to identify the team in the submission file name and inside the submission file as a header. This name should **not** contain white spaces (" "), tabs (" ") or underscores ("_"), but it can be different from the Synapse registered name.

2. Models

Each team can have up to three different submissions, namely three models. All models submitted will be evaluated.

3. Keywords

Each model should identify the methods it uses by providing a selection of keywords from the attached list (keywords.txt).

4. Team Designation

A person can only belong to one team, with the exception of a principal investigator. A principal investigator may lead up to three teams.

5. Submission Deadline

Deadline for submission: **Friday, April 20, 2018 11:59pm Baker Island Time Zone (UTC-12)**. All submissions should be made before the deadline via Synapse. No late submission accepted. Submission via other means (such as email) is not accepted.

6. Contacts

Contact for questions regarding submission rules, please email Naihui Zhou or Iddo Friedberg ({nzhou,idoerg}@iastate.edu)

Submission Format

1. Filename Format

Use team name, model number, taxon IDs and GO IDs as follows:
teamName_modelNo_taxonID_geneOntologyID.{txt/zip}
Three model numbers allowed.
Two *taxonID*s allowed (208963 and 237561)
Two *geneOntologyID*s allowed (0042710 and 0001539)
Please do not include the string "GO:" in *geneOntologyID*
Please do not include the string "model" in *modelNo*
Example: DoeGroup_1_208963_0042710.txt

2. File Header

Each submission file should contain the following header.
AUTHOR teamName
MODEL modelNo
KEYWORDS someKeyword, anotherKeyword
Keywords should be comma-separated and in one single line.
Please double check before submission that information in the file header agrees with the file name.
Note that the Gene Ontology term being predicted for a submission file is only present in the file name, and not in the file content.

3. Predictions

Following the header, the submission file should contain a list of protein targets that the team think are associated with the function designated in the filename, followed by a probabilistic estimate of the association (score). Allowed delimiters are tab and whitespace only.
The target protein name must correspond to the CAFA ID listed in the target files (in the FASTA header for each sequence). 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 predictions. If your method does not output a score associated with predicted terms, but rather just a set of terms, the team should set scores for all such predictions to the same value (e.g. 1.00).

4. Ending

Please append the string
END
to the end of the file in a line of its own.

5. Format Checker

The CAFA organizers 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 CAFA organizers will not analyze submissions that are in incorrect format.

6.Example submission file

Example submission files were provided in *sampleSubmissions.zip*

```
AUTHOR  DoeGroup
MODEL   1
KEYWORDS        autoencoder, neural network
T2089630003203  0.66
T2089630000379  0.95
T2089630004237  0.57
T2089630003566  0.17
T2089630003116  0.89
T2089630000157  0.39
T2089630002951  0.64
T2089630002588  0.45
T2089630003498  0.77
END
```