FindSubtypes / PredictSubtypes

Jeffrey T. Chang, Joseph E. Lucas, Michael L. Gatza, and Joseph R. Nevins
28 December 2010

Summary: This pair of modules divides samples into subtypes (FindSubtypes) and, for a different set of samples, predicts their subtypes (PredictSubtypes). Typically, the samples are grouped into subtypes based on the predictions of pathway activation from the ScoreSignatures module. The groupings are computed using a Bayesian mixture model developed by Joe Lucas.

For more information on its use, see:
    Gatza *et al*. A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA*. 107(15):6994-9, 2010.

FindSubtypes

Parameter: dataset
This is the data set that contains the samples to be grouped into subtypes. This file typically contains the predictions of pathway activation from ScoreSignatures (*probabilities.pcl*). It can be in PCL, GCT, and some other data formats.

Parameter: penalty
This is a parameter that controls the number of subtypes. This score should be less than 0. Higher numbers will generate more subtypes, and lower numbers will generate fewer. You can provide a comma-separated list of parameters (e.g. "-5,-10,-20,-30") if you wish to try multiple values.

PredictSubtypes

Parameter: dataset
This module will predict the subtypes for the samples in this file. This file typically contains the predictions of pathway activation from ScoreSignatures (*probabilities.pcl*). It can be in PCL, GCT, and some other data formats.

Parameter: model_file
This is the *subtype_model.zip* file that was generated by FindSubtypes.

For help with these modules, please email: jeffrey.chang@duke.edu.