ESPPredictor Documentation

Description: Predict high responding peptides in electrospray mass spectrometry (ESI-MS)

Author: Vincent Fusaro (Broad Institute), gp-help@broad.mit.edu

Summary:
The Enhance Signature Peptide (ESP) predictor is a computational model to predict high responding peptides (i.e., peptides with a high intensity) from a given protein in ESI-MS. A feature set consisting of 550 physicochemical properties is calculated for each peptide. The feature set is then analyzed with a Random Forest (RF) model to calculate the probability of high response for each peptide. It is important to note that the probability of high response is on a per protein basis and is relative to other peptides within the same protein. The probability can be used to rank peptides in order of their response in order to select the highest responding peptides.

Usage:
The ESPPredictor module requires a list of peptide sequences. When starting with protein sequences they can be digested *in silico* using Peptide Selector (http://proteomics.broad.mit.edu/millhtml/mssluice.htm). We tested the ESP predictor using the following settings:

Digest: trypsin (*Note: not tested with any other enzyme*)
Maximum # basic residues: 4
Minimum peptide MH+: 600
Maximum peptide MH+: 2800
Clear all “Peptide exclusion criteria” checkboxes
Delete amino acids from “AA Composition Filtering”

You must save the output (copy & paste usually into Excel) and then save peptide sequences as a separate text file. This text file can be used as input into the ESPPredictor module.

Reference:
Vincent A. Fusaro, D.R. Mani, Jill P. Mesirov, Steven A. Carr. Computational Prediction of High Responding Peptides for Development of Targeted Protein Assays by Mass Spectrometry. *Nature Biotechnology* (2009).

Parameters:

Name	Description
input.file	A list of tryptic peptide sequences. One sequence per line. Exclude the following non-standard amino acids: J, U, Z, B, O, X.

Output File:

Name	Description
Predictions.txt	A list of peptide sequences with their associated predicted probability of high response.
PeptideFeatureSet.csv	A peptide feature file that contains 550 physicochemical properties for each peptide. The ESPPredictor module uses this file as input to the Random Forest model.

Note: Depending on the number of peptide sequences the module may execute in a few seconds (<20 peptides) or many hours (>1,000 peptides).

Example input:

AYLETEIK
ANFQGAITNR
LAFTGSTEVGK
TVGAALTNDPR
NAGQICSSGSR
LHFDTAEPVK

Example output (Prediction.txt):

Sequence	ESP_Prediction
AYLETEIK	0.44658
ANFQGAITNR	0.77478
LAFTGSTEVGK	0.79398
TVGAALTNDPR	0.9486
NAGQICSSGSR	0.31292
LHFDTAEPVK	0.63772

Platform dependencies:

Module type: Proteomics
CPU type: any
OS: any
Language: Matlab (bioinformatics toolbox), R (Random Forest Library)