

[INTRODUCTION]

This document describes the command line options for the ProbeSelection executable program. The program reads the background corrected probe intensities for multiple exon arrays from the ProbeEffects executable program, and calculates gene expression indexes using a subset of selected probes.

Details of the probe selection algorithm are described in Xing, Y., Kapur, K. and Wong, W.H. *Probe selection and expression index computation of Affymetrix Exon Arrays*. PLoS ONE, in press.

The program requires installations of R (tested on version 2.2.0), the affy package of Bioconductor, and Python (tested on version 2.4.3). Python Rpy module is optional.

[RUNNING THE PROGRAM]

Before running the program, the user needs to make an empty sub-directory named ‘out’, which will store some of the outputs. The program is executed using,

```
python ProbeSelect.py DATA_DIR -t [T/F] -r [T/F]
```

DATA_DIR: the sub-directory should have background corrected probe intensities from the ProbeEffects executable program, one file per each array.

-t: whether negative background-corrected probe intensities are trimmed to 0. Default option is TRUE [T].

-r: whether the program calls R through Rpy module to calculate expression indexes. Default option is TRUE [T], and the program will complete the computation of expression indexes. If Rpy is not installed, please set the option as FALSE [F], and run *source("ProbeSelect.R")* in R after ProbeSelect.py is completed.

[OUTPUT OF THE PROGRAM]

The program outputs several files.

summary.selected: expression indexes computed from selected core probes for multiple arrays in tab-delimited format.

summary.core: expression indexes computed from all core probes for multiple arrays in tab-delimited format.

Probes.Selected: core probes selected for expression indexes computation for each transcript cluster.

In addition, the sub-directory ‘out’ stores the background corrected probe intensities, one file per transcript cluster. Users can inspect probe-level data for individual genes of interest.

[Remarks]

Our program is in active ongoing development. For questions, comments and suggestions, please contact yi-xing@uiowa.edu or whwong@stanford.edu.