Samster
A reformatting tool for SAM to Cluster
Version 1.4
April 8, 2002

cckim@stanford.edu

Summary

You've done a SAM analysis and have your lists of significantly upregulated and downregulated genes. You now want a more visual representation, or you want to see if there is even more detailed substructure within these genes by using Cluster. Samster will take an Excel spreadsheet or SAM output text files and extract the raw data into a text output file in pcl format, which can be fed directly into Cluster for further analysis or opened directly in Treeview to preserve the order of the SAM output. Typically, most people achieve this by creating databases to link the SAM output to the raw data. This program circumvents this tedious procedure.

Availability, Copyright, and Citing

The program is available as a Windows executable at http://falkow.stanford.edu/whatwedo/software. The source is also available at the same website, and is written in Perl. The script requires Tk and Spreadsheet::ParseExcel::Simple.

If you would like to make a citation to Samster, please cite this document as a website reference. (e.g., http://falkow.stanford.edu/whatwedo/software/programs/samster.pdf, Samster documentation).

Known Limitations

1) If an Excel spreadsheet is analyzed (as opposed to text files), workbooks which contain multiple non-empty datasheets are not supported. In other words, you should begin with only one worksheet that has data; if any other worksheets exist, they should be empty. The additional worksheets generated by SAM are ignored by the program.

2) The ID's for certain datasets are a problem. For certain mouse arrays, the clone ID's are huge numbers, e.g. 1.6×10^{21}. This is not a good unique identifier. When SAM does the analysis, this is often converted to 1.60×10^{21}. In the computer world, 1.6×10^{21} and

1.60×10^{21} can be very different entities. I have not yet created support for this problem, as it is likely to be specific only to a limited number of microarrays.

Version History

v1.4 completed March 26, 2002

- Error handling for numerical ID's which are converted to different format in SAM output. These lines are excluded from the output, but counted and reported to the user.
- Support for workbooks which contain additional worksheets, but these must still be empty (no ability to select the appropriate worksheet)