Log Ratio Tails
Author: Christophe Lambert and Bryce Christensen, Golden Helix, Inc.

Overview

This script calculates percentile values for the upper and lower tails of the log ratios using two user-specified thresholds. Missing values are skipped. A log ratio call rate is returned with the results. This script may also be used to identify percentiles for real-value data other than log ratios.

Assumes:
- Samples are columns of the spreadsheet
- A marker map is applied or active data is from autosomes only
- There are columns of real-valued log ratios.

Recommended Directory Location

Save the script to the following directory:
*..**\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Numeric\CNV_QA**

Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Vista. The easiest way to locate this directory on your computer is to open SVS and go to **Tools > Open Folder > UserScripts Folder**. If saved to the proper folder, this script will be accessible from the spreadsheet **Numeric** menu using the **CNV QA** submenu.

Purpose of the Script

Identifying the user-specified percentile thresholds for the upper and lower tails of the observed log ratios for each subject and comparing these values across samples can help to identify samples with poor signal-to-noise ratios or samples with a systematic upward or downward shift in LR values. More generally, this script may be used as a means to calculate percentiles (quartiles, quintiles, etc) of any column or columns containing real-valued numeric data.

Using the Script

1. Open the spreadsheet containing samples in columns and genetic markers in rows. A marker map can be applied to the rows or not. If a marker map is not applied to the genetic markers, then it is assumed that the active markers only come from autosomes.
2. Inactivate any rows or columns not to be included in the calculation of the sample-wise Log Ratio Tails.
3. Go to **Numeric > CNV QA > Log Ratio Tails**.

4. The spreadsheet is scanned and any inactive rows or columns are identified.
5. A prompt dialog will appear requesting thresholds; suggested values for filtering samples are 0.01 for the first percentile, and 0.001 for the second percentile. See **Figure 1**.
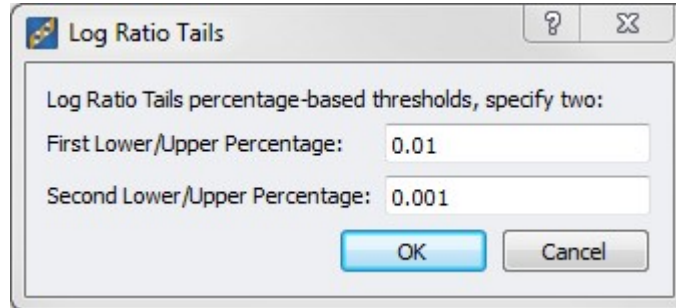

Figure 1: Percentile-based threshold dialog

The script will process the active data and output a spreadsheet containing a row for every active sample and columns for the log ratio tails values for the two specified percentiles, as well as the sample-wise log ratio call rate.

Figure 2: Log ratio tails results spreadsheet