Import PennCNV

Author: Sam Gardner, James Grover, Golden Helix, Inc.

Overview

This script imports PennCNV input signal intensity files, where each file contains data for a single sample. The files are expected to have real valued data arranged with a row for each marker and columns containing the marker names, chromosomes, positions, and data values (e.g. Log Ratios and B Allele Frequencies). One spreadsheet is generated for each column of values in the input files.

Note: If the header for the first data value column is of the format **sample.field**, **sample** will be used as the sample name corresponding to the sample file. Otherwise, the file name itself will be used as the sample name.

Note: If the data for a field is found to be non-numeric, it will be assumed that the field represents genotypes in a two-letter format, such as **AA**, **AB**, or **BB**.

More information about the file format can be found here:
http://www.openbioinformatics.org/penncnv/penncnv_input.html

Recommended Directory Location

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

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 select the **Tools >Open Folder > UserScripts Folder** menu option. If saved to the proper folder, this script will be accessible from the project navigator **Import** menu.

Using the Script

1. From an open project select **Import** > **Import PennCNV**

2. Choose a base dataset name and select the PennCNV files to import. Click **Import**.

The resulting spreadsheets have a row for each sample or file, sorted by import order. A marker map is applied to the columns providing Chromosome and Position fields at each probe.