Activate Variants by Genotype Count Threshold
Author: Autumn Laughbaum, GHI

Overview

This tool scans genotypic columns and activates columns based on a user-specified count or percentage threshold of user-specified genotypes. For example, you could use this tool to activate all genotypic columns that contain at least 20% homozygous alternate variants.

If a dependent (binary or categorical) column is set then the script allows the user to specify a count or percentage threshold that all rows in a given category must pass to remain active.

This tool requires a marker-mapped genotypic spreadsheet. The marker map must have a field that contains the reference allele.

Recommended Directory Location

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

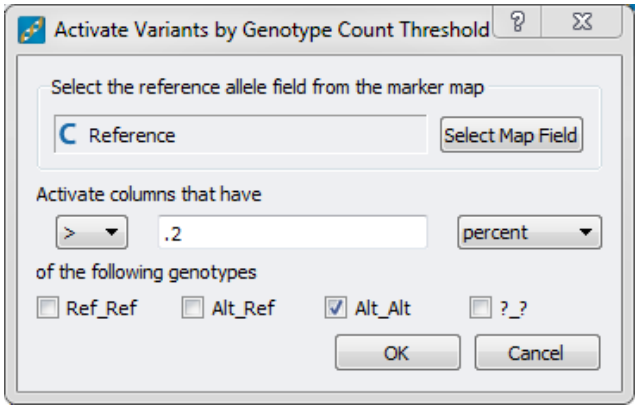
Note: The **Application Data** folder is a hidden folder on Windows operating systems and its location varies between XP and Vista. The easiest way to locate this directory on your computer is to open SVS and go to **Tools > Open Folder > User Scripts Folder**. If saved to the proper folder, this script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

Using the Script

1. Set dependent column by right-clicking on a binary of categorical column. (Optional)
2. From a marker-mapped genotypic spreadsheet, choose **Select >Activate Variants by Genotype Count Threshold**.
3. Choose the marker map field that contains the Reference allele. An attempt is made to find an appropriate default based on the marker map field names.
4. Specify the threshold direction, threshold value and type.
    a. Threshold direction: One of <, <=, >, >=, != or ==
    b. Threshold value: If Occurrences is selected as the type, enter an integer value (although real-valued numbers are also accepted). If percent is selected, enter a decimal value between 0 and 1.
    c. Threshold type: Either *Occurrences* or *Percent*.
5. Choose the genotypes that you wish to include in the calculation.
6. If optional dependent column is chosen in step 1 repeat steps 4 and 5 for each category that a genotype count threshold is required.

Example

The following screenshot shows an example that would activate all genotypic columns that contained greater than 20% homozygous alternate calls (Alt_Alt).



In this example, in each mapped genotypic column, the percent is calculated as the number of Alt_alt values found in the column over the length of the column. Or,

$$percent_c = count(Alt\_Alt)/count(Ref\_Ref,\ Alt\_Ref,\ Alt\_Alt,\ ?\_?)$$

Example

The following screenshot shows an example that would activate all genotypic columns that contain greater than 50% homozygous alternate calls (Alt_Alt) in the cases with greater than 50% reference (Ref_Ref) or missing (?_?) calls in the controls.