

Subset Informative Genotypes by Category

Author: Autumn Laughbaum, Golden Helix, Inc.

Overview

This tool scans genotypic columns to find informative genotypes defined by having at least one non-missing, non-reference allele. Informative genotype column sets are found for each unique category in a user-defined categorical column.

This tool requires a spreadsheet with several mapped genotypic columns. The marker map must contain a reference allele field and may contain a categorical column. The row labels may also be used as the categorical column.

Recommended Directory Location

Save the script to the following directory:

***..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\Quality Assurance**

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 **Tools >Open Folder > UserScripts Folder**. If saved to the proper folder, this script will be accessible from the spreadsheet's **Quality Assurance** menu.

Using the Script

1. Open a spreadsheet containing several mapped genotypic columns and choose **Quality Assurance >Subset Informative Genotypes by Category**.
2. In the prompt, choose the categorical column and reference allele field (may be auto-detected).

N output spreadsheets are created representing N unique categories. The genotypic columns included in each spreadsheet contain at least one non-missing, non-reference allele over all rows within the category. If no informative columns are found, the output spreadsheet is not created.