Report Samples with Unique Genotypes
Author: Autumn Laughbaum, Golden Helix, Inc.

Overview
This tool scans a genotype spreadsheet and determines samples that have unique genotypes, or are not found in any other sample at that loci. A report is created with binary columns representing the unique genotypes per sample per variant.

Optionally the user can set a categorical or binary dependent column. The output report generated will then contain the percent of samples in each group that contain a unique genotype (or not found in any other group of samples).

Recommended Directory Location
Save the script to the following directory:
For SVS 8.0.0+
*..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\DNA_Seq

For SVS 7.7.8 and earlier releases
*..\Application Data\Golden Helix SVS\User Scripts\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 > User Scripts Folder. If saved to the proper folder, this script will be accessible from the spreadsheet’s DNA-Seq or Quality Assurance menus depending on the version of SVS being used.

Using the Script
1. Open a spreadsheet containing several genotypic columns and optionally set a dependent column.
2. Choose DNA-Seq > Report Samples with Unique Genotypes (for SVS8) or choose Quality Assurance > Report Samples with Unique Genotypes (for SVS7)
3. The spreadsheet is then scanned and a reported is generated.

The report contains one row per variant and one binary column per sample. The binary values represent the following:
- 1: This sample had a unique genotype for this variant. The genotype in this cell was not found in any other sample at this variant.
- 0: The genotype in this cell was found in other samples at this variant.
- ?: The genotype was missing (no-call) at this variant.

If a dependent column was set, the output contains a percentage value representing the # samples that had unique genotypes (or not found in any other group).
A natural next step would be to use **Select > Activate Rows Based on Multiple Column Criteria** (or **Select > Filter on Multiple Columns** for SVS7) or **Activate by Category** in the column menu. These tools allow you to subset only to variants that contain a unique genotype for a given sample(s), such as a proband or affected individual.