Filter by PolyPhen2 Score
Author: Gabe Rudy and Mike Thiesen, Golden Helix, Inc.

Overview

This filter inactivates mapped markers that are either predicted as tolerated or have low confidence (do not pass filters) or are predicted as damaging (pass filters), depending on the inactivation option selected.

Note: This script has been deprecated by **Filter on NS Functional Predictions**, shipped in SVS version 7.6.7. PolyPhen2 filtering can now be applied simultaneously with other filtering tools such as SIFT and MutationTaster.

Recommended Directory Location

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

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 >Filter by Annotation** menu.

Using the Script
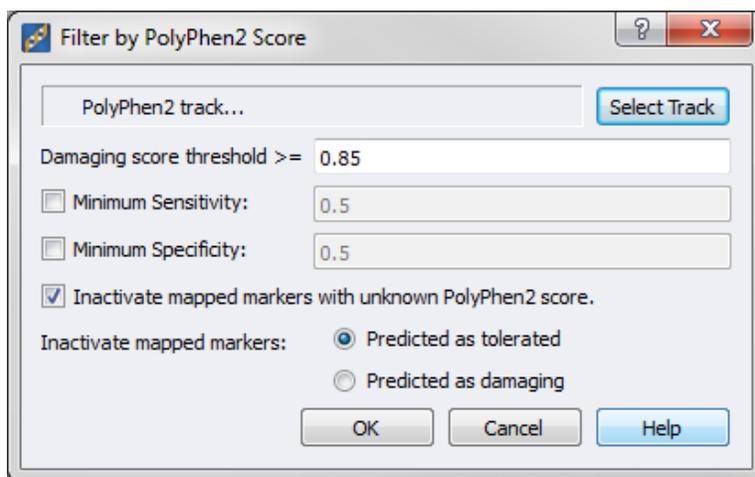
1. From a spreadsheet with a marker map applied to the columns, go to **Select >Filter by Annotation >Filter by PolyPhen2 Score.**
2. Select the PolyPhen2 track (required to be a local track)
   a. If the PolyPhen2 track has not yet been downloaded to your local machine, download it from the network using the annotations track manager.
3. Enter a **damaging score threshold**.

a. The PolyPhen2 score ranges from 0 to 1. A high score represents a variant that is more likely to be damaging. Any markers below this score will be treated as non-damaging, while those above will be treated as damaging.
4. Enter a **Minimum Sensitivity** threshold (optional).
   a. Sensitivity is the probability that a damaging mutation will be correctly classified. If the sensitivity is below this limit, then the polyphen score will be treated as unknown. This option can be used to reduce the chance of a false negative (Type II error).
5. Enter a **Minimum Specificity** threshold (optional).
   a. Specificity is the probability that a non-damaging mutation will be correctly classified. If the specificity is below this limit, then the polyphen score will be treated as unknown. This option can be used to reduce the chance of a false positive (Type I error).
6. Specify whether mapped markers with an unknown score should be inactivated or not.
   a. Many markers will not have a score due to the marker being non-coding, or because the prediction failed the sensitivity/specificity checks. This option specifies whether these markers should be inactivated.
7. Indicate which markers should be inactivated, either:
   a. Those predicted as tolerated
   b. Those predicted as damaging

In the original spreadsheet, the columns that meet the inactivation criteria are inactivated. The other columns are left active.

A marker mapped filtering results spreadsheet is also created as a child of the original spreadsheet. This spreadsheet contains the following columns:

- **Found in Track?** Indicates whether the marker was found in the PolyPhen2 track.
- **Passed QA?** Indicates whether the sensitivity and specificity values were above the specified thresholds.
- **Damaging?** Indicates whether the marker is predicted to be damaging.
- **Drop?** Indicates whether the marker was inactivated.
- **PPH2 Rsid** The dbSNP rsid for the marker used by polyphen.
- **Ref/Alt Alleles** The reference and alternate nucleotide bases used for the calculations of the PolyPhen2 prediction.
- **Description** A text description of the PolyPhen2 prediction.
- **Probability** The PolyPhen2 probability that a marker describes a damaging variant.
- **Sensitivity** The sensitivity of the prediction as reported by PolyPhen2.
- **Specificity** The specificity of the prediction as reported by PolyPhen2.