

Add Annotation Data to Marker Map From Spreadsheet

Author: Sam Gardner, Greta Peterson, Golden Helix, Inc.

Overview

This function takes the marker map applied to the current spreadsheet and adds specified annotation data from overlapping interval(s) to each marker in the marker map. It then saves a copy of the new map with the additional information in the users Marker Map Folder as well as applies the new map to the current spreadsheet.

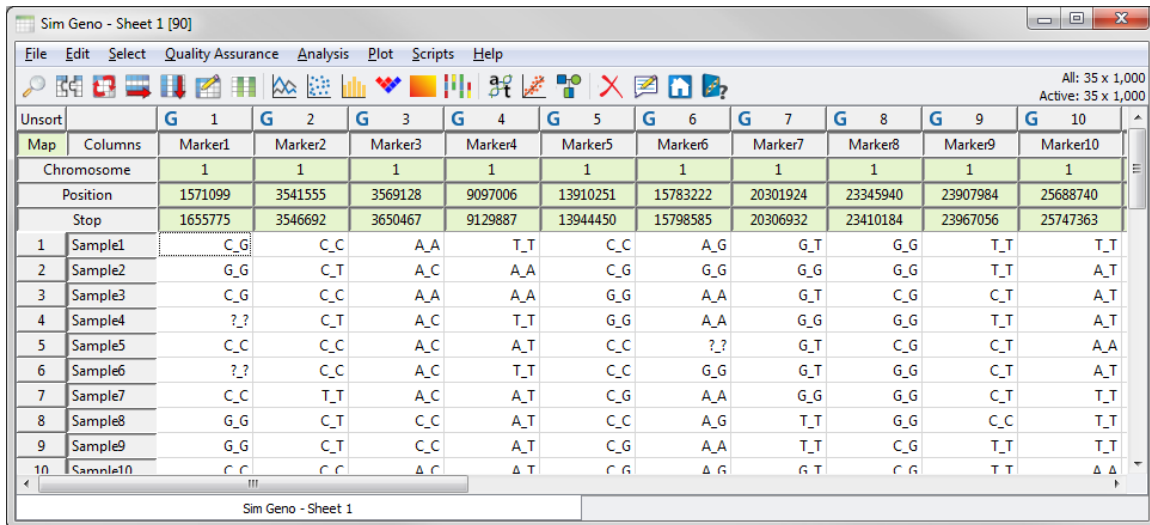
Recommended Directory Location

Save the script to the following directory:

*..\Application Data\Golden Helix SVS\UserScripts\Spreadsheet\File\

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 click on the **AppData** shortcut in your **C:\Program Files\Golden Helix SVS** directory. If saved to the proper folder, this script will be accessible from the spreadsheet **File** menu.

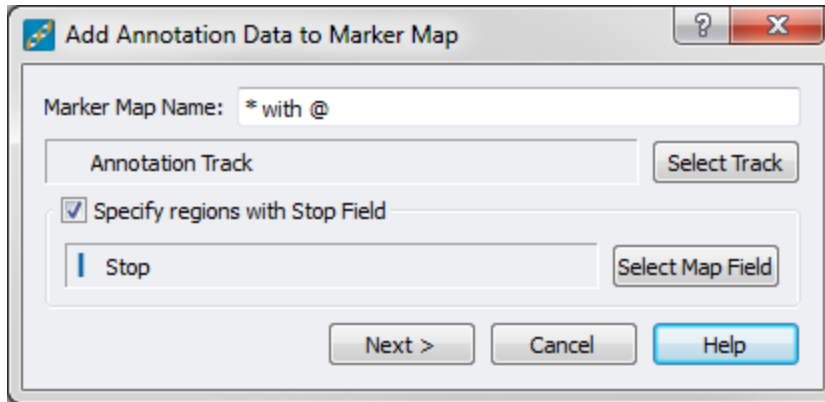
Using the Script



The screenshot shows the Sim Geno software interface with a spreadsheet titled 'Sim Geno - Sheet 1 [90]'. The spreadsheet has columns for 'Map', 'Columns', 'Marker1' through 'Marker10', and 'Stop'. The data is as follows:

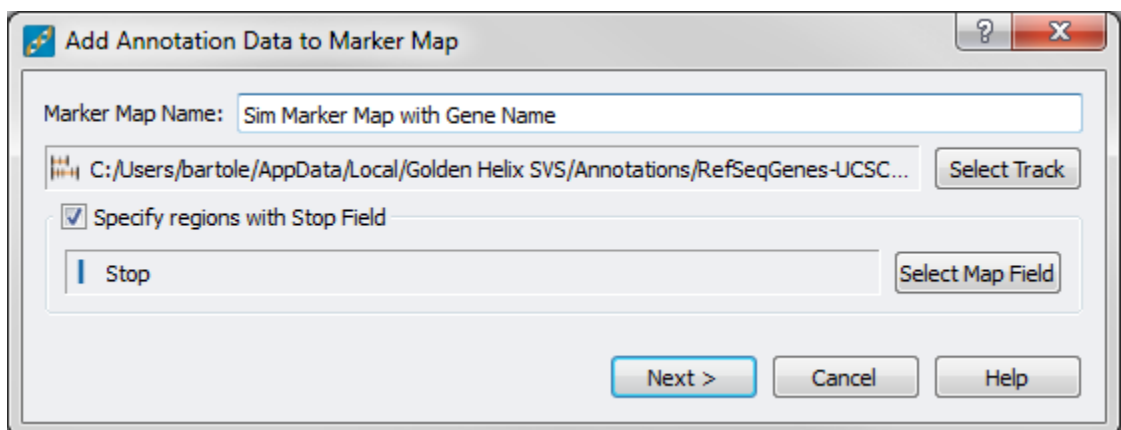
Map	Columns	Marker1	Marker2	Marker3	Marker4	Marker5	Marker6	Marker7	Marker8	Marker9	Marker10
Chromosome	1	1	1	1	1	1	1	1	1	1	1
Position	1571099	3541555	3569128	9097006	13910251	15783222	20301924	23345940	23907984	25688740	
Stop	1655775	3546692	3650467	9129887	13944450	15798585	20306932	23410184	23967056	25747363	
1	Sample1	C_G	C_C	A_A	T_T	C_C	A_G	G_T	G_G	T_T	T_T
2	Sample2	G_G	C_T	A_C	A_A	C_G	G_G	G_G	G_G	T_T	A_T
3	Sample3	C_G	C_C	A_A	A_A	G_G	A_A	G_T	C_G	C_T	A_T
4	Sample4	?_?	C_T	A_C	T_T	G_G	A_A	G_G	G_G	T_T	A_T
5	Sample5	C_C	C_C	A_C	A_T	C_C	?_?	G_T	C_G	C_T	A_A
6	Sample6	?_?	C_C	A_C	T_T	C_C	G_G	G_T	G_G	C_T	A_T
7	Sample7	C_C	T_T	A_C	A_T	C_G	A_A	G_G	G_G	C_T	T_T
8	Sample8	G_G	C_T	C_C	A_T	C_C	A_G	T_T	G_G	C_C	T_T
9	Sample9	G_G	C_T	C_C	A_T	C_G	A_A	T_T	C_G	T_T	T_T
10	Sample10	C_C	C_C	A_C	A_T	C_G	A_G	G_T	C_G	T_T	A_A

1. From a mapped spreadsheet with a minimum of chromosome and start position information, choose **File > Add Annotation Data to Marker Map From Spreadsheet**.

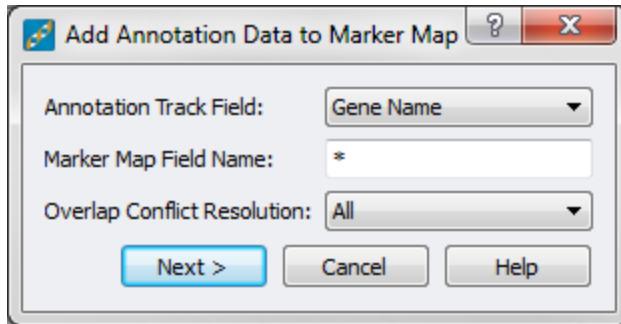


2. Give the map an informative name under **Marker Map Name**, if the default selection is left the script will append the name of the new marker field onto the original map name.
3. Click on **Select Track** to choose the appropriate annotation track that contains the annotation information you would like added. For example if you would like to add Gene name to your marker map then any of the available Gene tracks for your species and build will work.
4. If your marker map has a stop position defined and you would like to add all annotation data defined in the region between the start and stop positions of your defined markers then select **Specify Regions with Stop Field** and select the corresponding Stop position field from the existing map.

If you do not have a stop position field in your map or you only want to add the annotation information that matches the start position of the marker, then uncheck this option.



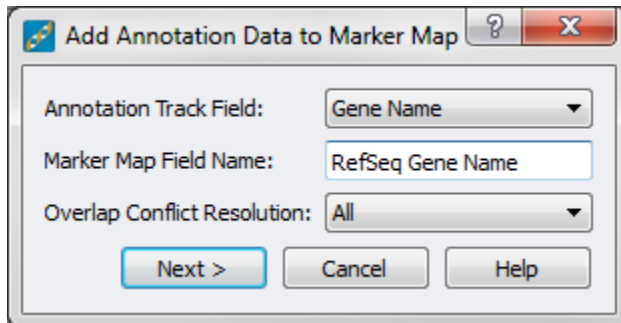
5. Click **Next**



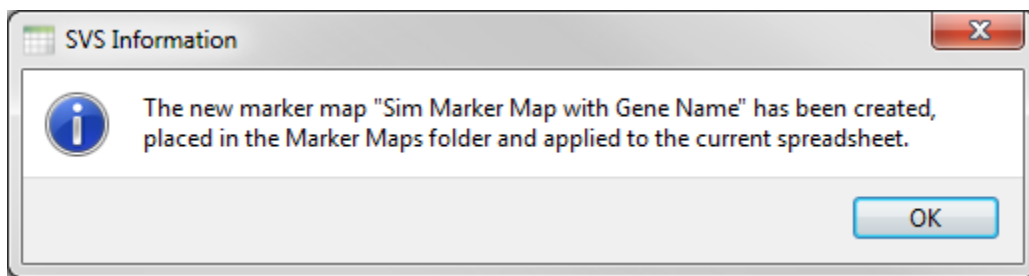
6. Select the annotation field from the track you would like added to the existing marker map from the drop down under **Annotation Track Field**.
7. If you would like to label the annotation field as something different than what is listed in the track you can add the name under **Marker Map Field Name**.

SVS requires that marker map fields be unique so for example if you will be adding Gene name from two different sources you will want to create unique field names from each source.

8. Select how you would like to resolve any conflicts from the drop down menu after **Overlap Conflict Resolution**.



9. Click **Next**.
10. You should get a similar message to the following if everything worked correctly.



11. A new spreadsheet should be created with the new map applied.

The screenshot shows a spreadsheet window titled "Sim geno - Mapped Sheet 1 [96]". The spreadsheet has columns labeled G 2 through G 10, which correspond to Marker2 through Marker10. The rows include a "Map" section with columns for "Columns", "Marker2", "Marker3", "Marker4", "Marker5", "Marker6", "Marker7", "Marker8", "Marker9", and "Marker10". Below this, there are rows for "Chromosome", "Position", "Stop", and "RefSeq Gene Name". The "RefSeq Gene Name" row lists: TPRG1L, TP73, SLC2A5, PDPN, CELA2A, PLA2G2A, MIA, MIR31, MDS2, MEM50A, RHC, and TM. The bottom section of the spreadsheet contains 9 rows of sample data, labeled "Sample1" through "Sample9", with columns for each marker (G 2 to G 10) containing genotype pairs like "C_C", "A_A", "T_T", "C_G", "A_G", "G_G", "G_T", "C_G", "C_T", "A_A", "T_T", "A_G", "T_T", "C_G", "T_T", "C_G", "T_T".

Map	Columns	Marker2	Marker3	Marker4	Marker5	Marker6	Marker7	Marker8	Marker9	Marker10
Chromosome		1	1	1	1	1	1	1	1	1
Position		3541555	3569128	9097006	13910251	15783222	20301924	23345940	23907984	25688740
Stop		3546692	3650467	9129887	13944450	15798585	20306932	23410184	23967056	25747363
RefSeq Gene Name		TPRG1L	TP73	SLC2A5	PDPN	CELA2A	PLA2G2A	MIA, MIR31	MDS2	MEM50A, RHC, TM
1	Sample1	C_C	A_A	T_T	C_C	A_G	G_T	G_G	T_T	T_T
2	Sample2	C_T	A_C	A_A	C_G	G_G	G_G	G_G	T_T	A_T
3	Sample3	C_C	A_A	A_A	G_G	A_A	G_T	C_G	C_T	A_T
4	Sample4	C_T	A_C	T_T	G_G	A_A	G_G	G_G	T_T	A_T
5	Sample5	C_C	A_C	A_T	C_C	?_?	G_T	C_G	C_T	A_A
6	Sample6	C_C	A_C	T_T	C_C	G_G	G_T	G_G	C_T	A_T
7	Sample7	T_T	A_C	A_T	C_G	A_A	G_G	G_G	C_T	T_T
8	Sample8	C_T	C_C	A_T	C_C	A_G	T_T	G_G	C_C	T_T
9	Sample9	C_T	C_C	A_T	C_G	A_A	T_T	C_G	T_T	T_T