Export LongGen File

Author: Alison Figueira, Golden Helix, Inc.

Overview

This script will export genotypic spreadsheet data into Plink's LongGen file format. For more information on this format, please see: http://pngu.mgh.harvard.edu/~purcell/plink/data.shtml#long

Recommended Directory Location

Save the script to the following directory:

*..\AppData\Local\Golden Helix SVS\UserScripts\Spreadsheet\File\Save As\

Note: The AppData (or Application Data) folder is a hidden folder on Windows operating systems and its location varies between various versions. The easiest way to locate this directory on your computer is to open SVS and select the Tools > Open Folder > UserScripts Folder menu option and save the script in the \Spreadsheet\File\Save As\ folder. If saved to the proper folder, this script will be accessible from the spreadsheet menu.

Using the Script

 Open the spreadsheet containing the data to be exported. The data should be by column, as the example below. The Family ID column is optional and if not present, there is be no Family ID column in the resultant text file. In addition, the sample IDs can be in the row labels.

📅 Edited 500K HapMap - Row Subset - Column Subset [203]										
<u>F</u> ile	<u>File Edit Select DNA-Seq Genotype Numeric RNA-Seq GenomeBrowse Plot Scripts H</u> elp									
£€	All: 10 x 1 Active: 10 x 1									
Unsort		1 1	2	G 3	G 4	G 5	G 6	G		
Мар	NSP_STY	Family ID	Sample ID	SNP_A-1909444	SNP_A-2237149	SNP_A-4303947	SNP_A-1886933	SNP_A-2		
1	CEU_NA06985.CEL	1	1	B_B	B_B	A_A	B_B			
2	CEU_NA06991.CEL	3	1	A_B	B_B	A_B	A_B			
3	CEU_NA06993.CEL	4	1	A_B	B_B	A_B	A_B			
4	CEU_NA06994.CEL	2	1	B_B	B_B	A_A	B_B			
5	CEU_NA07000.CEL	2	2	B_B	B_B	A_A	B_B			
6	CEU_NA07019.CEL	1	2	B_B	B_B	A_A	B_B			
7	CEU_NA07022.CEL	1	3	B_B	B_B	A_A	B_B			
8	CEU_NA07029.CEL	4	2	B_B	B_B	A_A	B_B			
9	CEU_NA07034.CEL	4	3	A_B	B_B	A_B	A_B			
10	CEU_NA07048.CEL	3	2	B_B	B_B	A_A	B_B			
•		III						F.		
	Edited 500K HapMa	ip - Row Subset - C	Column Subset							

Figure 1: Example Spreadsheet

- 2. While in the spreadsheet window, select **File > Save As... > Export LongGen File**.
- 3. In the prompt window, type in the desired file name, select the directory the file should be saved in, and select the sample ID column. If a family ID column is available, it can be chosen here too.

🔀 Export Spreadsh	eet to LongGen Format	? ×					
The LongGen file exp their Individual or Sa	The LongGen file exporter will export genotypes for samples based on their Individual or Sample IDs and Family IDs. Please select a file name as well as the output directory. Then indicate the columns that should be used for the Family ID (not required) and the Individual or Sample ID (required). The spreadsheet will first be sorted by Family ID (if entered) or the Individual ID, then exported.						
Please select a file n that should be used Sample ID (required) or the Individual ID,							
File Name: Edited 500K HapMap - Row Subset - Column Subset							
Choose the output directory:							
Family ID:							
Select the column to use for family identifiers: Select Column							
Sample ID:							
C Row Labels (NSF	P_STY)	Select Column					
		Save Cancel					

Figure 2: Prompt Dialog

- 4. Click **Save** to export the spreadsheet.
- 5. For exporting the spreadsheet will be sorted by family ID first (if available) then sample ID.
- 6. The exported text file will look similar to the example below. The columns are: Family ID, Sample ID, SNP Name, Allele 1, Allele 2. If no family ID column was chosen this field will be left out.

😑 Edited	Edi	ted	500K HapMap - Row Subs	et -	Column Subset - Sheet 1.Igen 🛛
1	1	1	SNP_A-1909444	в	В
2	1	1	SNP_A-2237149	в	В
3	1	1	SNP_A-4303947	A	A
4	1	1	SNP_A-1886933	в	В
5	1	1	SNP_A-2236359	в	В
6	1	1	SNP_A-2205441	Α	A
7	1	1	SNP_A-2116190	Α	A
8	1	1	SNP_A-4291020	Α	A
9	1	1	SNP_A-1902458	Α	A
10	1	1	SNP_A-2131660	Α	В
11	1	2	SNP_A-1909444	в	В
12	1	2	SNP_A-2237149	в	В
13	1	2	SNP_A-4303947	Α	A
14	1	2	SNP_A-1886933	в	В
15	1	2	SNP_A-2236359	в	В
16	1	2	SNP_A-2205441	Α	A
17	1	2	SNP_A-2116190	Α	A
18	1	2	SNP_A-4291020	Α	A
19	1	2	SNP_A-1902458	Α	A
20	1	2	SNP_A-2131660	Α	В
21	1	3	SNP_A-1909444	в	В
22	1	3	SNP A-2237149	в	В

Figure 3: Example Text File