Extracting Data from GenomeStudio Using the Golden Helix DSF Export Plug-in

Introduction

In collaboration with Illumina, Golden Helix has developed a set of custom plug-ins to streamline the transfer of genotype, log ratio and bookmark files between Golden Helix's SNP & Variation Suite (SVS) applications and Illumina's GenomeStudio Data Analysis software. These instructions lead you through extracting genotype, log ratio, b-allele frequency, computed CNV values, X/Y value pairs and raw X/Y value pairs data from GenomeStudio as Golden Helix DSF files and importing the created DSF files into SVS versions 7 and 6 (a.k.a HelixTree 6).

Installation Instructions

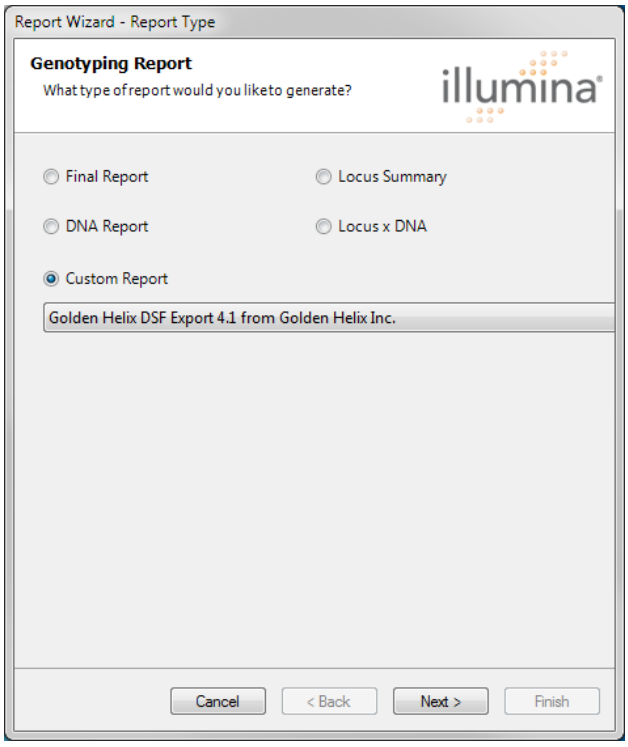
Before beginning the installation process, make sure all instances of GenomeStudio are not running.

Open the **Golden Helix GenomeStudio DSF Plugin-4.1.exe** file and follow the instructions in the installation wizard. Upon completion of the wizard, three plug-ins, including the Golden Helix DSF Export plug-in, will be extracted to the appropriate GenomeStudio folders. By default these folders are located in **C:\Program Files\Illumina\GenomeStudio**.

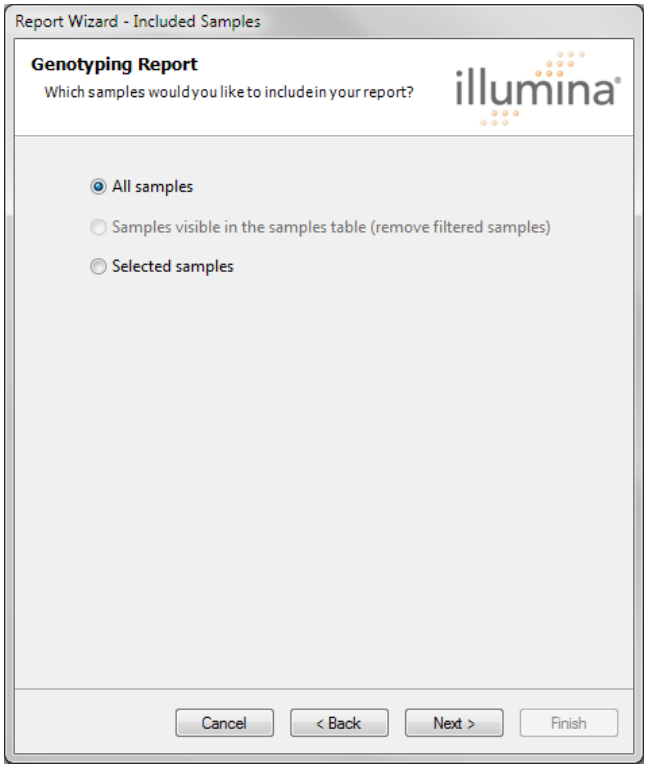
Extracting Data from GenomeStudio

If the installation was performed successfully, the Golden Helix DSF Export Plug-in can be accessed from an open project in GenomeStudio using the Report Wizard. The following instructions will guide you through this process.

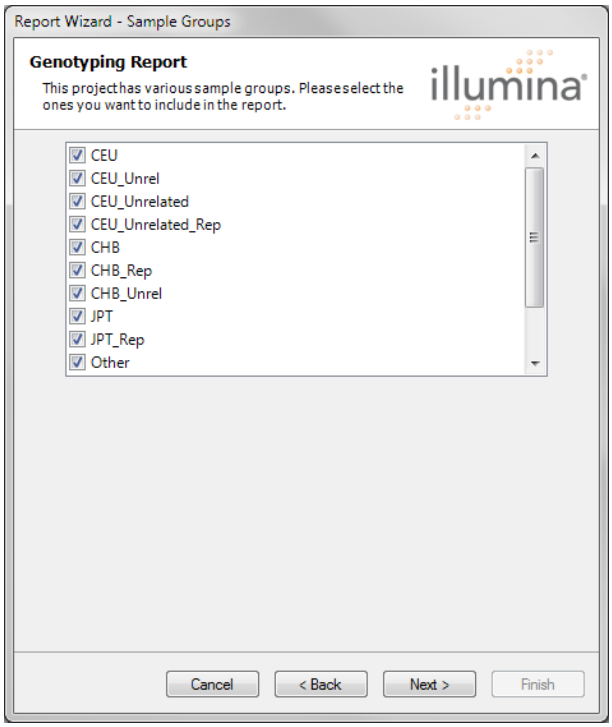
- Open a GenomeStudio project and select **Analysis >Reports >Report Wizard**.
- From the first window of the wizard choose **Custom Report** and then select **Golden Helix DSF Export 4.1 from Golden Helix, Inc.** (below). Click **Next** to continue.



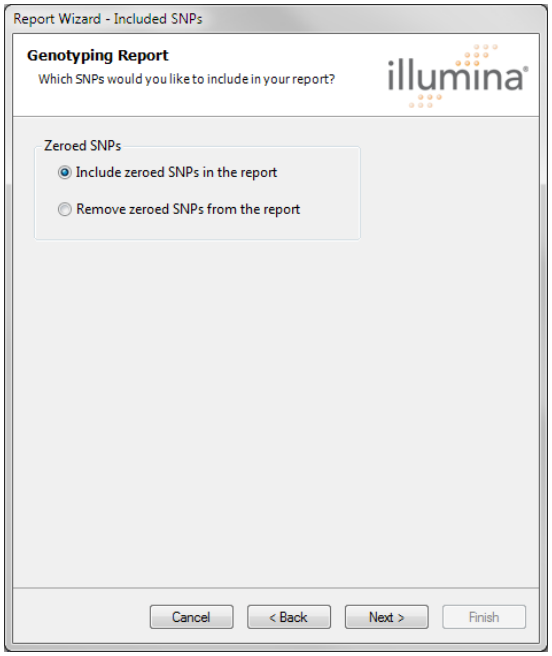
- The next window asks which samples you would like to include in your report. Select the appropriate sample set and click **Next**.



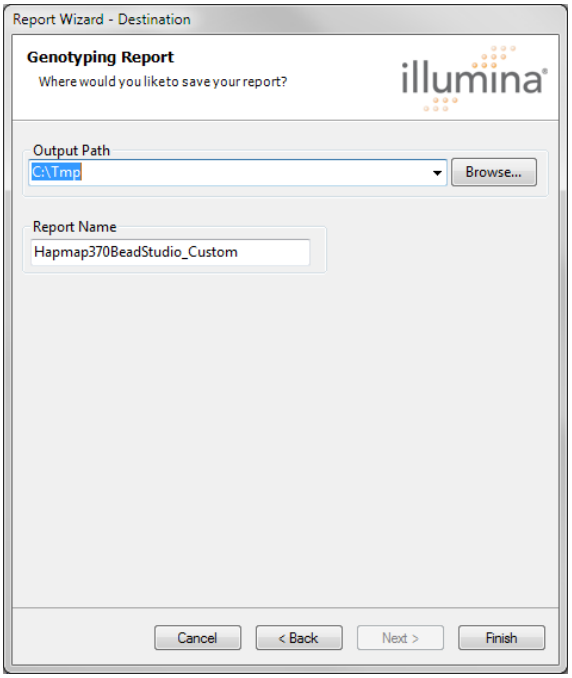
- If you have multiple sample groups in your project you will be asked to select the ones you want to export. Select the sample groups and click **Next**.



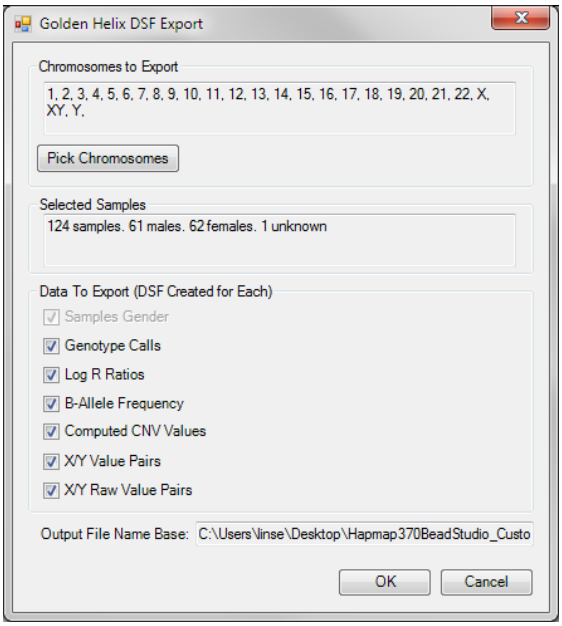
- If you have zeroed SNPs in the report you will be asked if you want the zeroed SNPs to be included in the report or not. Select the appropriate action and click **Next**.



- On the following window **Browse** to an appropriate directory where you want the resulting DSF file to be saved. Give the report a name and click **Finish**.



This will activate the Golden Helix DSF Export Plug-in, opening the following window:



Chromosomes to Export: By default all the chromosomes represented in your project are listed. If you only want to export a subset of chromosomes click **Pick Chromosomes** and check the chromosomes you want to export.

Selected Samples: Information about the selected samples is displayed in the **Selected Samples** box. If the information in this box does not appear to be correct, cancel the export and adjust the selected samples in the GenomeStudio project. A DSF is automatically created for the samples to export gender information.

Data to Export: There are six optional DSF output options to choose from. Multiple selections are allowed for simultaneous export, any one or all of these DSF files can be exported. The data will be exported in the optimal export format with samples in columns and genetic markers in rows. The options include:

- Genotype Calls
- Log R Ratios
- B-Allele Frequency
- Computed CNV Values
- X/Y Value Pairs
- X/Y Raw Value Pairs

Output File Name Base: The file name is set by the output directory chosen in the GenomeStudio Report Wizard and cannot be changed in this window. Each DSF file created will have this base and the specific DSF file name appended to the base.

When you are finished choosing the desired parameters click **OK** to begin the export process.

Upon completion a DSF file for sample gender information plus any additional DSF files will be created in the directory location you chose.

Importing DSF File into SVS 7

The method for importing an Illumina DSF is the same for all DSF files created by the plug-in.

- From an open project select **Import >Illumina DSF**.
- Select the DSF file to import.
- When the import is complete, a dialog box will appear asking if you would like to save the marker map included with the DSF file (for all DSF files other than the Samples Gender DSF file) to your permanent MarkerMaps folder for use on a different dataset. Click **Yes** if you want to save the included marker map and **No** if you don't.
- The Samples Gender DSF file will import with samples in rows and one column containing gender information.
- All optional DSF files will import with samples in columns and genetic markers in rows. Some quality control measures can be performed on log ratio data in this orientation. Plotting individual sample's log ratio and b-allele frequency data can also be done with the data in this orientation. However, most analyses require that a spreadsheet have markers in columns and samples in rows. This requires

transposing the data. See:
http://www.goldenhelix.com/SNP_Variation/Manual/svs7/working_with_a_single_spreadsheet.html#tailworking_with_a_single_spreadsheet.html

Importing DSF File into HelixTree 6

HelixTree 6 does not natively support transposing data, so to import data into this version of the software, first import into SVS 7 to transpose the data. Once the data is transposed, the method used to import data into HelixTree 6 depends on the data to import. The only two useful DSF files to import into this version are the Genotype Call and Log R Ratio files.

Genotype Calls:

- From an open project, select **File > Import Data > Import DSF**.

Log R Ratio Data: There are several analysis operations you can perform directly on the output LogR DSF file: import the Log R Ratio data as a spreadsheet, perform LogR association tests, or perform CNAM optimal segmenting on the LogR data.

- From an open project, select:
    - **CNAM > Import LogR DSF**: to import the data as a spreadsheet.
    - **CNAM > LogR Association Tests and PCA**: to perform LogR association tests or principal component analysis on the data.
    - **CNAM > Copy Number Segmentation**: to perform CNAM optimal segmenting on the data.