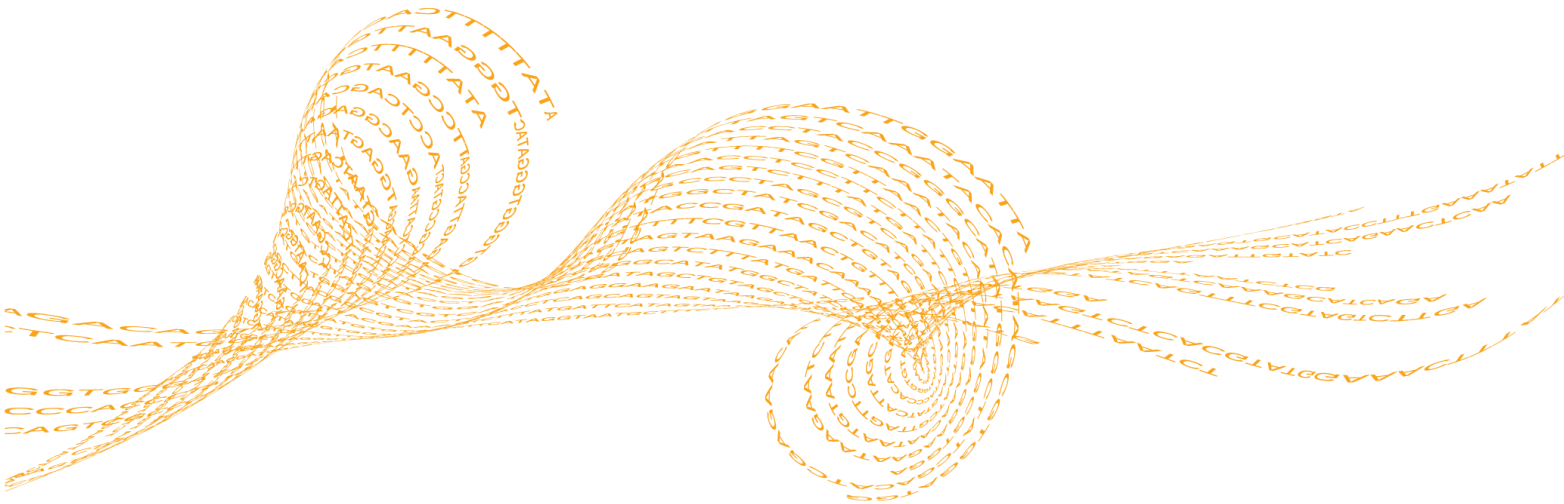
Performing Clustering in the GenomeStudio Polyploid Clustering Module

FOR RESEARCH USE ONLY



ILLUMINA PROPRIETARY
Part # 15044760 Rev A
September 2013

illumina®

Overview

The GenomeStudio Polyploid Clustering Module identifies clusters for samples for which the standard diploid clustering algorithm is not appropriate or otherwise useful, such as for polyploid organisms like wheat and potato.

NOTE
The Polyploid Clustering Module performs cluster assignment, but does not call genotypes. This is because the assignment of genotypes in polyploid species is highly dependent on the population and biology of the organism. Any downstream genotype assignment should be done with the biology and evolutionary history of the population taken into consideration.

The Polyploid Clustering Module uses two published population-based detection algorithms to assign samples to meaningful clusters in genotyping data: Density Based Spatial Clustering of Applications with Noise (DBSCAN)[1] and Ordering Points to Identify the Clustering Structure (OPTICS)[2]. After you find appropriate settings for most loci and samples in the data set, you can identify problematic loci and cluster positions at specific SNPs. These can be adjusted manually for the population.

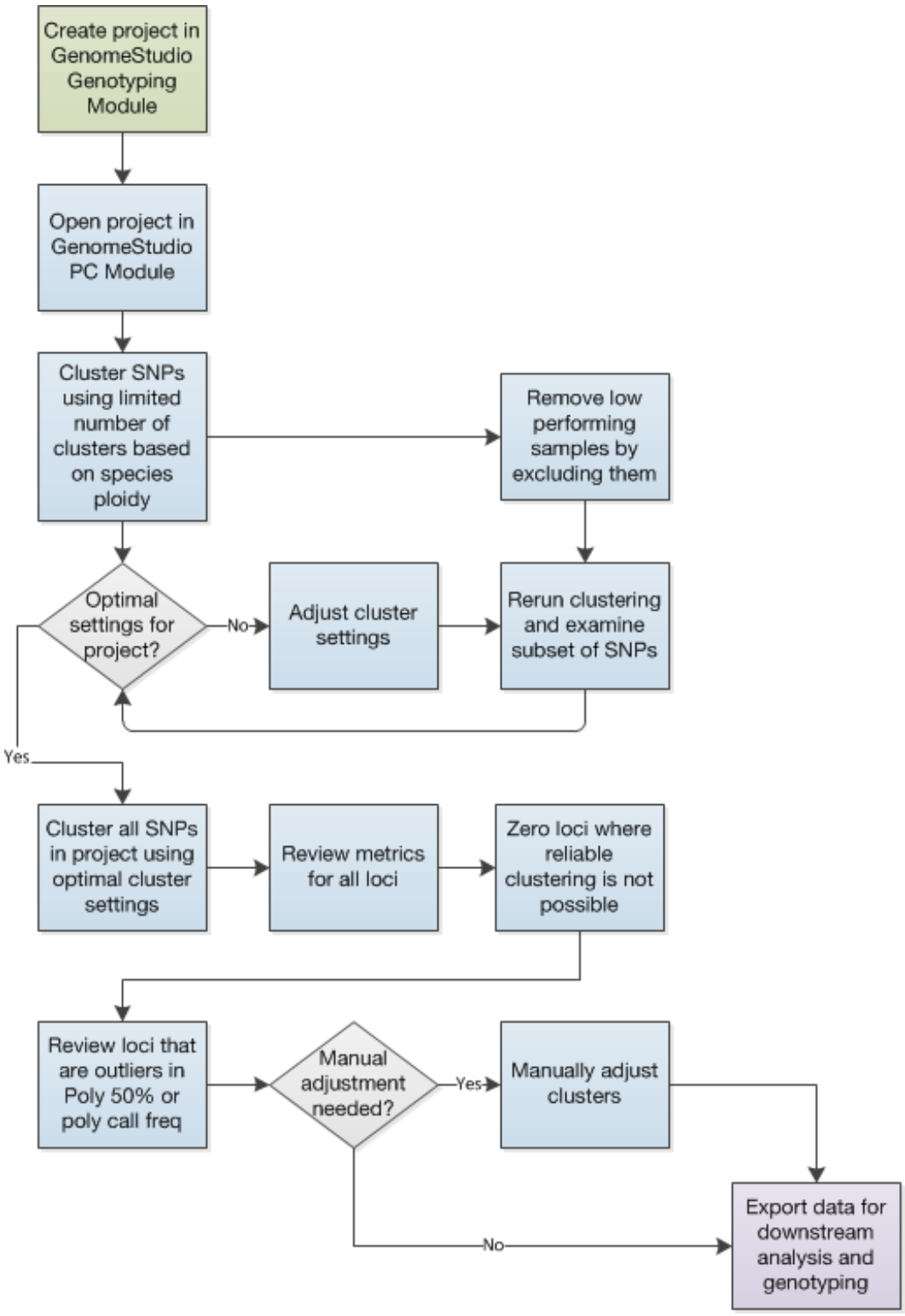
This guide provides instructions for the following:

- Using the Polyploid Clustering Module to perform clustering using these algorithms.
- Adjusting cluster settings to optimize the sensitivity of cluster detection for a project.
- Manually adjusting clusters on a per-SNP basis (if desired).
- Importing cluster positions from a polyploid cluster file.
- Exporting cluster data.

1 Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han, Usama M. Fayyad. Proceedings of the Second International Conference on Knowledge Discover and Data Mining (KDD-96). AAI Press. pp. 226–231. ISBN 1-57735-004-9.

2 Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander (1999). OPTICS: Ordering Points to Identify the Clustering Structure. ACM SIGMOD International Conference on Management of Data. ACM Press. pp. 49–60.

Recommended Workflow



Getting Started

The GenomeStudio Polyploid Clustering Module (Polyploid Clustering Module) is standalone software that must be installed separately from the other GenomeStudio modules.

This section describes the software installation, how to create a Polyploid Clustering project, and how to set up the Polyploid Clustering Module workspace to get the best use from this software.

Install the GenomeStudio Polyploid Clustering Module

NOTE
The GenomeStudio Polyploid Clustering Module has the same computing requirements as other GenomeStudio BeadArray modules. For more information, go to support.illumina.com/array/array_software/genomestudio/computing_requirements.ilmn.

Install the GenomeStudio Polyploid Clustering Module after installation of the GenomeStudio Genotyping Module (see *GenomeStudio Genotyping Module v1.0 User Guide PN 11318815*). Illumina recommends that you install GenomeStudio v2011.1 before installing the Polyploid Clustering Module.

Complete the following steps to install the GenomeStudio Polyploid Clustering Module.

1. Double-click the downloaded folder to unzip the contents of the GenomeStudioPolyploid Clustering Module Setup Wizard.
2. Follow the prompts in the installation wizard.

NOTE
You can use the Disk Cost feature on the Select Installation Folder dialog box to ensure that you select an installation folder that has enough free space.

Create a Project

The Polyploid Clustering Module can cluster data only in projects created initially from the standard (diploid) GenomeStudio Genotyping Module. After the diploid project is saved, the project can be opened and converted to a polyploid project using the Polyploid Clustering Module.

Create and Save a Diploid Genotyping Project

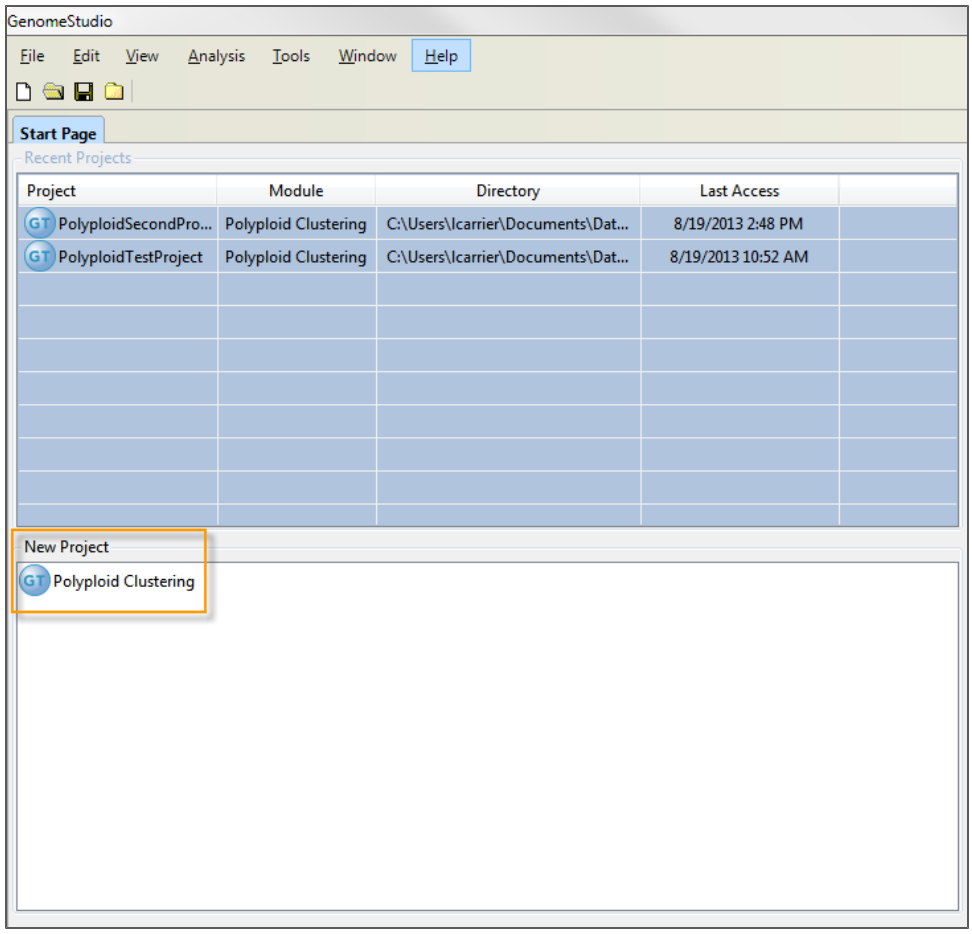
1. Open the Genotyping Module.
2. In the New Project area of the GenomeStudio Start page, click **Genotyping**.
3. Proceed through the GenomeStudio Project Wizard to create a diploid genotyping project.
4. Save and close the project.

For more information on any of these steps, see the GenomeStudio Genotyping Module v1.0 User Guide.

Convert a Diploid Genotyping Project to a Polyploid Project

1. Open the GenomeStudio Polyploid Clustering module.
2. In the New Project area of the GenomeStudio Start page, click **Polyploid Clustering**.

NOTE
The Polyploid Clustering option does not appear in the New Project area on the Start page of the standard GenomeStudio software. It is only available in the standalone Polyploid Clustering Module version of the software.



3 Navigate to the genotyping project (*.bsc) containing the data that you want to cluster and click **Open**.

NOTE
Do not use the New Project function from the File menu. This method does not work for initiating a polyploid clustering project.

The project data is loaded into the Full Data Table, SNP Table, and Samples Table. If the samples have not yet been clustered using the Polyploid Clustering Module, the Cluster Distance and cluster frequency (C1 Freq, C2 Freq, C3 Freq, ...) columns contain the value NaN.

NaN is an acronym for "Not a Number". It indicates that the SNP has not yet been clustered or that one of the clusters was adjusted manually.

4 Do one of the following to save the project:

- To overwrite the diploid genotyping version with the polyploid version, select **File | Save Project**.
- To preserve the diploid version of the project, select **File | Save Project Copy As**.

To continue working on the newly created project, close the current project, navigate to the project copy that you just saved, and follow instructions for opening a project.

Set Up the Workspace

When you open the Polyploid Clustering Module for the first time, the workspace is set up the same as the standard (diploid) GenomeStudio Genotyping Module. Before you start using the Polyploid Clustering Module, Illumina recommends reconfiguring the workspace as follows.

1 From the **Windows** menu, make sure that the **Full Data Table**, **SNP Table**, **Samples Table**, and **SNP Graph** are visible.
No other windows are needed.

2 In the Full Data Table, click the Column Chooser button and set the following columns related to polyploid clustering to appear:

Displayed Columns	Displayed Subcolumns
Cluster Distance	Cluster
# Clusters	Poly Score
# EGT Clusters	
Sample columns	

For information about the data in each of these columns, see *Polyploid Clustering Columns in Tables* on page 11.

3 In the SNP Table, click the Column Chooser button and set the following columns related to polyploid clustering to appear:

Displayed Columns	Displayed Subcolumns
Cluster Distance	Cluster T Mean columns
# Clusters	Cluster T Dev columns
# EGT Clusters	Cluster R Mean columns
Cluster frequency columns (C1, C2, etc.)	Cluster R Dev columns
# Calls	
# No Calls	
Call Freq	
Poly 10%	
Poly 50%	
[ManifestFileName].bpm column	

For information about the data in each of these columns, see *Polyploid Clustering Columns in Tables* on page 11.

4 In the Samples Table, click the Column Chooser button and set the Cluster, Poly Call Rate, Poly 10%, and Poly 50% columns to appear.
For information about the data in this column, see *Polyploid Clustering Columns in Tables* on page 11.

5 Display the legend for the SNP Graph and the Sample Graph by right-clicking each graph and selecting **Show Legend**.

Perform Polyploid Clustering

Polyploid clustering is an iterative process, not a single-step operation. Illumina recommends performing initial clustering on SNPs or a small group of representative SNPs and using the data to assess and remove low performing samples. After removing low performing samples, re-cluster and evaluate the results on a subset of loci. Adjust cluster distance to determine the optimal settings for most loci in the project to arrive at the best settings to use for each project. After identifying the appropriate cluster settings, the full project can be clustered using the best settings for automated clustering.

Illumina recommends that you perform clustering analysis on high performing samples. To remove low performing samples, Illumina recommends the following steps:

1. Select all SNPs in the SNP Table.
2. Using the Analysis menu, cluster selected SNPs using a cluster number based on the known ploidy of the population being studied. For example, if dealing with a tetraploid, initially designate 5 clusters at all loci and use them for the first-pass to characterize samples.
3. Use the Scatter plot icon to in the Samples Table to view sample metrics by Poly Call rate on the X axis and Poly 50% on the Y axis.
4. Select low performing samples (low Poly Call Rate and low Poly 50%): Hold down the control key and use the mouse to create a box around samples that show as outliers in the plot.
5. Low performing samples are now selected within the Samples Table. Using the context menu, select **Exclude Selected Samples**.
6. Return to the SNP Table and recluster SNPs or a subset of SNPs for evaluation.

After poor performing samples are removed, complete following steps:

1. Configure the cluster settings based on your knowledge of the biology and diversity of the samples. For example, highly diverse populations may have more spread in the clusters and therefore require a larger cluster distance.
2. Cluster a subset of the SNPs in your project.
3. Adjust the cluster settings to achieve the best distribution of clusters for your project based on observations.
4. Cluster all the SNPs in your project using the new cluster settings. Make sure that you include the subset of SNPs you used to determine the best cluster settings for your project.
5. Review Poly scores and call frequencies in the SNP table for outliers.

Configure Cluster Settings before Clustering

1 Select **Tools** | **Clustering Options**.
The Clustering Options dialog box opens.

2 In the **Minimum Number of Points in Cluster** field, specify how many samples must be within a certain distance of each other before they can be declared a cluster. Determine the appropriate minimum number of points in a cluster for this project based on your knowledge of the diversity of the samples in the project. Usually, you set this number higher when there are many samples in the project and lower when there are fewer samples.

3 In the **Cluster Distance** field for each algorithm, specify the maximum distance that samples can be away from each other to be still considered part of a cluster.

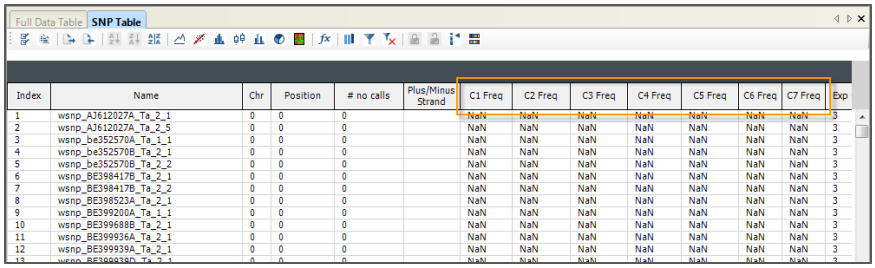
4 [Optional] In the OPTICS Algorithm Settings area, set the **Cluster Distance Increment** to enable cluster distance adjustments later through a keyboard shortcut.

TIP
By setting a Cluster Distance Increment, you can decrease the cluster distance while one or more SNPs are selected in the SNP Table. This prevents you from having to reopen the Clustering Options dialog box to adjust the cluster distance later. Type either a left square bracket ("[") or the L key (for "lower") or increase the cluster distance by typing a right square bracket ("]") or the H key (for "higher").

5 In the **Maximum Number of Clusters in SNP Table** field, enter the maximum number of clusters you expect to find in the SNPs in the current project.
It is unusual to see more than nine clusters.

NOTE
This setting controls the maximum number of Cluster Frequency columns that are displayed and limits how many total clusters can be found at a locus in the SNP table in the SNP Table as well as the maximum number of clusters shown in the Cluster Selected SNPs menu. It does not, however, prevent you from finding more clusters than defined here.

For example, if you set the maximum number of columns to 7, you can see up to 7 Cluster Frequency columns (C1 Freq, C2 Freq,... C7 Freq) in the SNP Table, but you can still designate 8 or more clusters if desired.



6 Click **OK**.

NOTE
Changing the cluster settings as described here does not change cluster positions that have already been determined for other SNPs in the project. The setting changes affect future clusters only. If you want to apply new cluster settings to SNPs that have already been clustered, select and recluster the SNPs using one of the menu or shortcut methods.

Cluster SNPs in your Project

1 Select one or more SNPs in the SNP Table or Full Data Table.
Illumina recommends performing initial clustering on a small group of representative SNPs, then clustering the full project later, after the best settings have been determined. For information on adjusting cluster settings, see *Adjust Cluster Settings* on page 14.

2 From the menu, select **Analysis | Cluster Selected SNPs As**.
The Cluster Selected SNPs submenu opens.

3 Select one of the following analysis methods:

Setting	Description
DBSCAN	Uses the selected clustering algorithm to determine the number of clusters automatically, without knowledge of the biology of the species.
OPTICS	
(Number) Clusters	Enables you to specify how many clusters GenomeStudio calculates for each SNP, based on your knowledge of the biology of the species. When clustering by this method, GenomeStudio uses the OPTICS algorithm to calculate a cluster distance that works to achieve the specified number of clusters. NOTE The number of clusters that appear in this drop-down menu is determined by the **Maximum Number of Clusters in SNP Table** setting in the Clustering Options dialog box.

TIP
You can also cluster SNPs by right-clicking the SNP Table or right-clicking the SNP Graph when a single SNP is selected in the SNP Table. Additionally, you can use shortcut keys to initiate clustering. Type a letter D to initiate DBSCAN clustering on the selected SNPs. Type a letter O to initiate OPTICS clustering. Type number 1 through 9 to cluster samples targeting the designated number of clusters.

GenomeStudio assigns the samples at each locus into clusters based on the number and relative distance of samples. The cluster data appears in the Full Data Table, SNP Table, Samples Table, and SNP Graph.

Review Cluster Data

Cluster data is reported in the Full Data Table, SNP Table, and Samples Table, as well as in the SNP Graph and Sample Graph in the Polyploid Clustering Module. The type of data reported is described in the following sections.

Polyploid Clustering Columns in Tables

Column	Description
Cluster Distance	Reported for each SNP in absolute numerical values. This number indicates the distance used for clustering. The number matches the distance value shown in the title above the SNP Graph. This value can be adjusted by changing the cluster distance settings in the Clustering Options dialog box, then reclustering the selected SNPs. NOTE If clusters at this SNP are adjusted manually, this value is NaN.
# Clusters	Reported for each SNP. This number indicates how many included clusters were calculated for the SNP. The number is represented visually on the SNP graph by distinct colored groups of samples.
# EGT Clusters	Reported for each SNP. This number indicates how many clusters were reported for the SNP from an imported polyploid cluster file (*.egtp). If no cluster data was imported, then this value is 0.
# Calls	Reported for each SNP. This number indicates how many samples GenomeStudio was able to assign into a cluster for the SNP.
# No Calls	Reported for each SNP. This number indicates how many outlying samples were not clustered for the SNP, not including excluded samples.
Call Freq	Reported for each SNP. This number indicates the ratio of samples that were assigned to clusters (as compared to no-calls). For example, a value of 0.9583 indicates that 95.83% of the samples at that locus were assigned to a cluster.
C1 Freq C2 Freq C3 Freq ...	Reported for each SNP. This number indicates the ratio of samples that were assigned to each cluster for the SNP (as compared to total samples assigned to all clusters). Outlying samples (no-calls) are not included in the calculation. For example, a value of 0.4946 in the C1 Freq column indicates that 49.46% of the called samples were part of Cluster 1.
Poly 50%	Reported for each SNP and for each sample; shows the 50th percentile of Poly Score values. This number ranges from 0 to 1. The higher the number, the closer samples are to the centroid of a cluster. For example, in the SNP Table, a Poly 50% value of 0.9974 indicates that half of the samples have a Poly Score greater than 0.9974 at that SNP.
Poly 10%	Reported for each SNP and for each sample; shows the 10th percentile of Poly Score values. This number ranges from 0 to 1. The higher the number, the closer most samples are to the centroid of a cluster. For example, in the SNP Table, a Poly 10% value of 0.9937 indicates that only 10% of the samples have a Poly Score lower than 0.9937 at that SNP.

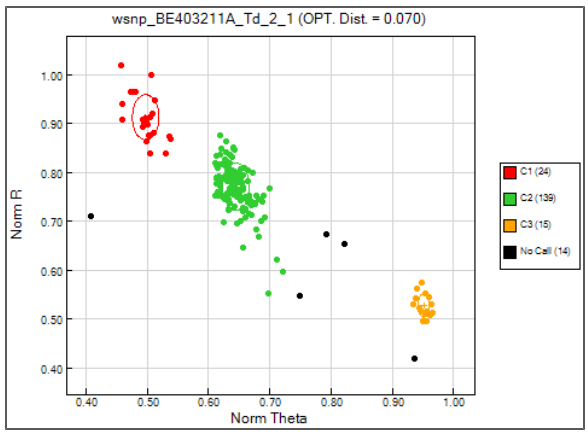
Column and Subcolumn		Description
[Sample] Columns	Cluster	Reported for each sample at each locus. These subcolumns appear only if the [Sample] columns are showing in the Full Data Table. This value indicates to which cluster the sample belongs. C1 is Cluster 1, C2 is Cluster 2, and so on. NC means that the sample is an outside of boundaries for any clusters (no call).
	Poly Score	Polyploid clustering score, reported for each sample at each locus. These subcolumns appear only if the [Sample] columns are showing in the Full Data Table. This is a relative population-based value, ranging from 0 to 1 and represents the relative population-based confidence of cluster assignment, with samples near the centroid of the cluster with higher values. The higher the number, the closer the sample is to the centroid and the more confident you can be that the sample is assigned to the appropriate cluster.

Column and Subcolumn		Description
[ManifestFileName].bpm column	C1 T Mean C2 T Mean C3 T Mean ...	Reported for each cluster at a SNP. These subcolumns appear only if the [ManifestFileName].bpm column is showing in the SNP Table. This value indicates the theta value of the center of each cluster, in normalized polar coordinates.
	C1 T Dev C2 T Dev C3 T Dev ...	Reported for each cluster at a SNP. These subcolumns appear only if the [ManifestFileName].bpm column is showing in the SNP Table. This value indicates the standard deviation in theta of each cluster, in normalized polar coordinates.
	C1 R Mean C2 R Mean C3 R Mean ...	Reported for each cluster at a SNP. These subcolumns appear only if the [ManifestFileName].bpm column is showing in the SNP Table. This value indicates the R value of the center of each cluster, in normalized polar coordinates.
	C1 R Dev C2 R Dev C3 R Dev ...	Reported for each cluster at a SNP. These subcolumns appear only if the [ManifestFileName].bpm column is showing in the SNP Table. This value indicates the standard deviation in R of each cluster, in normalized polar coordinates.

For information about the other columns in the Full Data Table, SNP Table, or Samples Table, see the *GenomeStudio Genotyping Module v1.0 User Guide*.

Polyploid Clustering Data in the SNP Graph

The SNP Graph shows cluster data for all samples at the locus currently selected in the SNP Table or Full Data Table.



- The title above the graph shows the SNP name followed in parentheses by the clustering algorithm and cluster distance used.
- Samples are colored to represent the cluster to which they belong. No-calls appear black on the graph where excluded samples appear gray.
- A plus sign (+) within each cluster indicates the centroid. The oval around the centroid is two standard deviations away from the centroid, based on the currently designated population of sample. The closer a sample is to the centroid, the higher the confidence that the sample is accurately assigned as part of that cluster.

  NOTE
  You might need to zoom in to the graph to see the plus sign and oval.

- Clusters are numbered in increasing order, based on the Norm Theta value of their centroids.

  In the example above, the centroid of cluster 1 (C1) has a Norm Theta value around 0.50; the centroid of cluster 2 (C2) has a Norm Theta value around 0.65; the centroid of cluster 3 (C3) has a Norm Theta value around 0.95.
- The name and confidence score of each sample appears when you hover over the sample on the graph.
- The legend indicates how many samples are grouped in each cluster for all loci and can be made visible or hidden via the context menu of the SNP Graph.

NOTE
Clusters cannot be edited by dragging and dropping the edges of the oval on the graph. For information on manually adjusting clusters, see *Adjust Clusters Manually* on page 16.

Adjust Cluster Settings

Modifying the cluster settings enables you to optimize the way that samples are clustered in your project, based on your knowledge of the biology and diversity of the samples.

1. Select **Tools** | **Clustering Options**.
   The Clustering Options dialog box opens.
2. In the **Minimum Number of Points in Cluster** field, specify how many samples must be within a certain distance of each other before they can be declared a cluster. Determine the appropriate minimum number of points in a cluster for this project based on your knowledge of the diversity of the samples in the project. Usually, you set this number higher when there are many samples in the project and lower when there are fewer samples.
3. In the **Cluster Distance** field for the algorithm being used, specify the maximum distance that samples can be away from each other to be still considered part of a cluster. If you have high diversity, you may expect a larger spread in point at each SNP, and you can designate a higher cluster distance value.

   TIP
   If you are using the OPTICS algorithm, setting the optional Cluster Distance Increment enables you to adjust the cluster distance with a keyboard shortcut. You can decrease the cluster distance by typing a left square bracket or increase it by typing a right square bracket while one or more SNPs are selected. When you use a keyboard shortcut, GenomeStudio automatically reclusters based on the new cluster distance.

4. Click **OK** to close the dialog box, and then recluster the desired SNPs to see the impact of your changes.
5. Repeat these steps until no further adjustments are required.

Clear All Cluster Data from a SNP or Set of SNPs

1 Select one or more SNPs in the SNP Table or Full Data Table.

2 Right-click the SNP table and select **Zero Selected SNPs**.

Adjust Clusters Manually

After you determine the best cluster settings for a project and recluster all the SNPs, some manual adjustment might still be needed to add or remove outlying samples from a cluster. Additionally, you can designate no-call samples to be part of a new cluster that the clustering algorithm did not detect.

Identify and Omit Low-Scoring Samples from a Cluster

After clustering, you can review the cluster positions at each locus. The higher the Poly Score, the closer the sample is to the centroid and the more confident you can be that the sample is correctly assigned. You can identify and exclude samples with lower confidence scores from a cluster, as follows:

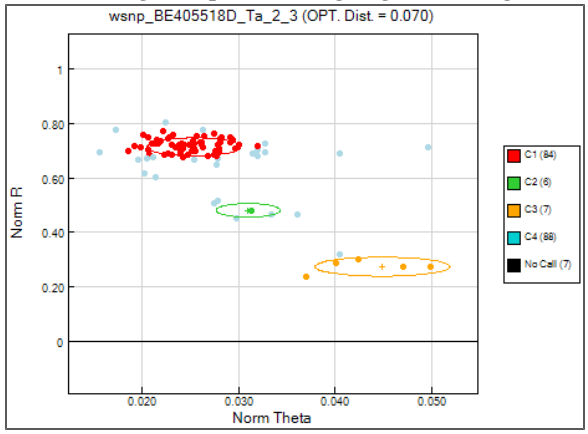
1 Select **Tools** | **Clustering Options**.
The Clustering Options dialog box opens.

2 In the **Confidence Score Limit** field, specify a value below which a sample should probably not be part of a cluster.

TIP
Start with a high confidence level (90%) and adjust the limit up or down, depending on what results you see on the SNP Graph.

3 Select the **Highlight Samples Below Limit** checkbox.

4 Click **OK** and review the SNP Graph.
Low-scoring samples are highlighted in light blue on the graph.



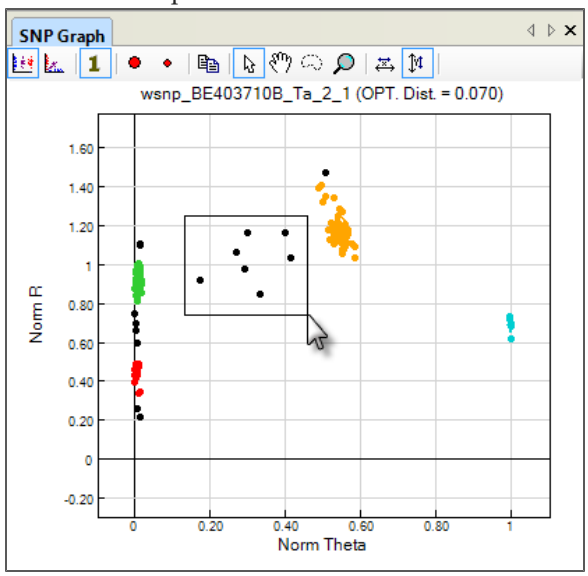
5 Repeat steps 2 through 4 until all questionable samples are highlighted.

6 [Optional] Right-click the graph and select **Uncluster Samples below Score Limit**.
The highlighted samples are changed to no-calls and appear black on the graph. Cluster data are removed and the clusters are recalculated accordingly.

NOTE
When the clusters are recalculated, new samples might fall below the confidence score limit.

Add or Remove a Sample from a Cluster

1 On the SNP Graph, use your cursor to draw a box around one or more samples to select the samples.



The selected samples are highlighted in yellow.

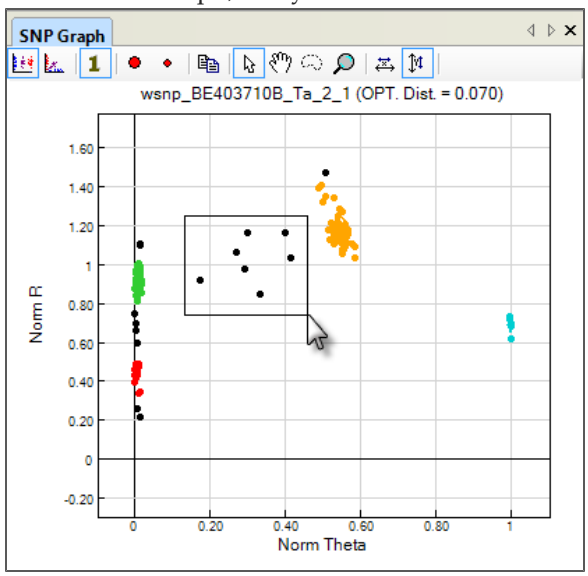
2 Right-click the SNP Graph, and select **Set Cluster for Selected Samples** from the menu; then, select one of the following:

- **C1**, **C2**, **C3**, ... adds the selected samples to the cluster you select.
- **NC** sets the cluster assignment to No Call, removing the selected samples from a cluster.

The affected clusters are recalculated when samples are added or removed from a cluster.

Create a Cluster Manually

1 On the SNP Graph, use your cursor to draw a box around one or more samples.



The selected samples are highlighted in yellow.

2 Right-click the SNP Graph, and select **Create Cluster from Selected Samples**.

Exclude Samples from the Clustering Algorithm

1 In the Samples Table or on the SNP Graph, select one or more samples.

2 Right-click the SNP Graph, and select **Exclude Selected Samples**.
Excluded samples are shown on the SNP Graph in gray and in the Samples Table in gray italics.

3 Recluster the SNPs in the project.
Cluster data are removed and the clusters are recalculated accordingly.

NOTE
If you want to re-include samples that have been excluded, select the samples you want to include, right-click and select **Include Selected Samples**. From the SNP Graph, you can also include all excluded samples without first having to select the samples. After including the samples, recluster the SNPs in the project.

Import Cluster Positions

To save time when calling cluster positions for SNPs or to apply information from one population of samples to another, import the cluster positions from a polyploid cluster file saved previously for the same gene pool.

1 Select **File** | **Import Cluster Positions**.
The Clustering Options dialog box opens.

2 Click **Browse** and navigate to the desired polyploid cluster file (*.egtp).

NOTE
The *.egtp format for polyploid projects is different than the *.egt format for diploid projects. The cluster files are not interchangeable.

3 In the **Cluster Distance Limit** field, specify the maximum distance (in standard deviations from the centroid) that samples can be away from the centroid to be still considered part of a cluster.
Illumina recommends a setting of 2 standard deviations from the centroid and adjusting as necessary based on how the data clusters.

4 Select or clear the **Selected SNPs Only** checkbox, as desired.
If the checkbox is selected, GenomeStudio clusters only for the SNPs currently selected in the SNP Table or Full Data Table if there are clusters available for the SNPs in the imported *.egtp file.

5 Click **OK**.

As SNP cluster positions are imported from the cluster file, GenomeStudio automatically clusters the equivalent SNPs in your project. Only SNPs that have cluster positions reported in the cluster file are clustered.

Review Imported Cluster Data

Imported cluster data are reported in the Full Data Table, SNP Table, and SNP Graph in the Polyploid Clustering Module. The type of data reported is described in the following sections.

Imported Cluster Data in Tables

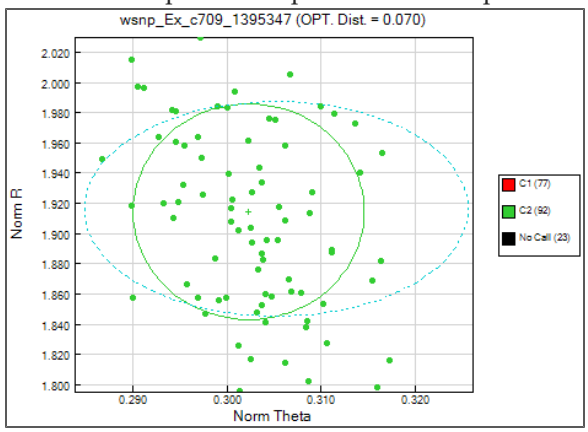
The number of clusters reported for a SNP in the imported cluster file is shown in the **# EGT Clusters** column in the Full Data Table and SNP Table. If the value is 0, it means that cluster data for that SNP were not reported in the imported cluster file.

Imported Cluster Data in the SNP Graph

The SNP Graph can be configured to overlay ovals representing the clusters reported in the imported file. This enables you to see how the cluster positions compare from project to project. To overlay cluster positions, do the following:

1 Right-click the SNP Graph and select **Show Imported Clusters Ovals**.

2 Zoom in on the SNP Graph to see the current cluster position (solid line) compared to the cluster position reported in the imported cluster file (dotted line).



Export Data

Cluster positions, as well as any information displayed in the SNP Table, Full Data Table, or Samples Table can be exported from the Polyploid Clustering Module for downstream analysis and genotype assignment.

Export Data from Tables

1. In the Full Data Table, SNP Table, or Samples Table, click the Column Chooser button and set the columns you want to export to show. Hide any columns of data that you do not want to export.
2. From the table toolbar, click the Export Displayed Data to a File button.
3. From the Save As Type drop-down menu, select **Tab Delimited (*.txt)** or **Data File (*.csv)**, as desired.
4. Specify a name for the exported file and navigate to the folder where you want to save the file.
5. Click **Save**.
6. Data is ordered in a matrix format with all sub-columns converted to columns with a single row header at the top of the file.

Export Cluster Positions

1. [Optional] Select one or more SNPs in the SNP Table or Full Data Table.
2. From the **File** menu, select one of the following export options:

Setting		Description
Export Cluster Positions	**For Selected SNPs**	Creates a cluster file containing the cluster positions for the SNPs currently selected in the SNP Table or Full Data Table.
	For All SNPs	Creates a cluster file containing the cluster positions for all the SNPs in the project.
Export and Add Cluster Positions	**For Selected SNPs**	Adds the cluster positions for the SNPs currently selected in the SNP Table or Full Data Table to an existing cluster file. If the cluster file already contained cluster position data for a particular SNP, the cluster positions are overwritten.
	For All SNPs	Adds the cluster positions for all the SNPs in the project to an existing cluster file. If the cluster file already contained cluster position data for a particular SNP, the cluster positions are overwritten.

3 Navigate to the folder where you want to save the cluster file (or to the cluster file that you want to update) and save the file with a *.egtp extension.

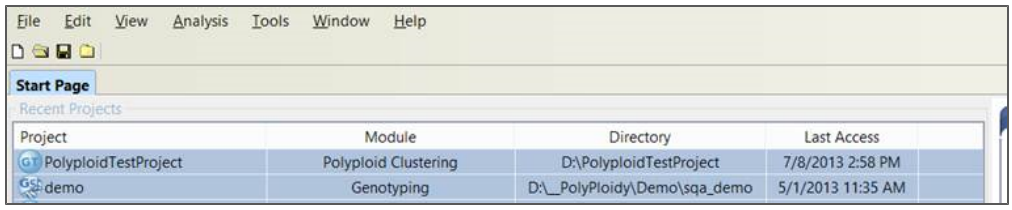
NOTE

Make sure that the file extension ends with a letter p. The letter p indicates that the cluster positions are for polyploid species. If the file does not include the letter p at the end, it is interpreted as a diploid cluster file.

Open a Polyploid Project in the Genotyping Module

On the GenomeStudio start page, each project listed has an associated default module. When you click a project in the list, the default module opens the project.

Figure 1 The GenomeStudio Start Page Displaying a Project in the Polyploid Clustering module



The polyploid module becomes default when you open a genotyping project in polyploid module and then save the project. To open the project in the genotyping module, set the genotyping module as default for the project as follows:

1 Select **Main menu | Tools | Set Default Project Module As | Genotyping**.

2 Close the project in the polyploid module.

3 Open the project in the genotyping module.

Similarly, you can assign a polyploid module to the project without saving the project.

Notes

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 1 Illumina General Contact Information

Illumina Website	www.illumina.com
Email	techsupport@illumina.com

Table 2 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Italy	800.874909
Austria	0800.296575	Netherlands	0800.0223859
Belgium	0800.81102	Norway	800.16836
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000

MSDSs

Material safety data sheets (MSDSs) are available on the Illumina website at www.illumina.com/msds.

Product Documentation

Product documentation in PDF is available for download from the Illumina website. Go to www.illumina.com/support, select a product, then click **Documentation & Literature**.

Illumina
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com