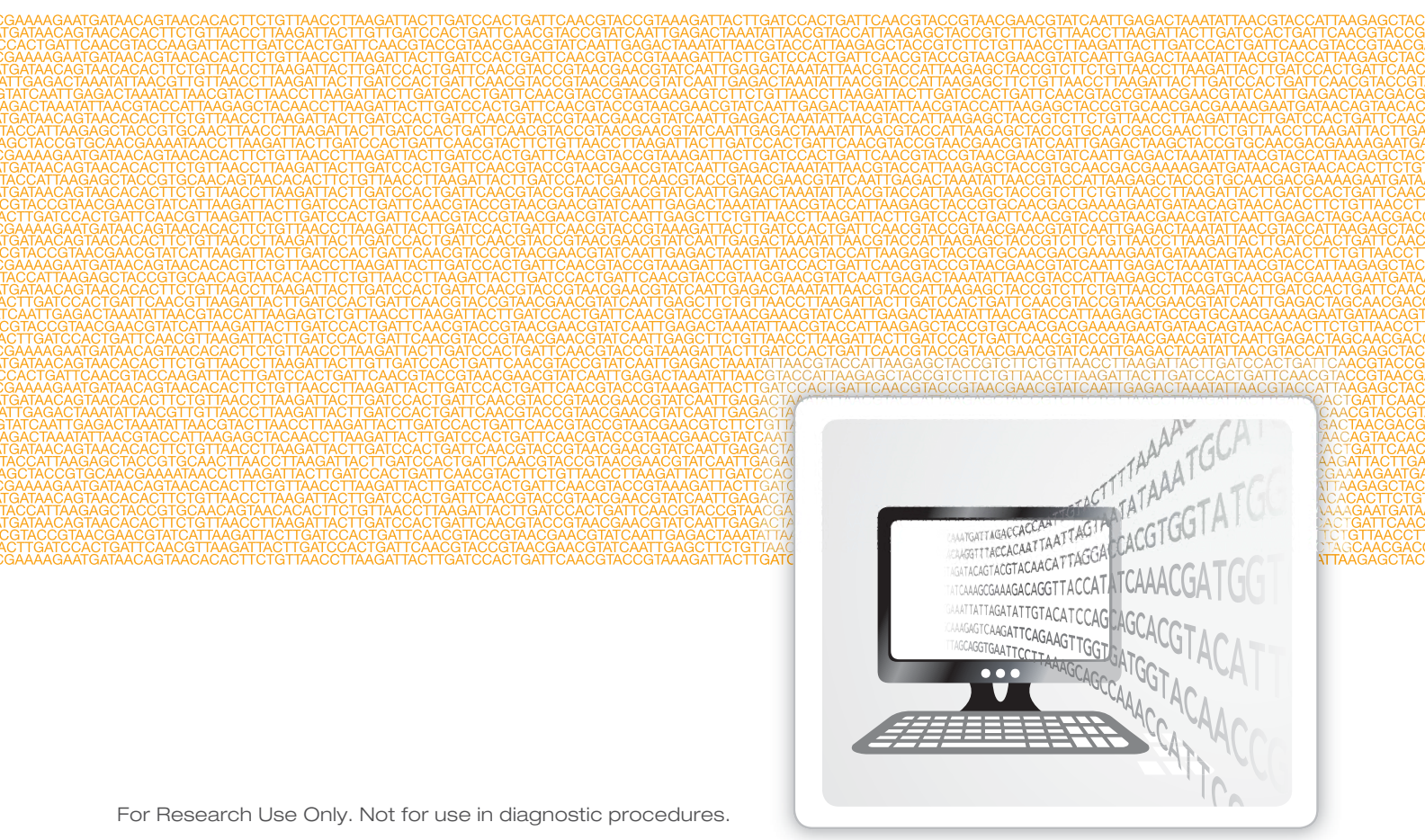
illumina®

GenomeStudio® Polyploid Genotyping Module v2.0 Software Guide



For Research Use Only. Not for use in diagnostic procedures.

ILLUMINA PROPRIETARY

Document # 10000001247 v00
September 2016

This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY.

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2016 Illumina, Inc. All rights reserved.

Illumina, GenomeStudio, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Table of Contents

Overview

Chapter 1

Introduction

The Illumina® GenomeStudio™ Polyploid Genotyping Module Software v2.0 analyzes genotyping intensity from a polyploid organism. Users can select the following algorithms for clustering samples:

- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- PolyGentrain

In addition to the automated clustering options, the Polyploid Genotyping Module provides robust capabilities for manually manipulating cluster definitions and genotypes. The module outputs reports and cluster files for downstream analysis.

Before installing the Polyploid Genotyping Module, you must have access to the Internet and a MyIllumina account. If you do not have access to the Internet, see *Technical Assistance* on page 59.

Features

You can perform the following tasks with the Polyploid Genotyping Module:

- Cluster autopolyploid and allopolyploid organisms with PolyGentrain and DBSCAN algorithms.
- Set genotyping and clustering intensity thresholds.
- Import and export cluster positions.
- Manually adjust cluster positions.

Install the GenomeStudio Software

1. Navigate to the GenomeStudio support page, and then select **Downloads**.
2. Select **GenomeStudio Software 2.0**, and then select **GenomeStudio Genotyping Module 2.0 Installer.**
3. Accept the software terms and license agreement, and then click **Install**.

Start the GenomeStudio Software

1. From your desktop, double-click the GenomeStudio  icon.
2. In the Platform Login Form dialog box, enter your email and password from your MyIllumina account.
3. Click **Log In**.

Getting Started

Chapter 2

Introduction

You can create a Polyploid Genotyping Module project from one of the following sources of sample data:

- A sample sheet
- Directories containing intensity data files

Create a Project From a Sample Sheet

1. In the new project pane, click **Polyploid Genotyping**, and then click **Next**.
2. In the Create field, enter the project name, and then click **Next**.
3. Click **Use sample sheet to load sample intensities**, and then click **Next**.
4. In the Sample Sheet field, browse to your sample sheet (*.csv) file.
5. In the Data Repository field, browse to the directory that contains your raw intensity (*.idat) files.
6. In the Manifest Repository field, browse to the directory that contains your SNP manifest (*.bpm) file.
   The software uses this directory to locate the names of the SNP manifests.
7. Click **Next**.
8. [Optional] Click the **Import cluster positions from a cluster file** checkbox. See *Import Cluster Positions* on page 9.
   a. Select **Browse**, highlight the cluster file, and select **Open**.
9. Click **Finish**.
   Polyploid Genotyping Module loads the files from the sample sheet and displays the data.

Create a Project From Directories

1. In the new project pane, click **Polyploid Genotyping**, and then click **Next**.
2. In the Create field, enter the project name, and then click **Next**.
3. Click **Load sample intensities by selecting directories with intensity files**, and then click **Next**.
4. At the Manifest Repository field, browse to the file that contains your SNP manifest (*.bpm) file.
5. At the Data Repository field, browse to the directory that contains your raw intensity (*.idat) files.
6. Select one or more directories, and click **Add**, and then click **Next**.
7. [Optional] Click the **Import cluster positions from a cluster file** checkbox. See *Import Cluster Positions* on page 9.
   a. Select **Browse**, highlight the cluster file, and select **Open**.
8. [Optional] In the Project Settings area, select the following options:
   - **Cluster SNPs**
   - **Calculate Sample and SNP Statistics**
   - **Calculate Reproducibility**
   - **Gen Call Threshold**
9. [Optional] In the Ploidy area, select the ploidy level and 1 of the following ploidy types:
   - **Autopolyploid**—Software uses the automatic clustering algorithm to fit the number of clusters that are equal to ploidy + 1.
   - **Allopolyploid**—Software uses the automatic clustering algorithm to fit 3 clusters, with the ploidy value controlling the separation between clusters.
10. [Optional] At the **Optional Script File** field, browse to the file you want to use.
11. Click **Finish**.

Import Cluster Positions

You can import a cluster (*.egtp) file when creating a project in the Polyploid Genotyping Module.

1. In the GenomeStudio Project Wizard, click the **Import cluster positions from a cluster file** checkbox.
2. At the Cluster File field, browse to the cluster file you want to use.
3. [Optional] In the Project Creation Actions settings, select from the following options:
   - **Cluster SNPs**
   - **Calculate Sample and SNP Statistics**
   - **Calculate Heritability**
4. At the GenCall Threshold field, specify the score cutoff for this project.
   We recommend the 0.15 GenCall score cutoff for Infinium products.
5. Click **Finish**.

Viewing Data

Chapter 3

Introduction

The Polyploid Genotyping Module has many views, such as graphs, tables, and project information. The views are detachable docking windows. Click and drag the windows to where you want to dock them on the main window. The views show the following data:

Table 1 Views in Polyploid Genotyping Module

View	Description
Sample Graph	All SNPs for the selected sample.
SNP Graph	All samples for the selected SNP in the SNP table and Full Data table.
SNP Graph Alt	All samples for the selected SNP in the SNP table and Full Data table. (Use to compare 2 different views of the samples.)
Full Data Table	Data for all samples.
SNP Table	Statistics for each SNP.
Samples Table	Data for each sample.
Errors Table	Discrepancies between replicate or related samples
Log Window	A list of the activities that are created by the software.
Project Window	Manifests and sample barcodes loaded in the project.

Full Data Table

The Full Data table contains data for all the samples.

Table 2 Full Data Table Columns

Column Heading	Description
Index	The SNP index row.
Name	The SNP name.
Address	The bead type identifier.
Chr	The SNP chromosome.
Manifest	The manifest name of the SNP.
Position	The chromosomal position of the SNP.
GenTrain Score	The score for a SNP from the GenTrain clustering algorithm.
FRAC A	The fraction of the A nucleotide in the top genomic sequence.
FRAC C	The fraction of the C nucleotide in the top genomic sequence.
FRAC G	The fraction of the G nucleotide in the top genomic sequence.
FRAC T	The fraction of the T nucleotide in the top genomic sequence.

Table 3 Full Data Table Subcolumns

Column Heading	Description
Genotype	The genotype for the sample.
Score	The SNP call score for the sample.
Theta	The normalized theta value for the sample.
R	The normalized R value for the sample.
X Raw	The raw intensity of the A allele.
Y Raw	The raw intensity of the B allele.
X	The normalized intensity of the A allele.
Y	The normalized intensity of the B allele.
Log R Ratio	$\text{Log2}(R_{observed}/R_{expected})$, where $R_{expected}$ is interpolated from the observed allelic ratio
Top Alleles	The Illumina-designated top strand genotype.
Import Calls	The genotype calls for an imported sample.

Column Heading	Description
Concordance	The numeric correlation between the top allele call for a SNP in the project and the imported allele call of a SNP from another project.
Plus/Minus Alleles	The nucleotide genotype for the sample on the plus strand.

SNP Table

The SNP table shows the statistics for each SNP.

NOTE
The frequency, mean, and deviation value for genotypes depend on the user-defined ploidy type and level.

Table 4 SNP Table Columns

Column Heading	Description
Index	The SNP index row.
Name	The SNP name.
Chr	The SNP chromosome.
Position	The SNP chromosomal position.
SNP	The nucleotide substitution for the SNP on the Illumina ILMN strand.
ILMN Strand	The design strand designation.
AA Freq	The frequency of AA calls.
AB Freq	The frequency of AB calls.
BB Freq	The frequency of BB calls.
Call Freq	The call frequency.
Minor Freq	The minor allele frequency.
Aux	The user-defined auxiliary value for the SNP.
Rep Errors	The number of reproducibility errors for the SNP as allele comparisons between replicates.
Customer Strand	The customer strand designation.
Comment	The user-defined comment.
Concordance	The measurement between 2 genotypes from the same SNP locus.

Table 5 SNP Table Subcolumns

Column Heading	Description
Address	The SNP index row.
Address 2	The unidentified bead type for the second allele. It is only used for Infinium I.
Intensity Threshold	The intensity threshold value.
Manifest	The manifest name of the SNP.

Column Heading	Description
Edited	Whether the SNP was edited after clustering positions were identified. **1**—Edited **0**—Unedited
Cluster Sep	The cluster separation measurement for the SNP, between 0 and 1.
Norm ID	The normalization ID for the SNP.
AA T Mean	The theta value of the AA cluster center in normalized polar coordinates.
AA T Dev	The standard deviation in theta of the AA cluster in normalized polar coordinates.
AB T Mean	The theta value of the AB cluster center in normalized polar coordinates.
AB T Dev	The standard deviation in theta of the AB cluster in normalized polar coordinates.
BB T Mean	The theta value of the BB cluster center in normalized polar coordinates.
BB T Dev	The standard deviation in theta of the BB cluster in normalized polar coordinates.
AA R Mean	The R value of the center of the AA cluster in normalized polar coordinates.
AA R Dev	The standard deviation in R of the AA cluster in normalized polar coordinates.
AB R Mean	The R value of the AB cluster center in normalized polar coordinates.
AB R Dev	The standard deviation in R of the AB cluster in normalized polar coordinates.
BB R Mean	The R value of the BB cluster center in normalized polar coordinates.
BB R Dev	The standard deviation in R of the BB cluster in normalized polar coordinates.

Samples Table

The Samples Table shows the statistics for each sample.

Table 6 Samples Table Columns

Column Heading	Description
Index	Sample index row.
Sample ID	Sample identifier.
Call Rate	Percentage of SNPs that have a GenCall score greater than the specified threshold.
Gender	User-specified gender for the sample.
p05 Grn	5th percentile of B allele intensity.
p50 Grn	50th percentile of B allele intensity.
p95 Grn	95th percentile of B allele intensity.
p05 Red	5th percentile of A allele intensity.
p50 Red	50th percentile of A allele intensity.
p95 Red	95th percentile of A allele intensity.
p10 GC	10th percentile GenCall score for all SNPs.
p50 GC	50th percentile GenCall score for all SNPs.
Rep Error Rate	Reproducibility error rate is calculated as 1 - sqrt(1 - errors/max_possible_errors). Errors and max_possible_errors do not include genotype calls that fall below the no-call threshold.
Call Rate	Percentage of SNPs that have a GenCall score greater than the specified threshold.
Aux	Arbitrary number you can use to differentiate and sort samples. Right-click in the Samples Table to set this value.
Subset	Grouping of samples into a subset.
Array Info	Position on the slide for this sample in terms of the sentrix ID and sentrix position.
Genotype	Genotype for this sample for the SNP currently selected in the SNP Table.
Score	GenCall score for this sample for the SNP currently selected in the SNP Table.
Sample Name	Sample name.
Sample Group	User-entered sample group.
Sample Plate	Sample plate identifier.

Column Heading	Description
Sample Well	Well within the sample plate.
Gender Est	Estimated gender.
Requeue Status	Sample requeue status in LIMS. Blank status indicates that the sample does not need requeuing.
Concordance	Concordance across all SNPs for this sample. The value is populated when alleles calls are imported for the same sample from another project.
Ethnicity	Ethnicity of the individual from the acquired sample.
Age	Age of the individual from the acquired sample.
Weight	Weight in kg of the individual from the acquired sample.
Height	Height in meters of the individual from the acquired sample.
Blood Pressure Systolic	Systolic blood pressure of the individual from the acquired sample.
Blood Pressure Diastolic	Diastolic blood pressure of the individual from the acquired sample.
Blood Type	Blood type of the individual from from the acquired sample.
Phenotype Pos 1	Positive phenotype 1 of the individual from the acquired sample.
Phenotype Pos 2	Positive phenotype 2 of the individual from the acquired sample.
Phenotype Pos 3	Positive phenotype 3 of the individual from the acquired sample.
Phenotype Neg 1	Negative phenotype 1 of the individual from the acquired sample.
Phenotype Neg 2	Negative phenotype 2 of the individual from the acquired sample.
Phenotype Neg 3	Negative phenotype 3 of the individual from the acquired sample.
Comment	User-entered comments.
Tissue Source	Tissue source of the individual from which this sample was acquired.
Calls	Number of loci called.
No Calls	Number of loci not called.
Excluded	Whether the sample is excluded. **1**—Excluded **0**—Included

Table 7 Samples Table Per-Manifest Subcolumns

Column Heading	Description
Sentrix ID	Barcode number of the Universal Array Product to which this sample was hybridized.
Sentrix Position	Section on the product.
Imaging Date	Imaging date.
Scanner ID	Name of the scanner.
PMT Green	Green PMT setting of the scanner.
PMT Red	Red PMT setting of the scanner.
Software Version	Version of the iScan software that scanned the sample.
User	User name of the individual that scanned the sample.
p05 Grn	5th percentile of B allele intensity.
p50 Grn	50th percentile of B allele intensity.
p95 Grn	95th percentile of B allele intensity.
p05 Red	5th percentile of A allele intensity.
p50 Red	50th percentile of A allele intensity.
p95 Red	95th percentile of A allele intensity.
p10 GC	10th percentile GenCall score over all SNPs.
p50 GC	50th percentile GenCall score over all SNPs.
Call Rate	Percentage of SNPs that have a GenCall score greater than the specified threshold.

Errors Table

The Errors table shows reproducibility errors or parent-child heritability errors.

Table 8 Errors Table Columns

Column Heading	Description
Error Index	The error index row.
Error Type	Rep—Reproducibility.
Replicate 1 Index	The sample index of the replicate 1 sample involved in the error.
Replicate 1	The sample ID of the replicate 1 sample involved in the error.
Replicate 1 GType	The genotype of the replicate 1 sample.
Replicate 2 Index	The sample index of the replicate 2 sample involved in the error.
Replicate 2	The sample ID of the replicate 2 sample involved in the error.
Replicate 2 GType	The genotype of the replicate 2 sample.
SNP Index	The SNP index number where the error occurred.
SNP Name	The SNP name where the error occurred.

Editing Data

Chapter 4

Introduction

You can edit the data in the tables and graphs with the following tools:

Tools	Description
Main Window Menus	The main window contains the following menus: • File • Edit • Analysis • Tools • Window • Help
SNP Graph Toolbar	Edits SNPs in the SNP, SNP Alt, and Sample graphs.
Data Table Toolbar	Selects, copies, filters, and sorts data in tables.
Context Menus	Right-click in graphs and tables to select additional functions.

Main Window Menus

File Menu

Table 9 File Menu Functions

Function	Description
New Project	Creates a new project.
Open Project	Opens a previously saved project.
Save Project	Saves all information in this project.
Save Project Copy As	Saves a copy of the current project.
Close Project	Closes the current project.
Import Cluster Positions	Imports cluster positions from an *.egtp file.
Export Cluster Positions	Exports selected SNPs or all SNPs to an *.egtp file.
Export Manifest	Exports a manifest (*.csv) file.
Page Setup	Sets up page properties.
Print Preview	Opens the Print Preview window.
Print	Sets up printing options.
Recent Project	Selects a recent project to open.
Exit	Closes GenomeStudio Polyploid Genotyping Module.

Edit Menu

Table 10 Edit Menu Functions

Function	Description
Cut	Cuts the selected content.
Copy	Copies the selected content.
Paste	Pastes the selected content.
Select All	Selects all data in the table.

View Menu

Table 11 View Menu Functions

Function	Description
Save Current View	Saves the current window configuration.
Restore Default View	Restores the default window configuration.
Save Custom View	Saves a custom window configuration.
Load Custom View	Loads a saved window configuration.
Log	Shows or hides the Log window.
Project	Shows or hides the Project window.

Analysis Menu

Table 12 Analysis Menu Functions

Function	Description
Cluster Selected SNPs As	Clusters selected SNPs based on the samples in a project and determine the genotype score for each locus. The clustering function overrides cluster files that are clustered at project creation. Choose 1 of the following options: • PolyGentrain • DBSCAN
Edit Replicates	Edits, includes, or excludes replicates for a sample.
Update Reproducibility Errors	Updates reproducibility information in columns and reports.
Reports	Generates the following reports: • Reproducibility Report • Final Report • DNA Report • Locus Summary Report • Locus x DNA Report
View Controls Dashboard	Allows you to view intensity data associated with various control probes.
Import Allele Calls	Imports allele calls.
Export Allele Calls	Exports allele calls.
Remove Imported Allele Calls	Removes imported allele calls from the project.

Tools Menu

Table 13 Tools Menu Functions

Function	Description
Options	• Project—Changes the project settings. • GenomeStudio—Modifies settings and attributes. • Module—Modifies module properties for either Genotyping or Polyploid Genotyping.
New Data Track Table	Adds a user-defined name for the new data track table.
Clustering Options	Shows the clustering options, including the PolyGentrain and DBSCAN algorithm options.

Window Menu

Click the checkbox to show or to hide the following windows:

- SNP Graph
- Heat Map
- SNP Graph Alt
- Samples Table
- Full Data Table
- SNP Table
- Errors Table
- Sample Graph

Help Menu

Table 14 Help Menu Functions

Function	Description
About GenomeStudio	Shows the GenomeStudio version and the software copyright information.
Help	Opens the GenomeStudio support page.

SNP Graph Toolbar

Use the buttons in the graph toolbar to change the way your data appears in the graphs. These tools do not modify the data in your project.

Icon	Tool Name	Description
	Polar Coordinates	Displays the plot in polar coordinates. X-axis represents normalized theta. Y-axis represents the distance of the point to the origin.
	Cartesian Coordinates	Displays the plot in Cartesian coordinates. X-axis represents the intensity of the A allele. Y-axis represents the intensity of the B allele.
	Plot Normalization Values	Views samples in raw format. This option toggles between raw and normalized values.
	Make Dots Larger	Enlarges the dot size.
	Make Dots Smaller	Reduces the dot size.
	Copy Plot to Clipboard	Copies the plot to the clipboard.
	Default Mode	Draws a rectangle area that includes the samples you want to analyze.
	Pan Mode	Drags the graph in the direction you want to analyze.
	Lasso Mode	Draws a region that includes the samples you want to analyze.
	Zoom Mode	Zooms in or out by scrolling your mouse wheel up and down.
	Auto Scale X-Axis	Scales SNPs to the X-axis.
	Auto Scale Y-Axis	Scales SNPs to the Y-axis.

Data Table Toolbar

Use the buttons in the table toolbar to change the appearance of your data in the statistics tables. These tools do not modify the data in your project.

Icon	Name	Description
	Calculate	Updates all statistics in the table. The button is active when statistics are out of sync.
	Select all	Highlights all the rows in the data table.
	Copy selected rows to the clipboard	Copies selected rows or columns to the clipboard.
	Export displayed data to a file	Exports selected rows or columns to a file.
	Import columns into the table	Imports sample or loci data from a tab-delimited file.
	Sort column (Ascending)	Sorts the data table in ascending order by the selected column.
	Sort column (Descending)	Sorts the data table in descending order by the selected column.
	Sort by multiple columns	Sorts the data table by more than one column.
	Line plot	Displays a line plot of the sample or loci statistics.
	Scatter plot	Displays a scatter plot of the sample or loci statistics.
	Histogram plot	Displays a histogram of the sample or loci statistics.
	Box plot	Displays a box plot of the sample or loci statistics.
	Frequency plot	Displays a frequency plot of the sample or loci statistics.
	Pie chart	Displays a pie chart of the sample or loci statistics.
	Calculate new column	Calculates data for a new column.
	Column chooser	Adds, removes, or moves columns in the data tables.
	Filter rows	Hides or shows rows of data based on the criteria you specify. This feature does not remove samples or loci from your data set. It is for viewing information in the data table only.
	Clear filter	Removes filter settings.
	Lock selected columns	Locks the selected column so it cannot scroll with other columns.
	Unlock selected columns	Unlocks the selected column so that it scrolls with other columns.

Context Menus

Right-click in the graph or table to view the context menus.

Table 15 Graph Window Context Menu

Function	Description
Cluster Selected SNPs As	Clusters the selected SNPs using the PolyGentrain or DBSCAN algorithm.
Create Cluster from Selected Samples	Creates a cluster from selected samples.
Delete Cluster	Deletes the cluster.
Set Cluster Colors	Changes cluster colors.
Associate Cluster with Genotype	Associates the cluster with the AA, AB, BB genotype. The output depends on the ploidy level and type.
Configure Mark	Marks selected samples in a color you choose.
Mark Selected Points - <Add New>	Creates a new mark.
Clear Marks - <All>	Clears all marks.
Exclude Selected Samples	Excludes selected samples from the genoplot.
Include Selected Samples	Includes selected samples in the genoplot.
Include All Samples	Includes all samples in the genoplot.
Show Legend	Displays the genoplot marks legend.
Show Excluded Samples	Shows excluded samples.
Auto Scale Axes	Scales the axes.
Show Only Selected	Shows the selected samples.
Copy Image to File as	Copies the image as one of the following file types: • Bitmap • GPEG • PNG • GIFF • TIFF

Table 16 Full Data Table Context Menu

Function	Description
Cluster Selected SNPs As	Clusters the selected SNPs using the PolyGentrain or DBSCAN algorithm.
Zero Selected SNPs	Zeroes a selected SNP.
Set Genotyping Intensity Threshold for Selected SNPs	Enters the threshold value.
Configure Marks	Configures marks.
Mark Selected Rows | <Add New>	Creates a mark and mark selected rows.
Select Marked Rows	Selects marked rows.
Clear Marks | <All>	Clears all marks.

Table 17 SNP Table Context Menu

Function	Description
Cluster Selected SNPs As	Clusters the selected SNPs using the PolyGentrain or DBSCAN algorithm.
Zero Selected SNP	Zeroes a selected SNP.
Set Genotyping Intensity Threshold for Selected SNP	Enters the threshold value.
Set Aux Value	Sets the aux value of a SNP.
Update Selected SNP Statistics	Updates selected SNP statistics.
SNP Properties	Edits SNP properties.
Show Only Selected Rows	Shows only selected rows in the SNP Table.
Configure Marks	Configures marks.
Mark Selected Rows | <Add New>	Creates a mark and marks selected rows.
Select Marked Rows	Selects marked rows.
Clear Marks | <All>	Clears all marks.

Table 18 Samples Table Context Menu

Function	Description
Exclude Selected Sample	Excludes the selected sample.
Include Selected Sample	Includes the selected sample.
Recalculate Statistics for Selected Sample	Recalculates the statistics for the selected samples.
Recalculate Statistics for All Samples	Recalculates the statistics for all samples.
Display Image	Displays the image if you have the *.idat file, the *.locs file, the *.xml file, and the *.jpg or *.tif image file for the sample or sample section.
Set Aux Value	Filters loci for your need.
Sample Properties	Changes values for sample data.
Show Only Selected Rows	Displays your preferred SNPs.
Configure Marks	Configures marks.
Mark Selected Rows | <Add New>	Creates a mark and marks selected rows.
Select Marked Rows	Selects marked rows.
Clear Marks | <All>	Clears all marks.

Table 19 Error Table Context Menu

Function	Description
Show Only Selected Rows	Configures the Samples table to show only selected rows.
Edit Replicates	Edits replicates.
Configure Marks	Allows you to configure marks.
Mark Selected Rows | <Add New>	Configures marks.
Select Marked Rows	Selects marked rows.
Clear Marks | <All>	Clears all marks.

Analyzing Data

Chapter 5

Select Samples in the SNP Graph

For the SNP Graph, selected rows in the Samples Table correspond with the samples in the SNP Graph.

1 In the main window, select the SNP Graph tab.

2 In the Default Mode, click-and-drag on the graph to draw a rectangle.
When you release the button, all points in the rectangle are selected.

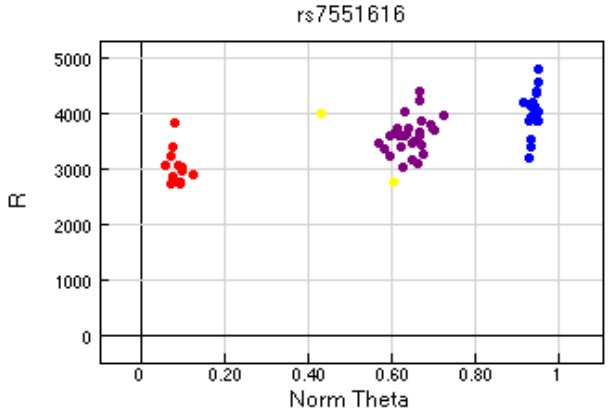
3 To add additional samples without losing your original selection, press and hold the **Ctrl** key and select additional samples.

4 [Optional] To change to (Pan Mode), position the cursor over an empty region of the plot (not on a cluster), then press and hold the **Shift** key.

5 [Optional] To change to (Lasso Mode), press and hold the **Z** key.

Selected samples are highlighted in yellow by default.

Figure 1 SNP Graph



Exclude Samples

You can exclude poor quality samples from clustering.

1 In the main window, right-click in the SNP Graph.

2 Select **Exclude Selected Samples**.

3 Click **Yes**.

Plot Excluded Samples

After excluding one or more samples from your sample group, you can plot the excluded samples in the genoplot.

1 In the main window, select **Tools** | **Options** | **Project**.

2 In Options, click the **Plot excluded samples** checkbox.

3 Click **OK**.

Document # 1000000012407 v00

Display Marked Samples

You can customize the color of the selected samples. Marked samples overwrite the default genotyping colors.

1. In the main window, right-click in the SNP Graph.
2. Select **Configure Marks**.
3. Click **Add** to create a mark.
4. Enter a name for your mark.
5. Select a color from the drop-down menu.
6. Click **OK**.

View Legend in Graph

1. In the main window, right-click in the graph.
2. Select **Show Legend**.
3. Click **Yes**.

Customize the SNP Table

Use the Column Chooser to select the columns you want to display in the SNP table and arrange the columns in any order.

1. In the SNP table, click the Column Chooser  tool.
2. Click the columns you want to display, and then click **Show**.
3. Click the columns you want to hide, and then click **Hide**.
4. [Optional] Select and drag a column back and forth between the Displayed and Hidden Columns sections to customize your table.
5. [Optional] Select a column and drag the column header up or down in the order that you want the columns to appear.
6. Click **OK**.

View Samples in the Controls Dashboard

View your samples in the Controls dashboard.

1. In the main window, select **Analysis** | **View Controls Dashboard**. The Controls dashboard does not show excluded samples.
2. Click **File**, and select 1 of the following options:
   - Export Data—Lets you save the data in a (*.csv) file.
   - Page Setup
   - Print Preview
   - Print
   - Close

Edit the Sample Properties

1. In the Samples table, select the samples you want to analyze.
2. Right-click anywhere on the selected samples.
3. Select **Sample Properties**.
4. Enter the information in the appropriate column.
5. Click **OK**.

Change the Genotyping Intensity Threshold

1. In the Full Data or SNP table, select the SNPs you want to analyze.
2. Right-click anywhere on the selected SNPs.
3. Select **Set Genotyping Intensity Threshold for Selected SNPs**.
4. In the Value field, enter the value you want to use.
5. Click **OK**.

Change the No-Call Threshold

In the SNP graph, samples are colored by their genotype call. Samples that are in black and are located in the lighter shaded regions indicate that they fall below the user-defined threshold value. As a result, these samples are assigned as no-call.

1 In the main window, select **Tools** | **Options** | **Project**.

2 In the No-call Threshold field, enter your value.
We recommend the 0.15 GenCall cutoff score for Infinium products.

3 Click **OK**.

View Concordance Calculations

The Polyploid Genotyping Module can show the concordance calculations when you import allele calls for the same sample from another project.

1. In the Samples table or in the Full Data table, click the Column Chooser  tool.
2. Select **Concordance**, and then click **Show**.
3. Click **OK**.

Export Allele Calls

You can compare the allele calls in your project to the allele calls in another project. When you export allele calls and import them into another project, the sample names in the project must be the same. If the sample names do not match, the allele calls cannot be compared.

1 In the main window, select **Analysis** | **Export Allele Calls**.

2 Browse to the location where you want to save the allele calls.

3 Click **OK**.

Import Allele Calls

You can import allele calls from another project to analyze in your project.

1 In the main window, select **Analysis** | **Import Allele Calls**.

2 Click **Browse** and navigate to the location of the allele calls you want to import.

3 Select the file you want from the Files Found in the Import Directory area.

4 Click **OK**.

Clustering SNPs

Chapter 6

Introduction

The Polyploid Genotyping Module employs the following algorithms for clustering samples:

- DBSCAN
- PolyGentrain

You can also set the following clustering options:

- The cluster distance and minimum number of points in a cluster for DBSCAN.
- Whether to model allele frequency in the PolyGentrain method.
- The clustering intensity threshold and whether the software considers an inbred population for both clustering algorithms.

Set Clustering Options

1. In the main window, select **Tools** | **Clustering Options**.
   The Clustering Options dialog box opens.
2. In DBSCAN Algorithm, set the following options:
   - **Cluster distance**—The distance controls the density of clusters. A larger cluster distance value allows for less dense clusters.
   - **Minimum number of points in cluster**—The minimum number of points required to form a cluster.
3. In PolyGentrain Algorithm, select the **Model Allele Frequency** checkbox to add a requirement to the model, which calculates the probability of each genotype based on the estimated minor allele frequency and the equilibrium assumptions.
4. In General, set the following options:
   - **Clustering intensity threshold**—When the threshold is set to a nonzero value, the software filters samples based on normR value before including them in the clustering algorithm.
   - **Inbred Population**—The clustering algorithm option only fit clusters corresponding to the homozygous genotypes.
5. Click **OK**.

Change the Clustering Intensity Threshold

1. In the main window, select **Tools** | **Clustering Options**.
2. At the Clustering Intensity Threshold field, enter the value you want to use.
3. Click **OK**.

Change the Color of Cluster Calls

In the SNP graph, the software assigns a color to samples based on their genotype call. You can customize the colors.

1 In the SNP table, right-click to open the context menu.

2 Select **Set Cluster Colors**.

3 Select the cluster colors you want for the specific genotype calls.

4 Click **OK**.

Cluster Selected SNPs

1 In the main window, select **Analysis** | **Cluster Selected SNPs As**.
- **PolyGentrain**
- **DBSCAN**

The software clusters the selected SNPs based on the selected algorithm.

Edit Clusters

Move the Cluster Location

1. In the graph, press and hold the **Shift** key.
2. Click at the center of the cluster. The ✣ move cursor appears.
3. Drag the cluster to a new location.

Change the Cluster Height and Width

1. In the graph, press and hold the **Shift** key.
2. Click at the edge of an oval. The ↕ resizing cursor appears.
3. Drag the edge of the oval to reshape the cluster.

Delete Cluster

1. In the graph, select the cluster you want to delete.
2. Right-click and select **Delete Cluster**.
   The software deletes the cluster from the graph.

Create a Cluster from Selected Samples

1. In the graph, select the samples you want to analyze.
2. Right-click and select **Create Cluster from Selected Samples**.

Set Custom Genotype

1. In the graph, select the cluster you want to analyze.
2. Right-click to open the context menu, and then select **Associate Cluster with Genotype | Other**.
   The Select Custom Genotype dialog box opens.
3. Select the genotype options from the drop-down list.
4. Click **OK**.

Export a Cluster File

1. In the main window, select **File** | **Export Cluster Positions**.
2. Select 1 of the following options:
   - **For Selected SNPs**
   - **For All SNPs**
3. Browse to the location where you want to save your cluster file.
4. [Optional] Rename your cluster file.
5. Click **Save**.

Document # 1000000012407 v00

Generating Reports

Chapter 7

Introduction

You can generate the following reports using the Report Wizard:

- Final Report
- DNA Report
- Locus Summary Report
- Locus x DNA Report

Also, you can generate a Reproducibility Report.

Document # 1000000012407 v00

Generate a Final Report

The Polyploid Genotyping Module produces a final report in a text (*.txt) or a comma-separated values (*.csv) file format.

1 In the main window, select **Analysis** | **Reports** | **Report Wizard**.
By default, the final report option is selected.

2 Click **Next**.

3 Select 1 of the following options:
- **All samples**
- **Selected samples**

4 Select 1 of the following options:
- **Include zeroed SNPs in the report**
- **Removed zeroed SNPs in the report**

5 Click **Next**.

6 [Optional] Select **Standard**.

a In the Available Fields area, select the fields you want to include in the report. Press **Ctrl** to select multiple fields.
b Click the **Show** button.
c At the Group by field, select whether you want to group by sample or by SNP.

7 [Optional] In the General Options area, select from the following:
- **Tab** —Generate the final report in a text (*.txt) file format.
- **Comma**—Generate the final report in a comma-separated values (*.csv) file format.

8 [Optional] Select **Create map files**.

9 [Optional] Specify the number of samples per file to include in the report.

10 Click **Next**.

11 At the Output Path field, browse to the directory you want to save the report or accept the default directory.

12 At the Report Name field, enter a report name or accept the default name.

13 Click **Finish**.
The software generates the report and saves it in the specified directory.

DNA Report

The Polyploid Genotyping Module produces a DNA report in a comma-separated values (*.csv) file format.

NOTE
The frequency value for genotypes depend on the user-defined ploidy type and level.

Table 20 DNA Report

Statistic	Description
Row	Row number.
DNA_Name	DNA name.
#No_Calls	Number of loci with GenCall scores below the call region threshold.
#Calls	Number of loci with GenCall scores above the call region threshold.
Call_Freq	Call frequency or call rate, calculated as #Calls/(#No_Calls + #Calls).
A/A_Freq	Frequency of homozygous allele A calls.
A/B_Freq	Frequency of heterozygote calls.
B/B_Freq	Frequency of homozygous allele B calls.
Minor_Freq	Frequency of the minor allele.
50%_GC_Score	50th percentile GenCall score for all loci.
10%_GC_Score	10th percentile GenCall score for all loci.
0/1	An algorithm to determine whether to include or exclude samples. **0**—Exclude sample **1**—Include sample

Locus Summary Report

The Polyploid Genotyping Module produces a locus summary report in a comma-separated values (*.csv file) format.

NOTE
The frequency, mean, and deviation value for genotypes depends on the user-defined ploidy type and level.

Table 21 Locus Summary Report

Statistic	Description
Row	Row number.
Locus_Name	Locus name from the manifest file.
IllumiCode_Name	Locus ID from the manifest file.
#No_Calls	Number of loci with GenCall scores below the call region threshold.
#Calls	Number of loci with GenCall scores above the call region threshold.
Call_Freq	Call frequency or call rate calculated as follows: #Calls/(#No_Calls + #Calls)
A/A_Freq	Frequency of homozygote allele A calls.
A/B_Freq	Frequency of heterozygote calls.
B/B_Freq	Frequency of homozygote allele B calls.
Minor_Freq	Frequency of the minor allele.
GenTrain_Score	Quality score for samples clustered for this locus.
50%_GC_Score	50th percentile GenCall score for all samples.
10%_GC_Score	10th percentile GenCall score for all samples.
Cluster_Sep	Cluster separation score.
AA_T_Mean	Normalized theta angles mean for the AA genotype.
AA_T_Std	Normalized theta angles standard deviation for the AA genotype.
AB_T_Mean	Normalized theta angles mean for the AB genotype.
AB_T_Std	Standard deviation of the normalized theta angles for the AB genotype.
BB_T_Mean	Normalized theta angles mean for the BB genotypes.
BB_T_Std	Standard deviation of the normalized theta angles for the BB genotypes.

Statistic	Description
AA_R_Mean	Normalized R value mean for the AA genotypes.
AA_R_Std	Standard deviation of the normalized R value for the AA genotypes.
AB_R_Mean	Normalized R value mean for the AB genotypes.
AB_R_Std	Standard deviation of the normalized R value for the AB genotypes.
BB_R_Mean	Normalized R value mean for the BB genotypes.
BB_R_Std	Standard deviation of the normalized R value for the BB genotypes.

Locus x DNA Report

The Polyploid Genotyping Module produces a locus x DNA report in a comma-separated values (*.csv) file format.

Table 22 Locus x DNA Report

Statistic	Description
instituteLabel	User-defined sample name for the DNA sample.
plateWell	Concatenation of the sample plate and sample well.
imageDate	Imaging date for the sample.
oligoPoolId	Manifest name.
bundleId	Bundle identifier.
status	Field is unused.
recordType	Record type shows 2 rows of data for each DNA sample. • **calls**—Depends on the user-defined ploidy level and type. • **Score_Call**—GenCall score for the call.
data	Calls or scores for the DNA sample and locus.

Generate a Reproducibility Report

The Polyploid Genotyping Module produces an error output report of the samples in a comma-separated values (*.csv file) format.

1 In the main window, select **Analysis** | **Reports** | **Create Reproducibility Report**.
   - **Without calculating errors**
   - **With calculating errors**

2 In the File Name field, enter a file name or accept the default name.

3 Click **Save**.
The software generates the report and saves it in the specified location.

Reproducibility Statistics

The Polyploid Genotyping Module produces a reproducibility report in a comma-separated values (*.csv file) format.

Table 23 Duplicate Reproducibility

Statistic	Description
Rep1_DNA_Name	Sample name designated as replicate #1.
Rep2_DNA_Name	Sample name designated as replicate #2.
# Correct	Number of loci with consistent replicate genotype comparisons.
# Errors	Number of loci with inconsistent replicate genotype comparisons.
Total	Number of total genotype comparisons (1 genotype comparison per locus per replicate pair). The report does not include genotypes with intensities that fall below the no-call threshold.
Repro_Freq	Reproducibility frequency. The error rate does not include genotype calls that fall below the no-call threshold.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 24 Illumina General Contact Information

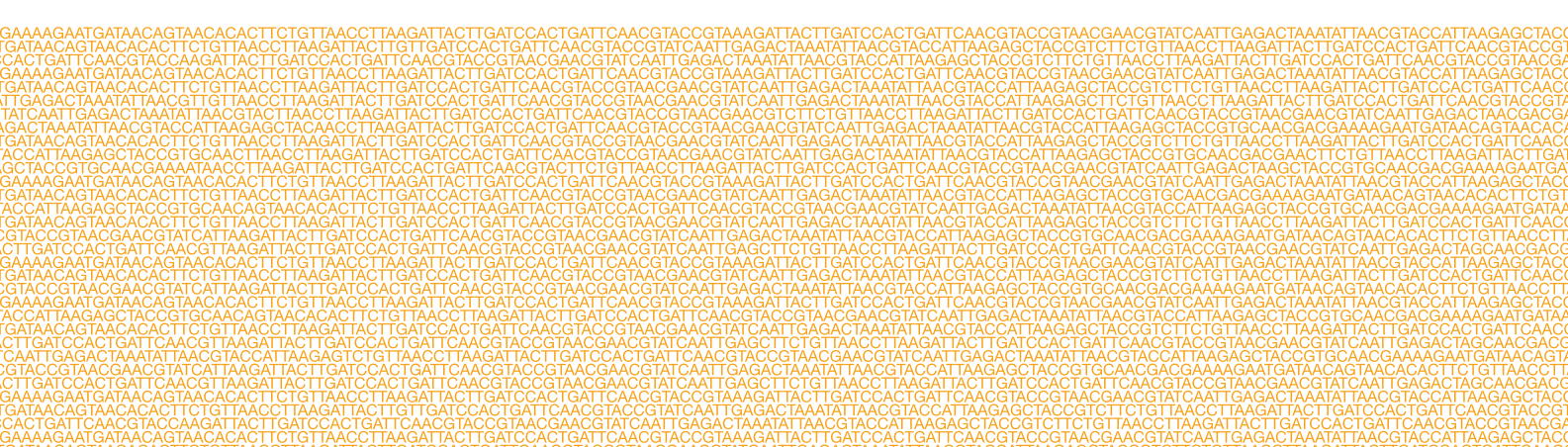
Website	www.illumina.com
Email	techsupport@illumina.com

Table 25 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Japan	0800.111.5011
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
China	400.635.9898	Singapore	1.800.579.2745
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	Taiwan	00806651752
Hong Kong	800960230	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000
Italy	800.874909		

Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **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