illumina®

GenomeStudio® Genotyping Module v2.0 Software Guide

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

ILLUMINA PROPRIETARY

Document # 11319113 v01
September 2016

Revision History

Document	Date	Description of Change
Document # 11371113 v01	September 2016	Supported GenomeStudio Genotyping Module v2.0
Document # 11371113 rev.A	November 2008	• Changed the name of BeadStudio Genotyping Module to GenomeStudio Genotyping Module • Supported GenomeStudio Genotyping Module v1.0
Document # 11284301 rev.A	December 2007	Supported BeadStudio Genotyping Module v3.2
Document # 11207066 rev.C	February 2007	Supported BeadStudio Genotyping Module
Document # 11207066 rev.B	March 2006	Supported BeadStudio Genotyping Module
Document # 11207066 rev.A	December 2005	Supported BeadStudio Genotyping Module

Table of Contents

Overview

Chapter 1

Introduction

The Illumina® GenomeStudio Genotyping Module™ Genotyping Module Software v2.0 analyzes Illumina genotyping data to optimize call rates.

Before installing the 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 69.

Features

You can perform the following tasks with the Genotyping Module:

- Identify poorly performing samples.
- Manually adjust cluster positions.
- Set genotyping and clustering intensity thresholds.
- Analyze genotyping data and create cluster files.
- Import and export 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 Genotyping Module project from one of the following sources of sample data:

- An Illumina LIMS database
- A sample sheet
- Directories containing intensity data files

Create a Project From LIMS

1. From your desktop, double-click the GenomeStudio  icon.
2. In the new project pane, click **Genotyping**.
3. Click **Next**.
4. Click **Select from LIMS**, and then click **Next**.
5. Click **Login**, and then enter your LIMS user name and password.
6. Click **OK**.
7. At the Institute field, select your institution.
8. At the Investigator field, select your investigator.
9. At the Project field, select your project.
10. At the Product field, select your product.
    If your product is multi-species, expand the product to select an individual species.
11. Click **Finish**.
12. [Optional] Click **Use Start Date** and select a start date from the calendar.
13. [Optional] Click **Use End Date** and select an end date from the calendar.
14. Click **OK**.
15. [Optional] Click **Yes** to calculate the heritability and reproducibility errors.
16. Click **OK**.

Create a Project From a Sample Sheet

1. In the new project pane, click **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 10.
   a. Select **Browse**, highlight the cluster file, and select **Open**.
9. Click **Finish**.
   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 **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 10.

   a Select **Browse**, highlight the cluster file, and select **Open**.

8 [Optional] In the Project Settings area, select the following options:
   - **Pre-Calculate**
   - **Cluster SNPs**
   - **Calculate Sample and SNP Statistics**
   - **Calculate Heritability**
   - **Gen Call Threshold**

9 Click **Finish**.

Import Cluster Positions

You can import a cluster (*.egt) file when creating a project in the 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] Click **Pre-Calculate** to optimize your project speed based on your computer memory capabilities.
4. [Optional] In the Project Creation Actions settings, select from the following options:
   - **Cluster SNPs**
   - **Calculate Sample and SNP Statistics**
   - **Calculate Heritability**
5. At the GenCall Threshold field, specify the score cutoff for this project.
   We recommend the 0.15 GenCall score cutoff for Infinium products.
6. Click **Finish**.

Viewing Data

Chapter 3

Introduction

The 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 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.
Paired Sample Table	Statistics for paired samples.
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	SNP index row.
Name	SNP name.
Address	Bead type identifier.
Chr	SNP chromosome.
Manifest	Manifest name of the SNP.
Position	Chromosomal position of the SNP.
GenTrain Score	Score for a SNP from the GenTrain clustering algorithm.
FRAC A	Fraction of the A nucleotide in the top genomic sequence.
FRAC C	Fraction of the C nucleotide in the top genomic sequence.
FRAC G	Fraction of the G nucleotide in the top genomic sequence.
FRAC T	Fraction of the T nucleotide in the top genomic sequence.

Table 3 Full Data Table Subcolumns

Column Heading	Description
GType	Genotype for the sample.
Score	SNP call score for the sample.
Theta	Normalized Theta-value for the sample.
R	Normalized R-value for the sample.
X Raw	Raw intensity of the A allele.
Y Raw	Raw intensity of the B allele.
X	Normalized intensity of the A allele.
Y	Normalized intensity of the B allele.
B Allele Freq	B allele frequency for this sample as interpolated from known B allele frequencies of 3 canonical clusters: 0, 0.5 and 1 if it is equal to or greater than the theta mean of the BB cluster. B Allele Freq is between 0 and 1, or set to NaN for loci categorized as intensity only.

Column Heading	Description
Log R Ratio	Base-2 log of the normalized R value over the expected R value for the theta value (interpolated from the R-values of the clusters). For loci categorized as intensity only; the value is adjusted so that the expected R value is the mean of the cluster.
Top Alleles	Illumina-designated top strand genotype.
Import Calls	Genotype calls for an imported sample.
Concordance	Numeric correlation between the top allele call for a SNP in the project and the imported allele call of a SNP from another project.
Orig Call	Genotype call of SNP and sample at the time the project was originally clustered.
CNV Value	Copy number estimate at each locus.
CNV Confidence	Level of confidence that the Copy Number Value (CNV) is correct based on the algorithm used.
Plus/Minus Alleles	Genotype for the sample, as related to the nucleotide alleles on the plus strand.

SNP Table

The SNP Table shows the statistics for each SNP.

Table 4 SNP Table Columns

Column Heading	Description
Index	SNP index row.
Name	SNP name.
Chr	SNP chromosome.
Position	SNP chromosomal position.
ChiTest 100	Normalized Hardy-Weinberg p value calculated using genotype frequency. The value is calculated with 1 degree of freedom and normalized to 100 individuals.
Het Excess	Excess of heterozygotes measurement for the SNP based on Hardy-Weinberg Equilibrium. **0**—No excess of heterozygotes **(-) Negative values**—A deficiency of heterozygotes
AA Freq	Frequency of AA calls.
AB Freq	Frequency of AB calls.
BB Freq	Frequency of BB calls.
Call Freq	Call frequency.
Minor Freq	Minor allele frequency.
Aux	User-defined auxiliary value for the SNP.
Rep Errors	The number of reproducibility errors for the SNP as allele comparisons between replicates.
SNP	Nucleotide substitution for the SNP on the Illumina ILMN strand.
ILMN Strand	Design strand designation.
Customer Strand	Customer strand designation.
Top Genomic Sequence	Sequence on the top strand around the SNP.
Comment	User-defined comment. Right-click in the column to set the value.
Norm ID	Normalization ID for the SNP.
HW Equil	Hardy-Weinberg Equilibrium score for the SNP.
Concordance	Measurement between 2 genotypes from the same SNP locus.

Column Heading	Description
CNV Region	SNPs and nonpolymorphic probes falling in known CNV regions. This column populates information from the product manifest and may not be current because the number of known CNV regions is constantly changing.
Exp Clusters	Number of expected clusters for a locus: **1**—Nonpolymorphic probes **2**—Mitochondrial DNA and Y loci **3**—Any other loci This column populates information from the product manifest.
Intensity Only	Indicates locus information. **1**—Locus with intensity information only that is not included in GenomeStudio statistics such as Call Rate. **0**—Locus with intensity and genotyping information that is included in GenomeStudio statistics such as call rate. The column populates information from the product manifest and can be edited.

Table 5 SNP Table Subcolumns

Column Heading	Description
Address	The SNP index row.
GenTrain Score	The SNP cluster quality.
Orig Score	The original GenTrain score for the SNP before edits.
Edited	The SNP was edited after identifying clustering positions. **1**—Edited **0**—Unedited
Cluster Sep	The cluster separation measurement for the SNP that ranges between 0 and 1.
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.

Column Heading	Description
AA R Mean	R value of the center of the AA cluster, in normalized polar coordinates
AA R Dev	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.
Intensity Threshold	The intensity threshold value.
ILMN Strand	The design strand designation.
Address 2	The bead type unidentified for the second allele. It is only used for Infinium I.
Norm ID	The normalization ID for the SNP.
Manifest	The manifest name of the SNP.

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.
PC Error Rate	Parent-child heritability error rate.
PPC Error Rate	Parent-parent-child heritability error rate.
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.

Column Heading	Description
Sample Name	Sample name.
Sample Group	User-entered sample group.
Sample Plate	Sample plate identifier.
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.

Column Heading	Description
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.

Paired Sample Table

The Paired Sample table shows statistics for paired samples.

Table 8 Paired Sample Table Columns

Column	Description
Index	SNP index row.
Name	SNP name.
SNP	SNP.
Address	Bead type identifier.
Chr	SNP chromosome.
Position	SNP chromosomal position.

Table 9 Per-Pair Sample Subcolumns

Columns Heading	Description
Theta Ref.	Theta value for the reference sample.
Theta Sub.	Theta value for the subject sample.
|dTheta sub-ref|	Absolute value of the difference between subject and reference theta values.
Allele Freq Ref.	Allele frequency of the reference sample.
Allele Freq Sub.	Allele frequency of the subject sample.
|dAlleleFreq sub-ref|	Absolute value of the difference between subject and reference allele frequency values.
R Ref.	R value for the reference sample.
R Sub.	R value for the subject sample.
Log2 (Rsub/Rref)	Log base 2 of the ratio of subject and reference R values.
GType Ref.	Genotype of the reference sample.
GType Sub.	Genotype of the subject sample.
LOH Score	Probability that there is loss of heterozygosity in a region of interest.
CN Estimate	Estimate of the copy number at an individual locus.
CN Shift	Statistical confidence level between 0 and 1 indicating whether a copy number change has occurred. **1**— No copy number change. **0**— Copy number change.

Errors Table

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

Table 10 Errors Table Columns

Column Heading	Description
Error Index	Error index row.
Error Type	Types of errors: • Rep—Reproducibility • P-C—Parent-Child heritability • P-P-C—Parent-Parent-Child heritability
Child/Rep Index	Sample index of the child sample involved in the error.
Child/Rep	Sample ID of the child sample involved in the error.
Child/Rep GType	For a parental relationship error, the genotype of the child.
Parent1/Rep Index	Sample index of the Parent1 sample involved in the error.
Parent1/Rep	Sample ID of the Parent1 sample involved in the error.
Parent1/Rep GType	For a parental relationship error, the genotype of Parent1. For a replicate error, the genotype of replicate 1.
Parent2 Index	Sample index of the Parent2 sample involved in the error.
Parent2	Sample ID of the Parent2 sample involved in the error.
Parent2 GType	For a parental relationship error, the genotype of Parent2. For a replicate error, the genotype of replicate 2.
SNP Index	SNP index number where the error occurred.
SNP Name	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 11 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.
Load Additional Samples	Opens the GenomeStudio Project Wizard. You can specify a sample sheet or directories to load intensity files.
Import Cluster Positions	Imports cluster positions from an *.egt file.
Export Cluster Positions	Exports selected SNPs or all SNPs to an *.egt file.
Export Manifest	Exports a manifest (*.csv) file.
Update Project from LIMS	Updates the project from LIMS.
Import Phenotype Information from File	Imports phenotype information for the samples from a 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 Genotyping Module Genotyping Module.

Edit Menu

Table 12 Edit Menu Functions

Function	Description
Cut	Cuts the selected content.
Copy	Copies the selected content.

Function	Description
Paste	Pastes the selected content.
Select All	Selects all data in the table.

View Menu

Table 13 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 14 Analysis Menu Functions

Function	Description
Exclude Samples by Best Run	Includes the sample with the highest GC10 or GC50 score for each sample ID. Excludes all the other samples.
Cluster All SNPs	Clusters all 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.
Update SNP statistics	Updates SNP statistics.
Edit Replicates	Edits, includes, or excludes replicates for a sample.
Edit Parental Relationships	Edits, includes, or excludes P-C and P-P-C relationships for a sample.
Update Heritability/Reproducibility Errors	Updates replicate P-C, and P-P-C heritability information in columns and reports.
Reports	Generates the following reports: • Reproducibility and Heritability Report • Final Report • DNA Report • Locus Summary Report • Locus x DNA Report

Function	Description
View Controls Dashboard	Allows you to view intensity data associated with various control probes.
Paired Sample Editor	Edits the list of paired samples.
Calculate Paired Sample LOH/CN	Calculates LOH and copy number-related scores for paired samples.
Show Genome Viewer	Shows the Illumina Genome Viewer (IGV).
Import Allele Calls	Imports allele calls.
Export Allele Calls	Exports allele calls.
Remove Imported Allele Calls	Removes imported allele calls from the project.
Create Plug-in Column	Selects an algorithm-based column plug-in. Use the column plug-in to create a subcolumn.

Tools Menu

Table 15 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.
Show Genome Viewer	Shows data in the Illumina Genome View (IGV).

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
- Paired Sample Table
- Errors Table
- Sample Graph

Help Menu

Table 16 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.
	Shade Call Region	Applies color to the gene plot calling regions. The size of the shaded area defines the cutoff gene call score.
	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 17 Graph Window Context Menu

Function	Description
Define AA cluster using selected sample	Determines the size and position of the AA genotype cluster.
Define AB cluster using selected sample	Determines the size and position of the AB genotype cluster.
Define BB cluster using selected sample	Determines the size and position of the BB genotype cluster.
Cluster this SNP	Determines cluster locations and score for each locus.
Cluster this SNP Excluding Selected Samples	Determines the cluster locations for each locus except the excluded locations.
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.
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 an image to 1 of the following files: • bitmap • GPEG • PNG • GIFF • TIFF

Table 18 Full Data Table Context Menu

Function	Description
Show Only Selected Rows	Shows your SNPs of interest.
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 SNP Table Context Menu

Function	Description
Cluster Selected SNP	Clusters a selected SNP.
Zero Selected SNP	Zeroes a selected SNP.
Set Genotyping Intensity Threshold for Selected SNPs	Opens the Set Genotyping Intensity Threshold dialog box. You can enter the threshold value.
Set Aux Value	Sets the aux value of a SNP.
Update Selected SNP Statistics	Updates selected SNP statistics.
SNP Properties	Opens the SNP Properties dialog box. You can manually edit the 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 mark selected rows.
Select Marked Rows	Selects marked rows.
Clear Marks | <All>	Clears all marks.

Table 20 Samples Table Context Menu

Function	Description
Exclude Selected Sample	Excludes the selected sample.
Include Selected Sample	Includes the selected sample.

Function	Description
Recalculate Statistics for Selected Sample	Recalculates statistics for selected samples.
Recalculate Statistics for All Samples	Recalculates statistics for all samples.
Estimate Gender for Selected Samples	Estimates gender for the selected samples.
Display Image	Displays image when 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	Sets the aux value of a sample.
Sample Properties	Changes values for sample data.
Upload Selected Samples to Illumina Controls Database	Uploads selected samples to the Illumina Controls Database.
Show Only Selected Rows	Shows only selected rows.
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 21 Error Table Context Menu

Function	Description
Show Only Selected Rows	Configures the Samples table to show only selected rows.
Edit Replicates	Edits replicates.
Edit Parental Relationships	Edits parental relationships.
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.

Common Tasks

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.

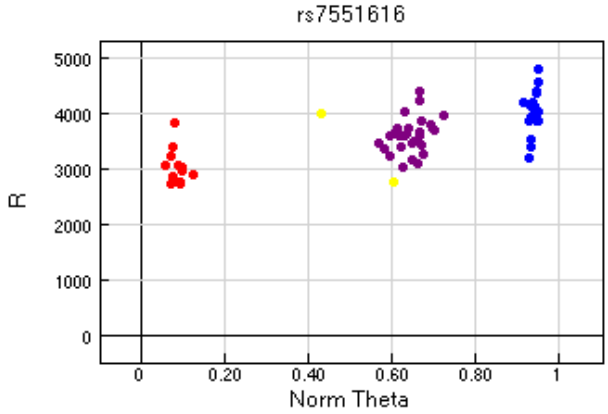
3 To add additional samples without losing your original selection, press and hold the **Ctrl** key and select additional samples.

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**.

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

Update SNP Statistics

When you adjust the loci or metrics, the SNP statistics require an update as indicated by the red highlighted rows in the SNP table.

The time to update increases with the size of the project. For best practices, make several edits before updating the SNPs statistics.

1 In the main window, select **Analysis** | **Update SNP Statistics**.

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**.

Generating Clusters

Chapter 6

Introduction

Genotype calls for genotyping assays are based on information from a standard or custom cluster file that provides statistical data from a sample set. The clustering process can generate the most accurate genotype for the data set loaded in the project because the locations of the heterozygotes and homozygotes for each SNP can vary from SNP to SNP. In the Genotyping Module, we use the Gentrain3 clustering algorithm to minimize erroneously clustered loci and deliver accurate genotyping data.

The software can automatically determine the cluster positions of the genotypes and estimate the missing cluster positions for SNPs that have 1 or 2 clusters that lack representation.

We recommend that you have a minimum of 100 or more diverse samples to achieve a representation of all clusters. The most desirable number of samples to use for cluster file generation varies depending on the minor allele frequency of your SNPs and the studied population.

Cluster SNPs

1. In the main window, select **Analysis** | **Cluster All SNPs**.
2. Click **Yes** to update SNP statistics for all SNPs.

Change the Clustering Intensity Threshold

You can change the clustering intensity threshold before reclustering all SNPs.

1. In the main window, select **Tools** | **Options** | **Project**.
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 main window, select **Tools** | **Options** | **Project**.
2. In the Colors section, click the drop-down arrow to select the colors you want for the specific genotype calls.
3. Click **OK**.

Edit Clusters

Redefine the Cluster

1. Select the samples in the graph.
2. Right-click to display the context menu.
3. Select **Define AB (or AA, or BB) cluster using selected samples**.

Exclude Samples From Clustering

1. Select samples in the graph.
2. Right-click to display the context menu.
3. Select **Cluster this SNP excluding selected samples**.

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.

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**.

Analyzing Data

Chapter 7

Import Phenotype Information

You can import a phenotype information file (*.csv) into a project. The file must contain an Index column that corresponds to the Index column in the Samples table.

You can import the following phenotypes in a Genotyping Module project:

- Gender
- Ethnicity
- Age
- Weight
- Blood Pressure Systolic
- Blood Pressure Diastolic
- Blood Type
- Phenotype Pos 1
- Phenotype Pos 2
- Phenotype Pos 3
- Phenotype Neg 1
- Phenotype Neg 2
- Phenotype Neg 3

1. In the main window, select **File** | **Import Phenotype Information From File**.
2. In the File name field, browse to the phenotype information file you want to use.
3. Click **Open** to import the file.

Estimate the Gender of Selected Samples

1. In the Samples table, select the samples you want to analyze.
2. Right-click on the selected samples.
3. Select **Estimate Gender for Selected Samples**.
4. Select 1 of the following options:
   - **Yes**—The Gender and Gender Est columns are populated with the estimated gender for the selected samples.
   - **No**—The Gender Est column is populated with the estimated gender for the selected samples.

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**.

Analyze Paired Sample Data

Paired sample data is useful for analyzing chromosomal aberrations. The Paired Sample table shows the differences in statistical measurements for a pair of samples, such as a subject sample and a reference sample.

Create paired samples by designating 1 of the following:

- Subject-and-reference pairs in the sample sheet used to create a project.
- Subject-and-reference samples using the paired samples editor.

1. In the main window, select **Analysis | Paired Sample Editor**.
2. After you designate paired samples, the Paired Sample table shows the paired sample data.
3. In the Paired Sample table, select the samples you want to analyze.
4. Analyze the paired sample with the following options:
   a. [Optional] Select **Analysis | Calculate Paired Sample LOH/CN Scores**.
   b. [Optional] View paired samples in the SNP graph.
   c. [Optional] View paired samples in the Integrative Genomics Viewer (IGV).

View Concordance Calculations

The 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**.

GenomeStudio Plug-ins

The GenomeStudio Genotyping Module provides plug-ins for copy number (CN) analysis, loss of heterozygosity (LOH) visualization, and other types of analysis. To download the plug-ins, see GenomeStudio Plug-ins. You can install the plug-ins after installing the Genotyping Module.

- Column plug-in—Creates subcolumns based on data in tables.
- CNV analysis plug-in—Creates a CNV analysis workflow in GenomeStudio,. See *CNV Analysis* on page 57.
- Report plug-in—Creates custom reports.

Create a Custom Subcolumn

With the Column Plug-in option, you can create custom subcolumns in the Full Data table. Install the column plug-in program to your desktop first, then copy the Dynamic Link Libraries (DLL) file for the column plug-in to the directory: C:\Program Files\Illumina\GenomeStudio 2.0

1. In the main window, select **Analysis** | **Create Plug-In Column**.
2. In the column plugins table, select a row from the list of available column plugins.
3. [Optional] At the New Subcolumn Name field, enter a name for the subcolumn.
4. [Optional] In the Column Plug-In Properties table, right-click in the column to enter new values.
5. Click **OK**.

CNV Analysis

A CNV Analysis computes the estimated copy number (CNV Value) and the accuracy score of the estimated copy number (CNV Confidence) for chromosomal regions in each sample.

Create a CNV Analysis

Install one or more CNV analysis plug-ins before you can perform the task.

1. In the main window, select **Analysis** | **CNV Analysis**.
2. Click the Create New CNV Analysis drop-down menu, and then, select a CNV algorithm.
3. [Optional] Click the **Calculate Only Selected Samples** checkbox.
4. [Optional] At the CNV Analysis Name field, enter a name.
5. [Optional] Modify parameters.
6. Click **Calculate New CNV Analysis**.
   When analysis is complete, the CNV Region Display window opens.
7. Click **OK** to close the CNV Analysis dialog box.

Edit a CNV Analysis

You can activate or delete a CNV analysis. Also, you can use the CNV Analysis Region Display and Full Data table to view the analysis.

1 In the main window, select **Analysis** | **CNV Analysis**.

2 Edit the analysis in Current CNV Analyses list:
- Activate—Select the CNV analysis, and then, click **OK**.
- Delete—Right-click on the CNV analysis, and then, click **Remove Analysis**.

View CNV Analysis Region Display

The CNV Analysis Region Display is a heat map that shows the copy number values for all samples. The heat map shows samples on the X-axis and chromosomal position on the Y-axis.

1 In the main window, select **Analysis** | **Show CNV Region Display**.

2 [Optional] Point the mouse over a region to view more information.

3 [Optional] Click the **Zoom** button to zoom in and out.

View CNV Analysis Data in the Full Data Table

You can view the CNV analysis that contains the estimated copy number and the confidence score of the estimated copy number in the Full Data table.

1 In the Full Data table, click the Column Chooser  button.

2 In the Hidden Subcolumns area, select **CNV Value** and **CNV Confidence**, and then, click the **Show** button.

3 Click **OK**.

Generating Reports

Chapter 8

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 and Heritability Report.

Generate a Final Report

The 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 Click **Next**.

5 Select 1 of the following report formats:
- **Standard**—All data are presented in rows in the final report. You can choose the statistics to include in the output.
- **Matrix**—Rows represent SNPs and columns represent samples. You can choose to include the GenCall score or the genotypes in the output.
- **3rd Party**—You can specify the output style of the final report based on the target application for downstream analyses.

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

Table 22 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 Genotyping Module produces a locus summary report in a comma-separated values (*.csv file) format.

Table 23 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.
Het_Excess_Freq	Heterozygote excess frequency, calculated as (Observed - Expected)/Expected for the heterozygote class. If f_{AB} is the heterozygote frequency observed at a locus, and p and q are the major and minor allele frequencies, then het excess calculation is the following: $$(f_{AB} - 2pq)/(2pq + \varepsilon)$$
ChiTest_P100	Hardy-Weinberg p-value estimate calculated using genotype frequency. The value is calculated with 1 degree of freedom and is normalized to 100 individuals.
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.

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

Table 24 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**—A, B, or H • **Score_Call**—GenCall score for the call.
data	Calls or scores for the DNA sample and locus.

Generate a Reproducibility and Heritability Report

The 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 and Heritability 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 and Heritability Statistics

The Genotyping Module produces a reproducibility and heritability report in a comma-separated values (*.csv file) format. The report contains the following sections:

- Duplicate Reproducibility
- Parent-Child Heritability
- Parent-Parent-Child Heritability

Table 25 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.

Table 26 Parent-Child Heritability

Statistic	Description
Parent_DNA_Name	Sample name designated as parent in a P-C relationship.
Child_DNA_Name	Sample name designated as child in a P-C relationship.

Statistic	Description
# Correct	Number of loci with consistent parent-child genotype comparisons.
# Errors	Number of loci with inconsistent parent-child genotype comparisons.
Total	Number of total genotype comparisons (1 genotype comparison per locus per parent-child pair). The report does not include genotypes with intensities that fall below the no-call threshold.
PC_Heritability_Freq	Heritability frequency calculated as (# Correct/Total).

Table 27 Parent-Parent-Child Heritability

Statistic	Description
Parent1_DNA_Name	Sample name designated as parent #1 in a P-P-C relationship.
Parent2_DNA_Name	Sample name designated as parent #2 in a P-P-C relationship.
Child_DNA_Name	Sample name designated as child in a P-P-C relationship.
# Correct	Number of loci with consistent Parent1-Child and Parent2-Child genotype comparisons.
# Errors	Number of loci with inconsistent Parent1-Child or Parent2-Child genotype comparisons.
Total	Total number of loci that contribute to the trio heritability analysis. The report does not include loci for Parent1, Parent2, or Child have genotypes with intensities that fall below the no-call threshold.
P-P-C Heritability Freq	Heritability frequency calculated as (# Correct / Total).

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 28 Illumina General Contact Information

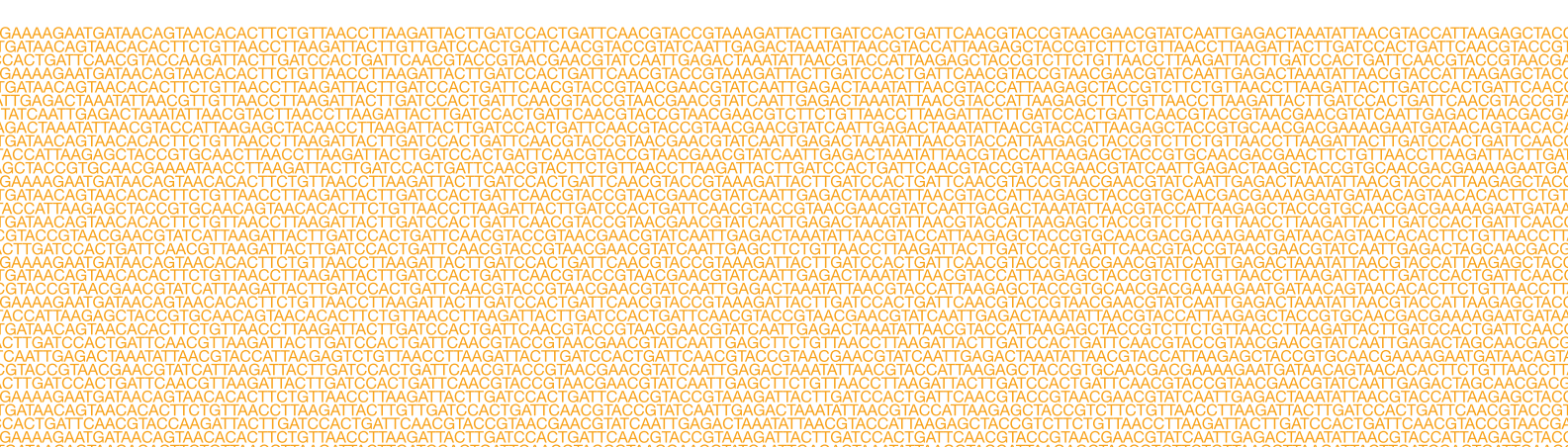
Website	www.illumina.com
Email	techsupport@illumina.com

Table 29 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