

Calling Polyploid Genotypes with GenomeStudio® Software v2010.3/v1.8

GenomeStudio Genotyping Module v1.8 supports calling genotypes for up to five clusters.

Introduction

GenomeStudio Data Analysis Software is a highly visual and intuitive platform for analyzing data generated from Illumina® assays. Illumina's GenomeStudio Software works seamlessly with Illumina's sequencing and genotyping platforms to support a diverse range of data analysis needs. Primary analyses, such as raw data normalization, clustering, and genotype calling of GoldenGate® and Infinium® Genotyping Assay data, are performed using algorithms in the Genotyping (GT) Module. This document provides best practices for using the GenomeStudio GT Module v2010.3/v1.8 to assign genotypes for up to five clusters in tetraploid genomes.

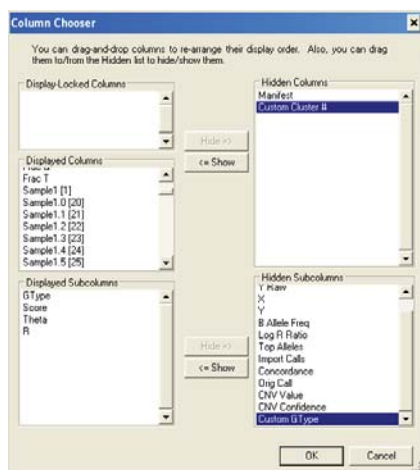
New Columns in the Full Data Table

GenomeStudio Genotyping Module v1.8 offers two new columns in the Full Data Table that support genotype calling of up to five clusters:

- Custom Cluster # enables users to specify whether a SNP has three or five clusters.
- Custom GType displays the genotype call for polyploid loci (AAAA, AAAB, AABB, BBBA, BBBB).

The two new columns are not displayed in the Full Data Table by default; they will need to be added manually.

Figure 1: Column Chooser Dialog Box



To view the new custom genotyping columns, open the Column Chooser dialog box by clicking the **Column Chooser** button in the Full Data Table toolbar. Select the **Custom Cluster #** column and **Custom GType** sub-column, then click **Show**.

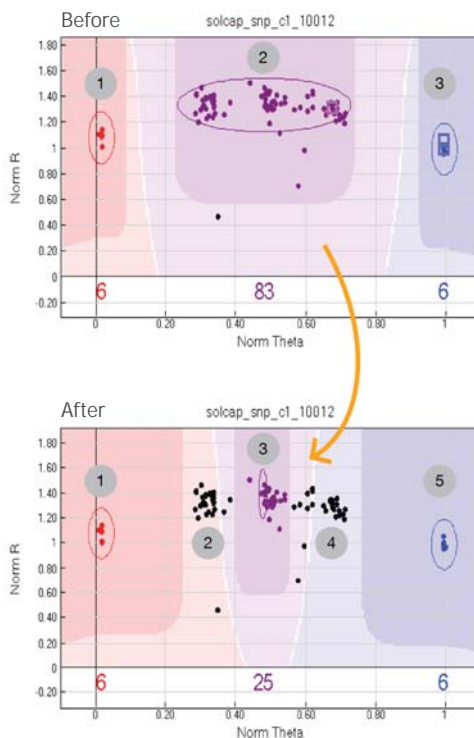
Custom Cluster # Column

The Custom Cluster # column has a value of **3** by default, reflecting the expectation that loci are diploid and have three possible genotypes or clusters. The user can change this default value to **5**, based on knowledge of the biology of the species of interest, to reflect the expectation that five clusters are present at a locus.

Before changing the Custom Cluster # column to 5, adjust the cluster positions within the SNP Graph so that the second and fourth clusters are in the no-call region and the first, third, and fifth clusters are within the called region of the SNP Graph.

After the first, third, and fifth clusters have been identified on the SNP Graph, right-click in the Custom Cluster # column of the Full Data Table and select **5** from the context menu.

Figure 2: SNP Graph, Before and After Adjustment



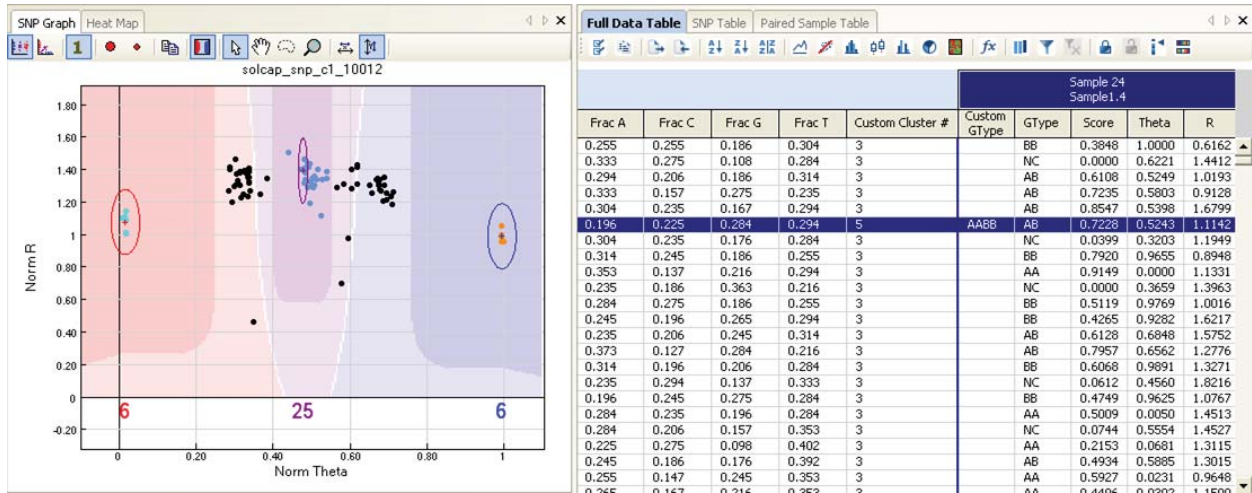
This example shows the cluster positions of a potato SNP from the SolCAP USDA-NIFA Consortium genotyping tool, before and after adjustment.

```

AAAGAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTAA
AATCAACGTACCGTAACGAAACGTATCATTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTAA
AACGACGAAAAGAATGATAACAGTAAACACACACTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGT
TTAACGTACCATTAAGAGCTACCGTTCGAAACAGTAACACACTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTTCGAAACAGTAAACACACTTCTGTAAACCTTAAG
AAAGAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTAA
AAGATTACTTGGCCACTGATTCAACGTAAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTAA
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTTGTAAACCTTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTTCGAAACGAAAAGAATGATAACAGTAA
AATCAACGTACCGTAACGAAACGTATCATTAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTTCGAAACGAAAAGAATGATAACAGTAAACACACTTCTGTAAACCTTAAG
AAGATTACTTGGCCACTGATTCAACGTAAAGATTACTTGGCCACTGATTCAACGTACCGTAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTTCGAAACGAAAAGAATGATAACAGTAAACACACTTCTGTAAACCTTAAG

```

Figure 3: SNP Graph and Full Data Table After Custom GType Value is Changed to 5



Changing the Custom Cluster # value to 5 results in three changes to the data handling:

- The colors of the genotype positions in the first, third, and fifth cluster of the SNP Graph will change from the default colors.
- The Custom GType column for any genotypes falling inside the call region for the first, third, and fifth clusters will be populated. For example, any samples at this SNP with a genotype in the first cluster will have the Custom GType populated with an AAAA.
- A new context menu, unique to loci with a Custom Cluster # of 5, will be available by right-clicking in the SNP Graph (not shown).

Custom GType Column

At this stage, the second and fourth clusters are in the no-call region of the SNP Graph, and the Custom GType column for these samples remains blank. The genotype dots are black to indicate the current status as a no-call. The next step is to call the genotypes for the second and fourth clusters manually. These calls will not be perpetuated in the cluster file nor will these cluster positions be applicable to future samples added to the same project; however, they can be saved within the current project for generating subsequent final reports (described later in this document).

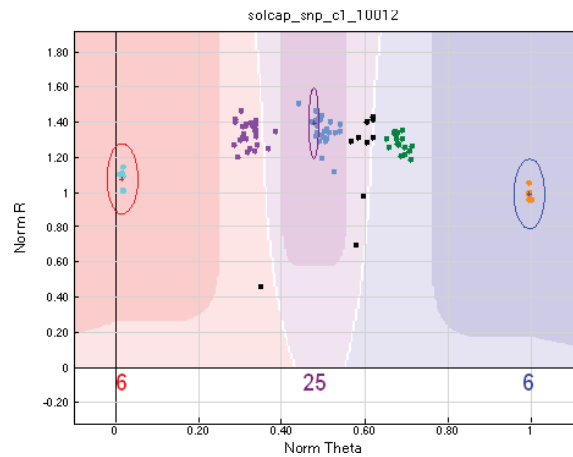
To call the genotypes on the second and fourth clusters, select the samples in the SNP graph for one of the two clusters. Using the default cursor, click the left mouse button and drag a selection square around the genotypes. When you release the mouse button, the selected genotypes will be highlighted yellow.

With the genotype positions highlighted in the SNP Graph, right-click to view the new context menu, which allows the user to define the Custom GTypes as AAAA, AAAB, AABB, BBBA or BBBB.

- For the second cluster, select **Define AAAB Genotype Using Selected Samples** from the context menu.
- For the fourth cluster, select **Define BBBA Genotype Using Selected Samples** from the context menu.

The Custom GType column for the selected samples will be populated using the assigned genotype. The color of the highlighted samples will change from black to a color that is specific to the assigned genotype, AAAB or BBBA, as shown in Figure 4. (Typically the user must click off the locus, then click back to the locus to view the color change.)

Figure 4: SNP Graph with Custom Genotype Calls



This example shows the cluster positions of the potato SNP from Figure 2 with all five clusters called. Note that the samples with genotypes that fall between the third and fourth cluster are left in the no-call region.

Without additional information about the biology of the organism, GenomeStudio v1.8 leaves these dots black by default. Individual researchers may decide to call genotypes for dots that show unusual spread or appear in between clusters, depending on the available background information and the experimental design. In the absence of any additional information, however, best practices favor a no-call rather than a potentially incorrect call.

Quickly Identifying 5-Cluster Loci

This section provides tips for moving quickly through your SNP Loci to identify 5-cluster loci within the project, so that you can spend the needed time on those loci after the apparently diploid loci are already called. This will be useful because the current solution is a manual one and applies on a per-project basis: genotypes for the second and fourth clusters will not be perpetuated to future projects nor to samples added to projects after calling has already been manually completed. For this reason, Illumina recommends completing analysis only after all samples have been scanned and added to the project. If this is not possible, then it is useful to have a quick way of identifying loci where manual intervention is needed.

The Aux column in the SNP Table can be used to quickly identify loci where you will need to go back and spend more time. By default, the Aux column is populated with 0 for all SNP loci. This can be changed manually by right-clicking in the SNP Table, selecting **Set Aux Value**, and entering any numeric value.

You can also change the value in the Aux column using the F1, F2 and F3 keys on your keyboard as you scroll through loci in the SNP Table:

- **F1 inserts a 1 into the Aux column and moves to the next locus in the SNP Table.**
- **F2 inserts a 0 into the Aux column, zeroes the locus, and moves to the next locus in the SNP Table.**
- **F3 inserts a -1 into the Aux column and moves to the next locus in the SNP Table.**

After completing a first pass through the SNP Table, sort on the Aux column, select all loci with a 1 or a -1, then move to the Full Data Table and change all loci at once to have a 5 in the Custom Cluster # column. Note that any sort done in the SNP Table is perpetuated in the Full Data Table.

Creating Custom GType Reports

Reports that include the Custom GType can be created from the Report Wizard in the Analysis menu.

1. **From the Genotyping Report menu, select Final Report and click Next.**
2. **Select choices for SNPs to be visible in the report and click Next.**
3. **On the Format page in the wizard, add Custom Allele1-AABB, Custom Allele2-AABB, Custom Allele3-AABB and Custom Allele4-AABB to the Displayed Fields.**
4. **Remove Allele1-Top, Allele2-Top, and GC Score from the Displayed Fields.**

The resulting report will show each allele in a separate column similar to the default GenomeStudio output Final Report for diploid loci. This report should be formatted to align with the genotyping software analysis tool for downstream QTL analysis.

Additional Information

For more information about GenomeStudio Software and supporting documentation, please refer to the GenomeStudio Portal or visit www.illumina.com/genomestudio.

GTAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
AAAATCAACGTACCCTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAACAGTAACACACTTCTGTTAACCTT
CGTAACGACGAAAAAGATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACC
CATTAACGTACCATTAAAGAGTACCCTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAA
GTAAAGATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAA
GAAAAAGATTACTTGATCCACTGATTCAACGTATTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAA
AACAGTATCAATTGAGACTAAATATTAACGTACCATTAAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAA
AAAATCAACGTACCCTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAACAGTAACACACTTCTGTTAACCTT
GAAAAATTACTTGATCCACTGATTCAACGTATTAAGATTACTTGATCCACTGATTCAACGTACCCTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTACCCTGCAACGACGAAAAAGATGATAA

ILLUMINA, INC. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2011 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, Eco, TruSeq, and MiSeq are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 970-2011-001 Current as of 13 January 2011



AAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTGTGTTAACCTTAAGATTACTTTGATCCACTGATTCACCGT
AATCAACGTACCGTAAACGAACTATCATTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAA
AACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAG
TTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAG
AAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAA
AAGATTACTTTGATCCACTGATTCACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAG
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTTTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAA
AATCAACGTACCGTAAACGAACTATCATTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAA
AAGATTACTTTGATCCACTGATTCACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAAGATTACTTTGATCCACTGATTCACCGTACCGTAAACGAACTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGAACGACGAAAAGAAATGATAACAGTAAACACACTTCTGTAAACCTTAA