illumina®

GenomeStudio™ Software 2009.1 Release Notes

1. GenomeStudio v2009.1 Framework

1.1. Project creation time for DNA sequencing and RNA sequencing projects is significantly improved compared to project creation time in GenomeStudio v2008.1. Due to performance increases, projects are now created in approximately half the time.

1.2. If you experience out-of-memory issues, Illumina recommends shutting down the GenomeStudio application and restarting before creating a new project.

    1.2.1. Setting the memory usage to Programs instead of System Cache can help reduce the chance of "out of memory" errors.

1.3. A Loci Plot feature has been added to the scatter plot. This feature allows you to plot all data points that include chromosomal positions in a scatter plot.

    1.3.1. Using the new **Loci** button in the Table menu, you can define which columns to use for Chromosome, Start, and End.

    1.3.2. The Loci End field is optional. This field is needed only in tables where chromosomal start and end positions are defined in separate columns.

    1.3.3. To use the Loci Plot feature, click **Scatter Plot** in the Table menu, then choose **Loci**, **Loci Start**, or **Loci End** from the list of available columns.

1.4. A Jump to Website feature has been added to the context menu for all tables with Chromosome and Position information.

    1.4.1. The following websites are available by default:

- UCSC Genome Browser
- Database of Genomic Variants
- DECIPHER-Sanger
- Ensembl
- OMIM
- PubMed

    1.4.2. You can configure the available websites by entering additional URLs in the tab-delimited *JumpToWebSiteConfig.txt* file (available in the main GenomeStudio directory).

1.5. The Create New Data Track Table option has been added to the Tools menu of all modules.

1.5.1.   You can create a new data track table in a GenomeStudio project using the data tracks files: Illumina Track File, Comma or Tab Delimited Text File, and WIG/BED file.

1.5.2.   The data track table is connected with the IGV. You can jump to the IGV from the data track table, or create a data plot in the IGV.

1.6.   A data track column can be added to a table in all GenomeStudio modules if the input file contains Chromosome and Start position (End position is optional).

1.6.1.   Only one column at a time can be inserted into the table from the data track file.

1.6.2.   The Chromosome, Start, and End positions (in the data track file to be imported) must be sorted before you include them in the table.

1.6.3.   A start position from the data track file within start/end range in the table is counted as a match.

1.7.   Heat maps have been updated with the following new features:

1.7.1.   The ability to link the heat map with a table selection has been added.

1.7.1.1.   Only the embedded heat map (the heat map next to the SNP Graph in the Genotyping module) is connected with both tables; the table heat map can only associate with the table itself.

1.7.2.   You can now save and load clustering files for an embedded heat map.

1.7.2.1.   **Known Issue:** You cannot save and load clustering files in the table heat map. To save or load a cluster file, please use the embedded heat map.

1.7.3.   The ability to make a single selection as well as a union selection in a heat map has been added.

1.7.4.   You can now zoom in and zoom out from specific areas of a heat map.

1.7.5.   Only the embedded heat map is saved with the project; a heat map created within the table is not saved.

1.7.6.   The Save Heat Map and Open Heat Map functions are no longer available.

1.7.7.   There is now a progress description in the bottom status bar of the heat map while it is processing data.

1.7.7.1.   You can now cancel an operation by clicking **Cancel**.

1.7.7.2.   Once the data has been processed (with * next to the items in the Data Value pulldown menu), you are not required to wait when reloading data. The statistics are cached and load immediately.

1.8.   The E-coli genome and reference FASTA sequence has been added as a default genome for use in the IGV, in addition to the Human, Rat, and Mouse genomes.

2.   Illumina Genome Viewer v1.1

2.1.   The performance of loading and scrolling through data has been greatly improved in all sequencing modules.

2.2.   Three zoomed-out display modes are now available for the sequences plot: AverageBaseCoverage, TotalBases, and MaxOverlap.

   2.2.1.   AverageBaseCoverage is now used by default whereas MaxOverlap was the only option available in previous versions of GenomeStudio software.

2.3.   You now have the ability to plot and view read type data on separate graphs in a DNA sequencing project.

   2.3.1.   For a paired end read, the forward strand is dark colored, and the reverse strand is lighter-colored.

      2.3.1.1.   The line linking the paired end reads is the same color as the read.

   2.3.2.   For single end reads, the forward strand is colored **blue**, and the reverse strand is colored **red**. This coloring scheme is also used in the RNA Sequencing and ChIP Sequencing Modules

2.4.   The ability to choose the read type (for PE projects) to display on the sequence plot of the IGV has been added. The choices are: All, Anomalous, Normal, AnomalousFF, AnomalousRR, AnomalousInsert, Chimera, and Orphan.

   2.4.1.   If both reads align to the reference sequence within the expected distance and orientation, and with the same chromosome, this is designated as a normal read (Normal) and colored in **green** in the IGV.

   2.4.2.   If only one read is aligned to the reference sequence, this is an orphan read (Orphan) and is colored in **pink** in the IGV.

   2.4.3.   If the read pair aligns to separate chromosomes, this is a chimera (Chimera) and is colored in **red** in the IGV.

   2.4.4.   If the paired reads have the same orientation (both designated as forward strand or reverse stand), this is a structural anomaly (AnomalousFF or AnomalousRR) and is colored in **purple** in the IGV.

   2.4.5.   If the aligned read pairs are not within a proper length, this is a length anomaly (AnomalousInsert) and is colored in **orange** in the IGV.

2.4.6.   You can discern the read type of a particular paired-end read in the Sequence Table by using the following logic:

Characteristic	Read Type
PE Score > 0	Normal
PE Score > 0 & PE Row Offset > 0	Read 1 (of a Paired End Read)
PE Score > 0 & PE Row Offset < 0	Read 2 (of a Paired End Read)
PE Score > 0 & PE Row Offset == 0	Single End Read

Note: Adding the current row Index to the PE Row Offset gives you the row describing Read2 of this PE read. You can calculate the insert length of this PE read by using the formula (Read2.Start – Read1.Stop).

PE Score == 0	Anomalous
PE Score == 0 & PE Row Offset == 0	Orphan
PE Score == 0 & SE Score < 0	Shadow Read (does not appear in IGV)
PE Score == 0 & PE Row Offset == -1000000	Chimera
PE Score == 0 & you find the mate and two reads have the same orientation	Anomalous FF or Anomalous RR
PE Score == 0 & insert length is not within reason for SIPE (~200-400 bp) or LIPE (2kb, 3kb or 4kb) (determined in CASAVA)	AnomalousInsert

2.5.   You now have the ability to view Read 1 vs. Read 2 data using the PERead1Read2 plot, which graphs Read 1 start vs. Read 2 start positions. (For singletons, Read 2 is set as equal to Read 1.)

2.6.   You now have the ability to jump from one paired end mate to another through the Jump to PE Mate function in the context menu for DNA Sequencing projects.

2.6.1.   Jump to PE Mate is disabled for chimera reads in the IGV.

2.7.   You now have the ability to jump from one end of a splice junction to the other through the Jump to Splice Mate function in the context menu for RNA Sequencing projects.

2.8.   You now have the ability to mark a read in the IGV through the context menu, which marks the same read in the Sequence table (and marks both reads for paired end reads).

2.8.1.   You can also delete marks through the context menu.

2.9. Jump from a read in the IGV to the Sequence table through the Jump to Sequence Table function in the context menu for all Sequencing projects.

2.9.1. The Sequence table is brought to the front after the jump with the read in the current view.

2.9.2. If the Sequence table is minimized or hidden behind the Samples or Lanes tables, this function does not bring the Sequences table to the front view.

2.10. The Zoom In/Out function is available through the context menu, which enables you to zoom in and out at the center of where they have right-clicked.

2.11. ChIP Sequencing Merged Regions can now be displayed as bookmarks in the ChIP Sequencing Module.

2.12. The ability to display multiple IGV graphs the genome viewer plot has been restored for microarray modules.

2.12.1. For genotyping projects, the data plot for Log R Ratio now scales from -2 to +2 and is centered at position 0 by default. B Allele Frequency scales from 0 to +1 by default.

2.13. The IGV genome viewer plot has been removed from all Sequencing modules. After creating a project, the ICB (Illumina Chromosome Browser) displays the reads for the first chromosome found in the Sequences table.

2.14. You can now load standard sequence FASTA files into the IGV with fewer restrictions. FASTA files are now parsed according to these rules:

2.14.1. The Header line starts with > and is only one line long.

2.14.2. All sequence lines except the last one have the same number of characters. No base length is required. (50bp long was required in the previous version of GenomeStudio software)

2.14.3. One sequence per file

2.14.4. The name of the file without the extension must match the chromosome name in the chromInfo.txt file.

2.15. The IGV now recognizes all of the following file extensions as potential FASTA files: .fa, .FA, .fasta, .FASTA, .fas, .FAS.

2.16. Tab-delimited text files can now be imported into the IGV as bookmarks.

3.   Genotyping Module v1.1

3.1.   GenomeStudio software is now compatible with data and projects created using the new Illumina LIMS workflow.

3.1.1.   To access an Illumina LIMS server, select the **Illumina LIMS** checkbox in the LIMS login window.

3.1.2.   You can now create projects using both data from GoldenGate® products (SAMs and BeadChips), and data from Infinium® BeadChips.

3.1.3.   If multiple product types are detected in the same Illumina LIMS project, you now have the option to load only one product type, or to load all product types (displaying only the common SNPs between products for Infinium projects and all SNPs for all products for Golden Gate projects).

3.2.   For projects containing multiple versions/lots, you can choose to cluster a specific version/lot by right-clicking a SNP in the Full Data table, and selecting **Cluster the Current Lot** from the context menu.

3.2.1.   You can change the current version/lot by using the dropdown menu in the SNP Graph.

3.3.   The Compute CN Metrics Report option is now available in **Analysis** | **Reports**.

3.4.   The minimum number of beads for Infinium I SNPs has been changed from three to two, as with the Infinium II assay.

3.4.1.   You may see an increase in call rates due to this change, as data from Infinium I SNPs that appeared only on two beads is now included in the GenomeStudio Genotyping Module.

3.5.   For Genotyping projects, the Data plot for Log R Ratio now scales from -2 to +2 and is centered at position 0 by default; B Allele Frequency scales from 0 to +1 by default.

4.   Gene Expression Module v1.1

4.1.   An issue with Cubic Spline normalization failing on data with many of the same raw intensity values has been fixed in this release.

5.   Methylation Module v1.1

5.1.   The Normalization options (Background and Average normalization) for Infinium Methylation have been removed in this release. Infinium Methylation Normalization will be addressed in a future release of GenomeStudio software.

6. Sequencing Modules v1.1

6.1.   The ability to use quality scores to reflect the quality of the read in the Sequences Table of all sequencing modules has been added.

    6.1.1.   You can adjust quality scores by going to **Tools** | **Options** | **Project**.

    6.1.2.   Bases with scores higher than the Quality Threshold are displayed as upper-case letters, and bases with quality scores equal to or lower than the quality threshold are displayed as lower-case letters.

    6.1.3.   Quality score adjustments are also reflected in the reads in the IGV.

6.2.   The Dynamic Filter feature in the Sequence Table allows you to filter reads by sample, chromosome, and chromosomal position, in all sequencing modules.

    6.2.1.   The **All** selection for Sample and Chromosome, and all available chromosome positions, are displayed by default. However, this may not show all data for very large projects.

    6.2.2.   All sequence data is available to display; however, a maximum row limit message appears when the Sequences table contains more than 134 million rows.

    6.2.3.   When dynamically filtered, the Final Report only contains the filtered sequences. To include all data, clear the dynamic filter in the Sequences table.

    6.2.4.   When the Sequences table is dynamically filtered, you are not able to Jump to PE Mate. Setting the dynamic filter back to **All** alleviates this issue.

    6.2.5.   When the Sequences table is dynamically filtered, you are not able to jump from a read in the IGV back to the Sequences table.

6.3.   For DNA Sequencing Projects, you can jump from one Paired End mate to the other using the Jump to PE Mate function, available in the context menu of the Sequences Table.

6.4.   dbSNP annotation information (human data only) now appears in the Alleles table for DNA Sequencing and RNA Sequencing projects.

    6.4.1.   The annotation information comes from dbSNP Build 129 and is available in the Human Build 36.1 folder in the GenomeStudio directory.

6.5.   You now have the ability to add and remove samples (CASAVA builds for RNA or DNA Sequencing projects, or Pipeline builds for ChIP Sequencing projects).

    6.5.1.   To add samples to a sequencing project, select **File** | **Add Sample**.

        6.5.1.1.   Only samples from the same genome are allowed to be added to an existing project.

6.5.2. Remove Sample is available in the Samples table context menu.

6.5.3. All table content, as well as the IGV, is updated after a sample is removed.

6.6. Chromosome Mapping

6.6.1. Chromosome mapping windows are included in the project creation wizards of the DNA Sequencing and RNA Sequencing modules. These allow you to manually map a chromosome from CASAVA to GenomeStudio software.

6.6.1.1. You can tell that a chromosome from Pipeline/CASAVA has not been mapped to a chromosome in GenomeStudio software if there are still items available in the Pipeline section of the Genomes and Chromosomes window.

6.7. In GenomeStudio v2008.1 software, if a sequencing project failed project creation, users were not notified of the reason. In GenomeStudio v2009.1 software, you are informed if you run into following issues:

6.7.1. GenomeStudio software now parses the *runs_summary.xml* file before loading the data. If possible errors are found, a text box appears, asking that you manually define the read length for each sample.

6.7.2. If you run out of the disk space on your hard drive while loading a project, an error message appears, informing you of the situation.

6.8. The Sequences table can display a limited number of reads for one chromosome from one sample. If a particular chromosome from a sample contains more than this limit (around 134 million reads), you are not able to see the reads beyond that limit in the Sequences table. However, all of the reads can be displayed in the IGV.