illumina®

KaryoStudio v1.4 User Guide

FOR RESEARCH USE ONLY

ILLUMINA PROPRIETARY

Part # 11328837 Rev. C
June 2011

Notice

Revision History

Title	Part Number	Revision	Date
KaryoStudio v1.4 User Guide	Part # 11328837	Rev. C	June 2011
KaryoStudio v1.2 User Guide	Part # 11328837	Rev. B	February 2010
KaryoStudio v1.0 User Guide	Part # 11328837	Rev. A	November 2008

Revision History

Table of Contents

List of Tables

List of Tables

Overview

Chapter 1

Introduction

Cytogenetics is an area of genetics that focuses on chromosomal aberrations resulting in various conditions and phenotypes. It includes the routine analysis of G-banded chromosomes (karyotyping), as well as molecular cytogenetics such as fluorescent *in situ* hybridization (FISH) and comparative genomic hybridization (CGH).

In recent years, many forms of cytogenetics have moved to array-based technologies. Because of the added benefit that SNP information can provide, the field has begun to incorporate such information into their research.

This user guide describes Illumina's KaryoStudio cytogenetics software. KaryoStudio allows researchers to analyze data generated from Illumina's Infinium® DNA Analysis BeadChips and generate *.pdf reports from this data.

NOTE

KaryoStudio software is for research use only.

Audience and Purpose

This guide is written for researchers who want to use Illumina's KaryoStudio software to analyze data generated from Illumina's Infinium assays, to identify cytogenetic or structural aberrations, and to generate *.pdf reports from this data.

KaryoStudio has been designed specifically for cytogeneticists who are using Illumina's Infinium products to detect aberrations in the genome. KaryoStudio accepts Infinium data, automatically performs normalization, scans data for aberrations, displays the analyzed data allowing interpretation by the use, and generates reports useful for interpreting results. The software is geared towards finding aberrations that are 75kb and larger, and cross-matching those aberrations with information from public databases.

NOTE

In the KaryoStudio software and throughout this manual, possible aberrations are referred to as *detected regions* or *found regions*.

KaryoStudio is standalone software application, separate from Illumina's GenomeStudio® software suite. If you would like to obtain comprehensive genotyping information from data generated using Infinium products, or perform a more customized data analysis, you might want to use the GenomeStudio Genotyping Module. For more information about this and other Illumina software products, contact your local account manager or visit http://www.illumina.com.

Installing KaryoStudio

When you install KaryoStudio v1.4, the cnvPartition v3.0.7 CNV Analysis plug-in is also installed automatically.

NOTE

If you have previous versions of KaryoStudio and the cnvPartition plug-in installed on your computer, uninstall them before continuing with the KaryoStudio v1.4 installation.

1 Download the appropriate installer for your system (32-bit or 64-bit) from iCom:
   a Log in to iCom.
   b Locate the KaryoStudio Software v1.4 download summary page.
   c Click the **Download** button next to the appropriate installer: 32-bit or 64-bit.
   d In the Security Warning dialog box, click **Save**.
   e Save the file on your desktop.

2 Double-click the KaryoStudio *.msi icon on your desktop.

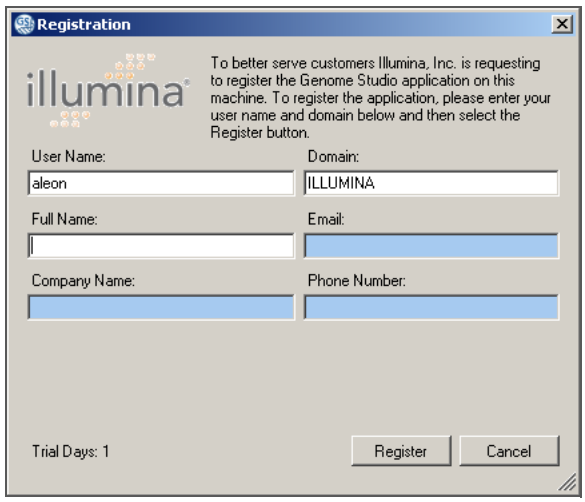
3 In the Security Warning dialog box, click **Run**.
The KaryoStudio Setup Wizard opens.

4 Click **Next** on the screens in the setup wizard to accept the default installation settings, and then click **Close** on the confirmation screen at the end.
The software is installed and the KaryoStudio icon appears on your desktop.

5 Double-click the KaryoStudio icon on your desktop.

If this is a new installation, the Registration window opens.

Figure 1 Registration Screen



If you have previously installed and registered another version of KaryoStudio, the Registration window does not appear. Instead, the KaryoStudio software opens to the main window, as shown in Figure 2.

NOTE
The Registration window has a GenomeStudio (GS) icon and refers to Illumina's GenomeStudio software suite. This is because KaryoStudio, while separate from GenomeStudio, uses the same registration database.

6 Complete the following information:

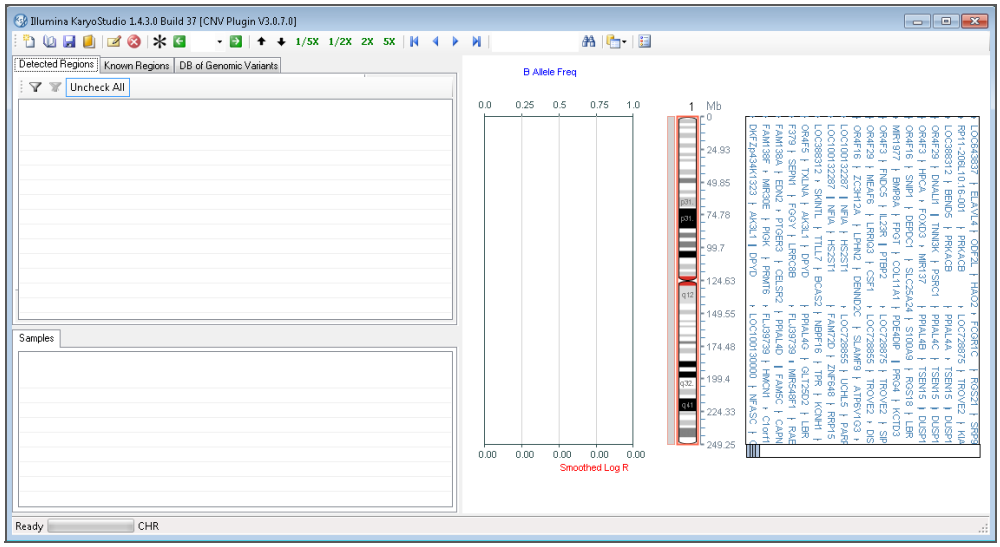
- **User Name**—Enter the user name you use to sign into your computer.
- **Full Name**—Enter your full name.
- **Email**—Enter your email address.
- **Company Name**—Enter the name of your organization (company, institute, university, etc.)
- **Phone Number**—Enter your phone number.

NOTE
KaryoStudio recognizes your network domain and populates the **Domain** field automatically.

7 Click **Register** to submit your registration information.
Illumina uses the information you provide for Technical Support purposes.

8 Click **OK** to complete the installation.
KaryoStudio opens to the main window.

Figure 2 Main Window



From the main window, you can create a new KaryoStudio project, open an existing project, view your data, generate reports, and more.

KaryoStudio Workflow

A typical data visualization and analysis workflow using KaryoStudio includes the following steps:

1. Use the project wizard to load data into KaryoStudio.
2. Scan the data for aberrations using KaryoStudio.
3. Adjust the filter settings to display detected regions of interest.
4. Display detected regions of interest in the chromosome browser and Detected Regions Table.
5. Cross-match detected regions to areas of the genome known to be associated with various conditions or phenotypes.
6. Select report settings.
7. Generate a KaryoStudio report.

Loading Your Data

Chapter 2

Introduction

Using the Project Wizard is an easy way to begin a new KaryoStudio project.

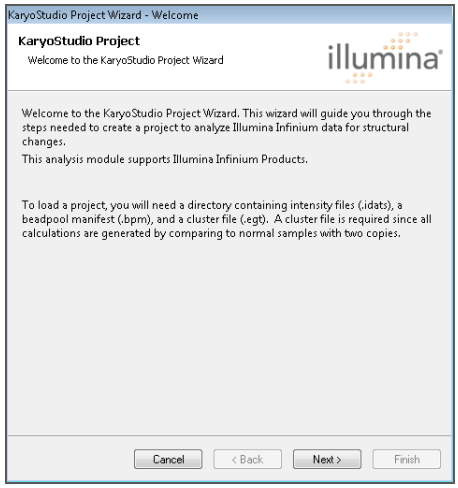
Before you create a project, make sure that you have the following files easily accessible in a known location on your computer:

- Intensity data files (*.idat files) from a BeadChip (or BeadChips) of interest
- A product manifest (*.bpm file) specific to the BeadChip product used
- A product cluster file (*.egt file) specific to the BeadChip product used

Creating a New Project

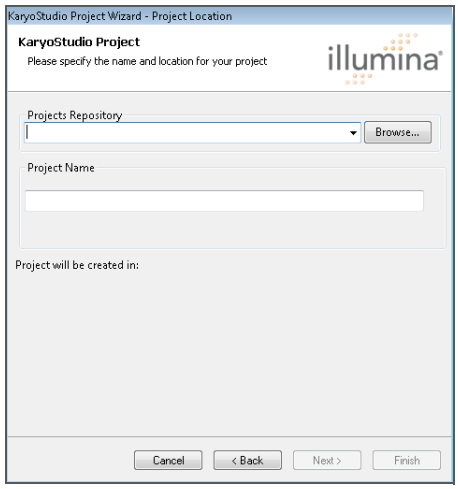
1 On the KaryoStudio main window, click **New Project**. The KaryoStudio Project Wizard opens.

Figure 3 KaryoStudio Project Wizard - Welcome Window



2 Click **Next** to advance to the Project Location window.

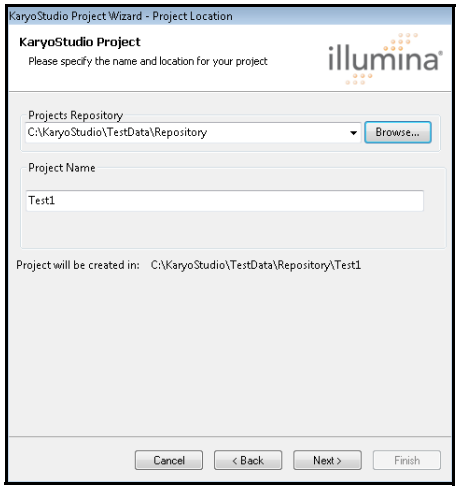
Figure 4 Wizard - Project Location Window



3 Browse to the folder where you would like to store your project.

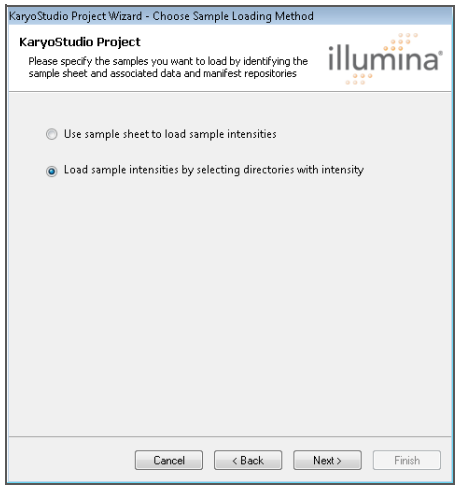
4 In the Project Name area, type a name for your project.

Figure 5 Wizard - Specifying Project Repository and Name



5 Click **Next** to advance to the Loading Sample Intensities window.

Figure 6 Wizard - Loading Sample Intensities Window



6 Do one of the following:

- If you have a sample sheet that you want to use with this project, select **Use sample sheet to load intensity data**.

  After selecting this method, continue to *Loading Sample Intensity Data Using a Sample Sheet* on page 11 to finish creating the project.
- If you want to load project data directly from intensity data files, select **Load sample intensities by selecting directories with intensity data**.

  After selecting this method, continue to *Loading Sample Intensity Data by Selecting Directories* on page 13 to finish creating the project.

NOTE

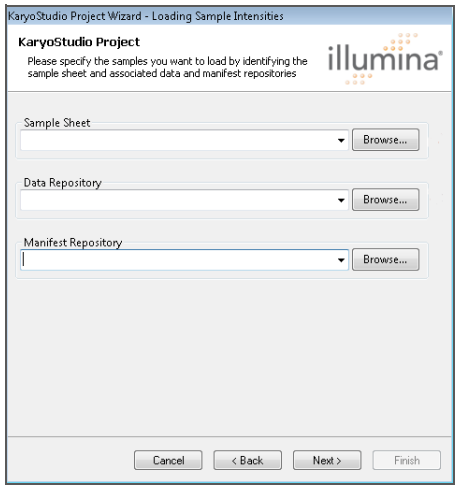
For optimal analysis of sex chromosomes, you must use a sample sheet. KaryoStudio uses the gender information in the sample sheet to correctly call detected regions or normal regions on the X and Y chromosomes. Without a sample sheet, data plots will be accurate, but detected regions might not be.

Loading Sample Intensity Data Using a Sample Sheet

Perform the following steps to load sample intensity data using a sample sheet.

1 From the Choose Sample Loading Method screen, click **Next** to advance to the Loading Sample Intensities window.

Figure 7 Wizard - Loading Sample Intensities Window



2 Browse tothe folder containing the sample sheet you want to use with this project.

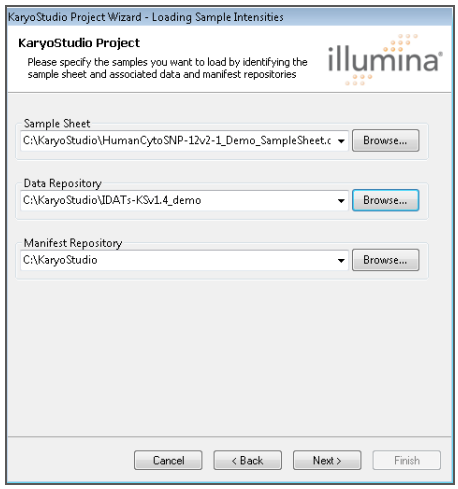
3 Browse to the folder containing the data you want to include in this project.

NOTE

You must specify a path in the wizard. However, if a different path is specified in the sample sheet, the path in the sample sheet overrides the path you provide in the wizard.

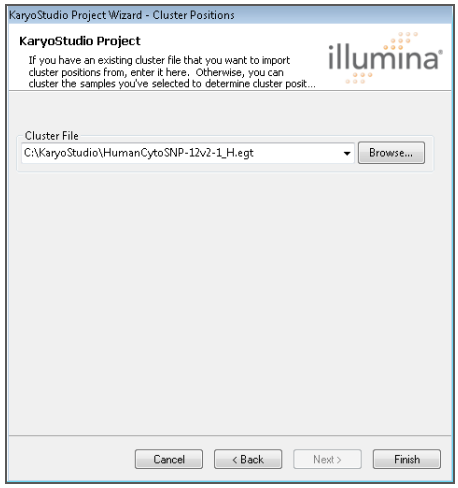
4 Browse to the folder containing the SNP manifest (*.bpm) you want to use with this project.

Figure 8 Wizard - Specifying Sample Sheet, Data Repository, and Manifest Repository



5 Click **Next** to advance to the Cluster Positions window.

Figure 9 Wizard - Cluster Positions Window



6 Browse to the cluster file (*.egt) you want to use with this project.

7 Click **Finish**.

KaryoStudio displays a progress bar while loading the intensity, manifest, and cluster files and running the cnvPartition algorithm on the data you included in the project.

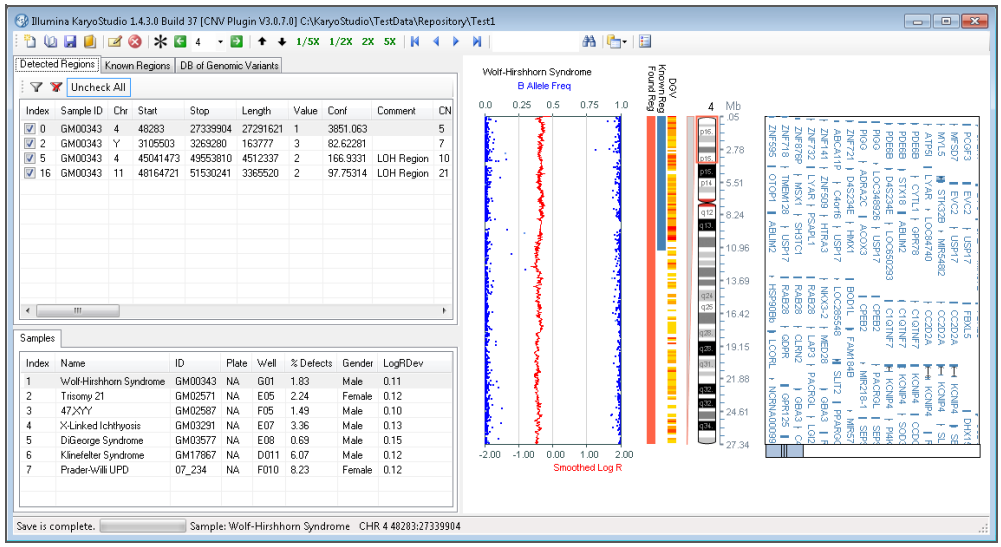
NOTE

The amount of time it takes to create your project is directly related to the Infinium product you are using and the number of samples you are processing. For example, it takes longer to load and analyze data from an Infinium HumanOmni1-Quad BeadChip than it does to load and analyze data from a HumanCytoSNP-12 BeadChip.

For more information about processing time, see the *KaryoStudio System Information and Benchmark Performance Technical Note* at http://www.illumina.com/Documents/products/technotes/technote_karyostudio_systems_benchmark_performance.pdf.

When KaryoStudio finishes processing and loading the data, the new project opens.

Figure 10 Project Created Using a Sample Sheet



8 Continue to *Assessing the Quality of Your Data* on page 16 for additional information.

Loading Sample Intensity Data by Selecting Directories

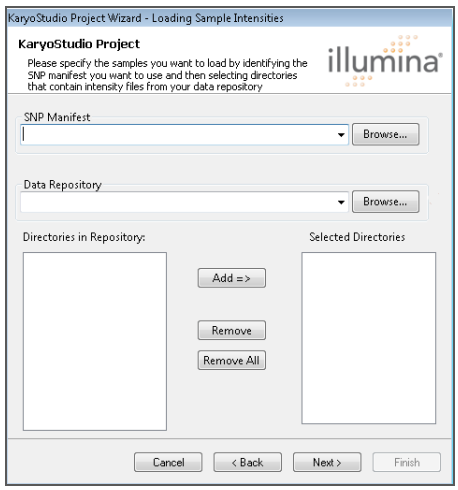
Perform the following steps if, instead of using a sample sheet, you prefer to load sample intensity data by selecting the directories that contain the data.

NOTE

If you want to include gender data for the samples in a KaryoStudio project, you must create the project using a sample sheet that includes gender data for the samples.

1 From the Choose Sample Loading Method screen, click **Next** to advance to the Loading Sample Intensities window.

Figure 11 Wizard - Loading Sample Intensities Window



2 Browse to the folder containing the SNP manifest (*.bpm) you want to use with this project.

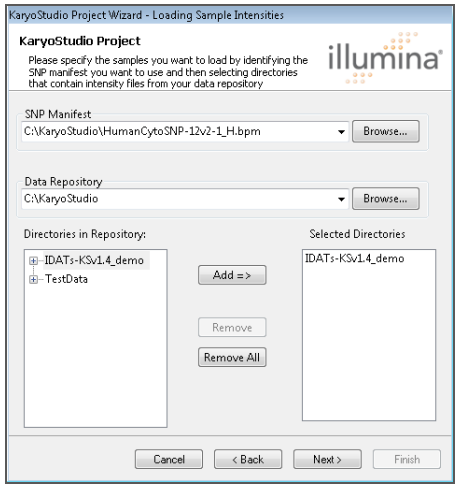
3 Browse to the folder containing the data you want to include in this project.
The subdirectories available for selection appear in the Directories in Repository area.

4 Add data to the project by selecting one or more directories from the list on the left and clicking **Add**.

NOTE

KaryoStudio accepts manifests only from a single Infinium product type. You cannot create a KaryoStudio project with multiple products or with different versions of the same product.

Figure 12 Wizard - Specifying SNP Manifest and Data



5 Click **Next** to advance to the Cluster Positions window.

6 Browse to the cluster file (*.egt) you want to use with this project.

7 Click **Finish**.

KaryoStudio displays a progress bar while loading the intensity, manifest, and cluster files and running the cnvPartition algorithm on the data you included in the project.

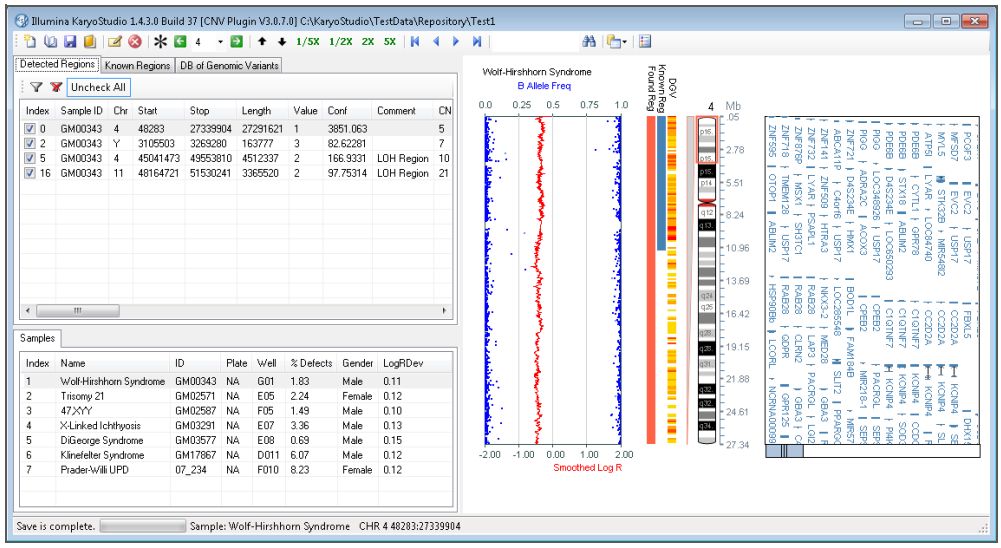
NOTE

The amount of time it takes to create your project is directly related to the Infinium product you are using and the number of samples you are processing. For example, it takes longer to load and analyze data from an Infinium HumanOmni1-Quad BeadChip than it does to load and analyze data from a HumanCytoSNP-12 BeadChip.

For more information about processing time, see the *KaryoStudio System Information and Benchmark Performance Technical Note* at http://www.illumina.com/Documents/products/technotes/technote_karyostudio_systems_benchmark_performance.pdf.

When KaryoStudio finishes processing and loading the data, the new project opens.

Figure 13 KaryoStudio Project Using Data from Directories



Opening an Existing Project

To open a project that was previously created and saved, perform the following steps:

1 From the KaryoStudio main window, click  **Open an Existing Project**.

2 Browse to the location of the project you want to open and click **Open**.
Your project is loaded into KaryoStudio and the project data appears in the tables and chromosome browser.

Assessing the Quality of Your Data

After your samples are loaded and analyzed in KaryoStudio, the software automatically calculates a percent defects score and a LogRDev for each sample. These metrics are useful when determining whether there are any issues with a specific sample. The percent defect score and LogRDev are displayed in the lower left in the Samples Table.

The percent defects score is calculated by summing the length of all of the detected regions within a particular sample of interest and dividing it by the length of the genome.

In control samples, such as non-constitutional samples, a percent defect score of less than 1% is expected. In most cases, the value is likely to be less than 0.5%. In samples that contain many aberrations, such as tumor samples, the percent defect score is expected to be much higher (in the 60-70% range). In addition, if there were any issues during the processing of your Infinium products, such as a misregistration or sample processing error, this score is expected to be abnormally high.

The LogRDev metric is a measure of the noise in the data, calculated as the standard deviation of the log R ratios for the SNPs on the autosomes. This metric is a simple indicator of the variation in the intensities measured for all markers from a particular sample. When analyzing relatively normal samples and following standard Illumina protocols for Infinium HD products, LogRDev is expected to be less than 0.3. For more information about interpreting LogRDev, see the *Interpreting Infinium Assay Data for Whole-Genome Structural Variation Technical Note* at http://www.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf.

NOTE

Illumina recommends analyzing control samples along with your experimental samples to ensure that there are minimal processing issues. This provides a good point of reference in the event that any issues do arise. For additional information about how to QC your data, refer to the KaryoStudio FAQs on the web or contact Illumina Technical Support.

User Interface

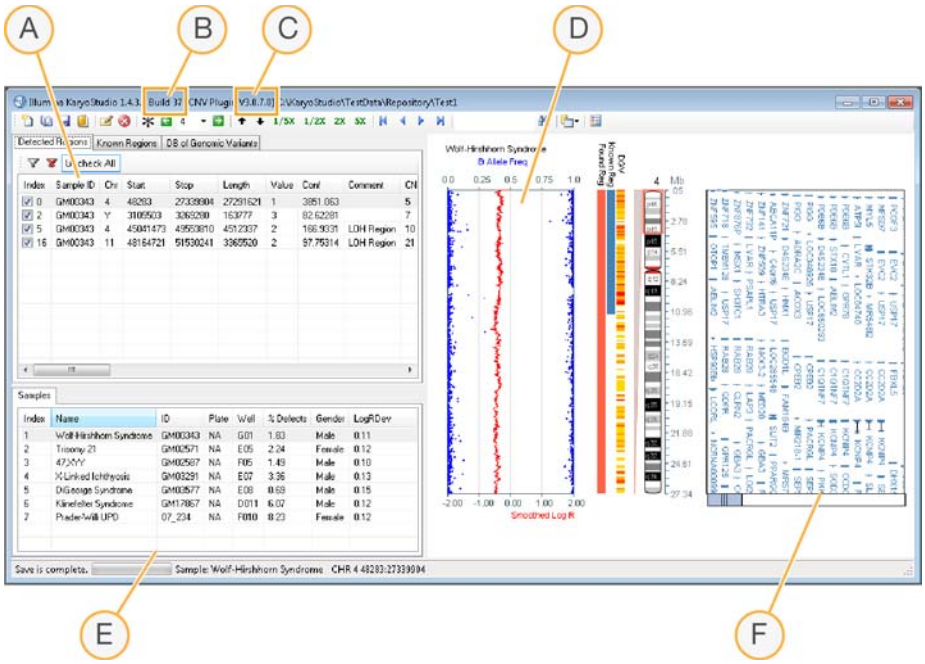
Chapter 3

Introduction

The KaryoStudio user interface provides tools for loading intensity files, browsing detected regions, comparing detected regions to known regions, and displaying them graphically.

Figure 14 shows KaryoStudio's screen configuration.

Figure 14 KaryoStudio Screen Configuration



- **A** Detected Regions Table
- **B** Current Genome Build
- **C** Current version of cnvPartition Algorithm
- **D** Chromosome Browser
- **E** Samples Table
- **F** Gene Information

Main Window Menus

Table 1 lists toolbar button options available from KaryoStudio's main window.

Table 1 KaryoStudio Toolbar Button Options

Toolbar Button	Name	Function
	Create New Project	Create a new project
	Open an Existing Project	Open a previously generated KaryoStudio project
	Save the Current Project	Save the current KaryoStudio project to your computer
	Close the Current Project	Close the current KaryoStudio project
	Write Reports to Project Directory	Generate a *.pdf report and save it to your computer
	Close Application	Close KaryoStudio
	Expand to Chromosome	Expands the current Chromosome Viewer view from selected detected region to entire chromosome
	Previous Chromosome	Display previous chromosome in the Chromosome Viewer
	Select Chromosome	Select chromosome to display in the Chromosome Viewer
	Next Chromosome	Display next chromosome in the Chromosome Viewer
	Pan Up 1/4 Step	Adjust view slightly towards the beginning of the chromosome
	Pan Down 1/4 Step	Adjust view slightly towards the end of the chromosome
	Zoom in 5X	Adjust chromosome view to five times closer
	Zoom in 2X	Adjust chromosome view to two times closer
	Zoom out 2X	Expand chromosome view by two times
	Zoom out 5X	Expand chromosome view by five times
	First Aberration	Jump to the first detected region in the Detected Regions Table.

Table 1 KaryoStudio Toolbar Button Options (Continued)

Toolbar Button	Name	Function
	Previous Aberration	Jump to the previous detected region in the Detected Regions Table.
	Next Aberration	Jump to the next detected region in the Detected Regions Table.
	Last Aberration	Jump to the last detected region in the Detected Regions Table.
	Settings	Adjust KaryoStudio settings for data plots and generating reports.
	Display Log	Display a log of all actions performed by KaryoStudio.

Table Window

The table window includes three tables:

- Detected Regions Table
- Known Regions Table
- Database of Genomic Variants

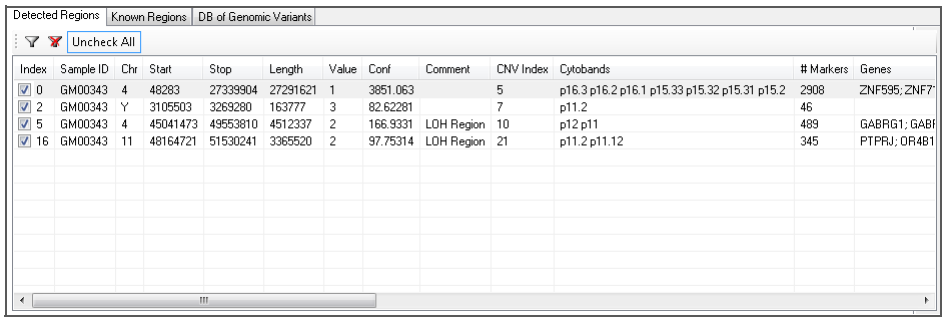
These tables are described in the following sections.

Detected Regions Table

The Detected Regions Table lists the regions passing the filter settings that were identified by the cnvParition algorithm ("detected regions") and provides the following information:

- Sample name in which a region is found
- Chromosome on which a region appears
- Start and stop positions of region
- Length of region
- Value (Copy Number)
- Conf (Confidence Score)
- Comment (if any)
- CNV Index
- Cytobands
- Number of Markers
- Genes

Figure 15 Detected Regions Table



The **Filter Table** and **Clear Filter** buttons allow you to filter the detected regions that appear in the Detected Regions Table and track. For more information, see *Filtering Detected Regions* on page 41.

The **Check All/Uncheck All** buttons allow you to include or exclude all displayed detected regions from cytogenetics reports.

The columns in the Detected Regions Table are listed and described in Table 2.

Table 2 Detected Regions Table Columns

Column	Description
Index	Identifier given to each region identified by the cnvPartition algorithm across the entire KaryoStudio project (all samples)

Table 2 Detected Regions Table Columns (Continued)

Column	Description
Sample ID	Sample name • If you use a sample sheet to load your data, sample names are the names assigned to each sample in the sample sheet. • If you load your data without a sample sheet, sample names are barcodes.
Chr	Chromosome number of the detected region
Start	Position of the first base of a detected region
Stop	Position of the last base of a detected region
Length	Length (in base pairs) of a detected region
Value	Estimate of the physical copy number of a detected region: • 0 indicates a homozygous deletion (loss of both copies) • 1 indicates a hemizygous deletion (loss of one copy) • 2 indicates a copy-neutral loss of heterozygosity (e.g., UPD or autozygosity) • 3 indicates a duplication (gain of one copy) • 4 indicates a copy number of 4 or above
Conf	Confidence score calculated for each detected region by the cnvPartition algorithm. The confidence score is relative within a specific sample. A higher score represents higher confidence. Values of ~50 or higher tend to reflect regions with high confidence. Larger regions tend to have higher confidence scores.
Comment	User-entered comment
CNV Index	Number assigned to a detected region within a specific sample. The range begins at 0.
Cytobands	List of cytobands in the detected region
#Markers	Number of markers (SNPs or intensity-only probes) in a detected region
Genes	RefSeq genes present within the detected region

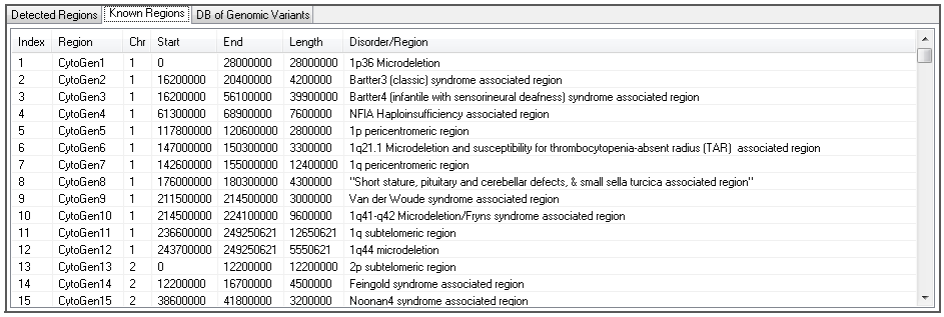
NOTE

Changing your selection in the Detected Regions Table changes what is displayed in the Chromosome Browser to the right of the table.

Known Regions Table

The Known Regions table lists all of the known regions that have been loaded into KaryoStudio.

Figure 16 Known Regions Table



Detected Regions | Known Regions | DB of Genomic Variants

Index	Region	Chr	Start	End	Length	Disorder/Region
1	CytoGen1	1	0	28000000	28000000	1p36 Microdeletion
2	CytoGen2	1	16200000	20400000	4200000	Bartter3 (classic) syndrome associated region
3	CytoGen3	1	16200000	56100000	39900000	Bartter4 (infantile with sensorineural deafness) syndrome associated region
4	CytoGen4	1	61300000	68900000	7600000	NFIA Haploinsufficiency associated region
5	CytoGen5	1	117800000	120600000	2800000	1p pericentromeric region
6	CytoGen6	1	147000000	150300000	3300000	1q21.1 Microdeletion and susceptibility for thrombocytopenia-absent radius (TAR) associated region
7	CytoGen7	1	142600000	155000000	12400000	1q pericentromeric region
8	CytoGen8	1	176000000	180300000	4300000	"Short stature, pituitary and cerebellar defects, & small sella turcica associated region"
9	CytoGen9	1	211500000	214500000	3000000	Van der Woude syndrome associated region
10	CytoGen10	1	214500000	224100000	9600000	1q41-q42 Microdeletion/Fryns syndrome associated region
11	CytoGen11	1	236600000	249250621	12650621	1q subtelomeric region
12	CytoGen12	1	243700000	249250621	5550621	1q44 microdeletion
13	CytoGen13	2	0	12200000	12200000	2p subtelomeric region
14	CytoGen14	2	12200000	16700000	4500000	Feingold syndrome associated region
15	CytoGen15	2	38600000	41800000	3200000	Noonan4 syndrome associated region

In this table, Illumina provides information about commonly affected regions by default. However, you can also edit this table based on the requirements of your project.

The columns in the Known Regions Table are listed and described in Table 3.

Table 3 Known Regions Table Columns

Column	Description
Index	Numeric identifier assigned to known regions
Region	Specific name assigned to a known region
Chr	Chromosome on which the known region appears
Start	Position of the first base pair of the known region
End	Position of the last base pair of the known region
Length	Length of the known region
Disorder/Region	Name of the region, or the syndrome associated with chromosomal aberrations in this region

NOTE

The default known regions file included with KaryoStudio is based mainly on affected chromosomal bands, so the base pair start and stop positions might be approximate.

Editing the Known Regions Table

Illumina has prepopulated the Known Regions table with a default list of regions of the genome that are of interest to cytogeneticists.

You can edit or add additional known regions files in order to better represent the regions of interest to your cytogenetics lab. You can also have multiple known regions files for different analysis workflows, or for use with different sample types.

To create a new Known Regions file

1 Create a new file using Microsoft Excel or a word processing program, or start with an existing Known Regions file.

2 Verify that the required column headers are included (see the bulleted list below).

3 Enter information in each row for each known region.

4 Save the file as a tab-delimited file (*.txt) in the KaryoStudio program directory.
The default KaryoStudio program directory is *C:\Program Files\Illumina\KaryoStudio*.

To edit a Known Regions file

1 Navigate to the KaryoStudio program directory.
The default KaryoStudio program directory is *C:\Program Files\Illumina\KaryoStudio*.

2 Open the known regions file to be edited (The default file is *KnownRegionsTable.txt*) in a text editing tool such as Microsoft Excel.

CAUTION
Any of the rows in this file can be edited; however, you cannot delete any columns, as this will crash the software.

3 Make changes to the file as desired, by adding rows, deleting rows, or adjusting the parameters for each row.
Some of the items you can adjust are:

- Region
- Disorder
- Chr #
- Start
- End
- Length
- Length minus overlap

For example, if you would like to more precisely identify the start and stop positions for a particular region, you can adjust them in this file. You can also add new rows to this file. This might be useful if you would like to cross-match regions in future samples to regions your lab has tracked in the past.

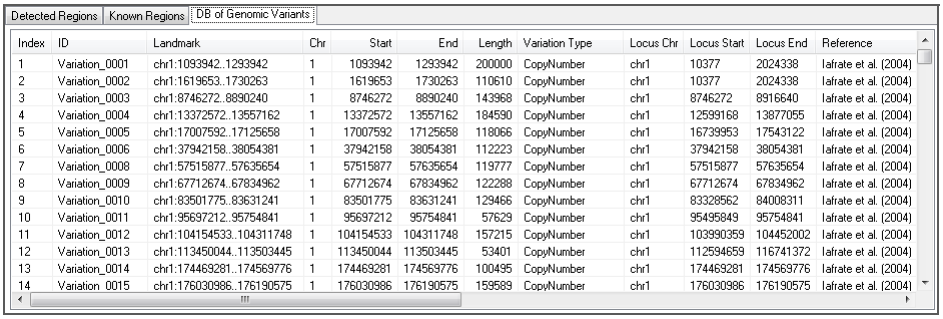
4 After you edit the file, you must save your changes as a *.txt file.

NOTE
Illumina recommends saving a copy of the original known regions table in the event that you would like to revert back to the default file that is originally provided with the software.

Database of Genomic Variants

The Database of Genomic Variants table is a list of all of the CNV regions found in the latest build of the DGV. Information from this database is loaded into KaryoStudio as a text file and included for cross-matching purposes.

Figure 17 Database of Genomic Variants


Detected Regions | Known Regions | DB of Genomic Variants

Index	ID	Landmark	Chr	Start	End	Length	Variation Type	Locus Chr	Locus Start	Locus End	Reference
1	Variation_0001	chr1:1093942..1293942	1	1093942	1293942	200000	CopyNumber	chr1	10377	2024338	Iafrate et al. (2004)
2	Variation_0002	chr1:1619653..1730263	1	1619653	1730263	110610	CopyNumber	chr1	10377	2024338	Iafrate et al. (2004)
3	Variation_0003	chr1:8746272..8890240	1	8746272	8890240	143968	CopyNumber	chr1	8746272	8916640	Iafrate et al. (2004)
4	Variation_0004	chr1:13372572..13557162	1	13372572	13557162	184590	CopyNumber	chr1	12599168	13877055	Iafrate et al. (2004)
5	Variation_0005	chr1:17007592..17125658	1	17007592	17125658	118066	CopyNumber	chr1	16739953	17543122	Iafrate et al. (2004)
6	Variation_0006	chr1:37942158..38054381	1	37942158	38054381	112223	CopyNumber	chr1	37942158	38054381	Iafrate et al. (2004)
7	Variation_0008	chr1:57515877..57635654	1	57515877	57635654	119777	CopyNumber	chr1	57515877	57635654	Iafrate et al. (2004)
8	Variation_0009	chr1:67712674..67834962	1	67712674	67834962	122288	CopyNumber	chr1	67712674	67834962	Iafrate et al. (2004)
9	Variation_0010	chr1:83501775..83631241	1	83501775	83631241	129466	CopyNumber	chr1	83328562	84008311	Iafrate et al. (2004)
10	Variation_0011	chr1:95697212..95754841	1	95697212	95754841	57629	CopyNumber	chr1	95495849	95754841	Iafrate et al. (2004)
11	Variation_0012	chr1:104154533..104311748	1	104154533	104311748	157215	CopyNumber	chr1	103990359	104452002	Iafrate et al. (2004)
12	Variation_0013	chr1:113450044..113503445	1	113450044	113503445	53401	CopyNumber	chr1	112594659	116741372	Iafrate et al. (2004)
13	Variation_0014	chr1:174469281..174569776	1	174469281	174569776	100495	CopyNumber	chr1	174469281	174569776	Iafrate et al. (2004)
14	Variation_0015	chr1:176030986..176190575	1	176030986	176190575	159589	CopyNumber	chr1	176030986	176190575	Iafrate et al. (2004)

The version of the DGV delivered with KaryoStudio v1.4 is:

http://projects.tcag.ca/variation/downloads/variation.hg19.v10.nov.2010.txt

For information about updating KaryoStudio with the latest DGV build, see page 55.

The columns in the Database of Genomic Variants are listed and described in Table 4.

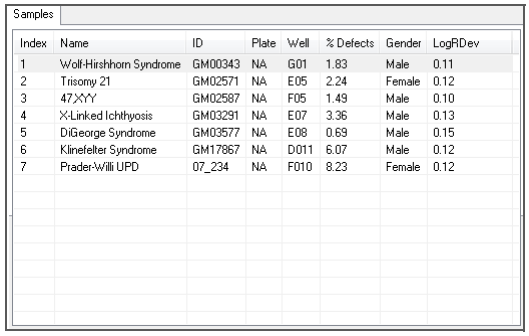
Table 4 Database of Genomic Variants Columns

Column	Description
Index	Number used as a simple identifier for regions
ID	Identifier assigned to a region by the Database of Genomic Variants; stays consistent over time
Landmark	BAC clone identifier for a specific region
Chr	Chromosome of a region
Start	Start position of a region
End	End position of a region
Length	Size of a region
Variation Type	Type of CNV
Locus Chr	General identifier that might contain multiple variation IDs
Locus Start	Start position of a locus region
Locus End	End position of a locus region
Reference	Publication reference showing where this region has been published
PubMed ID	Entrez PubMed ID of a publication linked to a specific region
Reference	Author of the study
Gain	Total number of patients in which a gain was seen
Loss	Total number of patients in which a loss was seen
Total Gain Loss	Total number of patients in which a gain or loss was seen
Sample Size	Total number of patients in the study

Samples Table

The Samples table lists all of the samples you have loaded into KaryoStudio for the current project.

Figure 18 Samples Table



Samples

Index	Name	ID	Plate	Well	% Defects	Gender	LogRDev
1	Wolf-Hirshhorn Syndrome	GM00343	NA	G01	1.83	Male	0.11
2	Trisomy 21	GM02571	NA	E05	2.24	Female	0.12
3	47,XYY	GM02587	NA	F05	1.49	Male	0.10
4	X-Linked Ichthyosis	GM03291	NA	E07	3.36	Male	0.13
5	DiGeorge Syndrome	GM03577	NA	E08	0.69	Male	0.15
6	Klinefelter Syndrome	GM17867	NA	D011	6.07	Male	0.12
7	Prader-Willi UPD	07_234	NA	F010	8.23	Female	0.12

When you select a sample in the Samples Table, the detected regions for that sample appear in the Detected Regions Table above.

TIP

You can choose more than one sample by pressing and holding the **Ctrl** button on your keyboard while selecting multiple samples of interest. As you select each sample, the Detected Regions Table above is dynamically updated.

The columns in the Samples Table are listed and described in Table 5.

Table 5 Samples Table Columns

Column	Description
Index	Number assigned to the sample for sorting purposes
Name	Name or barcode of a sample from a BeadChip
ID	Barcode of a BeadChip
Plate	Location on the sample preparation plate
Well	Well on the sample preparation plate
% Defect	Score given to each sample based on the number of detected regions. This value is the sum of the length of all detected regions per sample divided by the length of the genome.
Gender	Gender of the sample, as provided in the sample sheet. KaryoStudio uses the gender information to correctly interpret detected regions on the sex chromosomes.
LogRDev	Standard deviation of the log R ratios of the sample

Chromosome Browser

The Chromosome Browser includes representations of your data in the form of:

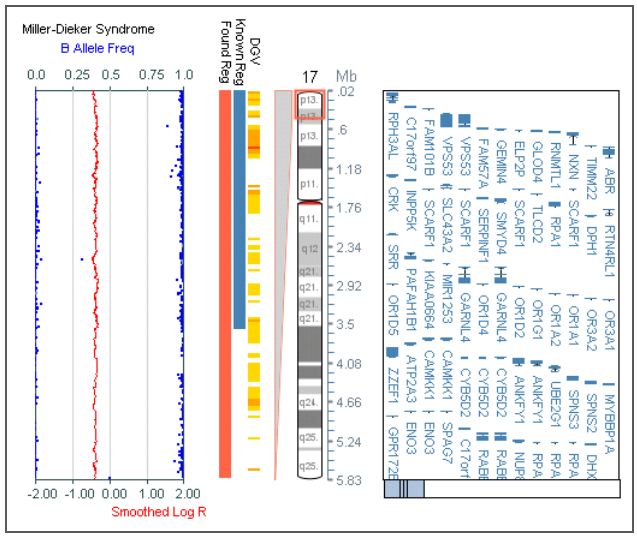
- B allele frequency (genotyping information) represented by blue dots
- Log R ratio (intensity information) represented by a grey line
- Smoothed log R ratio represented by a red line

NOTE

The B allele frequency for intensity-only probes is represented by light blue dots and is not used in copy number calculations.

In addition, there is an ideogram of the chromosome, found and known regions, information from the DGV (Database of Genomic) Variants, and gene information.

Figure 19 Chromosome Browser



You can adjust the display parameters from the **Settings** tab.

- To hide or display data types in the plot, go to **Settings** | **Data Plot** and clear or select **Smoothed LogR**, **B Allele Freq**, or **LogR**.
- To adjust the Log R Ratio axis to fit the data, select **Settings** | **AutoScale** | **LogR Axis**. By default, the axis range is -2.0 to 2.0.

Search Function

You can navigate to a particular location in the chromosome browser by entering a search term in the search field. Searches may be of the following types:

- **Chromosomal coordinate**

  Enter the chromosome number, a colon, and the start and stop coordinates separated by a hyphen.

  [Example] Enter `1:1-2000000` to display the first 2Mb of chromosome 1.
- **Cytoband**

  Enter the location in ISCN notation.

  [Example] Enter `1p22.1` or `1p22`.
- **Gene name**

Entering text into the search box displays the first gene that matches the search term, starting with the chromosome currently displayed in the browser. To search a particular chromosome, start your search with the chromosome number and a colon.

[Example] Enter `6:VEGF` to search for VEGFA rather than VEGFC on chromosome 6.

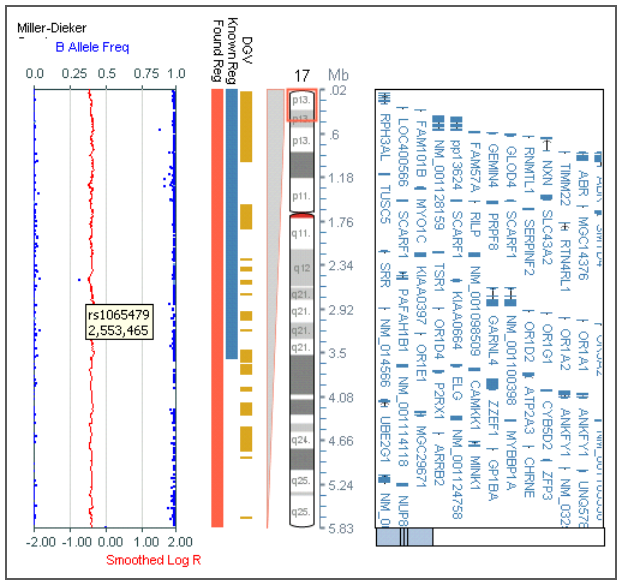
Navigating the Chromosome Browser

In addition to using the toolbar navigation buttons, you can zoom in and scroll by clicking and dragging the red box on the ideogram using your mouse.

- To scroll, click and drag the red box up or down along the ideogram.
- To zoom in our out, drag the top or bottom edge of the square up or down along the ideogram. The size of the box indicates the size of the chromosomal region displayed in the data plot.

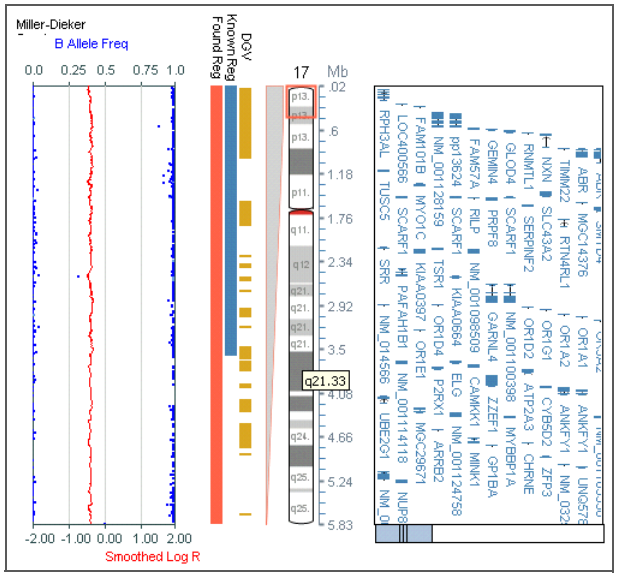
If you place the cursor over a data point, a tooltip featuring the SNP name and coordinates appears.

Figure 20 SNP Name Tooltip



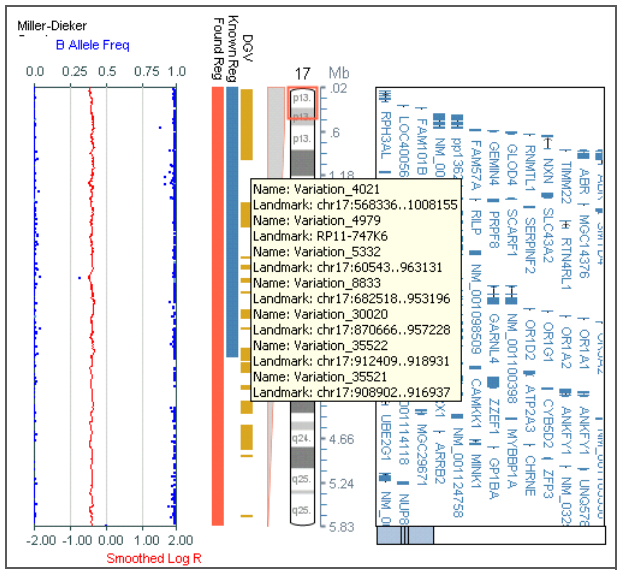
If you place the cursor over a cytoband on the ideogram, a tooltip featuring the cytoband name appears.

Figure 21 Cytoband Tooltip



If you place the cursor over a region found in the DGV track, a DGV tooltip appears.

Figure 22 DGV Tooltip



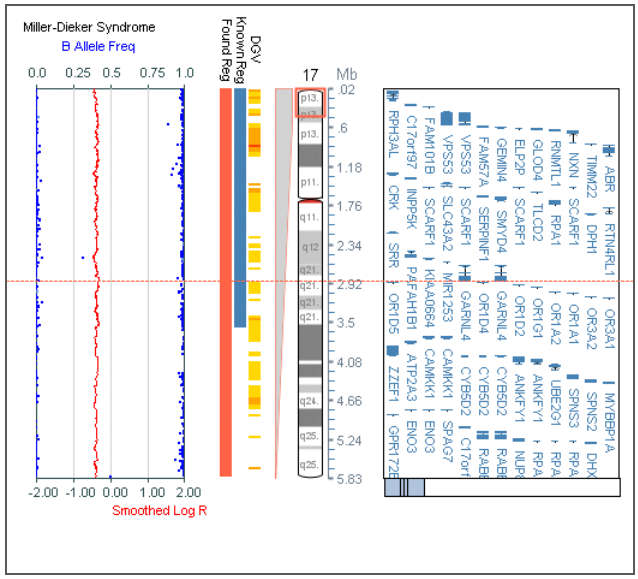
The DGV track is multicolored. The colors indicate the numbers of DGV variation regions overlapping at a given point, as described below:

Table 6 DGV Track Colors

Color	Number of DGV Variation Regions Overlapping at a Given Point
light gold	1-2
orange	3-10
orange-red	11-20
red	>20

If you click anywhere along the ideogram, an alignment cursor appears as a dashed red line.

Figure 23 Chromosome Browser Alignment Cursor



In the gene display area to the right, gene names and transcript locations appear, as in the RefSeq database. To open RefSeq, right-click a gene name and select **Open RefSeq**.

The elements of the Chromosome Browser are listed and described in Table 7.

Table 7 Chromosome Browser Elements

Element	Description
Sample Name	Shows the sample being displayed
B allele frequency plots	Shows B allele frequency data for intensity-only markers in blue
Log R ratio plot	Shows smoothed log R ratio data in red or log R ratio data in grey
Found Regions	Track next to the chromosome showing a detected region. Gains in copy number are indicated in green and losses are indicated in red.
Known Regions	Blue boxes represent known regions in this area. Directly correlates with the information in the Known Regions Table, populated from the Known Regions file currently selected in the Filter Settings table.
DGV	Track next to the chromosome shows correlation with information in the DGV (Database of Genomic Variants)
Chromosome	Vertical ideogram of the chromosome
Gene information	All available RefSeq gene information Note: For larger regions, such as whole chromosomes, the gene information appears very condensed. To refine the information use the zoom buttons.

NOTE

Pseudoautosomal region (PAR) markers and detected regions in the PAR are displayed as being on the X chromosome.

Karyotype View

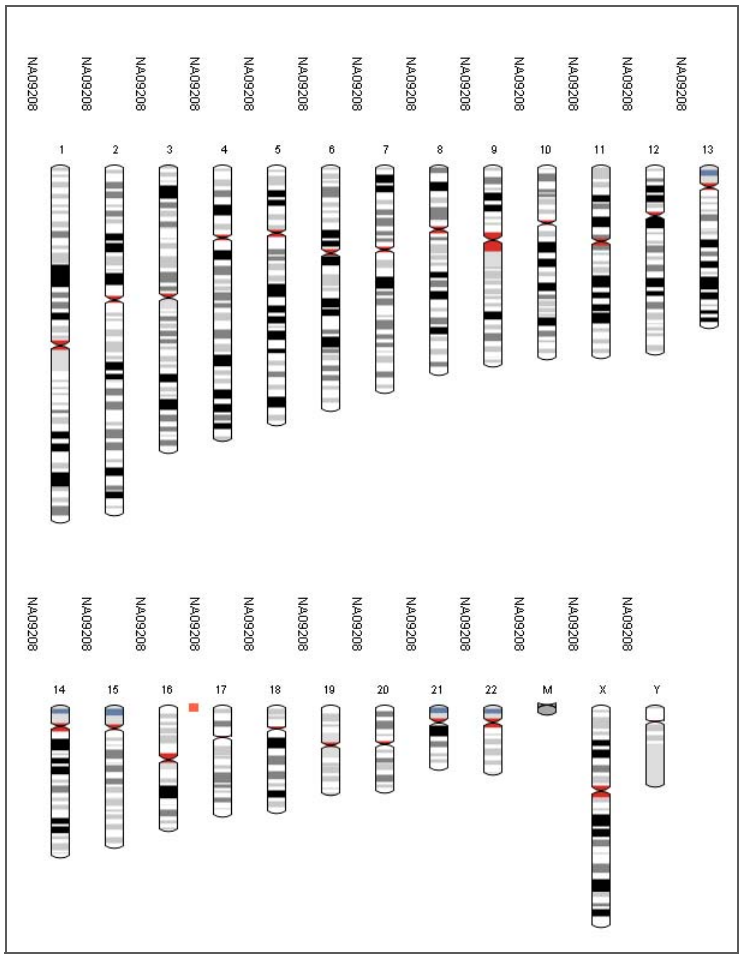
The Karyotype View window displays a whole-genome view of the detected regions for one or more samples. This view can be used to gain an initial appreciation for the variation found in a sample, or as a clickable visual table of contents for the detected regions in a sample.

To use the karyotype view, perform the following steps:

1 In the Samples table, right-click a sample and select **Show Karyotype**.
The Karyotype View window appears. Detected regions are displayed in a track for each sample selected. Regions are color coded:

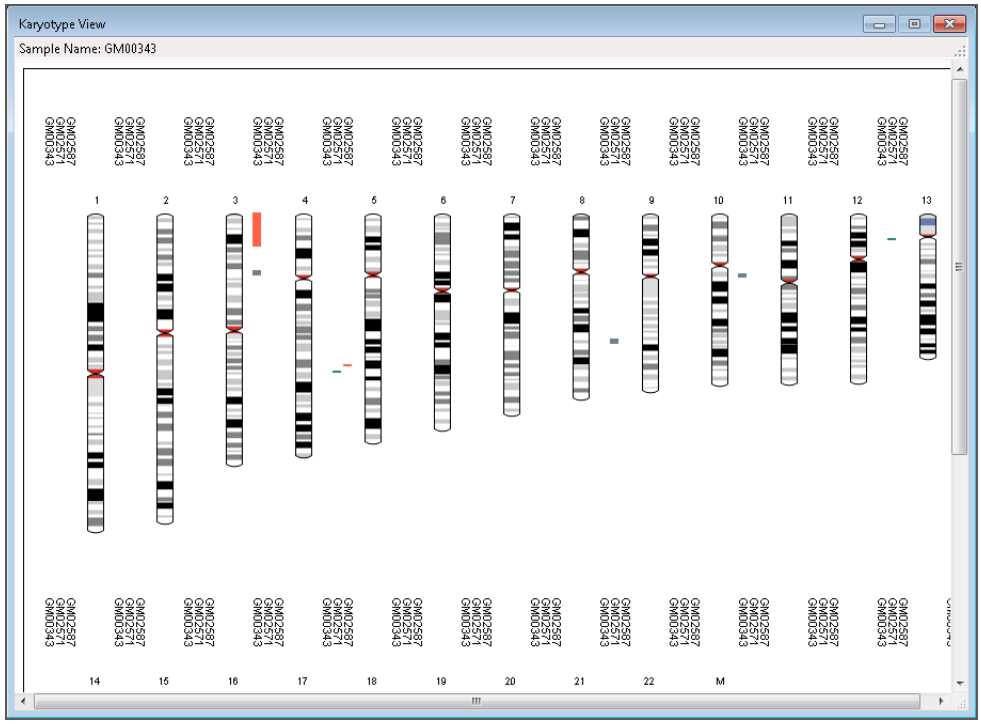
- green = gain
- red = loss
- purple = copy-neutral event

Figure 24 Karyotype View Window, One Sample Selected



2 To add or remove additional samples from the Karyotype View, press and hold the **Ctrl** key and click each sample in the Samples table.
Information for all selected samples appears in the KaryoStudio View window.

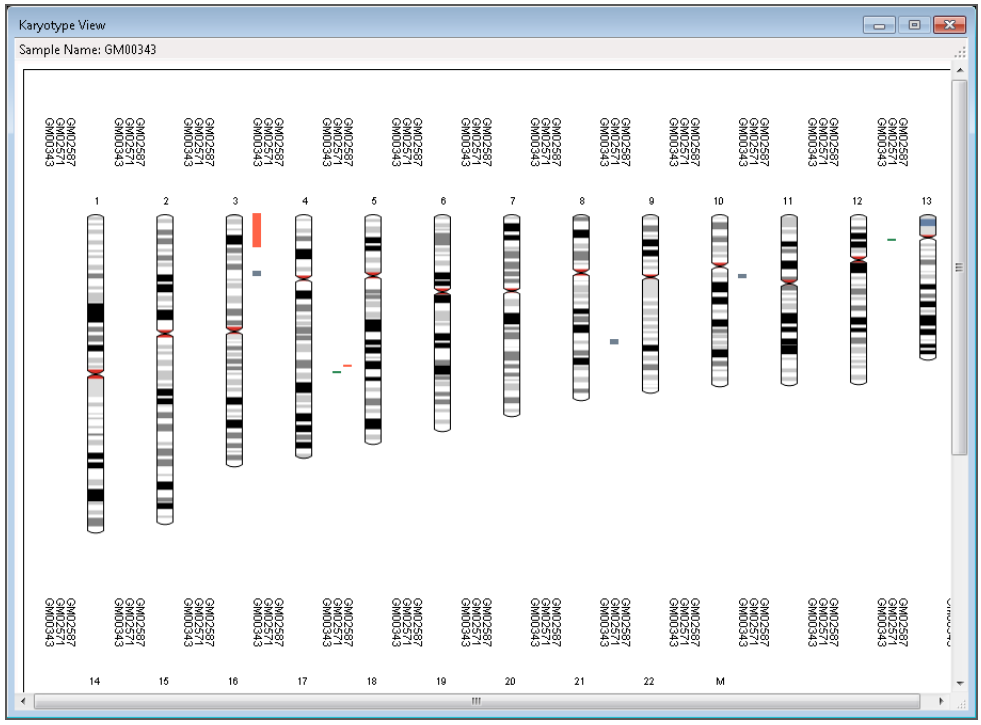
Figure 25 Karyotype View Window, Three Samples Selected



3 To display information about a detected region in the Found Region information bar at the top of the window, hover over a detected region.

The information displayed includes: sample ID, chromosome, start and end coordinates, and copy number value.

Figure 26 Karyotype View, Found Region Information Bar



4 To display a region in the chromosome browser, click a detected region.

Figure 27 Detected Region in Karyotype View Window

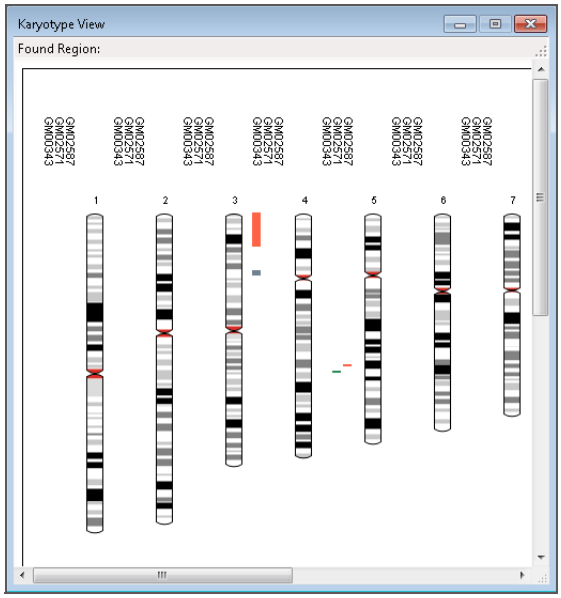
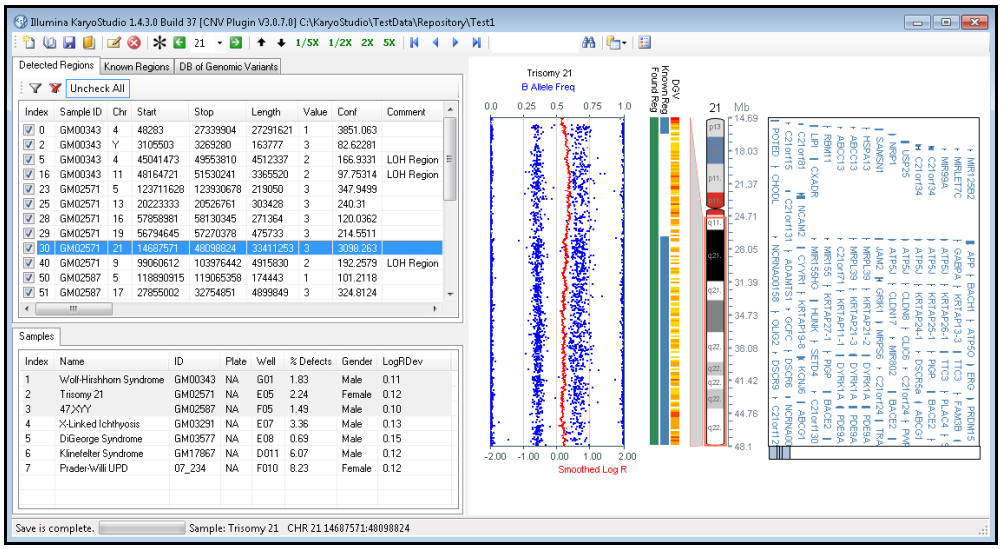


Figure 28 Detected Region in Chromosome Browser



Duo and Trio View

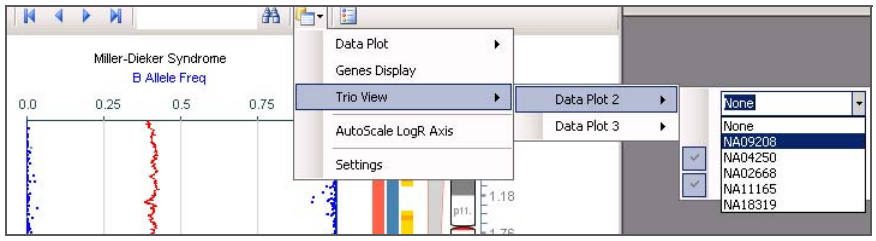
Data plots from multiple samples can be displayed simultaneously in the chromosome browser. Duo and trio view can be used to analyze, for example, a parent-parent-child trio, a control sample, or different display settings for two plots from the same sample.

To use the duo and trio view, perform the following steps:

1 In the Samples table, select the samples you want to display.

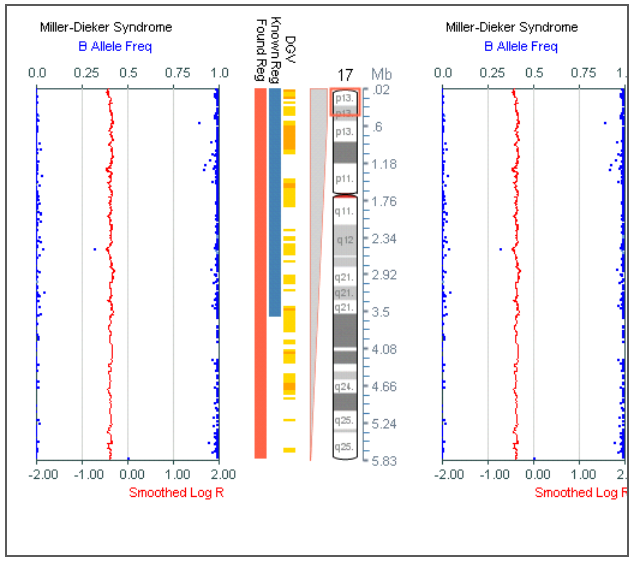
2 To add a second data plot, select **Settings** | **Trio View** | **Data Plot 2** | and select a sample from the dropdown menu.

Figure 29 Selecting Additional Data Plot to Display



The second data plot appears in chromosome browser to the right of the original data plot.

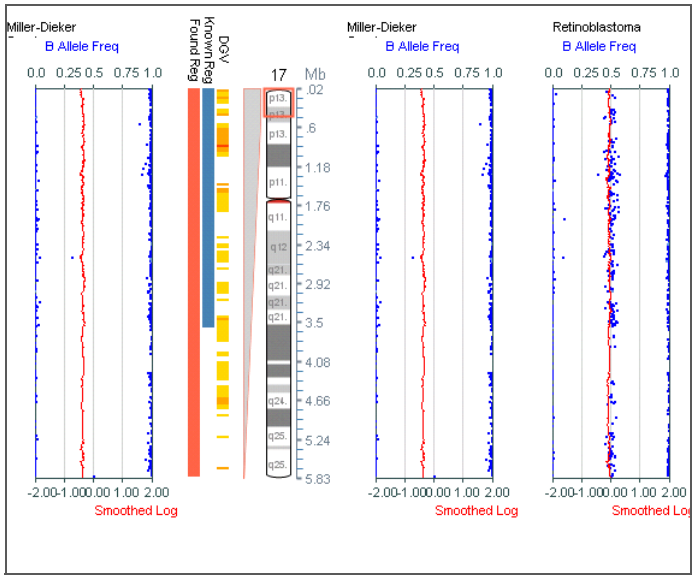
Figure 30 Chromosome Browser, Duo View



3 **[Optional]** To add a third data plot, select **Settings** | **Trio View** | **Data Plot 3** | and select a sample from the dropdown menu.

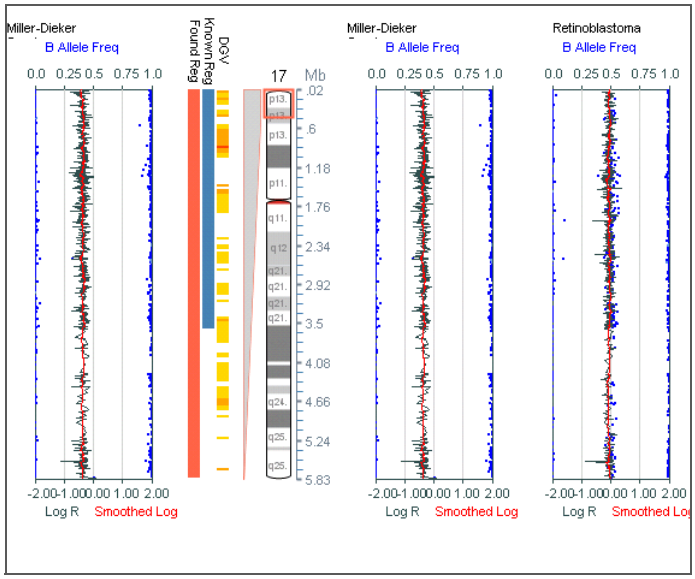
The third data plot appears in the chromosome browser to the right of the second data plot.

Figure 31 Chromosome Browser, Trio View



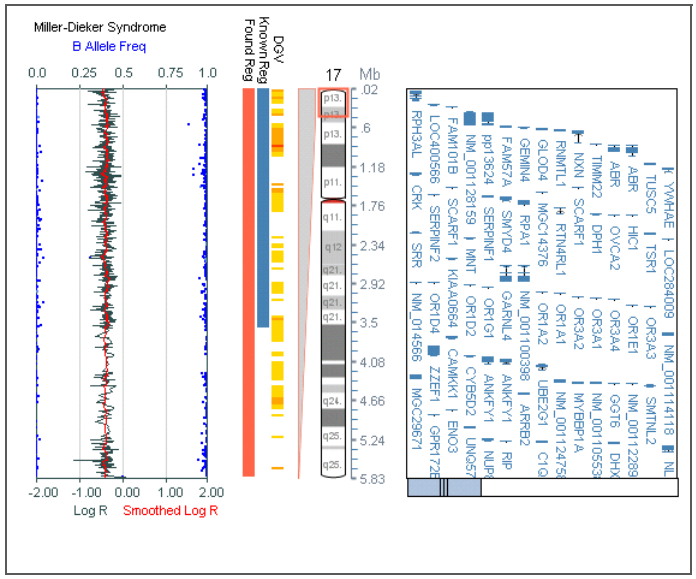
To select additional settings for each data plot, go to **Settings** | **Trio view** | **Data Plot**, then select or clear the **Log R**, **B Allele Freq**, and/or **Smoothed Log R** check boxes.

Figure 32 Trio View, Additional Plot Settings Selected



4 To return to the a single-plot view with the genes display visible, select **Settings** | **Genes Display**.

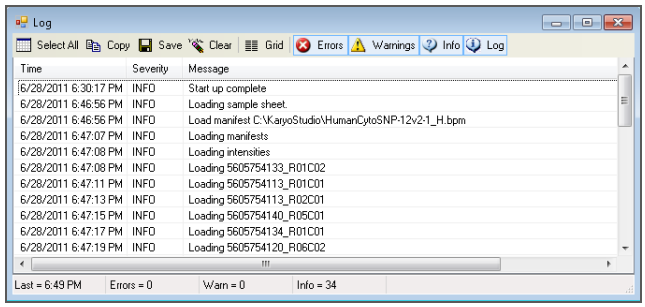
Figure 33 Single Sample View with Gene Display



Log Window

The Log window displays information about the projects and tracks for this analysis as they are being loaded into the software.

Figure 34 Log Window



The elements of the Log window are listed and described in Table 8.

Table 8 Log Window Elements

Element	Description	Toolbar Button (if used)
Select All	Selects all log entries	
Copy	Copies log entries to the clipboard	
Save	Saves all log entries	
Clear	Clears all log entries	
Grid	Toggles the grid on and off	
Time	Displays the time the log entry was generated	
Severity	Displays the severity of the log entry	
Message	Displays the text description of the log entry	
Source	Displays the source of the log entry	

Analyzing Detected Regions

Chapter 4

Introduction

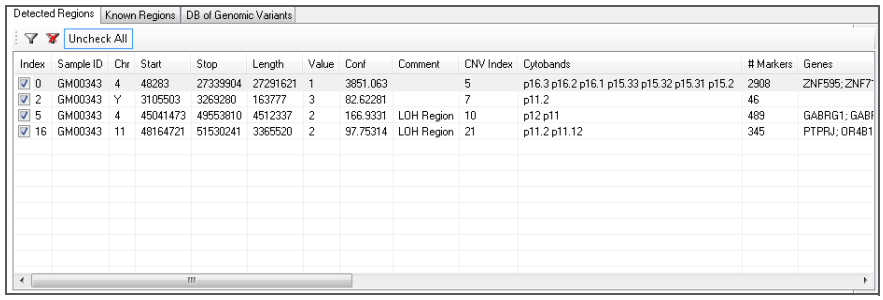
KaryoStudio includes Illumina's cnvPartition algorithm, which is designed to scan whole-genome Infinium data for intensity changes resulting from changes in copy number. Once you have completed entering information into the Project Wizard, your project data is automatically scanned using the cnvPartition algorithm.

Detected aberrations are displayed in the Detected Regions Table. Included in this table is information about each aberration including:

- the sample in which it was found
- the chromosome start and stop position
- its size (length in bases)
- an estimate of the copy number (**Value** column)
- a confidence value (relative score)
- its cytobands
- the number of SNPs it contains
- the genes present in the detected region

Once you have created a project, all detected regions are displayed in the Detected Regions Table and listed in the order in which they were found in each sample (Figure 35).

Figure 35 Detected Regions Table



This chapter describes how to view and analyze the Detected Regions in your data.

Filtering Detected Regions

When a project is created, the cnvPartition algorithm discovers all regions of aberrant copy number based on the cnvPartition configuration settings.

By default, cnvPartition identifies all regions with confidence value greater than 35 and all copy-neutral LOH regions larger than 1Mb. However, these confidence and size thresholds are likely not stringent enough for most uses.

More specific filtering parameters for size, number of markers, and confidence threshold can be entered in KaryoStudio. After entering the desired parameters, only regions of potential interest which meet these criteria are displayed in the Detected Regions Table and chromosome browser.

NOTE

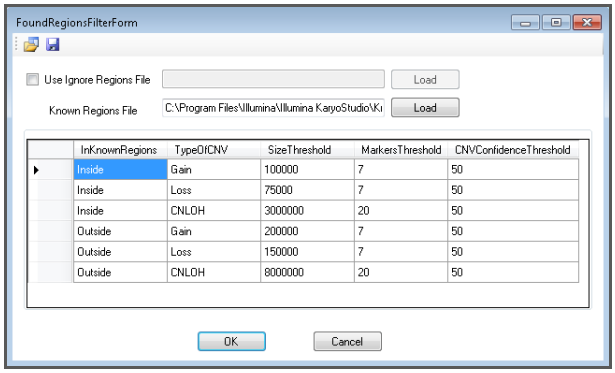
For information about adjusting the cnvPartition configuration file, see the cnvPartition documentation.

After a project is created, filter settings are applied to determine which regions display in the Detected Regions Table. Default settings are applied to a new project, but these parameters can be changed easily.

To change the filter settings, perform the following steps:

1 Click  (Filter Table).
The FoundRegionsFilterForm opens. The Known Regions File to be used in the project is listed in the second text field.

Figure 36 FoundRegionsFilterForm



2 To change the Known Regions File, click **Load** and browse to the file you would like to use.
See Appendix C of this document for information about formatting a Known Regions File.
Different classes of detected regions can be assigned different filter settings:
- Settings for regions that at least partially overlap with regions defined in the selected Known Regions File are labeled **Inside** in the InKnownRegions column.
- Settings for regions that do not overlap with regions defined in the selected Known Regions File are labeled **Outside** in the InKnownRegions column.
- Settings for Gains, Losses, and Copy-Neutral events (CNLOH) can be set independently in the TypeOfCNV column.

3 **[Optional]** You can define detected regions from which data should never be displayed by selecting the **Use Ignore Regions File** check box and loading a file.

In order to be ignored, a Detected Region must fall completely within an Ignored Region.

This feature has many potential uses, such as ignoring common polymorphic CNVs or constraining analysis to a portion of the genome. For example, you could ignore all chromosomes except 21 to analyze only aberrations on chromosome 21.

See Appendix D of this document for more information about formatting an Ignored Regions File.

4 To save the settings in the FoundRegionsFilterForm, click  (Save Filter Settings).

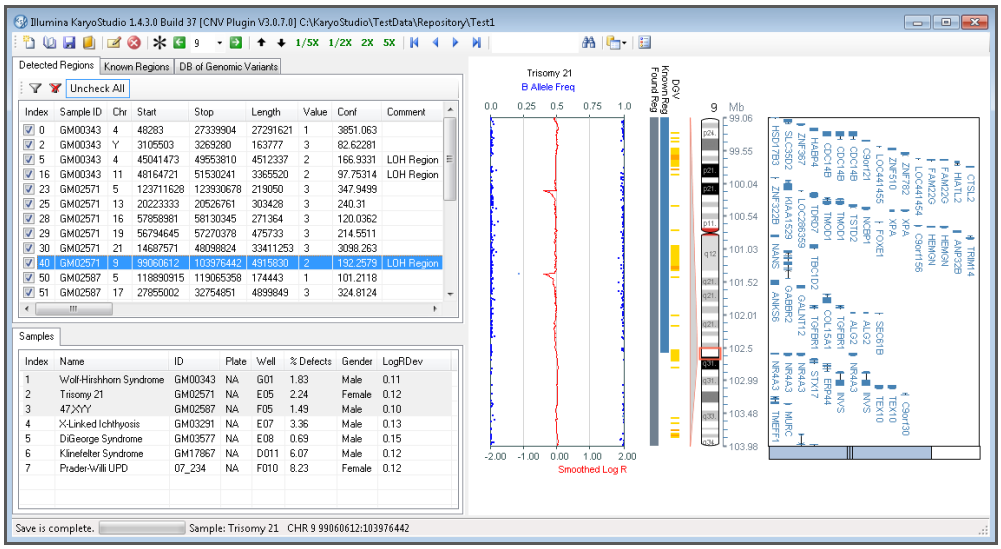
5 To load settings from a file into the FoundRegionsFilterForm, click  (Load Filter Settings) and browse to the settings file of interest.

6 Click **OK** to apply the new filter settings.

Displaying Detected Regions

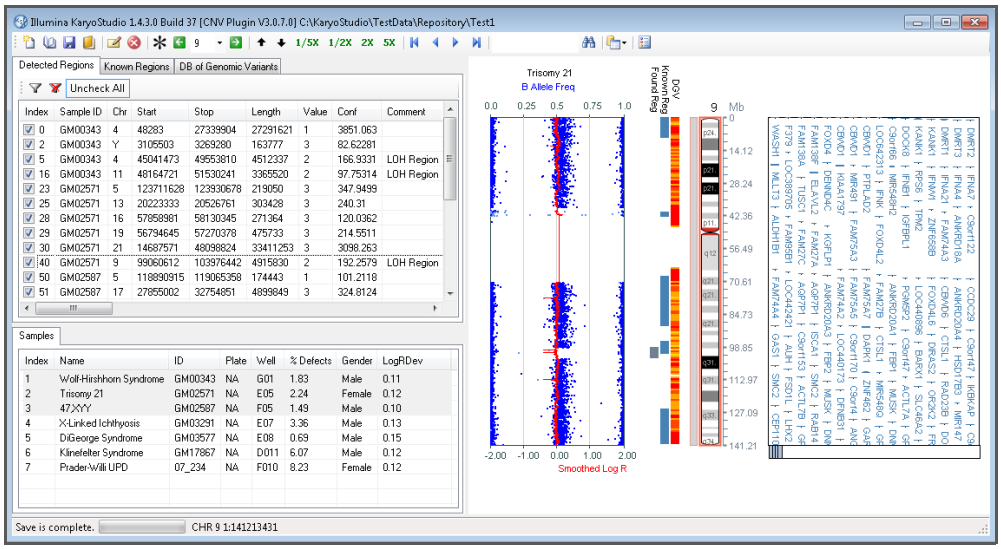
Notice that when you select a detected region, the entire region displays in the chromosome browser. By default, genotyping information is displayed as B-allele frequency (blue dots) and intensity information is displayed as smoothed log R ratio (red line).

Figure 37 Detected Region Displayed in Chromosome Browser



If you want to view the selected aberration in relation to the chromosome in which it appears, click ✱ (Expand to Chromosome). Intensity data for the whole chromosome is displayed.

Figure 38 Detected Region Displayed in Context of Whole Chromosome



Scrolling Through Detected Regions

Once your data has loaded, you can view each detected region one by one. There are multiple ways to select data to display in the chromosome browser:

- Click a detected region to display it in the chromosome browser.
- Use the blue arrow buttons  to move up or down in the Detected Regions Table.
- Use your mouse wheel to scroll up or down the Detected Regions Table.

Detected regions are only displayed for the sample(s) selected in the Samples table. You can choose multiple samples by Ctrl-clicking or Shift-clicking multiple rows in the Samples table.

Sorting Detected Regions

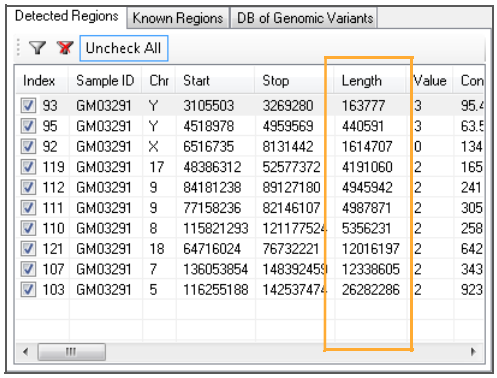
In addition to clicking and scrolling in the Detected Regions Table, you can sort data based on whichever column you like. To sort data, click a column header.

NOTE
If your Detected Regions Table contains hundreds of aberrations, it might take some time for KaryoStudio to sort the regions by column header.

Example 1

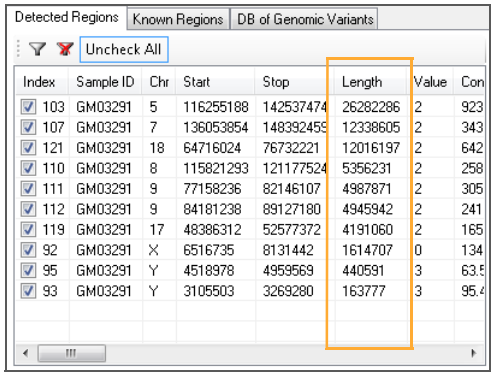
Click the column header **Length**, which represents the number of bases of an aberration. The aberrations listed in this table are now sorted from shortest to longest.

Figure 39 Aberration Length Sorted Short to Long



Click **Length** again to reverse the sort order. The aberrations now display from longest to shortest.

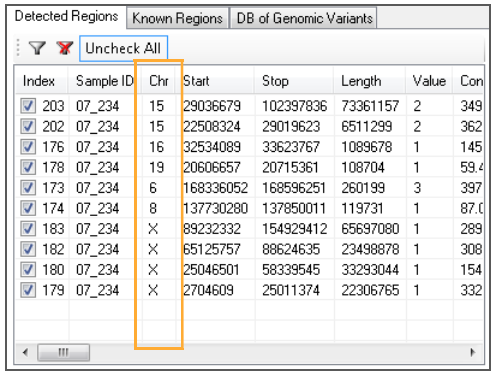
Figure 40 Aberration Length Sorted Long to Short



Example 2

Click **Chr** to sort aberrations based on the number of the chromosome on which they occur. Sorting by chromosome is useful if you want to view only aberrations occurring on a certain chromosome.

Figure 41 Aberrations Sorted by Chromosome



Index	Sample ID	Chr	Start	Stop	Length	Value	Con
203	07_234	15	29036679	102397836	73361157	2	349
202	07_234	15	22508324	29019623	6511299	2	362
176	07_234	16	32534089	33623767	1089678	1	145
178	07_234	19	20606657	20715361	108704	1	59.4
173	07_234	6	168336052	168596251	260199	3	397
174	07_234	8	137730280	137850011	119731	1	87.0
183	07_234	X	89232332	154929412	65697080	1	289
182	07_234	X	65125757	88624635	23498878	1	308
180	07_234	X	25046501	58339545	33293044	1	154
179	07_234	X	2704609	25011374	22306765	1	332

NOTE

When you choose to sort, the data for all samples are sorted together. Therefore, detected regions from multiple samples might be intermixed.

Adding or Editing Comments

To add comments or edit existing comments for a detected region, enter or change a comment in the **Comments** field.

For example, you can type a note for a colleague suggesting that he or she follows up on this region later. You might also want to enter additional information obtained from your cross-matching results. The comments you enter are included in the Comments section of the *.pdf reports you generate.

Adjusting Detected Region Parameters

Though it is not generally necessary to do so, KaryoStudio offers the ability to edit some parameters of a detected region such as confidence score and estimated copy number value. KaryoStudio also gives you the option to enter comments for a detected region. In addition, if necessary, you can edit the start and stop positions of a detected region.

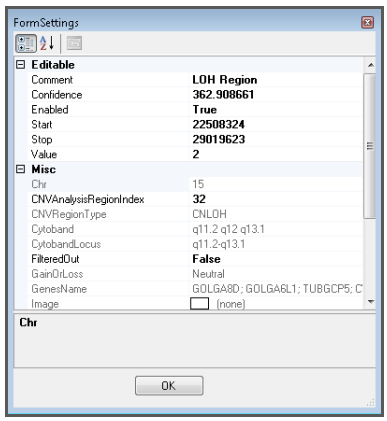
Editing Confidence Score or Estimated Copy Number

To adjust the confidence score or estimated copy number value of a detected region, perform the following steps:

1 Right-click a region of interest in the Detected Regions Table and select **Edit a Detected Region** from the context menu.

The Form Settings dialog box opens.

Figure 42 Form Settings Dialog Box



2 Do one of the following:

- If you would like to change the confidence value for this region, adjust the **Confidence** score.

NOTE
Although it is not recommend to change the confidence score of a detected region, you can change it to 0 if you want to exclude this detected region from analysis.

- If, on visual inspection, you feel that cnvPartition has not accurately estimated the copy number of the detected region, adjust the **Estimated Copy Number Value**.

NOTE
The parameters in the Misc area cannot be adjusted.

Editing Start and Stop Positions

You can verify the accuracy of the cnvPartition algorithm by examining the start and stop positions of the detected region. In some cases you might want to adjust start and stop positions based on visual inspection, though it is not common to do so. KaryoStudio also allows you to optionally adjust the information in the Detected Regions Table.

CAUTION

If you edit the start and stop positions of a detected region, the original positions identified by the cnvPartition algorithm are not stored with the project.

The confidence score, number of SNPs, and genes do not update after you have edited a region.

If you need to retain the original start and stop positions for any reason, copy the Detected Regions Table before you load a project and save it as a separate file, or save the project on your computer with a different file name.

To adjust the start and stop positions of a detected region of interest:

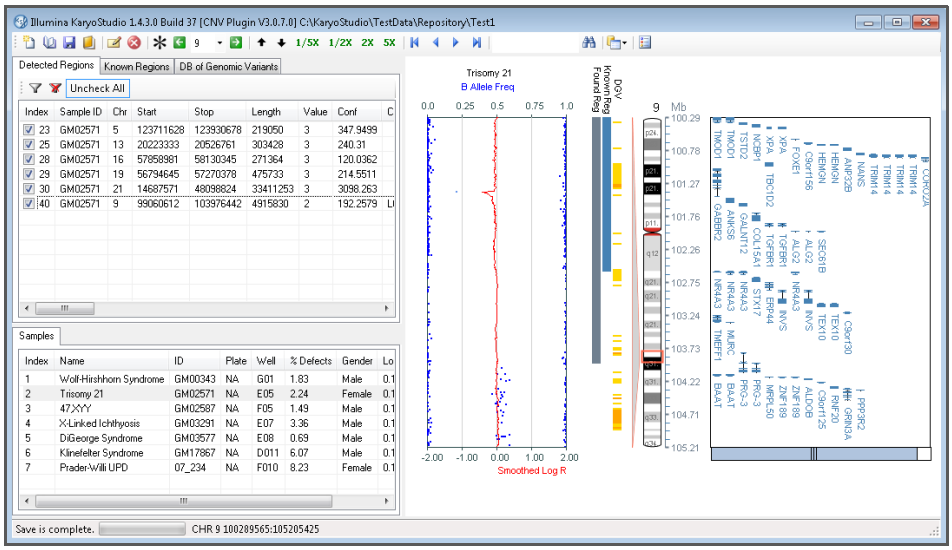
1 In the Detected Regions Table, select a region of interest.

KaryoStudio displays your detected region within the chromosome browser.

2 Click ↑ or ↓ to pan up or down until your detected region is centered in the chromosome browser.

This ensures that when you zoom in further, the SNPs of interest are visible in the window.

Figure 43 Panning Down to Examine the Boundaries of a Detected Region



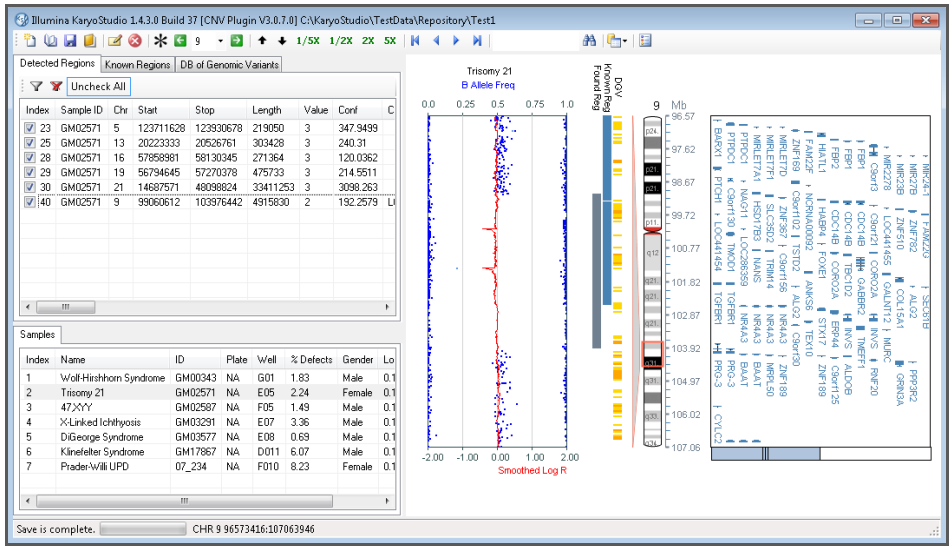
3 Click 1/2X (Zoom in 2x) or 1/5X (Zoom in 5x) to get a higher-resolution view of the boundaries of your detected region.

NOTE

If you want to view the boundaries of your detected region at the SNP level, to see each individual data point in the plot, you might need to zoom in multiple times.

Alternatively, you can pan and zoom by selecting and dragging the red box which denotes the detected region, as described in *Navigating the Chromosome Browser* on page 28.

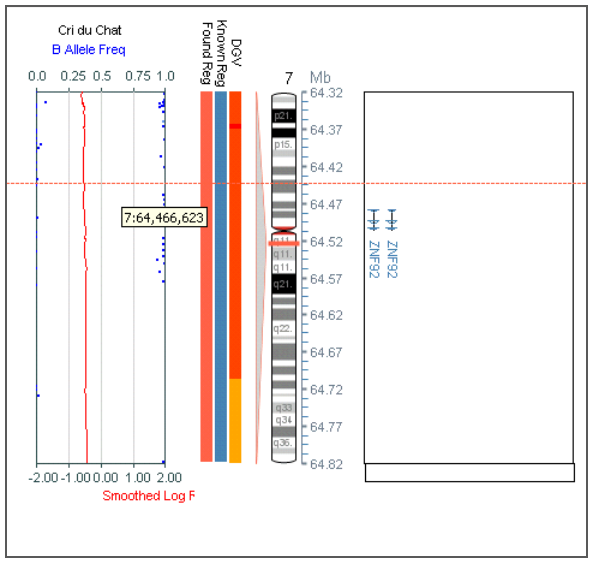
Figure 44 Centering a Detected Region



4 To display a horizontal ruler, click anywhere on the ideogram.
The horizontal ruler appears.

5 Mouse over the ruler at any location to display a tooltip with the location coordinates.

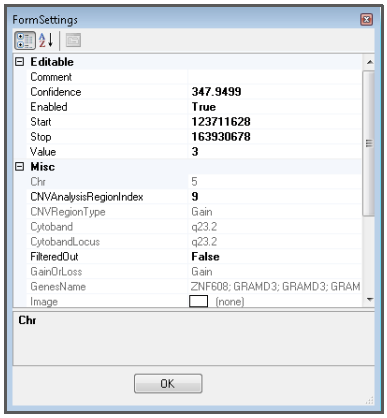
Figure 45 Ruler with Coordinates Tooltip



In most cases, the start and stop positions identified by the cnvPartition algorithm are very close to the positions you can identify by visual inspection. However, at this resolution, you have the option to edit the boundaries of the detected region.

6 To edit the boundaries of a detected region, right-click the detected region of interest in the Detected Regions Table and select **Edit a Detected Region** from the context menu.
The Form Settings dialog box opens.

Figure 46 Editing the Parameters of a Detected Region



NOTE

The Form Settings dialog box gives you the flexibility to adjust the Detected Regions Table a single row at a time. Repeat this process for each detected region you want to edit.

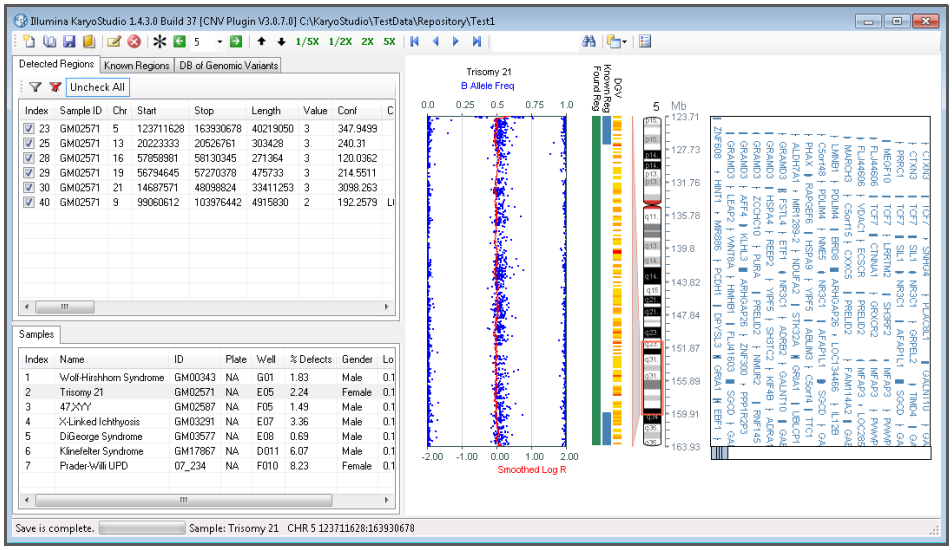
7 Click **Start** or **Stop** and enter a new value to the right, based on the start or stop position you see in the chromosome browser.

8 Click **OK**.

The results of your change are not immediately visible in the chromosome browser.

9 To see the change, click a different detected region in the Detected Regions Table.

Figure 47 Viewing the New Detected Region Positions



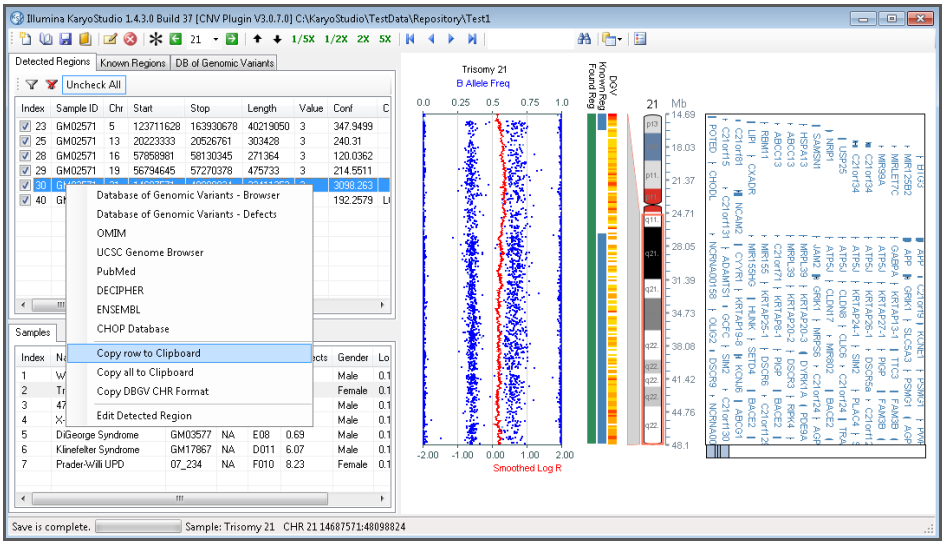
The new start and stop regions are included in the Detected Regions Table, in the Detected Regions track, and in all subsequent analyses.

Exporting Detected Regions

If you want to export data from the Detected Regions Table, you can export a single row of data, or the entire table.

- To export a single row, right-click in the Detected Regions Table and select **Copy Row to Clipboard**.
- To export the whole table, right-click in the Detected Regions Table and select **Copy All to Clipboard**.

Figure 48 Copying Detected Regions to the Clipboard



You can now paste this data into an Excel file or import it into other downstream, third-party applications.

Including Detected Regions in a Report

The check boxes to the left of the Index numbers indicate whether or not a detected region will be included in your report. All checkboxes are selected by default, which means that all detected regions will be included in your report by default.

- To deselect all regions, click **Uncheck All**.
- To select all regions, click **Check All**.
- To exclude certain detected regions from your report, deselect the checkbox(es) to the left of the region(s) you want to exclude.

Cross-Matching to Known Variants

Chapter 5

Introduction

Because a sample may potentially contain many aberrations, it is especially important to know what has already been identified as a normal variation or a region associated with a condition. For this reason, KaryoStudio allows you to check a detected region against a list of known regions as well as several different external sources. This function allows you to determine which regions found in your samples might be the most relevant.

This chapter describes how to link to these external information sources from within KaryoStudio, and how to use this additional information in your study.

KaryoStudio allows you to link to several databases, including:

- **DGV (Database of Genomic Variants)**—to determine whether your region has been identified as a copy number variant in normal, healthy individuals
- **OMIM (On-line Mendelian Inheritance of Man)**—to allow further mining of a particular phenotype, region, or gene of interest
- **UCSC Genome Browser**—to view genes or numerous other information tracks for a particular region in this popular public genome browser
- **PubMed**—to research publications relevant to a particular phenotype/gene of interest
- **DECIPHER**—to compare detected regions to the DECIPHER database of submicroscopic chromosomal imbalances
- **Ensembl**—to view a detected region in the graphical Ensembl genome viewer
- **CHOP Database**—to determine whether a detected region overlaps with copy number variations described in this database of samples from healthy individuals

Database of Genomic Variants (DGV)

Since the widespread adoption of high-density arrays for screening the genome for structural changes, many studies have identified various segments of the genome of differing sizes that range in copy number. From arrays to sequencing, the number of regions is growing dramatically and includes both common (>5%) and rare (1%) regions. Although many of these regions are not yet linked to specific conditions, the hope is that these studies will have a major impact on human health.

The DGV is one of the databases that is collating these regions in the hope of providing a comprehensive summary of structural variation human genomes. This database includes regions of the genome larger than >1kb that were identified in various studies of healthy control samples.

Updating KaryoStudio with the Latest DGV Build

The DGV currently includes many known CNV regions and covers the majority of the genome.

To update KaryoStudio with the latest DGV build, perform the following steps:

1. Find the latest version of the Genomic Variation Table at the Database of Genomic Variants web site (http://projects.tcag.ca/variation/) and click **Downloads**.
2. Save the latest version of the Genomic Variation Table as a file of name *GenomicVariation.txt* in the following location:
   C:\Program Files\Illumina KaryoStudio
3. Restart KaryoStudio to load the updated table.

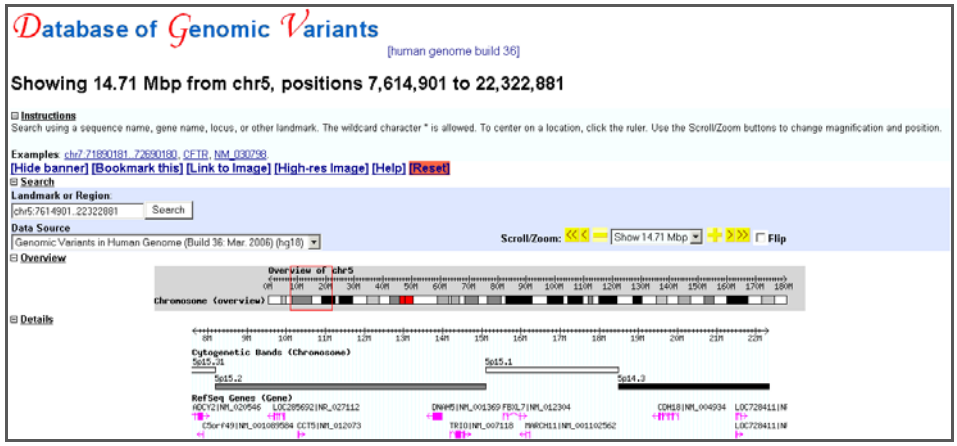
Using the DGV with KaryoStudio

You can use KaryoStudio to see if a region identified by KaryoStudio has already been identified and deposited in the DGV. Based upon chromosome and position, the DGV shows you information about what is known about the method used to identify the region, the reference (including details of the study), and a graphical reference to any CNVs found in the region.

Perform the following steps to use the DGV with KaryoStudio:

1. In the Detected Regions Table, select the region you want to search for in the DGV.
2. Right-click the region and select **Database of Genomic Variants - Browser** from the context menu.
   KaryoStudio sends the chromosome start and stop position information of the detected region to the DGV, which notifies you of how many matches were found, and displays a genome browser view of that region.

Figure 49 DGV Genome Browser View

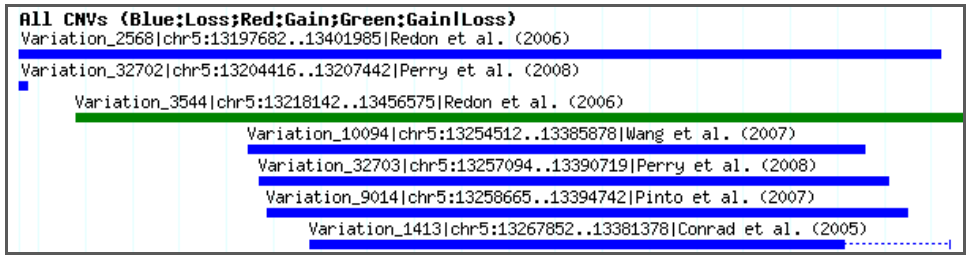


NOTE

Many optional tracks can be displayed in the DGV browser view. Refer to the documentation supplied with the Database of Genomic Variants for more information about using and interpreting information in the browser.

3 Scroll down until you get to the section labeled **All CNVs** (Figure 50).

Figure 50 All CNVs



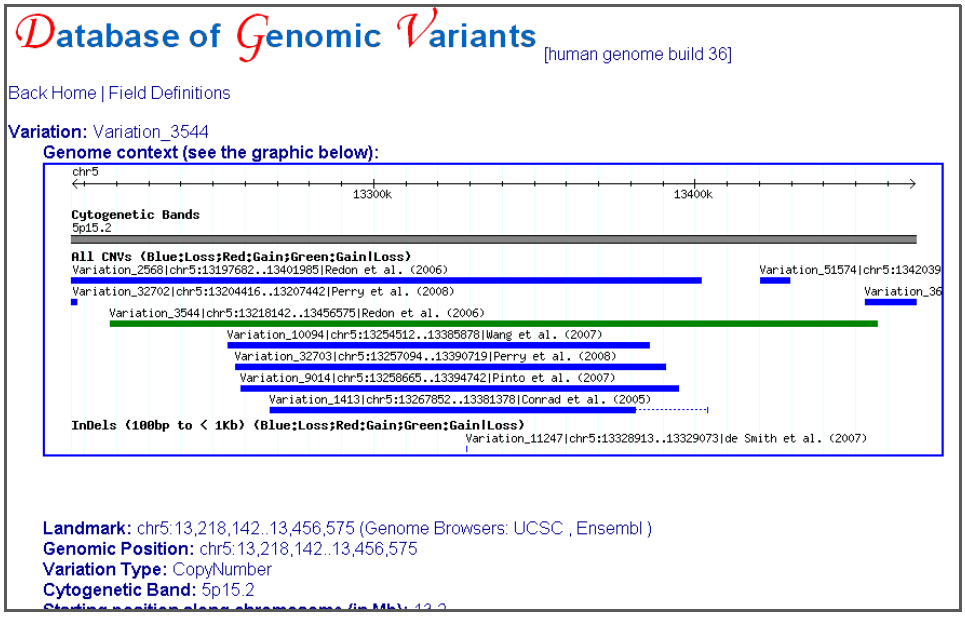
The All CNVs section shows colored bars representing all CNVs present in the database for the genomic region in the view.

4 Inspect a variation entry by clicking it.

A new window opens with details about the selected entry, including:

- variation number
- cytogenetic band
- genes
- other overlapping CNVs
- genomic coordinates
- nearby segmental duplications
- study details
- references to the study that reported the variant

Figure 51 Inspecting a Variation



- The variation number, shown in the upper-left of the screen, is a permanent number assigned to the region. If your lab is interested in this region, you can use this ID to refer to this region in the DGV.
- The Genome Context graphic in the middle of the screen illustrates the following:
  - The first section, Cytogenetic Bands, tells you which cytoband your region sits in. This is particularly important if you already know that a certain condition is associated with a change in a specific cytoband
  - The next section, All CNVs, tells you whether any other CNVs have been identified in this region. In addition, start and stop positions of each region, and the study in which the region was identified are shown. You can click a variation for more information.
  - The sections below the CNVs contain information about genomic inversions, indels, segmental duplications, etc. In Figure 51, only indels are shown.

  You can click each available section for more information that you might want to incorporate into your cytogenetics study.
- Below the graphic, there is additional information, most noticeably:
  - The Frequency Information might be particularly useful for interpretation. It includes the number of samples studied, and the number of times a loss or gain was seen.
  - You can click a link to PubMed (if available), which displays the publication reference for the study that identified this region.

The DGV is a resource provided by SickKids Hospital in Toronto, Canada. If you need more information about this database, go to http://projects/tcag.ca/, or send email to dgv-contact@sickkids.ca.

All of the CNVs in the DGV were identified in healthy control samples. The DGV is constantly being updated as new studies deposit information and as technology evolves. For the latest build, and latest information, please see the DGV web page: http://projects.tcag.ca/variation/.

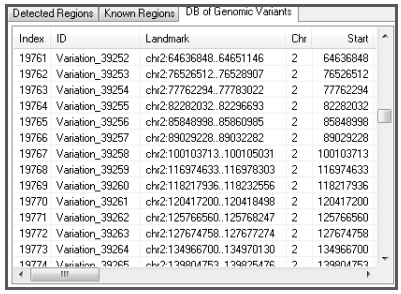
DGV Data

You can use the following option if you prefer to view CNV data within KaryoStudio rather than linking out to the DGV. The advantage of viewing your data in KaryoStudio is that you can look at a CNV region across multiple samples.

1 Click the **DGV** tab in KaryoStudio.
It might take a moment to load.

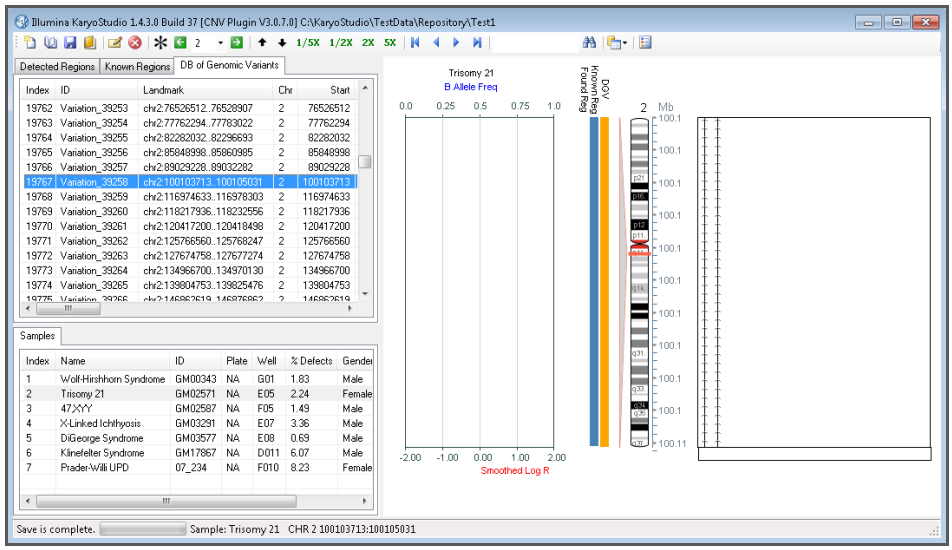
2 When it finishes loading, scroll to the bottom of the list.

Figure 52 KaryoStudio DGV Table



3 If you want to see whether any of the detected regions in your data match known DGV regions, click a region in the DGV table.
The data for the region you clicked appears in the chromosome browser.

Figure 53 Selecting a Known Region in the DGV Table



Updating the DGV Table and Track

To update the data used to populate the DGV table and track in KaryoStudio, perform the following steps:

1. Download the database build from the following location: http://projects/tcag.ca/variation/tableview.asp?table=DGV_Content_Summary.txt.
2. Download the variation file (*.txt format) in the Current Version section, based on Build 36 (hg 18) of the genome, to your computer.
   Example variation file: *variation.hg18.v8.aug.2009.txt*
3. Rename this file "*DatabaseOfGenomicVariation.txt*".
4. Copy the file *DatabaseOfGenomicVariation.txt* (replacing the existing file) to the directory *C:\Program Files\Illumina\Illumina KaryoStudio*.
5. Restart KaryoStudio to load the new data file.

OMIM (Online Mendelian Inheritance of Man)

The OMIM database, hosted by the National Center for Biotechnology Information (NCBI), contains information on human genes and genetic phenotypes on all known Mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype.

Using OMIM with KaryoStudio

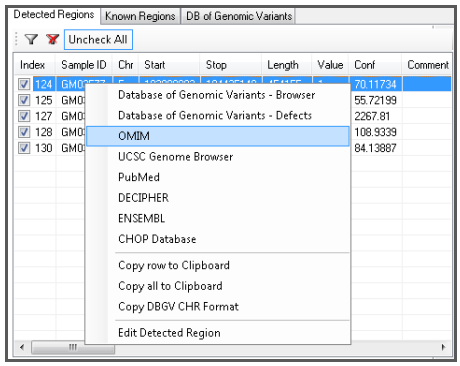
Because OMIM is based upon specific keywords and not basic chromosome start and stop positions, KaryoStudio only provides a link to the homepage of this database. If KaryoStudio has identified a specific phenotype within the Detected Regions Table, you can enter a key word (e.g., gene, cytoband, or phenotype) in the homepage of OMIM to obtain more information if available.

To use OMIM with KaryoStudio, perform the following steps:

1 Select a region of interest in the Detected Regions Table.

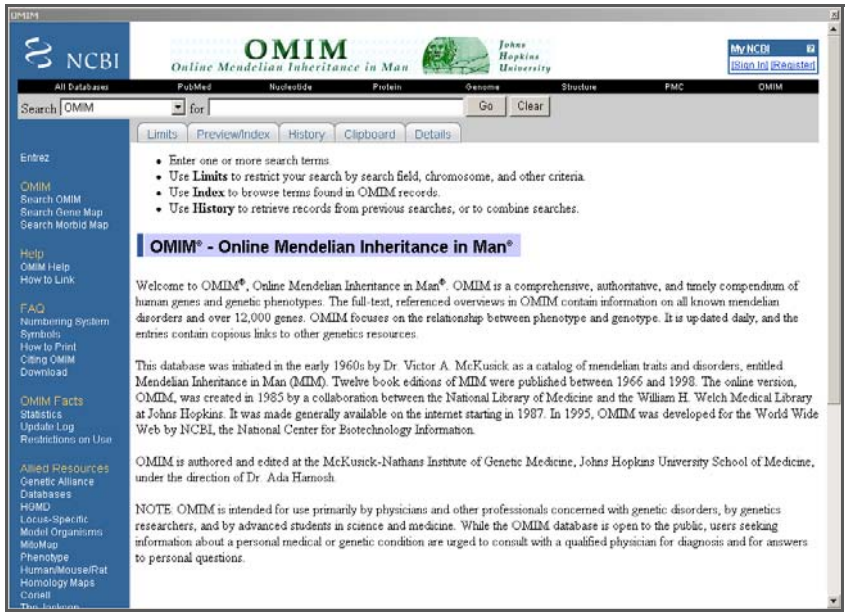
2 Right-click and select **OMIM** from the context menu.

Figure 54 Selecting OMM



OMIM opens to the homepage.

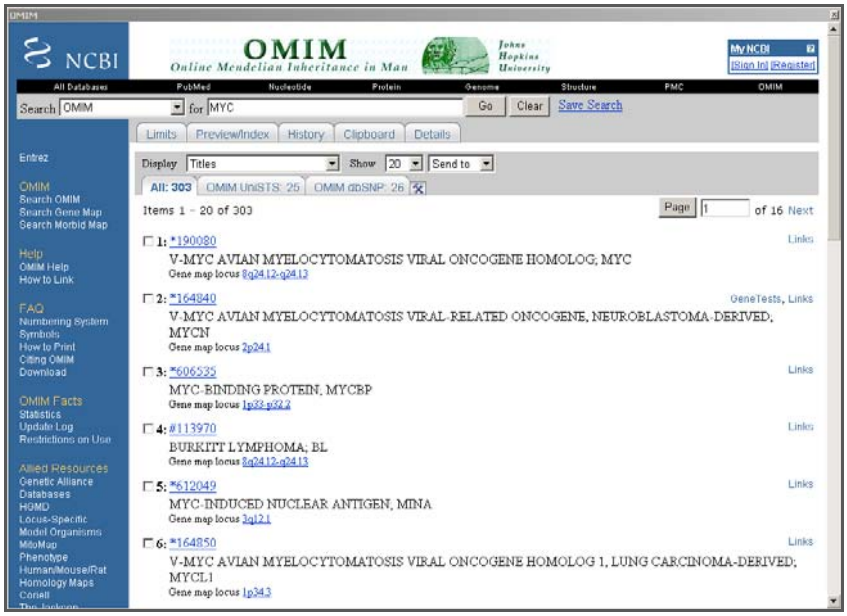
Figure 55 OMIM, Home Page



3 Type a key word in the search text field.

Key words can be gene names, etc. For example, if you had identified the MYC gene in your search, you might want to enter MYC as a key word. OMIM identifies your gene in multiple organisms and displays a list of the occurrences of this gene in all organisms.

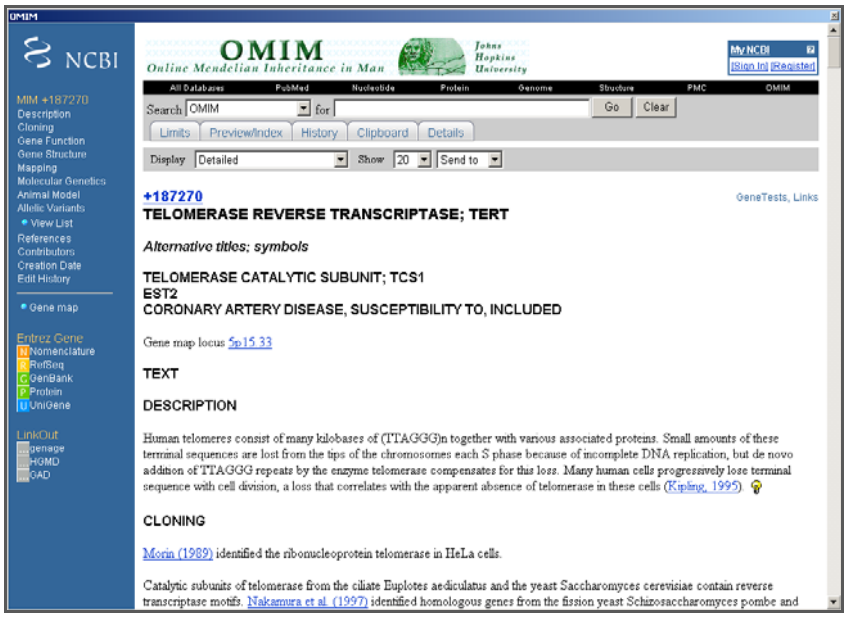
Figure 56 OMIM, MYC Gene



If you want to look at this gene in the human genome only, for example, you must scroll down to find it in the list.

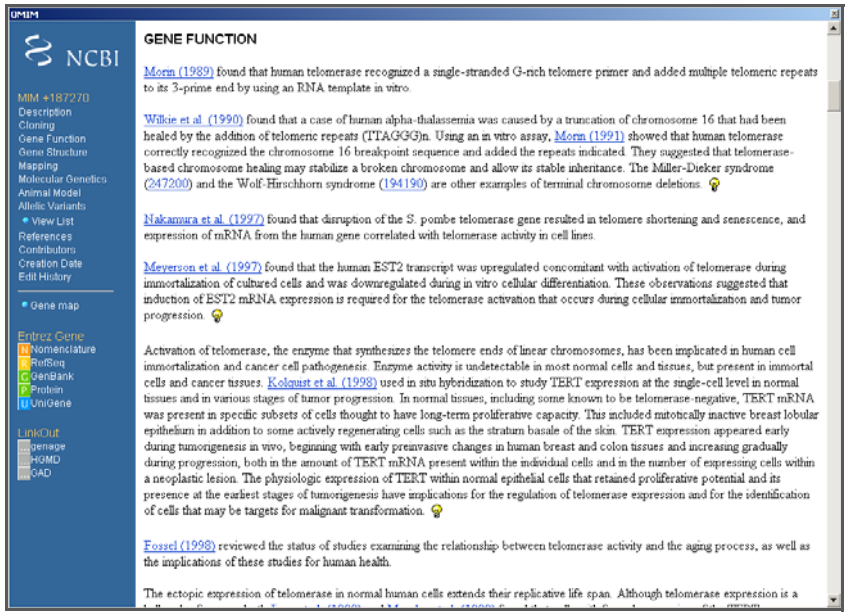
If you click on a hit, information about this gene is displayed, including the name, description, etc.

Figure 57 OMIM, Additional Information



Of particular interest is the Gene Function section, which describes any known molecular function of this gene.

Figure 58 OMIM, Gene Function



If you enter a key word that is a condition in the OMIM search field, OMIM provides information about that condition.

For example, enter “down syndrome.” A listing page comes up with the closest matches. Click the first entry, **#190685 DOWN SYNDROME**. A down syndrome information page appears.

Figure 59 OMIM, Down System Information Page

Information such as a description of the condition, clinical features, and more is provided.

UCSC Genome Browser

The UCSC Genome Browser is a popular tool that allows you to examine gene information, expression data, and many other parameters within your detected region. Yo can also upload custom data tracks to display in the UCSC Genome Browser.

Using the UCSC Genome Browser with KaryoStudio

You can use KaryoStudio to see what other genomic features sit within a detected region. All analyses are based on chromosome and position. The UCSC Genome Browser shows you information about:

- RefSeq genes
- mRNAs
- ESTs
- genes in other organisms

There are also other options within the page that allow you to add or remove additional data tracks of information.

NOTE

The UCSC Genome Browser might take a while to load. Be patient while your data is loading; it will eventually display.

- At the top, the chromosome and position are identified.
- Next, information about RefSeq genes is listed.
- Next, there is mRNA and EST information, which is useful for looking at locations of specific gene transcripts.
- Next, the conservation of this gene in mammalian species is shown on a graph.
- Next, the conservation of this gene in other species is shown on tracks.
- Finally, the location of SNPs in humans, and the location of repeated DNA sequences.
- Below all of this, you can select various additional information to include in this graph.
- The Phenotype and Disease Association Studies track might be of interest. Select these tracks to see if any known phenotypes have been associated with this specific region.

PubMed

PubMed, available via the NCBI Entrez retrieval system, is hosted by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) at the U.S. National Institutes of Health (NIH). PubMed provides access to citations from biomedical literature and is a streamlined way to determine if there are any peer-reviewed and published studies related to the region, condition, or phenotype under study.

Using PubMed with KaryoStudio

Since this database is based on specific key words, not on chromosome number and position, KaryoStudio provides only a link to the homepage of this database.

If KaryoStudio has identified a specific phenotype within the Detected Regions Table, you can enter the keyword in the homepage of PubMed to obtain more information.

This section is similar to OMIM in that it is based on keywords. Type in a gene name, condition name, etc. This brings up a list of scientific publications in which the key word has been found.

Note that the more common your inquiry, the more hits you will get. To reduce the number of hits, make your search term as specific as possible.

DECIPHER

The DECIPHER consortium is a network of clinical genetic centers. KaryoStudio allows you to examine a specific detected region of interest against this cytogenetics database of clinical cytogenetics cases by right-clicking on the detected region and sending it to the database.

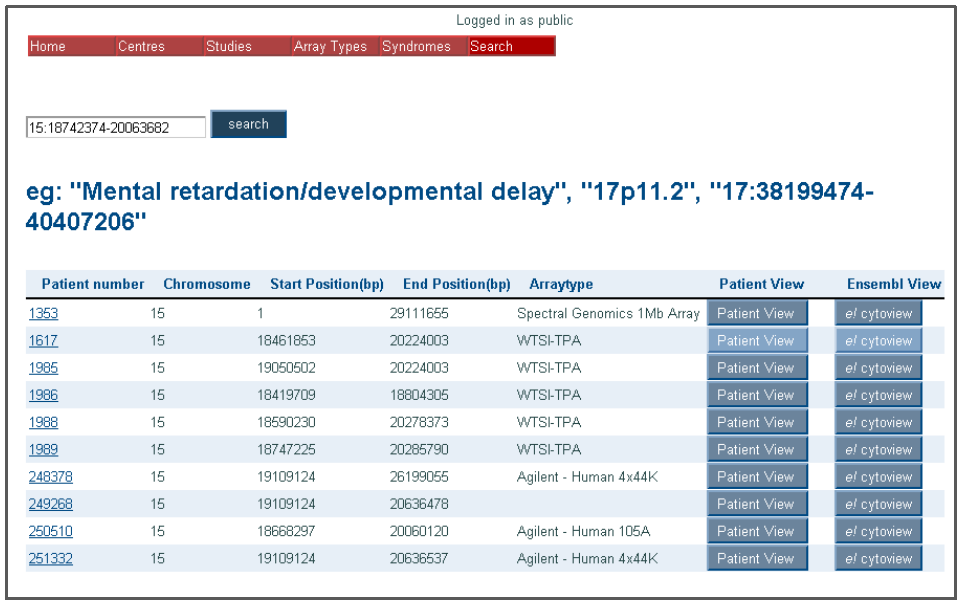
If a particular portion (or the whole region) of your detected region has been previously linked to a certain phenotype, you can view that here. From the DECIPHER database, you can view your detected region within the Ensembl Genome Browser, identify the type of array or technology used to identify the specific region, and obtain a DECIPHER Syndrome Report for a specific phenotype of interest, including a clinical description, references, affected genes, and a phenotype report.

Using DECIPHER with KaryoStudio

To use DECIPHER with KaryoStudio, do the following:

- Right-click a detected region and select **DECIPHER** from the context menu.
  KaryoStudio automatically sends information about the detected region to DECIPER, and opens the DECIPHER window for that region.

Figure 60 DECIPHER



Ensembl

KaryoStudio allows you to examine your detected region within the cytogenetics view of the Ensembl Genome Browser. Ensembl is a useful database for examining characterized disorders that might have been previously linked to your region.

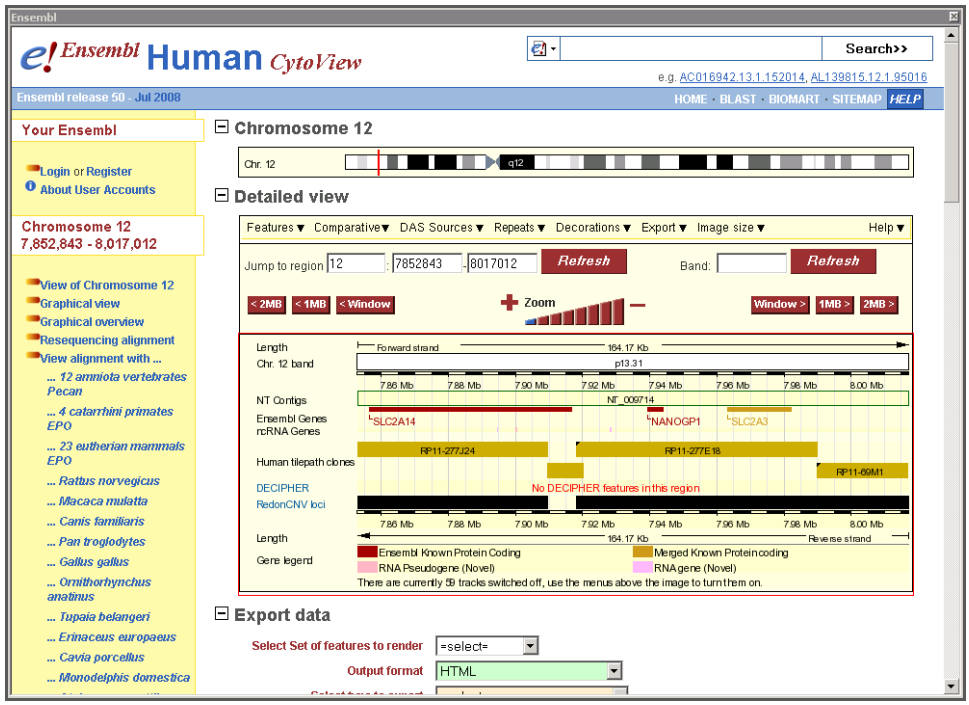
Using Ensembl with KaryoStudio

To use Ensembl with Karyostudio, do the following:

- Right-click a detected region and select **Ensembl** from the context menu.

  Karyostudio automatically sends information about your region to the Ensembl database, and the Ensembl Human CytoView window appears (Figure 61).

Figure 61 Ensembl



CHOP Database

The Copy Number Variation project at the Children's Hospital of Philadelphia (CHOP) represents an effort to identify all frequent copy number variations (CNVs) that exist in the human genome. The database currently consists of data from over 2,000 healthy individuals. This database was originally described in the following manuscript:

TH Shaikh et al. *(2009)* High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. *Genome Res 19: 1682-1690; doi:10.1101/gr.083501.108*

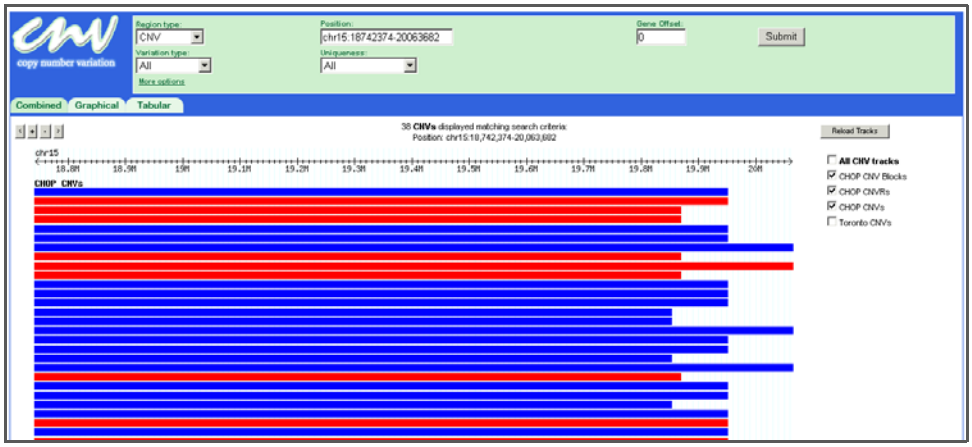
Using the CHOP CNV Database with KaryoStudio

To use the CHOP Database with KaryoStudio, do the following:

- Right-click a detected region and select **CHOP Database** from the context menu.

  KaryoStudio automatically sends information about the detected region to the CHOP Database and opens the CHOP CNV browser window to display any CNVs in the CHOP Database that map to the region you submitted (Figure 62).

Figure 62 CHOP CNV Database CNVs



Generating a Cytogenetics Report

Chapter 6

Introduction

KaryoStudio allows you to create a *.pdf Cytogenetics Report for each sample, summarizing the aberrations found in that sample and showing whether they cross-match with any known regions in the genome. Depending upon how many aberrations are found for a sample, and how many you choose to include in the report, it can be from one page to many pages long.

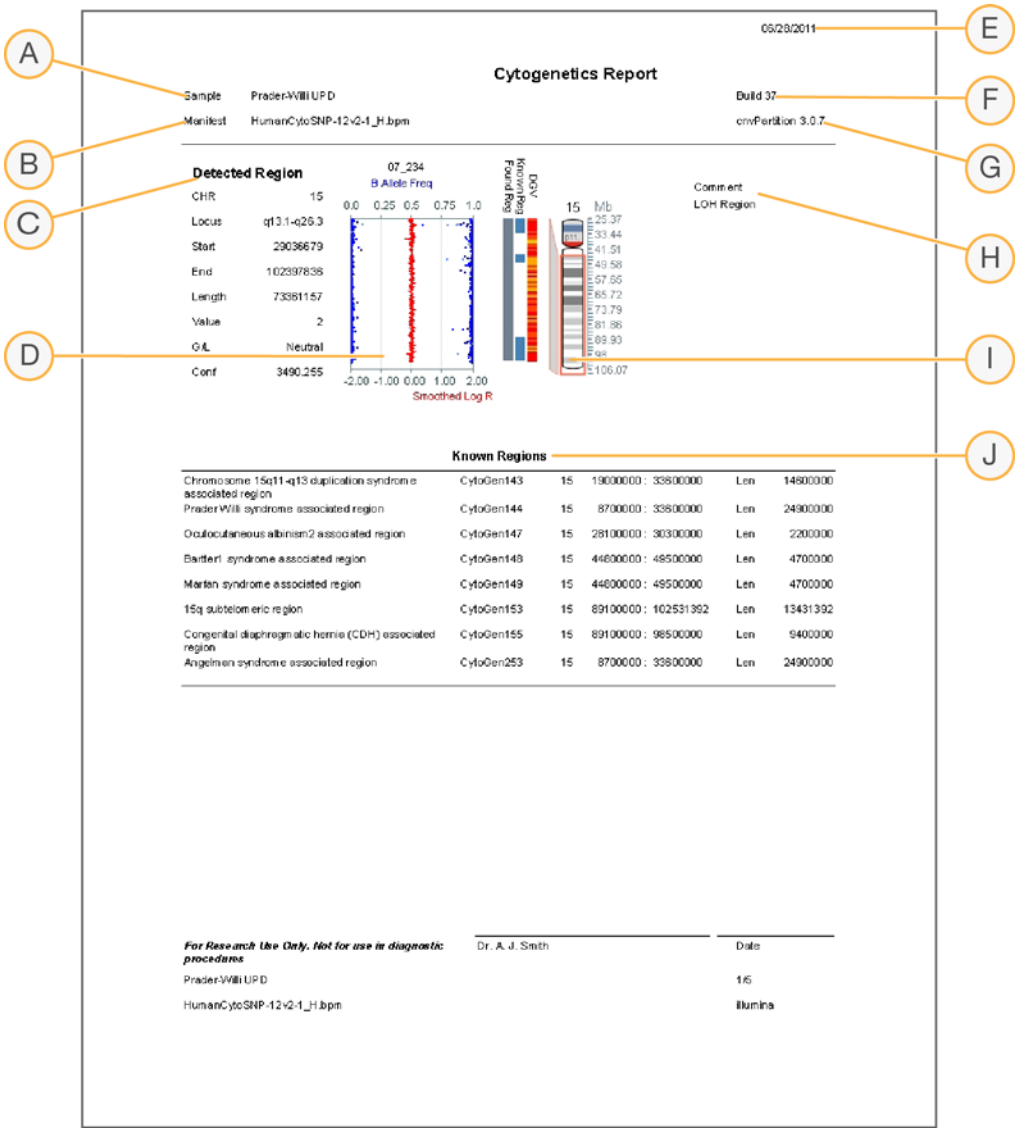
NOTE

You need a recent version of Adobe Reader to view *.pdf reports. Adobe Reader is available free from http://get.adobe.com/reader/.

What's in a Cytogenetics Report?

A Cytogenetics Report contains all of the essential information required to quickly summarize aberrations found within a specific sample. Within each report is an entry for each detected region, including an image of the aberration, its size, an estimate of the copy number, the confidence value (relative score), and a list of the disorders that have already been associated with that specific region.

Figure 63 Cytogenetics Report



- **A** Sample name
- **B** Product manifest
- **C** Detected region information
- **D** Detected region
- **E** Date of report
- **F** Current genome build
- **G** Current version of cnvPartition algorithm
- **H** User-entered comments
- **I** Chromosome view
- **J** Known regions

Table 9 Cytogenetics Report - Header

Field	Description
Date	Date the report was generated.
Sample Name	Name of the sample presented in the report.
Product Manifest	Product manifest used to create the project.
Genome Build	Genome Build version used for analysis.
cnvPartition Algorithm Version	Algorithm version used for analysis (not necessarily the most recently installed version).
Image of Detected Region	Can adjust number of images shown per report.

Table 10 Cytogenetics Report - Detected Region Information

Field	Description
CHR (Chromosome)	Chromosome on which the detected region is located.
Locus	Cytoband(s) on which the detected region is located.
Start	Start coordinate of the detected region.
End	End coordinate of the detected region.
Length	Length of the detected region, measured in base pairs.
Value	Copy number of the detected region.
G/L	Gain or loss.
Conf	Confidence score for the detected region, calculated by the algorithm.
Data plot	Plot of B Allele Freq and Log R of markers in the detected region.
Regions tracks	Found Region, Known Region, and DGV tracks from the chromosome browser.
Coordinate ruler	Chromosomal coordinates of the data plot in megabases.
Comment	User-defined comments for the detected region. Can be left blank.
Known regions	Regions known to overlap with the detected region. Includes Region Name, Region ID, Chromosome, Start coordinate, End coordinate, and Length.

Table 11 Cytogenetics Report - Footer

Field	Description
Sample Name	Name of the sample presented in the report.
Manifest	Product manifest used to create the project.
Signature and date	Area for the investigator to sign and date the report.
Username	Name of the user logged into the system when the report was generated.

Adjusting the Information in a Cytogenetics Report

You can adjust the information included in a Cytogenetics Report by using the KaryoStudio **Settings** tab.

You can adjust the number of aberrations included in a Cytogenetics Report by entering a value for *MaxReportedDefects*. On average, KaryoStudio includes roughly two aberrations per page, so a report with ten aberrations would be approximately five pages long.

You can limit the number of known regions to display for each detected region by entering a value for *MaximumNumberOfKnownDefects*. If you do not want to cross-match to any known regions, set *MaximumNumberOfKnownDefects* to 0.

Additionally, you can select which detected regions to include in a Cytogenetics Report by selecting the checkbox to the left of each detected region in the Detected Regions Table. Clear the checkbox to the left of a detected region to exclude it from a report.

Generating a New Cytogenetics Report

After you choose the parameters for the aberration filter and number of aberrations for each report, click **Write Reports to Directory**. A dialog box opens, allowing you to browse to and select a directory where you want the reports to be saved.

KaryoStudio generates the reports in the location you selected. This takes roughly 30 seconds to one minute per sample.

Additional Information

Appendix A

System Information

Use the following information as a guide to determine the minimum and recommended system configuration for the computer on which you plan to install KaryoStudio.

NOTE

The minimum and recommended system configurations vary depending on the products you plan to use to generate the data you load into KaryoStudio.

Table 12 System Information

	Human CytoSNP-12	Human 610-Quad	Human 1M-Duo	HumanOmni1-Quad
Minimum System Requirements	32-bit	32-bit	64-bit	64-bit
Operating System	WinXP SP2 or Vista	WinXP SP2 or Vista	WinXP SP2 or Vista	WinXP SP2 or Vista
Minimum / RecommendedProcessor Requirements	Pentium IV 1.5 GhZ / Pentium IV 2.0 GhZ	Pentium IV 1.5 GhZ / Pentium IV 2.0 GhZ	Pentium IV 1.5 GhZ / Pentium IV 2.0 GhZ	Pentium IV 1.5 GhZ / Pentium IV 2.0 GhZ
.NET Version	.NET 3.5	.NET 3.5	.NET 3.5	.NET 3.5
Network Connection	1 GB	1 GB	1 GB	1 GB
Minimum Memory Requirements	4GB	4GB	8GB	8GB
Minimum Hard Drive Requirements	100GB	100GB	100GB	100GB
Minimum / Recommended Video Display Requirements	1024x768 / 1280x1024	1024x768 / 1280x1024	1024x768 / 1280x1024	1024x768 / 1280x1024

cnvPartition Algorithm

The cnvPartition algorithm automatically scans your data when you create a new project. cnvPartition v3.0.7, provided with KaryoStudio v1.4, has been designed to scan for deletions, duplications, and amplifications on the autosomes and sex chromosomes.

The cnvPartition algorithm is based on a recursive partition method. cnvPartition, which has been optimized for speed and accuracy, uses log R intensity and B Allele Freq for identification of chromosomal aberrations, estimates copy number values, and calculates per locus confidence scores. After you load your data into KaryoStudio, cnvPartition will automatically scan the data for aberrations.

For more information about how this algorithm works, please see the *DNA Copy Number Analysis Algorithms Technical Note* at http://www.illumina.com/Documents/products/technotes/technote_cnv_algorithms.pdf.

Confidence Score

The confidence score that cnvPartition generates is defined as the sum of all logged likelihoods in the region for the assigned copy number minus the sum of all logged likelihoods of copy number equal to 2 for loci in the region. Thus the confidence scores provide a means to rank regions relative to their (dis)similarity to normal, copy number 2 segments. Higher values represent higher confidence in the aberration call.

For more information about recommended confidence score cutoff values, see the KaryoStudio FAQs on http://www.illumina.com/support/faqs.ilmn.

B Allele Frequency and Log R Ratio

B Allele Frequency and Log R Ratio data form the basis of the analysis of all copy number changes in Infinium data. B Allele Frequency data is derived from SNP genotypes, while Log R Ratio data is derived from intensity information.

Calculation

The B Allele Freq for a sample shows the theta value for a SNP, corrected for cluster position. Cluster positions are generated from a large set of normal individuals. The B Allele Frequency can also be referred to as "copy angle" or "allelic composition." It is easier to visualize genotyping data for all SNPs within a chromosomal region using B Allele Freq rather than theta values. This is true because B Allele Freq exhibits less locus-to-locus variation than the theta values for a given sample. The transformation of theta values to allele frequencies allows for improved measurements and better visualization of both LOH and copy number changes.

B allele freq is described by the following equations.

B allele freq

= 0 if theta < tAA

= 0.5 * (theta - tAA) / (tAB - tAA) if theta < tAB

= 0.5 + 0.5 * (theta - tAB) / (tBB - tAB) if theta < tBB

= 1 if theta >= tBB

where:

- tAA = mean theta value of all genotypes in the AA cluster plotted in polar normalized coordinates
- tAB = mean theta value of all genotypes in the AB cluster plotted in polar normalized coordinates
- tBB = mean theta value of all genotypes in the BB cluster plotted in polar normalized coordinates

The Log R Ratio is based on normalized intensity data and for a sample is the log (base 2) ratio of the normalized R value for the marker divided by the expected normalized R value. For loci included in GenomeStudio statistics such as Call Rate, the expected R value is computed by linear interpolation of the R value at the SNP's theta value for a sample, relative to the R values of the surrounding clusters. Because no clusters are generated for loci in the "Intensity Only" category, the Log R Ratio for these loci is adjusted so that the expected R value is based on the weighted mean of the cluster itself. Log R Ratio is displayed the same way for these loci as it is for loci included in GenomeStudio statistics in tools such as the IGV. Both SNPs and intensity-only loci such as nonpolymorphic probes (which usually have the identifier "cnv" in their names) are displayed in the log R ratio plots in KaryoStudio.

For example, if for a given sample and SNP with:

- A theta value of 0.2
- an AA cluster at theta = 0.1, R = 1.5
- an AB cluster at theta = 0.4, R = 2.5

The estimated R at theta for the sample is: 0.2 is 1.5 + (0.2-0.1) * (2.5-1.5) / (0.4-0.1) = 1.83. If the R value for the SNP is 1.6, the Log R Ratio is: log2 (1.6/1.83) = -0.196

Interpretation

cnvPartition automatically scans both the B allele frequency and log R ratio data for the presence of aberrations. In regions of the genome with two copies, the B allele frequency sits at 0, 0.5, and 1 representing the AA, AB, and BB genotype clusters. In regions of the genome that do not have two copies, various patterns may be seen. You must take into account the direction in which the log R ratio is deflected to determine if a gain or loss of DNA is present. Increases in the log R ratio indicate duplications (or amplifications) and decreases indicate deletions.

For more information about interpreting this data, see the *Interpreting Infinium Assay Data for Whole-Genome Structural Variation Technical Note* at http://www.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf.

Supporting Files

Appendix B

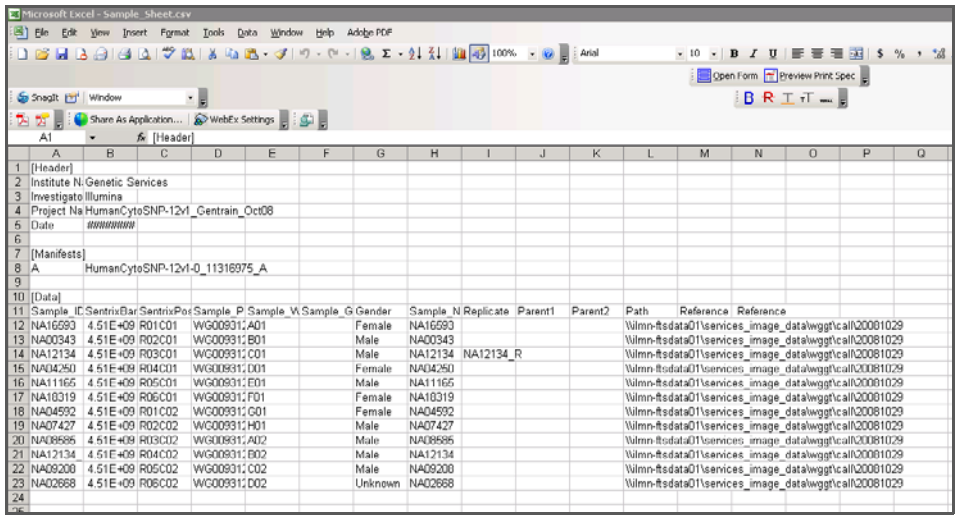
Sample Sheet

The figure below is an example Sample Sheet for KaryoStudio. The following columns are required:

- Sample_ID
- SentrixBarcode_A
- SentrixPosition_A
- gender

All other columns are optional. If path is not provided in the sample sheet, it will be required in the new project wizard.

Figure 64 Example Sample Sheet



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	[Header]																
2	Institute N	Genetic Services															
3	Investigato	Illumina															
4	Project Na	HumanCytoSNP-12v1_Centrain_Oct08															
5	Date	##########															
6																	
7	[Manifests]																
8	A	HumanCytoSNP-12v1-0_11316975_A															
9																	
10	[Data]																
11	Sample_ID	SentrixBar	SentrixPos	Sample_P	Sample_W	Sample_G	Gender	Sample_N	Replicate	Parent1	Parent2	Path	Reference	Reference			
12	NA16593	4.51E+09	R01C01	WG00931	A01		Female	NA16593				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
13	NA00343	4.51E+09	R02C01	WG00931	B01		Male	NA00343				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
14	NA12134	4.51E+09	R03C01	WG00931	C01		Male	NA12134	NA12134_R			\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
15	NA04250	4.51E+09	R04C01	WG00931	D01		Female	NA04250				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
16	NA11165	4.51E+09	R05C01	WG00931	E01		Male	NA11165				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
17	NA18319	4.51E+09	R06C01	WG00931	F01		Female	NA18319				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
18	NA04592	4.51E+09	R01C02	WG00931	G01		Female	NA04592				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
19	NA07427	4.51E+09	R02C02	WG00931	H01		Male	NA07427				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
20	NA08585	4.51E+09	R03C02	WG00931	A02		Male	NA08585				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
21	NA12134	4.51E+09	R04C02	WG00931	B02		Male	NA12134				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
22	NA09208	4.51E+09	R05C02	WG00931	C02		Male	NA09208				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
23	NA02668	4.51E+09	R06C02	WG00931	D02		Unknown	NA02668				\\ilmn-ftsdata01\services_image_data\wggt\call\20081029					
24																	

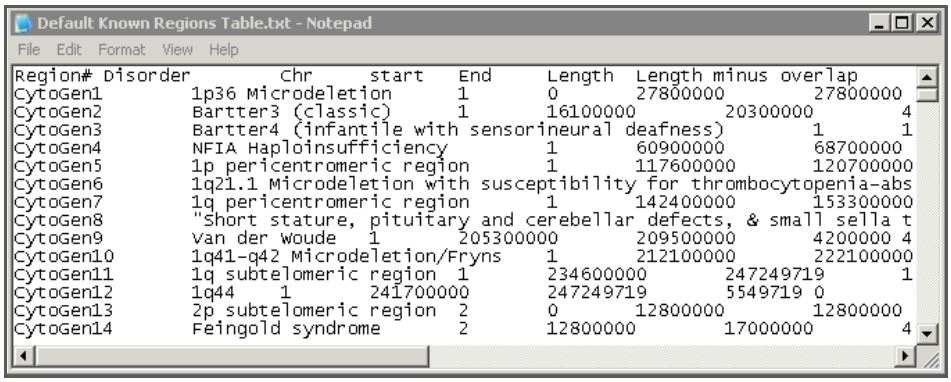
Known Regions File

The figure below is an example Known Regions File. Table 13 lists and describes the required columns of a Known Regions File for use with KaryoStudio.

Table 13 Known Regions File Column Descriptions

Column	Description
Region#	ID code for the region
Disorder	Text description of the significance of the region
Chr	Chromosome number of the region
Start	Start coordinate of the region
End	End coordinate of the region
Length	Integer indicating the length of the region in base pairs
Length minus overlap	No longer used by the KaryoStudio software; however, an integer value must be supplied (may be the same as the Length value)

Figure 65 Example Known Regions File



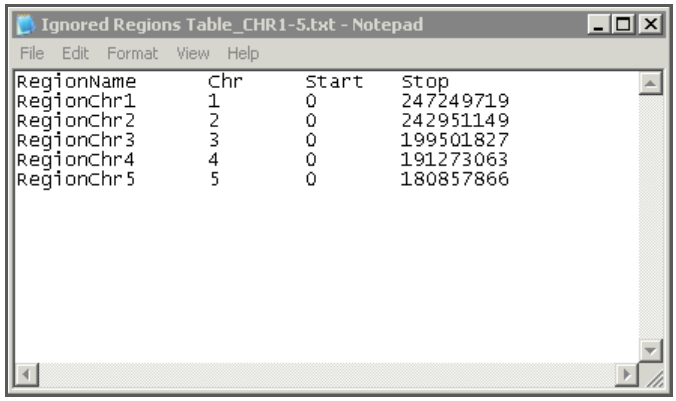
Ignored Regions File

The figure below is an example Ignored Regions File. Table 14 lists and describes the required columns of an Ignored Regions File for use with KaryoStudio.

Table 14 Ignored Regions File Column Descriptions

Column	Description
Region Name	Text description of the region
Chr	Chromosome number of the region
Start	Start coordinate of the region
Stop	Start coordinate of the region

Figure 66 Example Ignored Regions File



Technical Assistance

For technical assistance, contact Illumina Customer Support.

Table 15 Illumina General Contact Information

Illumina Website	http://www.illumina.com
Email	techsupport@illumina.com

Table 16 Illumina Customer Support Telephone Numbers

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

MSDSs

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

Product Documentation

If you require additional product documentation, you can obtain PDFs from the Illumina website. Go to http://www.illumina.com/support/documentation.ilmn. When you click on a link, you will be asked to log in to iCom. After you log in, you can view or save the PDF. To register for an iCom account, please visit https://icom.illumina.com/Account/Register.

Illumina, Inc.
9885 Towne Centre Drive
San Diego, CA 92121-1975
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com