



# **Affymetrix® Chromosome Analysis Suite 1.2**

## **User Manual**

**For research use only.**

**Not for use in diagnostic procedures.**

#### **Trademarks**

Affymetrix®, GeneChip®, NetAffx®, Command Console®, Powered by Affymetrix™, GeneChip-compatible™, Genotyping Console™, DMET™, GeneTitan®, Axiom®, CytoScan™, and GeneAtlas® are trademarks or registered trademarks of Affymetrix, Inc. All other trademarks are the property of their respective owners.

All other trademarks are the property of their respective owners.

This database/product contains information from the Online Mendelian Inheritance in Man® (OMIM®) database, which has been obtained under a license from the Johns Hopkins University. This database/product does not represent the entire, unmodified OMIM® database, which is available in its entirety at [www.ncbi.nlm.nih.gov/omim/](http://www.ncbi.nlm.nih.gov/omim/).

#### **Limited License Notice**

Limited License. Subject to the Affymetrix terms and conditions that govern your use of Affymetrix products, Affymetrix grants you a non-exclusive, non-transferable, non-sublicensable license to use this Affymetrix product only in accordance with the manual and written instructions provided by Affymetrix. You understand and agree that except as expressly set forth in the Affymetrix terms and conditions, that no right or license to any patent or other intellectual property owned or licensable by Affymetrix is conveyed or implied by this Affymetrix product. In particular, no right or license is conveyed or implied to use this Affymetrix product in combination with a product not provided, licensed or specifically recommended by Affymetrix for such use.

#### **Patents**

Software products may be covered by one or more of the following patents: U.S. Patent Nos. 5,733,729; 5,795,716; 5,974,164; 6,066,454; 6,090,555; 6,185,561; 6,188,783; 6,223,127; 6,228,593; 6,229,911; 6,242,180; 6,308,170; 6,361,937; 6,420,108; 6,484,183; 6,505,125; 6,510,391; 6,532,462; 6,546,340; 6,687,692; 6,607,887; 7,062,092 and other U.S. or foreign patents.

#### **Copyright**

© 2009-2011 Affymetrix, Inc. All Rights Reserved.

## Table of Contents

INTRODUCTION .....	6
ABOUT THIS UPDATE .....	6
ABOUT THIS MANUAL .....	7
TECHNICAL SUPPORT.....	7
<b>CHAPTER 1: INSTALLING CHROMOSOME ANALYSIS SUITE.....</b>	<b>9</b>
MINIMUM AND RECOMMENDED REQUIREMENTS AND PREREQUISITES .....	9
ZIP FILE CONTENTS .....	11
INSTALLING CYTOSCAN™ HD SOFTWARE FOR AGCC .....	12
INSTALLING CHAS .....	13
INSTALLING ANALYSIS FILES.....	18
UPDATES AND GENERAL INFORMATION .....	23
<b>CHAPTER 2: GETTING STARTED WITH CHROMOSOME ANALYSIS SUITE.....</b>	<b>25</b>
STARTING CHROMOSOME ANALYSIS SUITE .....	25
FILE TYPES AND DATA ORGANIZATION IN CHROMOSOME ANALYSIS SUITE .....	27
BASIC WORKFLOW FOR CYTOGENETICS ANALYSIS .....	29
WORKING WITH CHAS .....	33
<b>CHAPTER 3: CN/LOH/MOSAICISM ANALYSIS .....</b>	<b>36</b>
SINGLE SAMPLE ANALYSIS .....	36
INTRODUCTION TO SINGLE SAMPLE ANALYSIS .....	37
CREATING A REFERENCE FILE .....	48
USING THE ANALYSIS MANAGER .....	53
USING THE ANALYSIS PARAMETERS FUNCTIONS.....	57
<b>CHAPTER 4: LOADING DATA.....</b>	<b>66</b>
INTRODUCTION TO LOADING DATA .....	66
LOADING FILES .....	67
SEGMENT SMOOTHING AND SEGMENT JOINING (OPTIONAL) .....	71
SETTING QC PARAMETERS .....	76
<b>CHAPTER 5: VIEWING DATA IN CHAS.....</b>	<b>81</b>
OVERVIEW OF CHAS WINDOW COMPONENTS .....	82
FILES LIST .....	83
DATA TYPES LIST .....	84
NAMED SETTINGS .....	85
STATUS BAR.....	85
DISPLAY AREA .....	86
CHANGING THE CHAS BROWSER NETAFFX GENOMICS ANNOTATION FILE VERSION .....	88

<b>CHAPTER 6:     DISPLAYING DATA IN GRAPHIC VIEWS .....</b>	<b>89</b>
THE GRAPHIC DISPLAY VIEWS.....	89
CONTROLLING THE DISPLAY OF DATA .....	105
LEARNING MORE ABOUT FEATURES.....	122
<b>CHAPTER 7:     FILTERING SEGMENTS .....</b>	<b>130</b>
APPLYING SEGMENT PARAMETER FILTERS.....	130
<b>CHAPTER 8:     USING CYTOREGIONS .....</b>	<b>134</b>
SELECTING A CYTOREGIONS INFORMATION FILE .....	134
VIEWING CYTOREGIONS.....	136
SEARCHING CYTOREGIONS.....	141
USING FILTERS WITH CYTOREGIONS .....	141
USING RESTRICTED MODE .....	142
<b>CHAPTER 9:     USING THE OVERLAP MAP .....</b>	<b>144</b>
SELECTING THE OVERLAP MAP FILE .....	144
VIEWING OVERLAP REGIONS.....	146
SETTING THE OVERLAP FILTER .....	150
<b>CHAPTER 10:    CREATING AND EDITING AED FILES .....</b>	<b>152</b>
CREATING AN AED FILE OF ANNOTATIONS.....	152
CREATING AN AED FILE OF REGIONS .....	154
VIEWING OR EDITING AED FILE PROPERTIES .....	159
VIEWING AND EDITING ANNOTATIONS.....	162
AED/BED COLOR RULES .....	169
EXPORTING INFORMATION IN AED OR BED FORMAT .....	174
<b>CHAPTER 11:    DISPLAYING DATA IN TABLE VIEWS.....</b>	<b>177</b>
COMMON TABLE OPERATIONS.....	178
SEGMENTS TABLE.....	181
GRAPHS TABLE.....	191
QC AND SAMPLE INFO TAB.....	196
CHROMOSOME SUMMARY DATA .....	198
SEARCHING RESULTS .....	200
FINDING INTERSECTIONS.....	203
<b>CHAPTER 12:    REPORTING RESULTS .....</b>	<b>206</b>
EXPORTING GRAPHIC VIEWS .....	206
EXPORTING TABLE DATA.....	212
COMBINING PDFs INTO A SINGLE REPORT.....	219
<b>CHAPTER 13:    USER PROFILES AND NAMED SETTINGS .....</b>	<b>225</b>
TYPES OF SETTINGS .....	225



CREATING AND USING USER PROFILES.....	227
CREATING AND USING NAMED SETTINGS .....	228
EXPORTING AND IMPORTING PREFERENCES.....	230
<b>APPENDIX A: ANALYSIS PARAMETERS.....</b>	<b>233</b>
ANALYSIS PARAMETERS FOR SINGLE SAMPLE ANALYSIS.....	233
REFERENCE MODEL FILE CREATION .....	236
<b>APPENDIX B: AED FILE FORMAT .....</b>	<b>237</b>
AED FILE DESCRIPTION.....	237
PROPERTY NAME ELEMENTS.....	239
COMPATIBILITY .....	243
REFERENCES .....	245
<b>APPENDIX C: CHAS PROPERTIES AND TYPES .....</b>	<b>246</b>
IDENTIFYING PROPERTIES WITHIN CHAS.....	246
CHAS PROPERTY STYLE.....	247
FULL URI PROPERTY STYLE.....	248
IDENTIFYING VALUE TYPES WITHIN CHAS .....	248
AUTOMATIC CONVERSION OF CxCHP HEADERS TO PROPERTIES .....	248
<b>APPENDIX D: CONFIDENCE OF SEGMENT CALCULATION .....</b>	<b>251</b>
COPY NUMBER GAIN AND LOSS CONFIDENCE CALCULATION.....	251
LOH CONFIDENCE CALCULATION .....	251
MOSAICISM CONFIDENCE CALCULATION.....	251
<b>APPENDIX E: GENOMIC POSITION COORDINATES .....</b>	<b>253</b>
GENOME ASSEMBLIES.....	253
SNP AND MARKER POSITIONS.....	253
SEGMENT POSITIONS.....	253
DIFFERENCE BETWEEN CHAS 1.0 AND CHAS 1.1 .....	254
<b>APPENDIX F: EDITING BED FILES.....</b>	<b>255</b>
<b>APPENDIX G: CYTOSCAN™ HD ALGORITHMS AND QC METRICS .....</b>	<b>257</b>
ALGORITHM OVERVIEW .....	257
DUAL QUANTILE NORMALIZATION.....	259
COPY NUMBER WORKFLOW .....	259
SNP MARKER WORKFLOW.....	262
COPY NUMBER CONFIDENCE SCORE .....	263
ARRAY DATA QC METRICS.....	263
MAPD – DETAILED DESCRIPTION.....	264
WAVINESS-SD – DETAILED DESCRIPTION .....	265
SNPQC – DETAILED DESCRIPTION .....	266

## Introduction

The Chromosome Analysis Suite (ChAS) software for cytogenetic analysis enables you to view and summarize chromosomal aberrations across the genome. Chromosomal aberrations may include copy number gain or loss, mosaicism, or loss of heterozygosity (LOH).

ChAS provides tools to:

- Perform single sample analysis of CEL files from the Affymetrix® CytoScan™ HD Array or Cytogenetics Whole-Genome 2.7M Array.
- Analyze segment data at different levels of resolution
- View results data (CYCHP files) that summarize chromosomal aberrations in table and graphical formats
- Display CNCHP data from Affymetrix® Genome-Wide Human SNP Array 6.0
- Customize and load your own annotations and regions for focused analysis
- Apply separate filters to the entire genome and user-specified regions of interest to remove irrelevant information such as segments in areas that are not of interest
- Perform detailed comparisons between different samples
- Directly access external databases such as NCBI, UCSC Genome Browser, Ensembl, and OMIM
- Generate reports of user-selected data in graphical and tabular formats



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

Chromosome Analysis Suite is not a secondary analysis package. However, it does create CYCHP files and tab-delimited text files required for secondary analysis packages available from companies in the Affymetrix® GeneChip® Compatible Program.

## About this Update

ChAS 1.2 includes the following new features and enhancements:

- Genotype calling for Affymetrix® CytoScan™ HD Array results (page 191), exporting genotype calls (page 193)
- Status bar displays the NetAffx database build loaded in the ChAS Browser (page 85)
- Load a different NetAffx Genomic Annotations Browser database into the ChAS Browser without restarting the software (page 88)
- View sample and results information in the Sample Information window (page 69)
- New Named Settings (page 225)
- Easily identify User Profiles available for import (page 231)
- Apply user-selected colors to reference annotations (except DGV annotations, which contain a pre-defined set of differential color codes) in the Detail View (page 99)
- Segments table updates:
  - When multiple reference annotations overlap a segment, the genes in the Segments table row are listed in the order that they appear from left to right in the Detail View

- The “Types” column sorts segments in the same order that they appear in the Data Types windowpane
- ChAS verifies the hg version of AED or BED files and will not load files with hg version different from that currently loaded in the ChAS Browser. If the hg version information is not found in the AED or BED file, ChAS warns you before loading the file.
- AED files created in ChAS 1.2 include the hg version
- All loaded AED or BED files are automatically reloaded in a subsequent session with the same user profile
- Reports include CytoRegion and Overlap Map file names
- ChAS remembers the destination folder for exported reports for a user profile during a session and in subsequent sessions

## About this Manual

This manual provides step-by-step instructions for performing the procedures required to use ChAS. The content of this manual is also available in the online help for ChAS. The steps outlining procedures are frequently supplemented with screen captures to further illustrate the instructions. The screen captures in this manual may not exactly match the windows displayed on your screen.

## Technical Support

Affymetrix provides technical support to all licensed users via phone or E-mail. To contact Affymetrix Technical Support:

AFFYMETRIX, INC.

3420 Central Expressway

Santa Clara, CA 95051 USA

Tel: 1-888-362-2447 (1-888-DNA-CHIP)

Fax: 1-408-731-5441

[sales@affymetrix.com](mailto:sales@affymetrix.com)

[support@affymetrix.com](mailto:support@affymetrix.com)

AFFYMETRIX UK Ltd.,

Voyager, Mercury Park,

Wycombe Lane, Wooburn Green,

High Wycombe HP10 0HH

United Kingdom

UK and Others Tel: +44 (0) 1628 552550

France Tel: 0800919505

Germany Tel: 01803001334

Fax: +44 (0) 1628 552585

[saleseurope@affymetrix.com](mailto:saleseurope@affymetrix.com)  
[supporteurope@affymetrix.com](mailto:supporteurope@affymetrix.com)

AFFYMETRIX JAPAN K.K.

Mita NN Bldg. 16F

4-1-23 Shiba Minato-ku,

Tokyo 108-0014 Japan

Tel. 03-5730-8200

Fax: 03-5730-8201

[salesjapan@affymetrix.com](mailto:salesjapan@affymetrix.com)

[supportjapan@affymetrix.com](mailto:supportjapan@affymetrix.com)

## Chapter 1: Installing Chromosome Analysis Suite

Chromosome Analysis Suite (ChAS) is a stand-alone application.

To process CytoScan™ HD Arrays and analyze the data, you need to install:

- CytoScan™ HD Array software for AGCC for processing arrays
- Chromosome Analysis Suite 1.2 for data analysis
- Analysis files for data analysis



**Note:** Due to the amount of memory that ChAS needs to operate, Affymetrix very strongly recommends that you do not install the ChAS software on production AGCC computers being used for scanning and operating fluidics systems.

This chapter includes:

- [Minimum and Recommended Requirements and Prerequisites](#) (below)
- [Zip File Contents](#) (page 11)
- [Installing CytoScan™ HD Software for AGCC](#) (page 12)
- [Installing ChAS](#) (page 13)
- [Installing Analysis Files](#) (page 18)
- [Updates and General Information](#) (page 23)

### Minimum and Recommended Requirements and Prerequisites

A 64-bit system is recommended for all array types. System requirements are specific for the array type that you plan to analyze.



**Note:** Chromosome Analysis Suite requires AGCC 3.2 or higher to produce CytoScan™ HD CEL files.



**Note:** A 64-bit system is required to generate CytoScan™ HD CYCHP data files. CytoScan™ HD CYCHP files can be viewed on a 32-bit system.

Chromosome Analysis Suite has been verified on the following operating systems. ChAS may work on other Windows operating systems, but has not been verified on them.

- Operating system on 32-bit
  - Windows 7 Professional with service pack 1
  - Windows XP with service pack 3
- Operating system on 64-bit
  - Windows 7 Professional with service pack 1
  - Windows XP Professional with service pack 2

**Table 1.1 Minimum operating system and hardware requirements for ChAS software per array type**

Item	Array Type		
	CytoScan™ HD Array	Cytogenetics Whole-Genome 2.7M Array	Genome-Wide Human SNP 6.0 Array

<b>Processor</b>	At least 2 GHz Pentium Dual Processor	At least 2 GHz Pentium Dual Processor	At least 2 GHz Pentium Dual Processor
<b>64-bit Windows Operating System and Web Browser</b>	Windows 7 Professional SP1 and Internet Explorer 8.0 or above or Windows XP SP2 and Internet Explorer 7.0 or above	Windows 7 Professional SP1 and Internet Explorer 8.0 or above or Windows XP SP2 and Internet Explorer 7.0 or above	Windows 7 Professional SP1 and Internet Explorer 8.0 or above or Windows XP SP2 and Internet Explorer 7.0 or above
<b>32-bit Windows Operating System and Web Browser</b>	<b>( Data Viewing Only)</b> Windows XP SP3 and Internet Explorer 7.0 or above or Windows 7 Professional SP1 and Internet Explorer 8.0	Windows 7 Professional SP1 and Internet Explorer 8.0 or above or Windows XP SP3 and Internet Explorer 7.0 or above	Windows 7 Professional SP1 and Internet Explorer 8.0 or above or Windows XP SP3 and Internet Explorer 7.0 or above
<b>Disk Space</b>	150 GB HD + data storage*	150 GB HD + data storage	150 GB HD + data storage
<b>Free Disk Space Required at Install</b>	≥ 5 GB	≥ 5 GB	≥ 5 GB
<b>RAM</b>	3 GB (32-bit for data viewing only) 8 GB (64-bit)	3 GB (32-bit) 8 GB (64-bit)	3 GB (32-bit) 8 GB (64-bit)
<b>CYCHP file generated in:</b>	ChAS 1.2	ChAS 1.2, 1.1, 1.0.1	N/A
<b>CYCHP file viewed in:</b>	ChAS 1.2	ChAS 1.2, 1.1, 1.0.1	N/A
<b>CNCHP file generated in:</b>	N/A	N/A	Genotyping Console (GTC)
<b>CNCHP file viewed in:</b>	N/A	N/A	ChAS 1.2, 1.1, 1.0.1 GTC

\*The larger file sizes associated with the CytoScan™ HD Array should be taken into account when calculating the necessary free space requirement. A CytoScan™ HD Array CYCHP file is ~120 MB.

**Table 1.2 Recommended operating system and minimum hardware requirements for ChAS software per array type**

Item	Array Type
------	------------

	<b>CytoScan™ HD Array</b>	<b>Cytogenetics Whole-Genome 2.7M Array</b>	<b>Genome-Wide Human SNP 6.0 Array</b>
<b>Processor</b>	4 GHz Pentium Quad Core Processor	4 GHz Pentium Quad Core Processor	4 GHz Pentium Quad Core Processor
<b>Windows Operating System and Web Browser</b>	Windows 7 Professional SP1 and Internet Explorer 8.0 or above	Windows 7 Professional SP1 and Internet Explorer 8.0 or above	Windows 7 Professional SP1 and Internet Explorer 8.0 or above
<b>System</b>	64-bit	64-bit	64-bit
<b>Disk Space</b>	250 GB HD + data storage*	250 GB HD + data storage	250 GB HD + data storage
<b>Free Disk Space Required at Install</b>	≥ 5 GB	≥ 5 GB	≥ 5 GB
<b>RAM</b>	16 GB	16 GB	16 GB
<b>CYCHP file generated in:</b>	ChAS 1.2	ChAS 1.2	N/A
<b>CYCHP file viewed in:</b>	ChAS 1.2	ChAS 1.2	N/A
<b>CNCHP file generated in:</b>	N/A	N/A	Genotyping Console (GTC)
<b>CNCHP file viewed in:</b>	N/A	N/A	ChAS 1.2

\* The larger file sizes associated with the CytoScan™ HD Array should be taken into account when calculating the necessary free space requirement. A CytoScan™ HD Array CYCHP file is ~120 MB.

## Zip File Contents

There are four items you will need to run the CytoScan™ HD Array and visualize the data in ChAS 1.2.

- 2 AGCC workstation files, from the CytoScan™ HD product web page at Affymetrix.com
  - CytoScanHD for AGCC.zip (see below for instructions)
  - CytoScanHD Fluidics Script (see array product page for details)
- 2 Data Analysis workstation files, from ChAS product web page ([www.affymetrix.com/chas](http://www.affymetrix.com/chas)):
  - Chromosome Analysis Suite 1.2.zip (see below for instructions)
  - CytoScanHD Analysis Files.zip (see below, or download within ChAS 1.2)

The ChAS software installer is in a zipped package with several files.

The zipped package can be downloaded from [Affymetrix.com/chas](http://Affymetrix.com/chas)

After you extract the files from the "Chromosome Analysis Suite 1.2.zip" zipped package, there is a folder named "Chromosome Analysis Suite 1.2". This folder contains:

- A "64Bit" ChAS installer folder and a "32Bit" ChAS installer folder
- Installation Instructions
- Release Notes
- ChAS User Manual

The CytoScanHD for AGCC.zip contains files that are required to install the CytoScanHD software on the computer running AGCC.

- CytoScanHD\_Array.exe
- Pedigree template
- AGCC\_CytoScanHD\_Install\_Instructions.pdf

## Installing CytoScan™ HD Software for AGCC

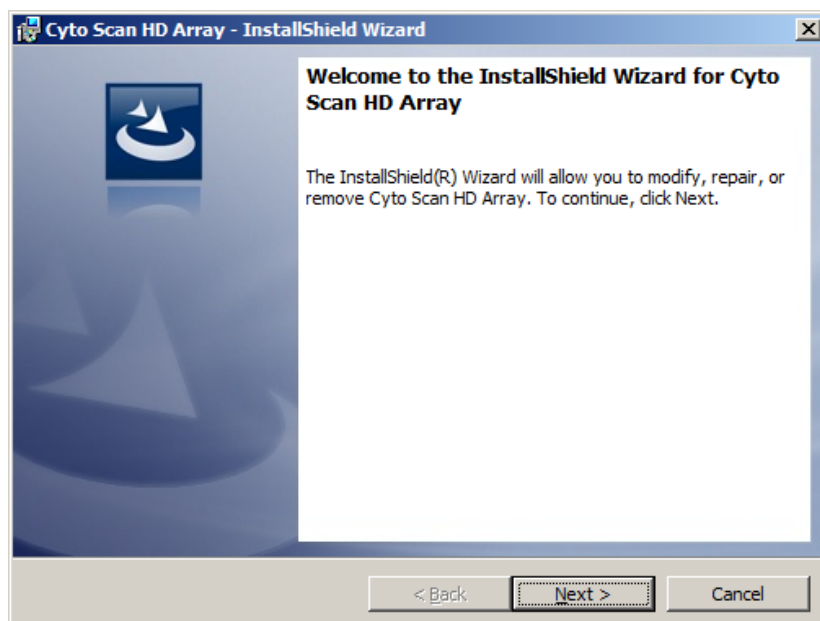
To process CytoScan™ HD Arrays on AGCC, the "CytoScanHD for AGCC.zip" package must be installed on the AGCC instrument control workstation. This module enables AGCC to scan and generate CEL files for the Affymetrix® CytoScan™ HD Array.

**To install the CytoScanHD software:**

 **Note:** The AGCC CytoScanHD library file installer requires AGCC 3.2 or higher. This installer installs the library files and gridding components that support the CytoScan™ HD Array.

1. Double-click the executable (CytoScanHD\_Array.exe).

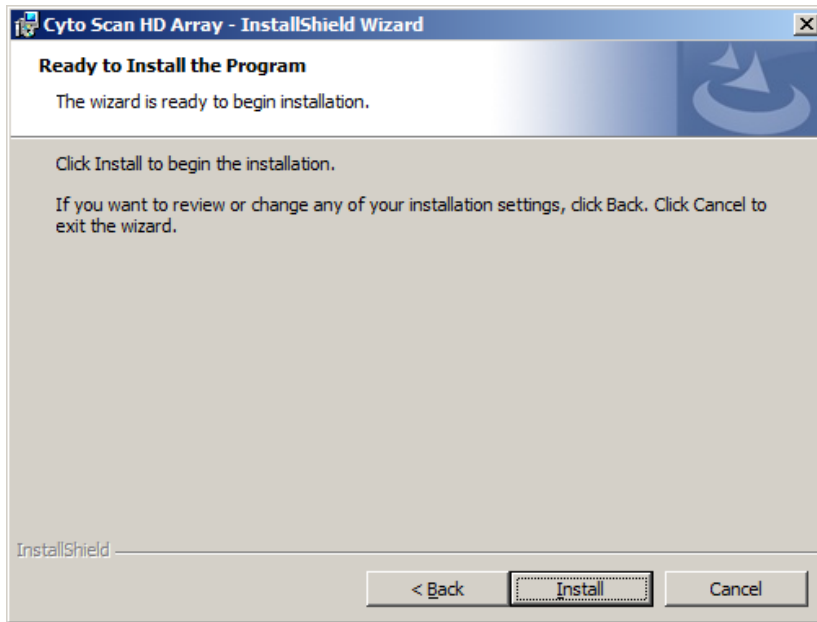
The InstallShield Wizard screen appears.



**Figure 1.1** CytoScanHD Array install wizard

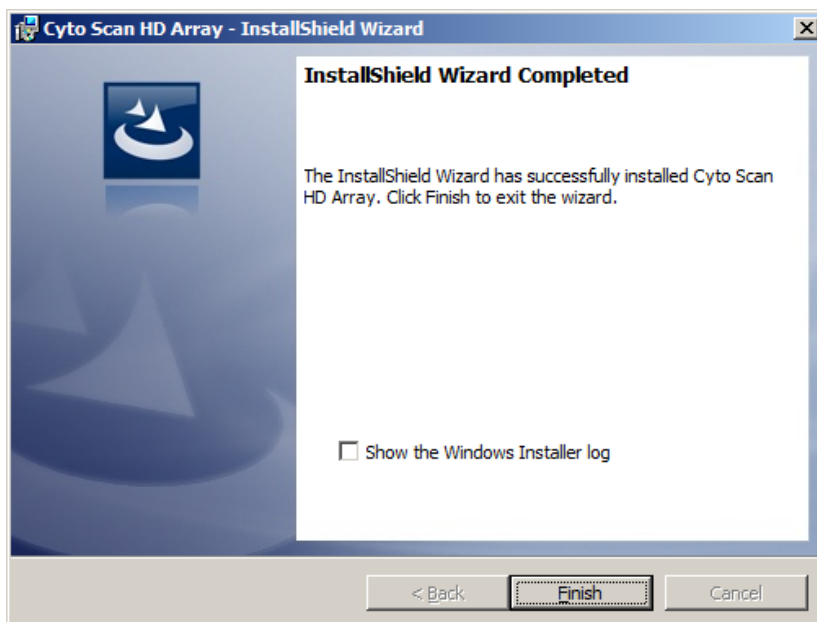
2. Click **Next**.





**Figure 1.2 Ready to Install**

3. Click **Install** to begin the installation.



**Figure 1.3 Install completed**


4. Click **Finish** to complete the installation.


## Installing ChAS



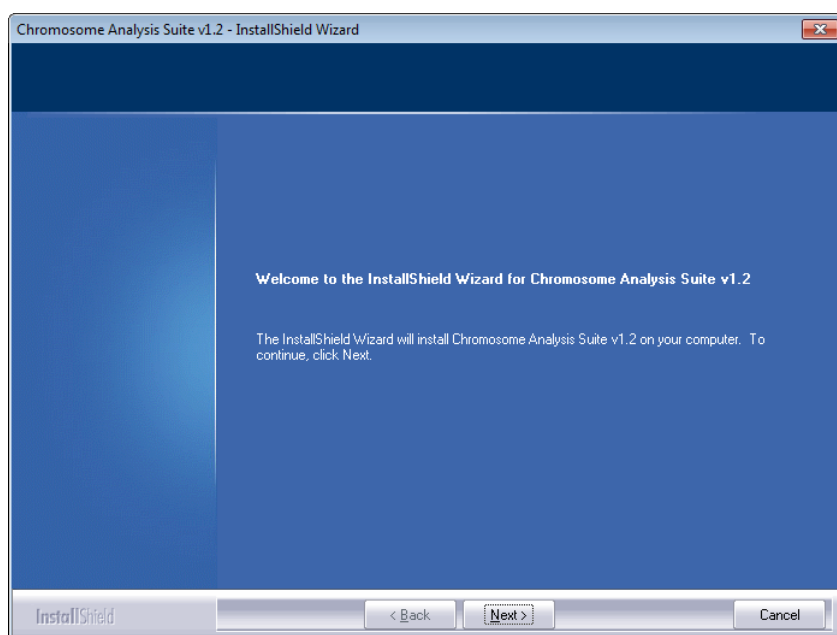
**Note:** The setup process installs additional required components, which includes Java components and Visual C++ runtime.

 **Note:** The ChAS 1.2 installer automatically archives outdated ChAS library files when installing ChAS 1.2 on workstations with ChAS 1.0 or 1.0.1. For workstations with ChAS 1.1, no files need to be archived.

1. In the download of ChAS, the folder contains a 32-bit folder and 64-bit folder. Select the appropriate folder based on your computer (32-bit system or 64-bit system). Double-click the Chromosome Analysis Suite setup icon  on the desktop.
2. Follow the directions in the next screen that appears to install the version of Java that ChAS requires.

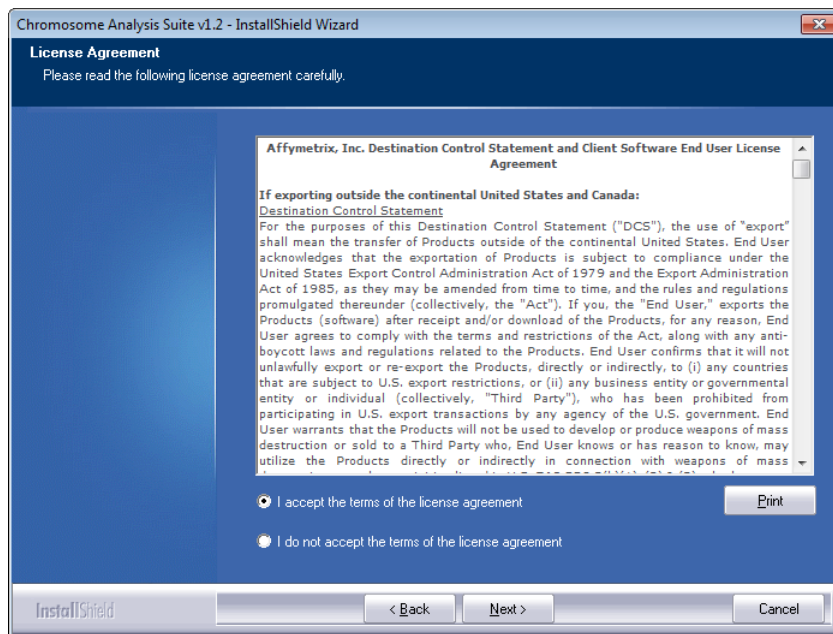
 **Note:** Starting with ChAS 1.2, the software installs a copy of Java JRE into the program directory for its own use. The installation proceeds regardless of whether there are other versions of Java JRE or Java SDK installed on your system. Adding or removing other instances of Java will not affect the copy of Java JRE in the ChAS program directory.

3. In the Welcome screen that appears, click **Next**.



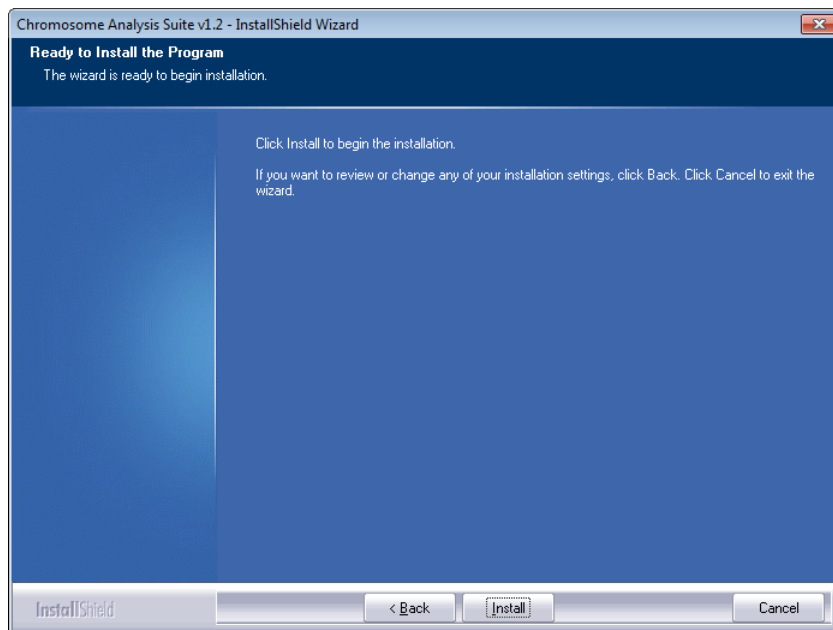
**Figure 1.4 InstallShield Wizard, Welcome screen**

4. Accept the terms of the license agreement and click **Next**.



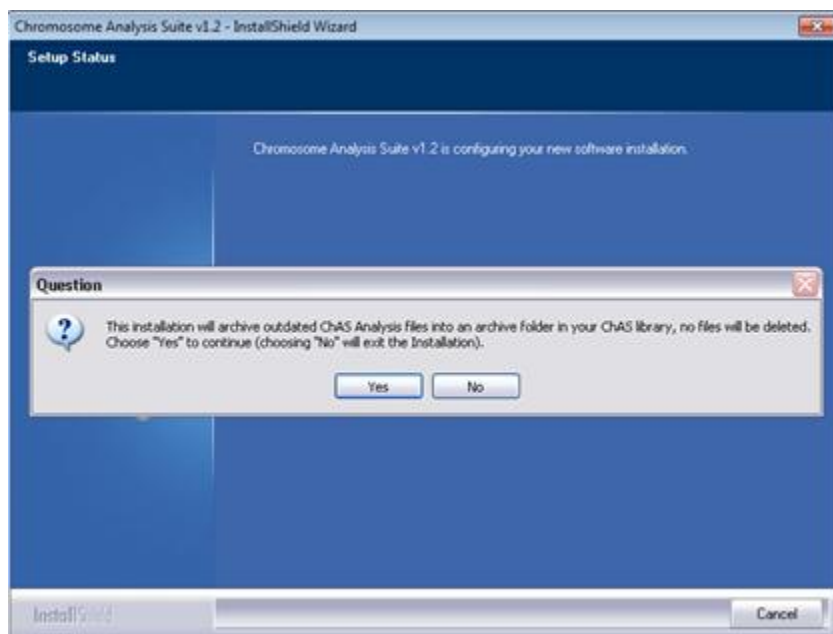
**Figure 1.5 InstallShield Wizard, License Agreement screen**

5. In the next screen that appears, click **Next** to begin the installation.



**Figure 1.6 InstallShield Wizard, Ready to Install the Program screen**

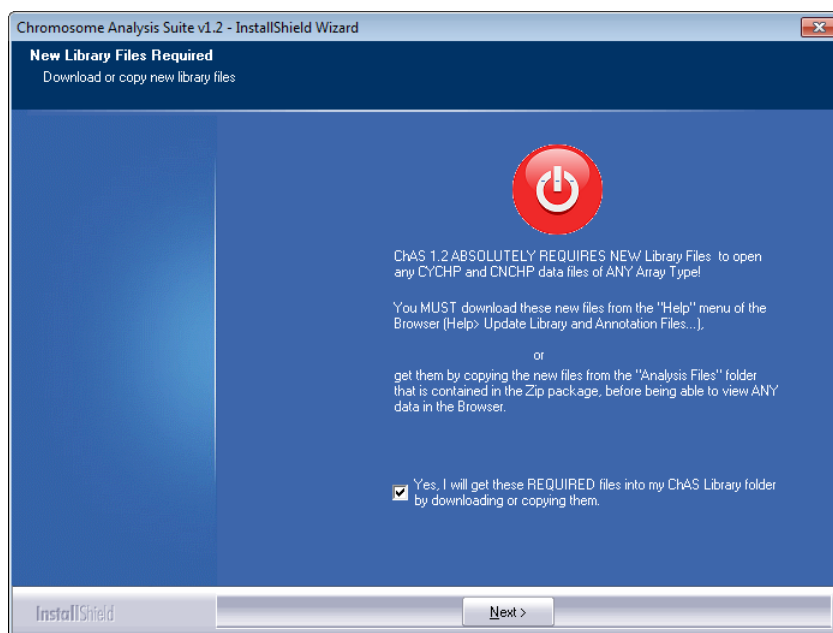
6. Click **Yes** in the message that appears to continue with the installation.



**Figure 1.7 Notification that outdated ChAS files will be archived**

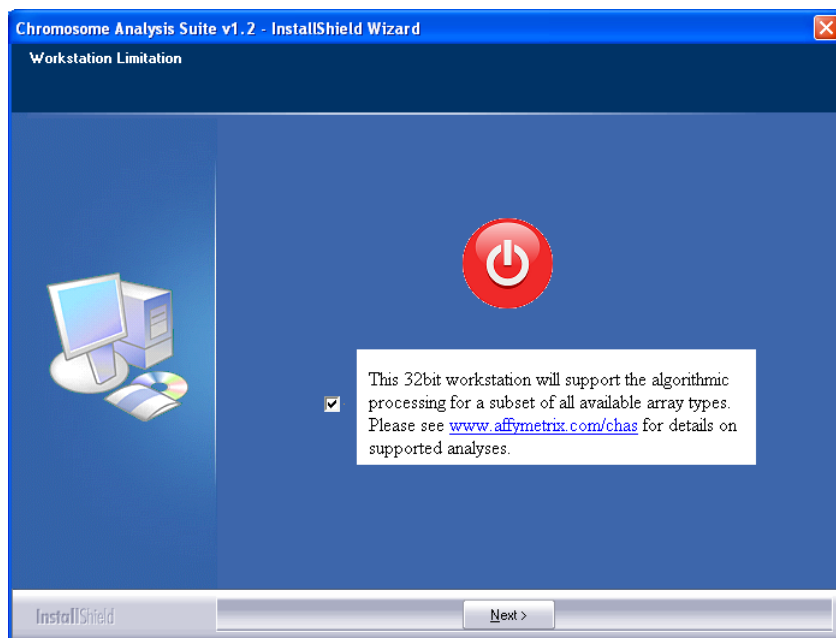
 **Note:** This message only appears when upgrading from ChAS 1.0 or 1.1 to ChAS1.2.

7. Put a check mark in the check box to confirm that you will obtain the new library files that ChAS requires to open CYCHP and CNCHP data files. Click **Next**.



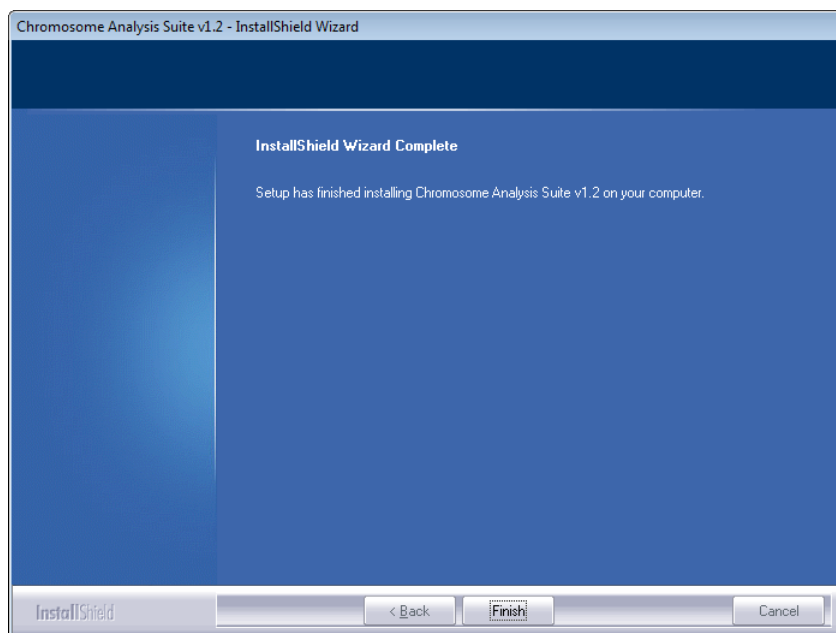
**Figure 1.8 Confirm that you will obtain the required library files**

8. If installing the 32-bit version of the software, the following screen appears. Acknowledge that you have read the information by putting a check mark in the check box, then click **Next**.



**Figure 1.9 Notification for 32-bit workstations**

9. When the install is complete, click **Finish**.



**Figure 1.10 Installation complete**

## **Un-installing ChAS**

1. From the Windows Start Menu, select **Settings > Control Panel**.
2. Select **Add or Remove Programs**.
3. Select Chromosome Analysis Suite in the programs list.
4. Click the **Change/Remove** button and following the on screen directions.

 **Note:** You can also uninstall ChAS by running the installation file.

## Installing Analysis Files


You need to install the required analysis files, also known as library files, before using ChAS.

This section covers:

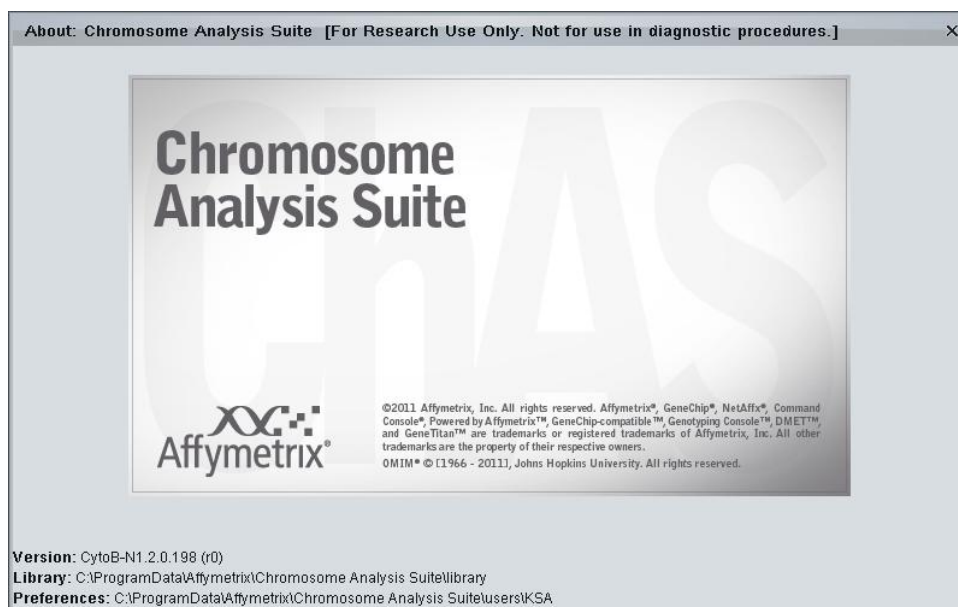
- [Analysis File Locations](#), below
- [Analysis File Download](#) (page 21)
- [Copying Analysis Files](#) (page 23)

### Analysis File Locations

To see the location of the analysis files:

- From the Help menu, click **About...**  
or
- Click the **About**  button in the main toolbar.

The About this Software screen opens. At the bottom it displays the path and location for the library files and preference files.



**Figure 1.11 About Chromosome Analysis Suite screen**

 **Note:** The location of the files is different in Windows XP and Windows 7.

For Windows XP:

- Library: C:\Documents and Settings\All Users\Application Data\Affymetrix\Chromosome Analysis Suite\Library
- Preferences: C:\Documents and Settings\All Users\Application Data\Affymetrix\Chromosome Analysis Suite\users

For Windows 7:

- Library: C:\ProgramData\Affymetrix\Chromosome Analysis Suite\Library
- Preference file: C:\ProgramData\Affymetrix\Chromosome Analysis Suite\preferences.xml
- All other user profile related preference files and saved settings: C:\ProgramData\Affymetrix\Chromosome Analysis Suite\users

The ChAS analysis and preference files may be placed in folders and files that are normally hidden from the user in Windows.

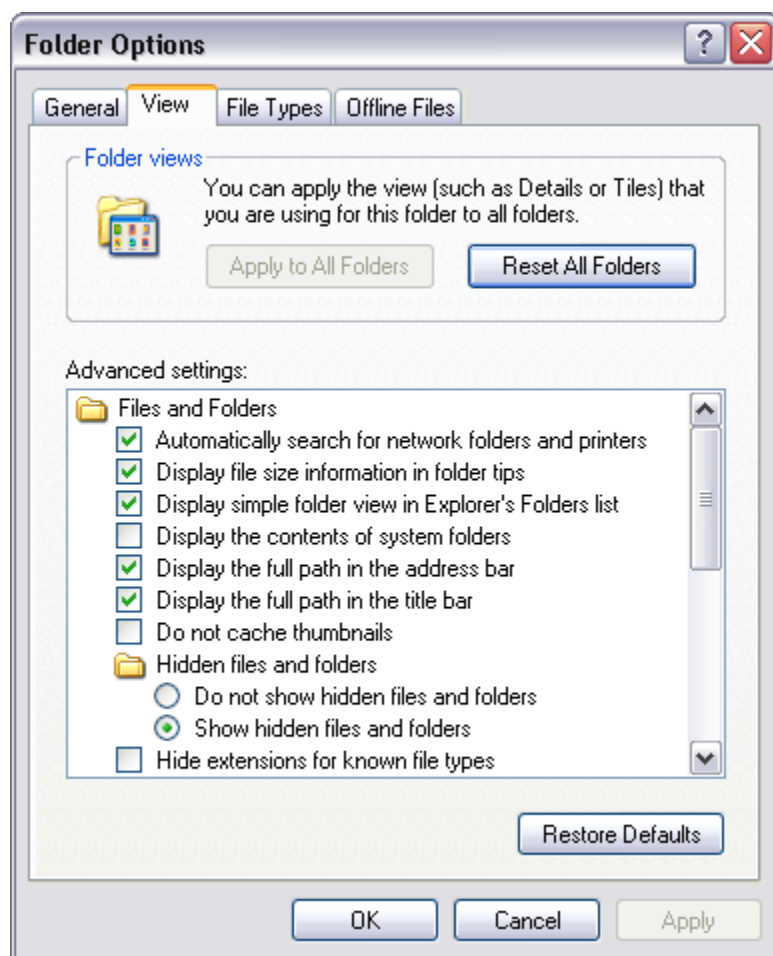
**To display hidden files and folders in Windows XP:**

1. From the Start button, select **Settings > Control Panel**.

The Control Panel window opens.

2. Double-click **Folder Options** icon in Control Panel.

The Folder Options dialog box opens.



**Figure 1.12 Folder Options dialog box**

3. On the View tab, under Hidden files and folders, click **Show hidden files and folders**.

Hidden files and folders are dimmed to indicate they are not typical items.

If you know the name of a hidden file or folder, you can search for it.

### To display hidden files and folders in Windows 7:

1. In the Start menu, click Control Panel.  
The Control Panel window opens.
2. Click **Appearance and Personalization** in Control Panel. Under Folder Option, click “Show hidden files and folders”.

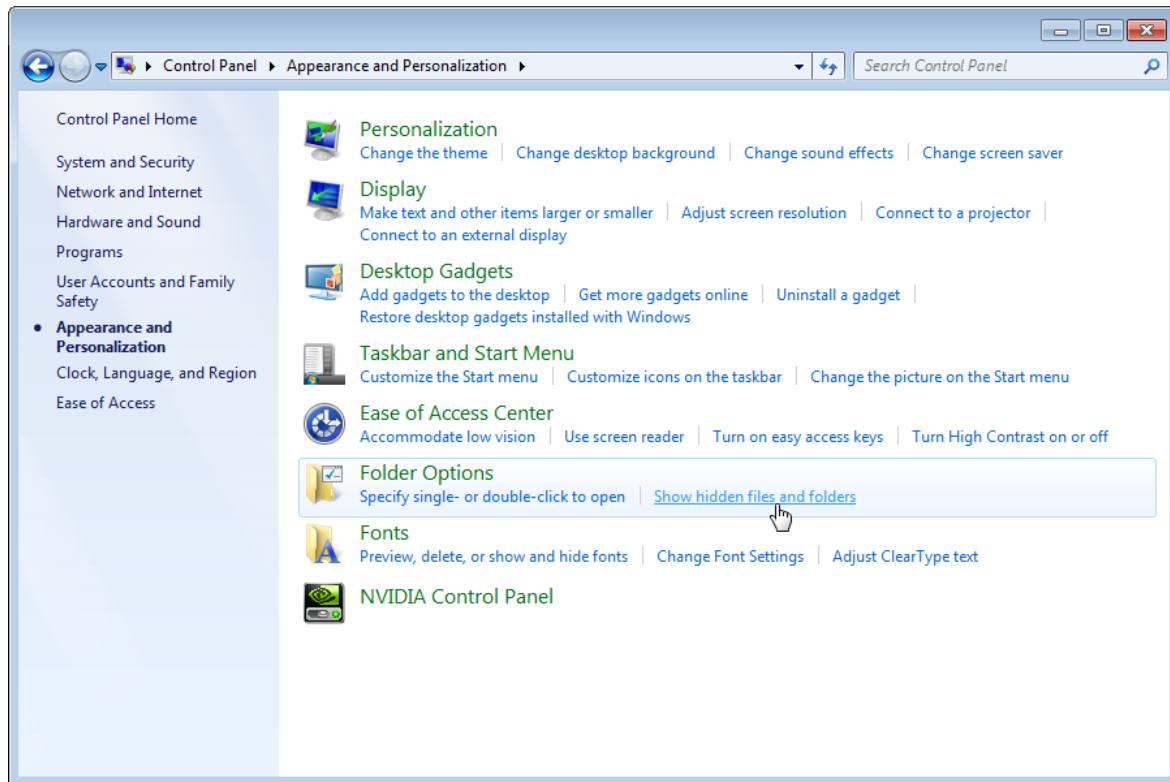
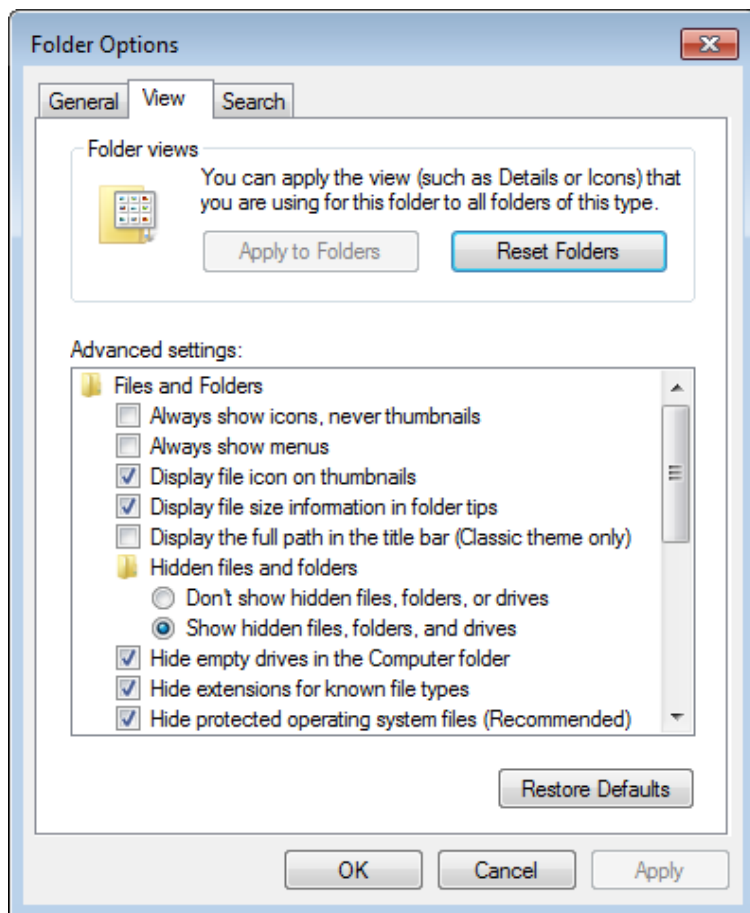


Figure 1.13 Control Panel

3. In the Folder Options dialog box that appears, click the View tab. Under Hidden files and folders, click **Show hidden files and folders**.

Hidden files and folders are dimmed to indicate they are not typical items. If you know the name of a hidden file or folder, you can search for it.





**Figure 1.14 Folder Options dialog box**

## **Analysis File Download**

When you start ChAS for the first time, you will be prompted to:

1. Create a user profile (see [Creating and Using User Profiles](#), page 227).
2. Download the library files from NetAffx that ChAS uses to analyze and annotate the data.

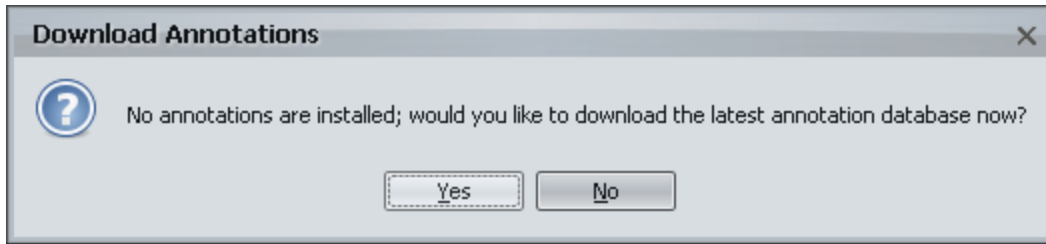


**Note:** To process the CytoScan™ HD Array in AGCC, you will also need to install the appropriate analysis files for AGCC (see [Installing CytoScan™ HD Software for AGCC](#) on page 12).

### **To download the ChAS analysis files:**

1. Create a NetAffx account with a user name and password (go to [www.affymetrix.com](http://www.affymetrix.com) and click “Register” at the top of the page).
2. Start ChAS.

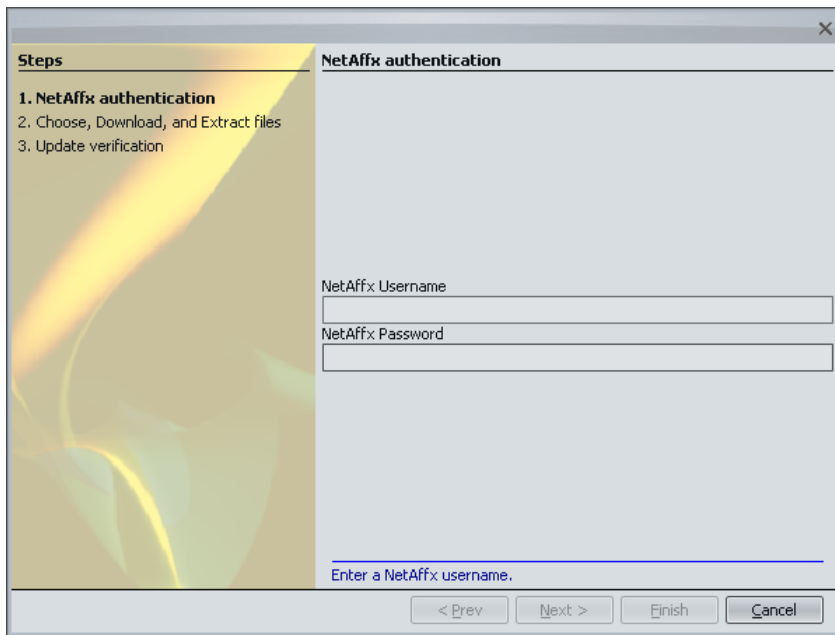
If no annotations are installed, the Download Annotations notice appears.



**Figure 1.15 Download Annotations notice**

3. Click **Yes**.

The NetAffx Authentication dialog box opens.

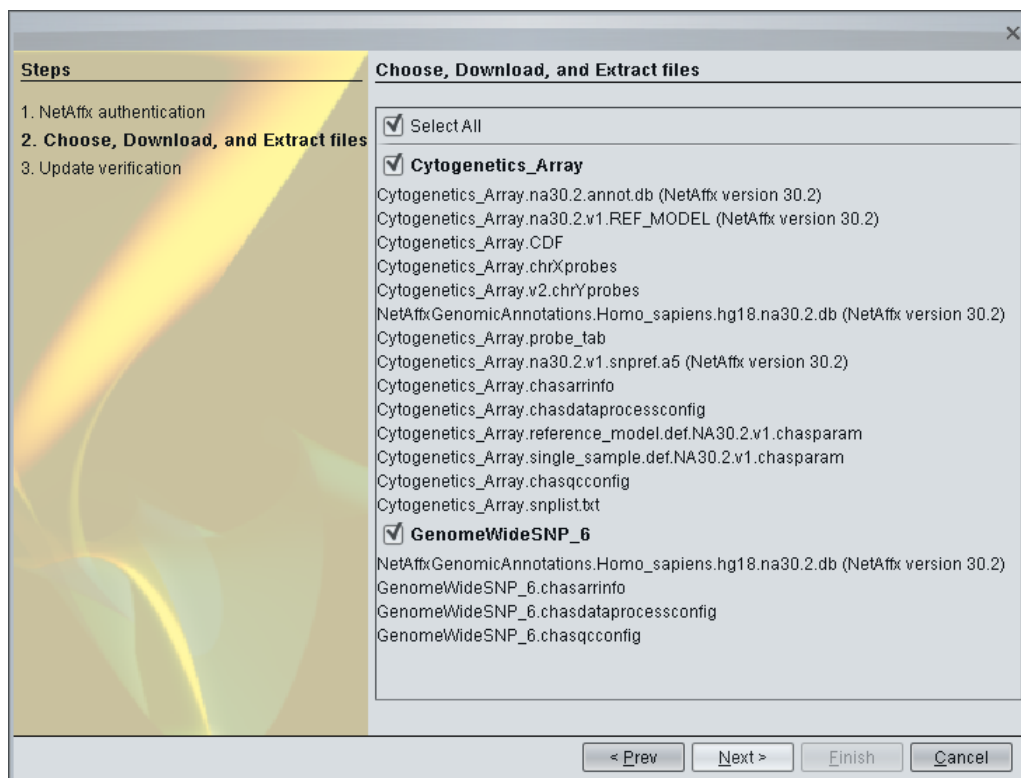


**Figure 1.16 NetAffx Authentication dialog box**

 **Note:** You can also open the NetAffx Authentication dialog box by selecting “Update Library and Annotation Files” from the Help menu.

4. Enter your NetAffx user name and password and click **Next**.

The Choose Files dialog box opens with a list of library files that your local hardware can support. Some array types require more memory or processing power. If your hardware does not meet the power requirements, the library files for that array will not appear.



**Figure 1.17 Choose Files dialog box**

5. Select the library and annotation files you wish to download.

You can choose **Select All** to download all the available files.

6. Click **Next**.

The Download Progress dialog box displays the progress of the downloading and unpacking of the files.

7. Click **Finish** when the download is complete.

## Copying Analysis Files

If you are unable to connect to the Internet to download the analysis files, copy the contents from the Analysis Files folder to the following location:

- If using Windows XP copy the files to:  
C:\Documents and Settings\All Users\Application Data\Affymetrix\Chromosome Analysis Suite\Library
- If using Windows 7, copy the files to:  
C:\ProgramData\Affymetrix\Chromosome Analysis Suite\Library

See [Analysis File Locations](#) (page 18) for more information on the location of the library files and on making hidden folders visible.

## Updates and General Information

New information about Chromosome Analysis Suite (ChAS) is available to customers through the Update Button on the main ChAS tool bar. There are three different options: Updates Available, No New Updates, or Updates (Offline).



Indicates updated information is available. Click the button to launch a web browser which indicates the new information that is available. Clicking this button merely provides you with a link to information; it does not cause any new software or files to be installed or any existing files to be modified.



Indicates there is no new updated information available. Clicking the button will launch a web browser that shows the current informational messages.



Indicates that the computer is offline and ChAS is unable to determine if there are any updates available.

## Chapter 2: Getting Started with Chromosome Analysis Suite

This chapter provides some context for using ChAS. It covers:

- [Starting Chromosome Analysis Suite](#) (below)
- [File Types and Data Organization in Chromosome Analysis Suite](#) (page 27)
- [Basic Workflow for Cytogenetics Analysis](#) (page 29)
- [Working with ChAS](#) (page 33)

### Starting Chromosome Analysis Suite

To start Chromosome Analysis Suite:

1. Double-click the ChAS icon  on the desktop.

Alternatively, from the Windows Start Menu, select **Programs > Affymetrix > Chromosome Analysis Suite > Chromosome Analysis Suite**.

The ChAS splash screen and the Select User dialog box open.

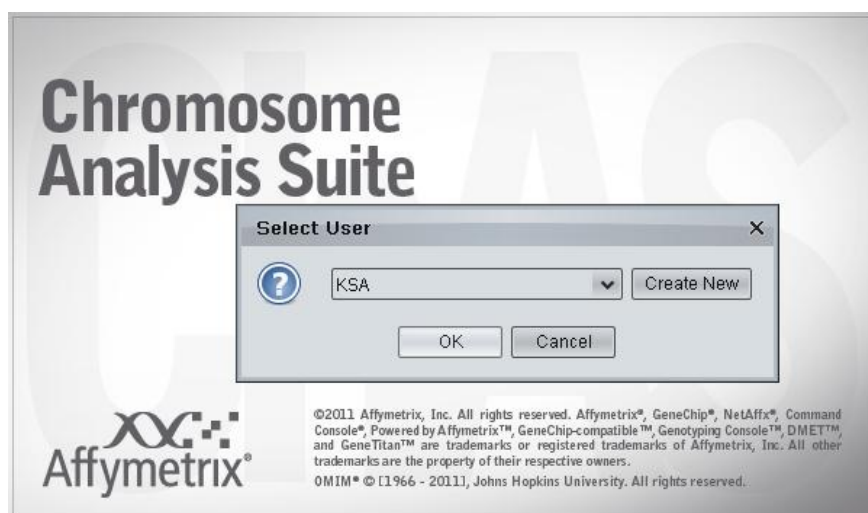
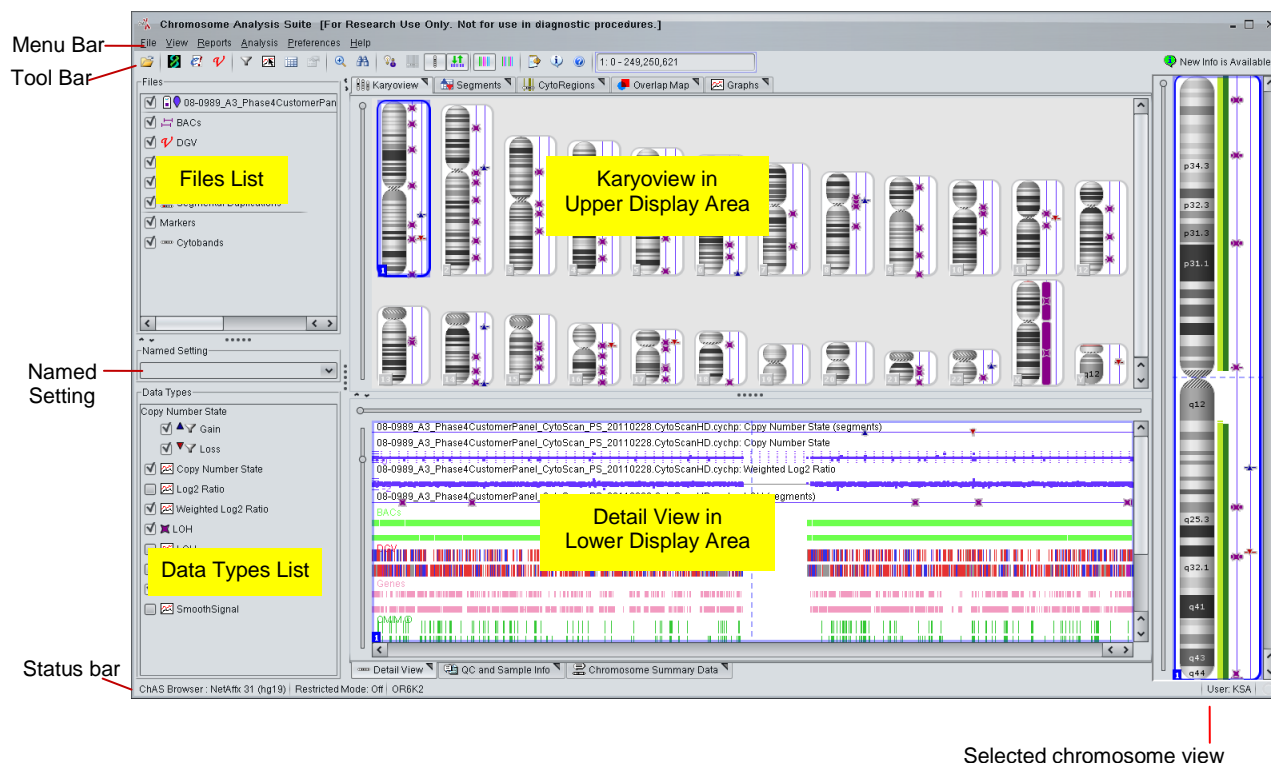


Figure 2.1 Splash screen and Select User dialog box

2. Select or create a User Profile (see [Creating and Using User Profiles](#), page 227).
3. Click **OK**.

The Chromosome Analysis Suite opens.



**Figure 2.2 Chromosome Analysis Suite when first opened**

The ChAS window has the following components:

- **Menu Bar** – Provides access to the functions of the software
- **Tool Bar** – Provides quick access to commonly used functions
- **Files List** (page 83) – Shows the data and annotation files that are available for display
- **Data Types List** (page 84) – Displays the type of data available in the files.
- **Named Settings** (page 85) – Displays a list of the previously saved display settings for ChAS.
- **Status Bar** (page 85) – Displays information on the status of the software, the ChAS Browser NetAffx Genomic Annotation file version, the hg version, information about the annotation or probe that the mouse pointer is nearest to in the Detail View, and the user profile name
- **Display Area** (page 86) – Displays the following data in graphical and table formats:
  - CYCHP and CNCHP marker data
  - Detected segments
  - Region information file data
  - Reference annotations
- **Status Bar** – Displays information on the status of the software, as well as the NetAffx version and hg version of the currently loaded NetAffx Genomic Annotation file.

See [Viewing Data in ChAS](#) (page 81) for more information.

## The Analysis Manager


The Analysis Manager tracks ongoing analysis tasks for ChAS and alerts you when the computer does not have sufficient memory to perform an analysis. When the Analysis Manager is running, it is indicated by an icon  in the right side of the Windows taskbar.



Figure 2.3 Analysis Manager icon in the Windows task bar

To open the Analysis Manager:

- Right-click the Analysis Manager icon  and select **Status**.

The Analysis Manager Status window opens.

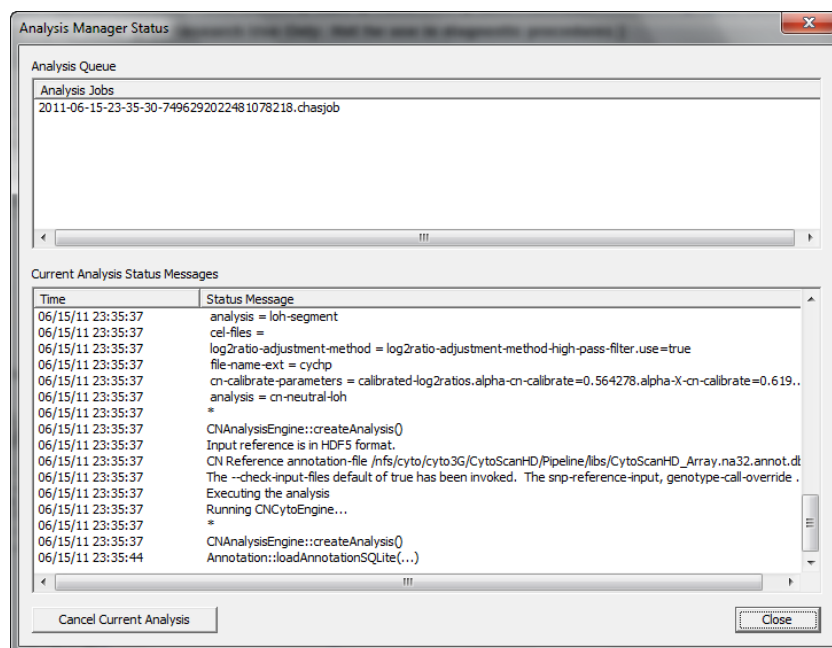


Figure 2.4 Analysis Manager

See [Using the Analysis Manager](#) (page 53) for more information.

## File Types and Data Organization in Chromosome Analysis Suite

To fully use the capabilities of ChAS, you need to understand the ChAS file types and data organization in ChAS.

### Chromosome Analysis Suite File Types

ChAS uses the following types of files.

- Data files
- Region Information files
- Support files


## Files Types Supported in ChAS

Some of the data files that ChAS uses are generated by other Affymetrix software.

**Table 2.1 File types used by ChAS**

File Type	Created In	ChAS...
Sample file (ARR)	AGCC	Uses this information to associate sample attribute information with CEL and CYCHP or CNCHP files.
Intensity Data file (CEL)	AGCC	Analyzes the intensity data and generates a CYCHP file.  Note: A 64-bit system is required to analyze CytoScan™ HD intensity data.
Analysis Results (CYCHP)  CytoScan™ HD array: CYCHP contains copy number, LOH, and genotype call information  Cytogenetics Whole-Genome 2.7M arrays: CYCHP includes copy number, LOH, and mosaicism information	ChAS	Displays results in graphical and tabular formats.
CopyNumber/LOH (CNCHP)	GTC	Can open this file and display probe-level analysis data and generate segment data on-the-fly.
Region Information File (BED or AED)	ChAS or Text Editor	Allows users to display their own custom data and optionally use the information to define CytoRegions or an Overlap Map. ChAS can export data in BED format for use with the UCSC Browser and other programs which understand this format.
Tab-separated values (TSV, TXT)	ChAS	Exports data in this format for use in a spreadsheet program or other user-defined uses. This format is for export only. ChAS does not import TSV or TXT files.

## Region Information Files

The region information files in Browser Extensible Data (BED) and Affymetrix Extensible Data (AED) format provide lists of regions in the genome with position information and other annotations. To open a BED or AED file, click the  button or select **File > Open** on the menu bar. All BED or AED files that are opened during a session will reload when you start a new session with the same user profile.



**Note:** You can also use the reference annotations to provide region information.



**Note:** You can use the Export feature to export data in existing BED files to an AED file. See [Exporting Information in AED or BED Format](#) (page 174).



## Support files

The support files are necessary to use all of the features of ChAS.

- Library file sets with files for Copy Number/LOH/Mosaicism analysis (Analysis files)
- Reference Model files for single sample Copy Number/LOH/Mosaicism analysis of Intensity Data files (CEL) from CytoScan™ HD Arrays and Cytogenetics 2.7M Arrays (downloaded with other Analysis files or generated in ChAS)
- Reference Annotation files (Browser Annotation files named with the following format: <NetAffxGenomicAnnotations.Homo\_sapiens.hgxx.naxx.db>)

## Data Organization in ChAS

ChAS allows you to keep your CEL and CYCHP files in any folder on your computer, instead of being kept in a workspace. As long as you know where the files are, you can load them from anywhere and move them around at your convenience.



**Important:** Affymetrix recommends that you perform analysis operations with all files stored on a local disk drive.

## Basic Workflow for Cytogenetics Analysis



**Important:** The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.




**Note:** When referring to steps that apply to both CYCHP and CNCHP files, the term “CxCHP” is used.

ChAS can be used to:

- Perform probe-level analysis of CEL file data for CytoScan™ HD Arrays and Cytogenetics Whole-Genome 2.7M Arrays
- Display probe-level analysis data (CxCHP) from:
  - CytoScan™ HD Arrays (CYCHP)
  - Genome-Wide Human SNP Array 6.0 (CNCHP)
  - Cytogenetics Whole-Genome 2.7M Arrays (CYCHP)

There are some differences in the way the ChAS handles these three types of arrays and how it treats the data from these two types of files.

**The basic cytogenetic analysis workflow includes the following steps:**

1. [Array Processing Workflow, done in AGCC](#) (see below).
  2. [Probe Level Analysis of CEL file data](#) (page 30).
    - For the CytoScan™ HD Array or Cytogenetics Whole-Genome 2.7M Array, this analysis is performed in ChAS and produces CYCHP files (See [CN/LOH/Mosaicism Analysis](#) on page 36).
-  **Note:** A 64-bit system is required to analyze CytoScan™ HD array CEL files.
- For the Genome-Wide Human SNP Array 6.0, this analysis is performed in Genotyping Console (GTC) software and produces CNCHP files (see the [GTC User Guide](#)).
  3. [Load CNCHP and CYCHP data into ChAS for display](#) (page 31).
  4. [View the data using the ChAS display controls to find features of interest](#) (page 32).

## Array Processing Workflow

Array processing is performed in AGCC 3.2 or higher for the CytoScan™ HD Array, Genome-Wide Human SNP 6.0 Array, and Cytogenetics Whole-Genome 2.7M Array.

 **Note:** You need to have the appropriate AGCC library files installed to perform these analyses for the different array types.

The array processing includes the following steps:

1. Registering samples and arrays
2. Performing fluidics processing on arrays
3. Scanning arrays and generating intensity (CEL) file data.

The following file types are produced:

- Sample (ARR files) (AGCC only)
- DAT Files
- CEL Files

See the *AGCC User's Guide* for more information.

## Probe-Level Analysis

Cytogenetic analysis data is handled differently from genome-wide genotyping data in this step.

 **Note:** You need to have the appropriate ChAS library files installed to perform these analyses for different array types. A 64-bit system is required to analyze CytoScan™ HD CEL files.

- CytoScan™ HD Array or Cytogenetics Whole-Genome 2.7M Array Data: The probe level analysis on CEL file data is performed in ChAS and produces CYCHP files which contain the data shown in Table 2.2. See [CN/LOH/Mosaicism Analysis](#) (page 36).
- Genome-Wide Human SNP Array 6.0 Data: The probe level analysis on CEL file data is performed in GTC and produces the data shown in Table 2.2. CNCHP files. See the *GTC User Guide* for more information.

**Table 2.2 Analysis results data by array type**

	Analysis Results <sup>1</sup>		
	CytoScan™ HD Array <sup>2</sup>	Cytogenetics Whole-Genome 2.7M Array <sup>3</sup>	Genome-Wide Human SNP Array 6.0 <sup>4</sup>
<b>Graph Data for the individual CN and SNP probes</b>			
Copy Number State	yes	yes	yes
Log2 Ratio	yes	yes	yes
Weighted Log2 Ratio	yes	yes	no
LOH	yes	yes	yes
Allele Peaks	yes	yes	no

Allele Difference	no	no	yes
Genotype Calls	yes	no	no
Smooth Signal	yes	yes	yes
<b>Segment data</b>			
Gain and Loss segments based on Copy Number State data based on runs of aberrant Copy Number State data	yes	yes	yes
Mosaicism segments of non-integer Copy Number States between CN=1 and CN=3	no	yes	no
LOH (Loss of Heterozygosity) based on runs of SNPs where heterozygote calls are absent	yes	yes	yes

<sup>1</sup>CYCHP for CytoScan™ HD Array or Whole-Genome Cytogenetics 2.7M Array. CNCHP for Genome-Wide Human SNP 6.0 Array.

<sup>2</sup>For more details on CytoScan™ HD Array data, see Table 6.1 on page 99.

<sup>3</sup>For more details on Cytogenetics Whole-Genome 2.7M Array data, see Table 6.2 on page 100.

<sup>4</sup>For more details on Genome-Wide SNP Array 6.0 data, see Table 6.3 on page 101.

## Loading CYCHP or CNCHP Data

You perform the same steps for the different types of array data (CYCHP or CNCHP), but ChAS handles the two types of data differently.

### CytoScan™ HD Array or Cytogenetics Whole-Genome 2.7M Array (CYCHP files)

When loading CYCHP files into ChAS for viewing, the software:

1. Selects the run-length encoded segments in the CYCHP file to display as segments.
2. Applies any smoothing or joining that would alter the length and other properties of segments.
3. Displays the segments and graph data:
  - Segment Data
    - Copy Number Gain/Loss
    - Loss of Heterozygosity (LOH)
    - Mosaicism (for Cytogenetics Whole-Genome 2.7M CYCHP only)
  - Graph Data
    - Copy Number State
    - Log2Ratio
    - Weighted Log2Ratio
    - Smooth Signal
    - Loss of Heterozygosity (LOH)

- Allele Peaks
- Genotype calls (for CytoScan™ HD CYCHP only)

### Genome-Wide SNP Array 6.0 (CNCHP files)

When loading CNCHP files into ChAS for viewing, the software:


1. Performs segment generation by analyzing the CN and LOH graph data in the CNCHP file.
2. Applies any smoothing or joining that would alter the length and other properties of Copy Number segments.  
In GTC software, these steps were performed in the Segment Reporting Tool.
3. Displays the segments and graph data:
  - Segment data
    - Copy Number Gain/Loss
    - Loss of Heterozygosity (LOH)
  - Graph Data
    - Copy Number State
    - Log2 Ratio
    - Allele Difference
    - SmoothSignal
    - Loss of Heterozygosity (LOH)

### Viewing Data

ChAS provides several options for viewing and studying the loaded CYCHP or CNCHP data:

- Graphic Displays  
See [Displaying Data in Graphic Views](#) (page 89).
- Tables  
See [Displaying Data in Table Views](#) (page 177).

After the data is loaded, you can:

- Filter the segments by Segment Parameters to hide segments that do not meet your requirements for significance.  
See [Filtering Segments](#) (page 130)
- Select a region information file for use as a CytoRegion file and:
  - Perform differential filtering for segments in CytoRegions and in the rest of the genome.
  - Display only segments that appear in CytoRegions using Restricted Mode.
 See [Using CytoRegions](#) (page 134).
- Select a region information file for use as an Overlap Map and use the Overlap filter to identify or conceal segments that appear in the Overlap Map regions.  
See [Using the Overlap Map](#) (page 144).
- Add selected features of the genome to new or existing Region (AED) files, and edit annotation data on existing annotations. (To open a BED or AED file, click the  button or select **File > Open** on the menu bar.)  
See [Creating and Editing AED Files](#) (page 152).

- Prepare reports on your findings by exporting graphics and table data in PDF and other formats.  
See [Reporting Results](#) (page 206).
- Save setups of ChAS for different tasks in user profiles and named settings.  
See [User Profiles and Named Settings](#) (page 225).

## Working with ChAS

This section covers:

- [Accessing Functions in ChAS](#) (page 33)
- [Changing the Size of Panes in ChAS](#) (page 33)
- [Opening Panes in Separate Windows](#) (page 34)

### Accessing Functions in ChAS



Commands in ChAS can be accessed in multiple ways. You can access commands from:

- Main menus
- Toolbar
- Right-click options in the:
  - Files list
  - Data Types list
  - Karyoview
  - Selected Chromosome View
  - Detail View
  - Table headers

The alternatives for a particular function are described in the step-by-step instructions for using the function.

### Changing the Size of Panes in ChAS



You can change the size of the panes in the ChAS window:

- Click and drag the dividers between panes
- Click the arrows in the dividers ( or ) to hide or maximize an entire pane



**Figure 2.5** Resize or show/hide panes

### Opening Panes in Separate Windows

You can display a pane in a separate window by clicking the  icon on the tab. To close the window and return the information to the tab panel, click the  icon in the window.

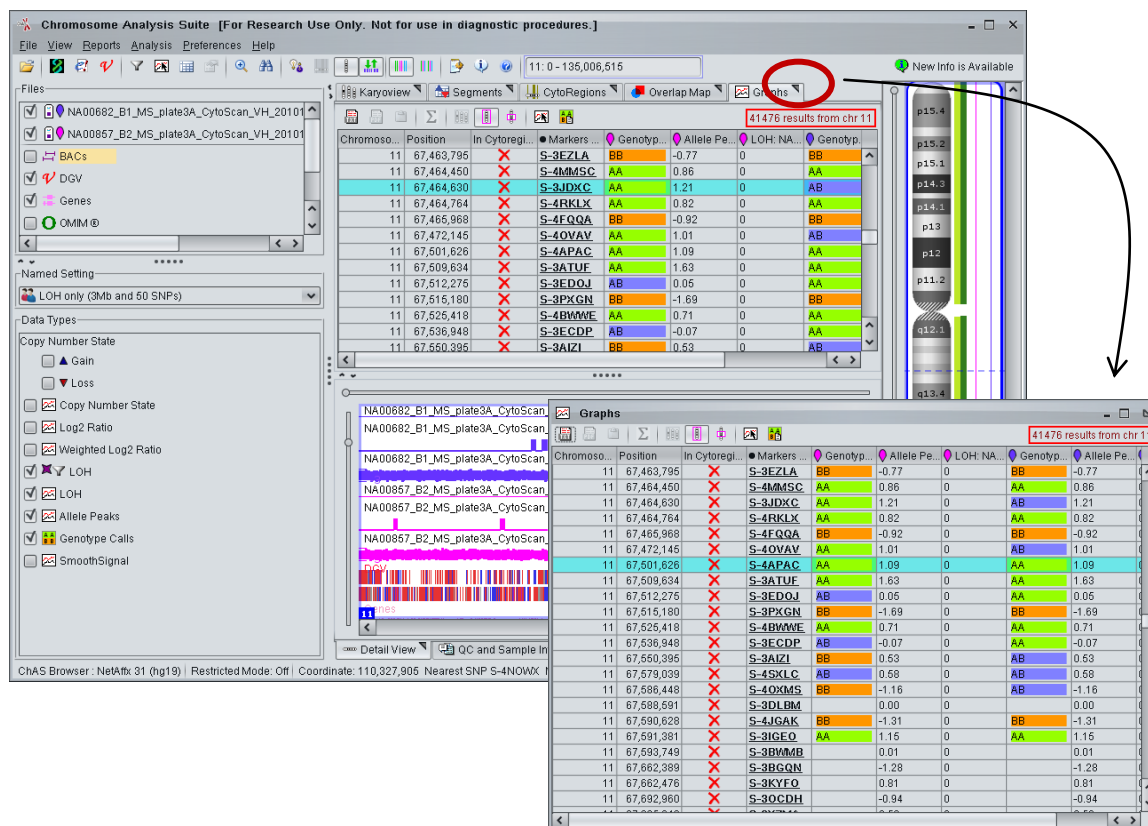


Figure 2.6 Graphs pane opened in new window

## Chapter 3: CN/LOH/Mosaicism Analysis

ChAS analyzes the intensity data (CEL file) from CytoScan™ HD Arrays or Cytogenetics Whole-Genome 2.7M Arrays. The software performs a single sample analysis which compares the data in a CEL file to a previously created reference file, using analysis parameters specified in the .chasparam file. The analysis generates a CYCHP data file that you load and view in ChAS. The analysis detects segments that exhibit:

- Copy number state gain or loss – Regions of integer copy number gain or integer copy number loss
- Mosaicism – Regions of non-integer copy number gain or loss (CN states between 1 and 3)



**Note: Mosaicism analysis is currently only available for Cytogenetics Whole-Genome 2.7M CEL files. However, CxCHP files for the other array types contain the SmoothSignal data type which displays non-integer copy number changes.**

- Loss of Heterozygosity (LOH) – Regions where the preponderance of SNPs do not display heterozygosity

For more details on loading and viewing CYCHP data, see [Loading Data](#) (page 66).



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

Affymetrix provides .chasparam and reference files. You can also create your own reference file or set custom analysis parameter values.



**Note: You can also load Genome-Wide Human SNP Array 6.0 CNCHP files into ChAS to display and detect Copy Number and Loss of Heterozygosity segments. See [Loading Data](#) (page 66).**



**Important: Affymetrix recommends that you perform analysis operations with all files stored locally.**

This chapter explains how to:

- [Perform single sample analysis](#) (page 36)
- [Create a reference file](#) (page 48)
- [Use the Analysis Manager](#) (page 53)
- [Customize and manage analysis parameters](#) (page 57)

### Single Sample Analysis

Single Sample Analysis compares the values in one or more user-selected CEL files with the values in a reference file that is created from a set of sample files. You can use the factory-loaded reference file provided by Affymetrix or create your own (for more details, see [Creating a Reference File](#) on page 48).

This section covers:

- [ChAS and NetAffx Analysis File Compatibility](#) (below)
- [Performing Single Sample Analysis](#) (page 42)

#### ChAS and NetAffx Analysis File Compatibility

Table 3.1 lists the compatibility between ChAS and NetAffx Analysis file versions for the CytoScan™ HD Array and the Cytogenetics Whole-Genome 2.7M Array.



**Note: ChAS automatically prevents you from selecting an incompatible NetAffx Analysis file version for analysis or when viewing analysis results.**



**Table 3.1 Compatibility between ChAS version and NetAffx analysis file set version**

	ChAS Software Version			
	ChAS 1.2	ChAS 1.1	ChAS 1.0.1	ChAS 1.0
<b>CytoScan™ HD Array NetAffx Analysis File Set Version</b>				
NA32(hg19)	yes	no	no	no
<b>Cytogenetics Whole-Genome 2.7M Array NetAffx Analysis File Set Version</b>				
NA32(hg19)	yes	yes	no	no
NA31(hg19)	yes	yes	no	no
NA30.2(hg18)	yes	yes	no	no
NA30.1(hg18)	no	no	yes	no
NA30(hg18)	no	no	yes	no
NA29(hg18)	no	no	yes	no
NA28(hg18)	no	no	no	yes

## Introduction to Single Sample Analysis

Single Sample Analysis requires:

- ChAS analysis files for the array  
See Installing Analysis Files (page 18).
- A previously created reference model file

You can use the reference model file provided by Affymetrix or create your own using your own CEL file data and the Reference File creation function. The Reference Model file in the CytoScan HD™ Array set of NA32 Analysis includes 380 microarrays which were run as part of a larger set of microarrays by nine operators processing ~48 unique samples in two rounds each, with randomization of the placement of sample DNAs across the PCR plates and randomization of the reagents and instruments used. The source DNA includes:

- 284 HapMap samples including at least one replicate of each of 270 HapMap samples: 90 from each of the Yoruban, Asian, and Caucasian ethnic groups, from cell-line derived DNAs from the Coriell Institute of Medical Research
  - 96 DNA samples from blood of phenotypically healthy male and female individuals obtained from BioServe Biotechnologies
- CEL file data

During the analysis, ChAS generates CYCHP files with:

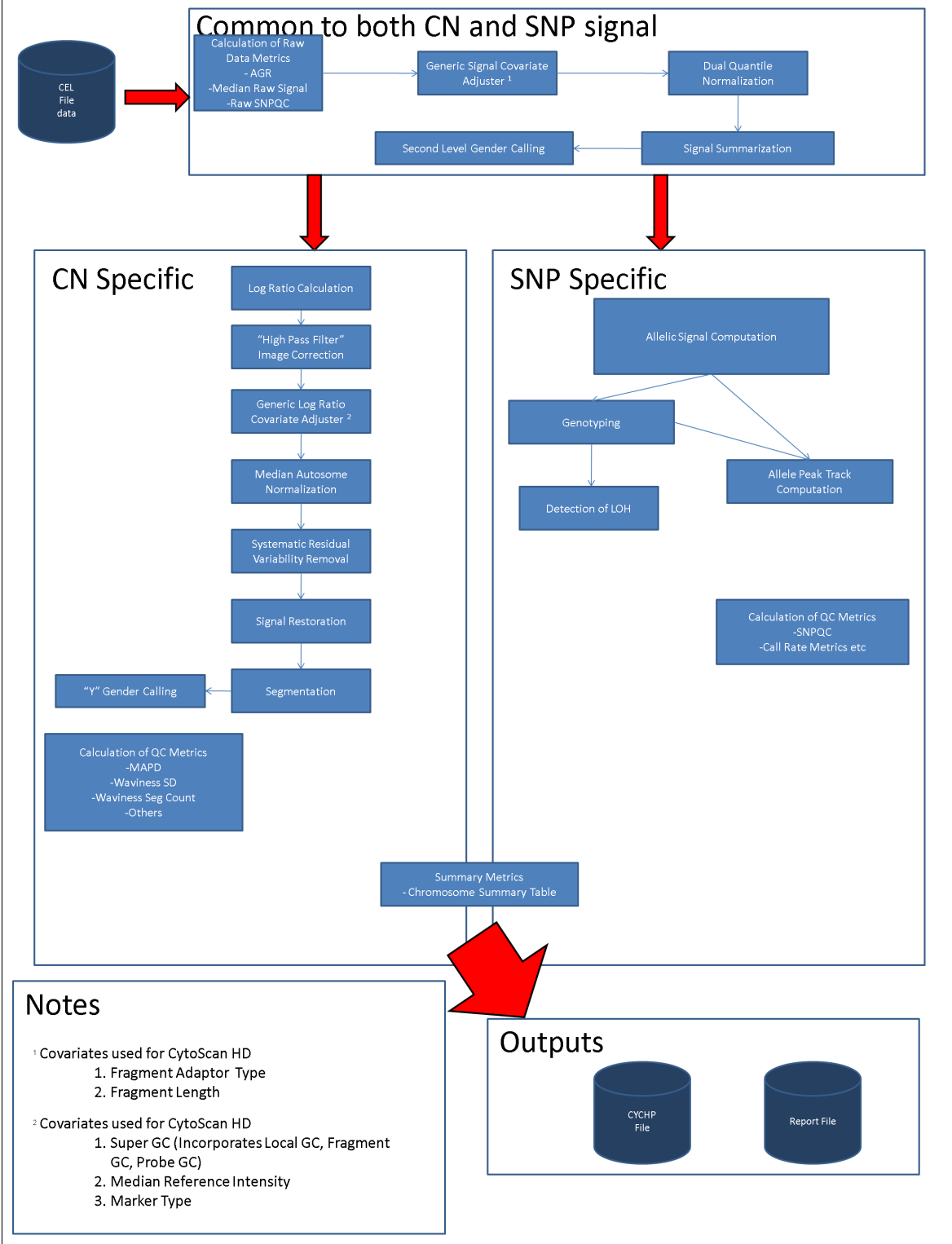
- Graph Data

- Copy Number State
- Log2 Ratio
- Weighted Log2 Ratio
- LOH
- Allele Peaks
- Smooth Signal
- Segment Data
  - Copy Number Gain/Loss
  - Mosaicism (currently available for Cytogenetics Whole-Genome 2.7M Array)
  - Loss of Heterozygosity (LOH)

The CYCHP files can be loaded into ChAS for viewing and study.

Figure 3.1 shows an overview of single sample analysis for the CytoScan™ HD Array.

## Rough Sketch of Analysis Pipeline (for Single Sample Analysis)



**Figure 3.1 CytoScan™ HD single sample analysis**

## **Copy Number Segments on the X and Y Chromosomes**

The expected copy number state on the X chromosome in normal males is not constant over its entire length. This is due to the structure of the sex chromosomes, and the fact that they share extensive homology with each other only in the Pseudo Autosomal Regions (PARs) that they each have at either end. PAR1 is at the top of the p-arm and PAR2 at the bottom of the q-arm.

Markers occurring in the PAR regions are mapped exclusively to the X Chromosome. Therefore, in normal males the PAR regions of the X are expected to be CN=2 (probes on the X and Y contribute to the signal), while the rest of the Chr X is expected CN=1 for normal males. As a result, we treat the two X PARs in males as independent units (CN=2 expected) from the rest of the X chromosome (CN=1 in males) when generating Copy Number Segments.

Aberrant segments that cross PAR/non-PAR boundaries may be normalized into one segment if they have equivalent type (Gain or Loss) and CN State. During this normalization process, ChAS will not combine an aberrant (Gain or Loss) segment with a normal segment across PAR/non-PAR boundaries, even if they have the same CN State. If smoothing is subsequently applied, aberrant segments with different copy number state may be combined. If joining is subsequently applied, aberrant segments separated by a non-aberrant segment may be combined.

Because only Y-specific probes are mapped to the Y chromosome, the expected state of the entire Y chromosome is 1 for males and is 0 for females.

## **LOH Segments on the X and Y Chromosomes**

### **CytoScan™ HD Array**

For normal XY male samples, the X chromosome will have single-copy based LOH calls (CN = 1). Male samples with more than one X chromosome (for example, XXY) may have LOH calls on the X chromosome, depending on the constitution of the X chromosomes' origins.

Table 3.2, Table 3.3, and Table 3.4 briefly describe how the array-specific algorithms call LOH segments for the X or Y chromosome.

**Table 3.2 Expected LOH calls on the X and Y chromosomes for the CytoScan™ HD Array**

	X Chromosome	Y Chromosome
Normal male sample (XY)	LOH calls that are single copy-based LOH call (CN = 1).	No LOH calls are made for the Y chromosome. Genotype calling is not performed on the Y chromosome.
Male sample with multiple X chromosomes (for example, XXY)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	
Normal female sample (XX)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	
Female sample with a single X chromosome (X0)	LOH calls on X regions which have only a single copy. Heterozygous SNP genotypes are possible, but are due to the low inherent Heterozygote call error rate noise, not the true presence of two alleles.	

**Table 3.3 Expected LOH calls on the X and Y chromosomes for the Genome-Wide Human SNP Array 6.0**

	X Chromosome	Y Chromosome
Normal male sample (XY)	LOH calls on the non-PAR region of the X chromosome resulting from "forced" homozygote-only calls due to the presence of the Y chromosome.  Heterozygous calls are ignored on the X chromosome in males.	LOH calls that are due to single copy genotyping calls (CN = 1).
Male sample with multiple X chromosomes (for example, XXY)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.  SNP genotypes are not constrained to homozygous calls. Heterozygous calls are ignored on the X chromosome in males.	
Normal female sample (XX)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	LOH analysis is not performed on the Y chromosome since it is assumed that there not substantial Y chromosomal material.
Female sample with a single X chromosome (X0)	LOH calls on X regions with only a single copy. Heterozygous SNP genotypes are possible, but are due to the low inherent Heterozygote call error rate noise, not the true presence of two alleles.	

**Table 3.4 Expected LOH calls on the X and Y chromosomes for the Cytogenetics Whole-Genome 2.7M Array**

	X Chromosome	Y Chromosome
Normal male sample (XY)	LOH calls that are single copy-based LOH call (CN = 1).	LOH Analysis is not performed because there are no SNP probes mapped to the Y chromosome for this array
Male sample with multiple X chromosomes (for example, XXY)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	
Normal female sample (XX)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	LOH Analysis is not performed because there are no SNP probes mapped to the Y chromosome for this array
Female sample with a single X chromosome (X0)	LOH calls on X regions which have only a single copy.	

## Performing Single Sample Analysis

1. From the Analysis menu, select **Single Sample Analysis...**

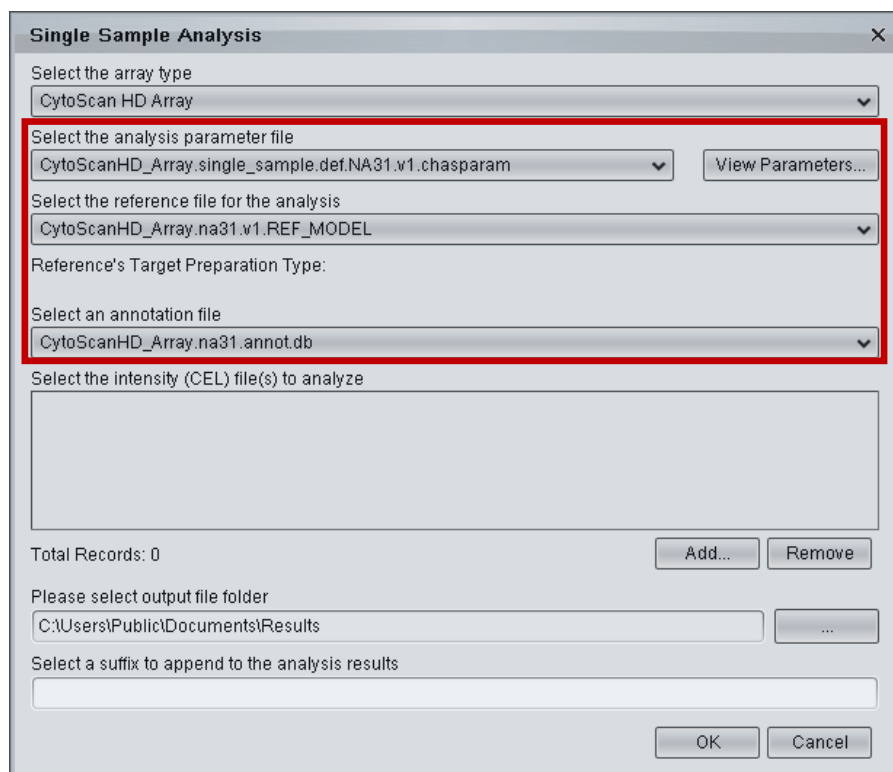
The Single Sample Analysis dialog box opens.

**Figure 3.2 Single Sample Analysis**

2. Select CytoScan™ HD Array from the Array Type drop-down list.

 **Note:** The “Select the array type” drop-down list includes only the array types for which library (analysis) files have been downloaded from NetAffx or copied from the Library package provided with the installation.

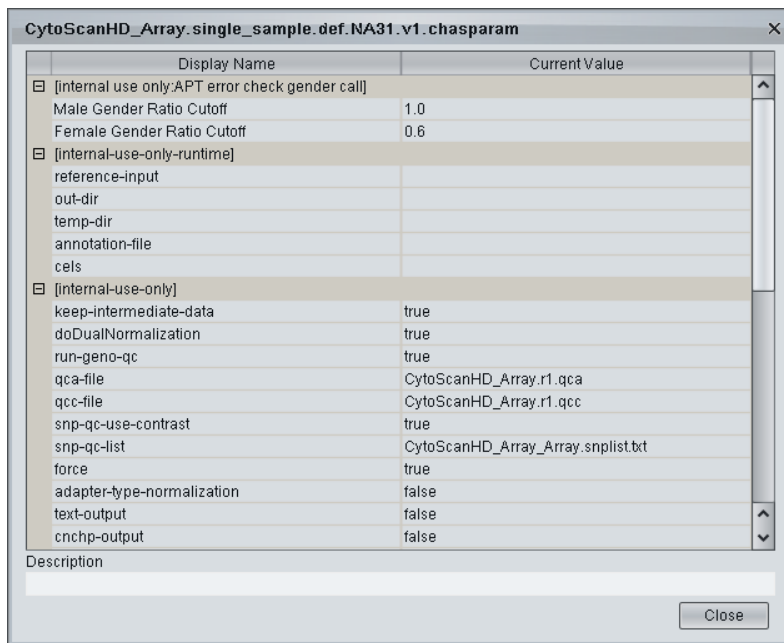
The analysis parameter file, reference file, and annotation file drop-down lists are automatically filtered after you choose an array type.



**Figure 3.3** Drop-down lists that are automatically filtered after you select the array type

3. Select an analysis parameter file from the drop-down list.

You can click **View Parameters** to display all of the parameters in the parameter file.



**Figure 3.4 Analysis parameters for CytoScan™ HD array**

Click a parameter to display a short description in the Description box at the bottom of the window.

See [Analysis Parameters](#) (page 233) for more information about the parameters that you can adjust.



**Important:** The analysis algorithm uses different terms to identify the parameters. You may see these terms in various error messages.

4. Select a reference file from the drop-down list. For more details on creating a reference file, see page 48.
5. Select an annotation file from the drop-down list.

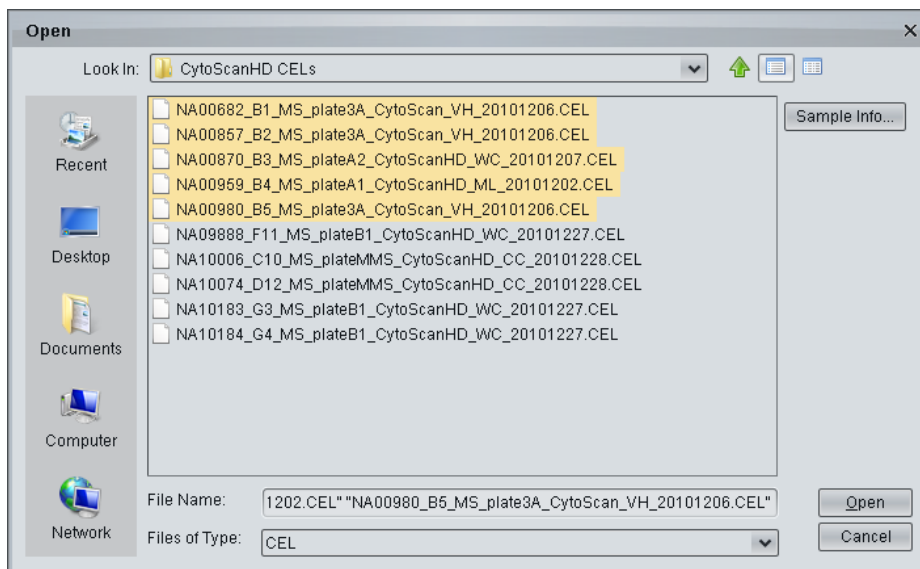


**Note:** The analysis must use the same annotation file that was used to create the reference model.

6. Select the CEL files for analysis.
  - a. Click **Add**.

The Open dialog box appears.





**Figure 3.5 Open dialog box**

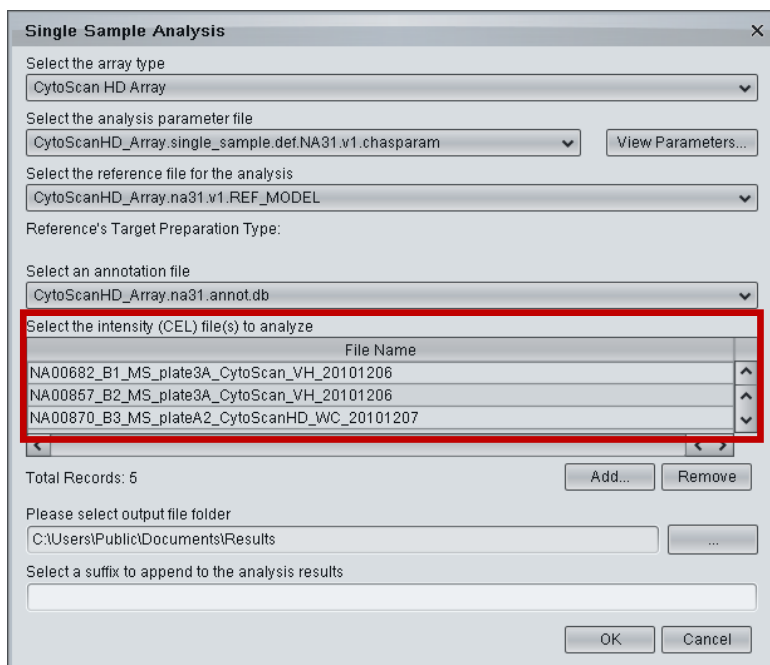
- b. Navigate to the folder with the CEL file(s) that you want to analyze.
- c. Select the CEL file(s) to be analyzed.



**Note:** You can load several CEL files at a time for a Single Sample Analysis. “Single Sample Analysis” refers to comparing one sample at a time against a reference model.

- d. Click **Open**.

The files appear in the intensity file list.



**Figure 3.6 CEL files for the analysis**

To remove a file from the list, select a file in the box and click **Remove**.

The total number of selected files is displayed below the box.

7. Select a folder for the output files.

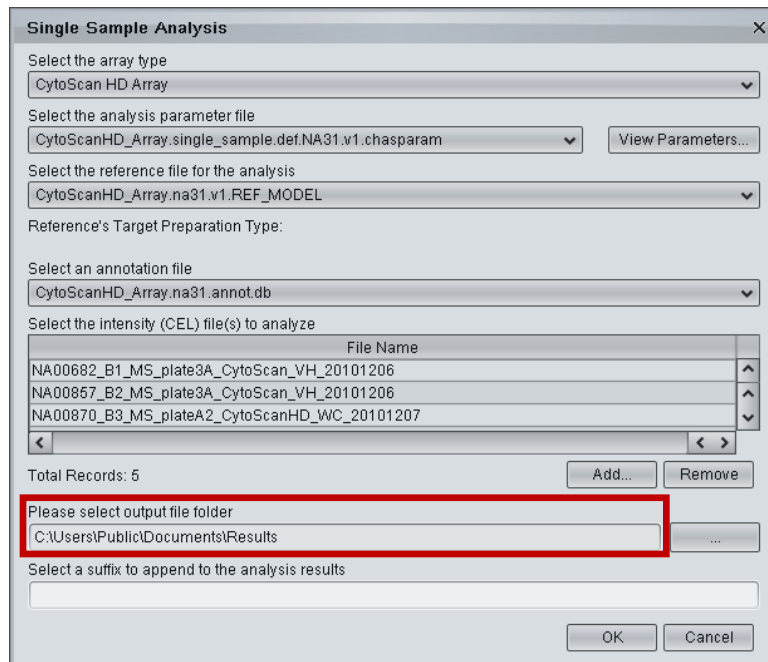


Figure 3.7 Output folder location

Enter a path to the output folder. Alternatively,

- a. Click the **Browse** button .

The Select Output Folder dialog box opens.

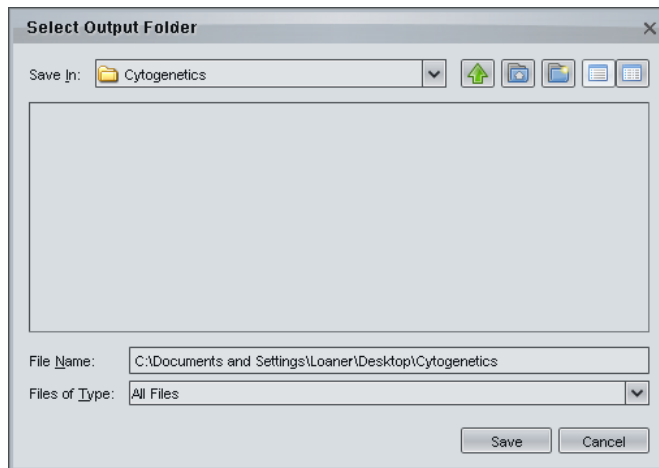



Figure 3.8 Select Output Folder dialog box

- b. Navigate to the folder where you want to save the resulting CYCHP files...

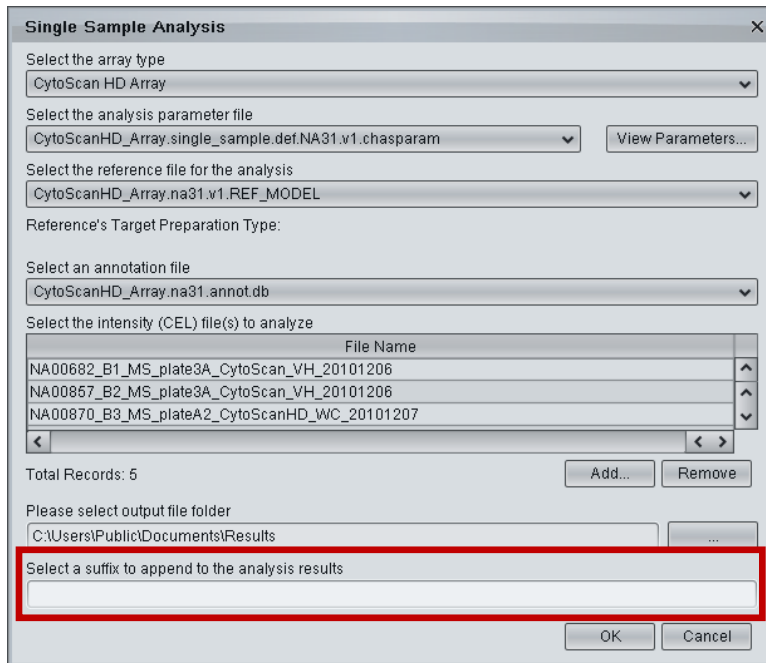
You can also create a new folder (click the  button).

- c. Click **Save**.

8. Enter a suffix to be added to the output file names (optional).

Adding a suffix to the CYCHP file name may help track results, for example, if you reanalyze CEL files using different parameters.

 **Important:** If you previously analyzed the selected CEL files and are saving the new CYCHP files to the same location as the previous files, you need to add a unique suffix to the new files, otherwise the original CYCHP files will be overwritten.



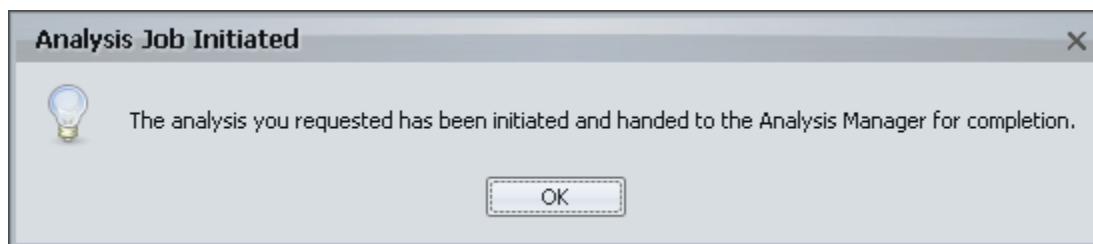
The dialog box titled "Single Sample Analysis" contains the following fields and controls:

- Select the array type:** CytoScan HD Array (dropdown)
- Select the analysis parameter file:** CytoScanHD\_Array.single\_sample.def.NA31.v1.chasparam (dropdown) with a "View Parameters..." button.
- Select the reference file for the analysis:** CytoScanHD\_Array.na31.v1.REF\_MODEL (dropdown)
- Reference's Target Preparation Type:** (label)
- Select an annotation file:** CytoScanHD\_Array.na31.annot.db (dropdown)
- Select the intensity (CEL) file(s) to analyze:** A list box containing three file names: NA00682\_B1\_MS\_plate3A\_CytoScan\_VH\_20101206, NA00857\_B2\_MS\_plate3A\_CytoScan\_VH\_20101206, and NA00870\_B3\_MS\_plateA2\_CytoScanHD\_WC\_20101207. Below the list are "Add..." and "Remove" buttons.
- Total Records:** 5
- Please select output file folder:** C:\Users\Public\Documents\Results (text field) with a browse button (...).
- Select a suffix to append to the analysis results:** An empty text field, highlighted with a red rectangle.
- Buttons:** OK and Cancel at the bottom right.

**Figure 3.9 Suffix field**

9. Click **OK** to begin the analysis.

If there are no errors, the analysis proceeds and the following notice appears.



**Figure 3.10 Analysis job message**

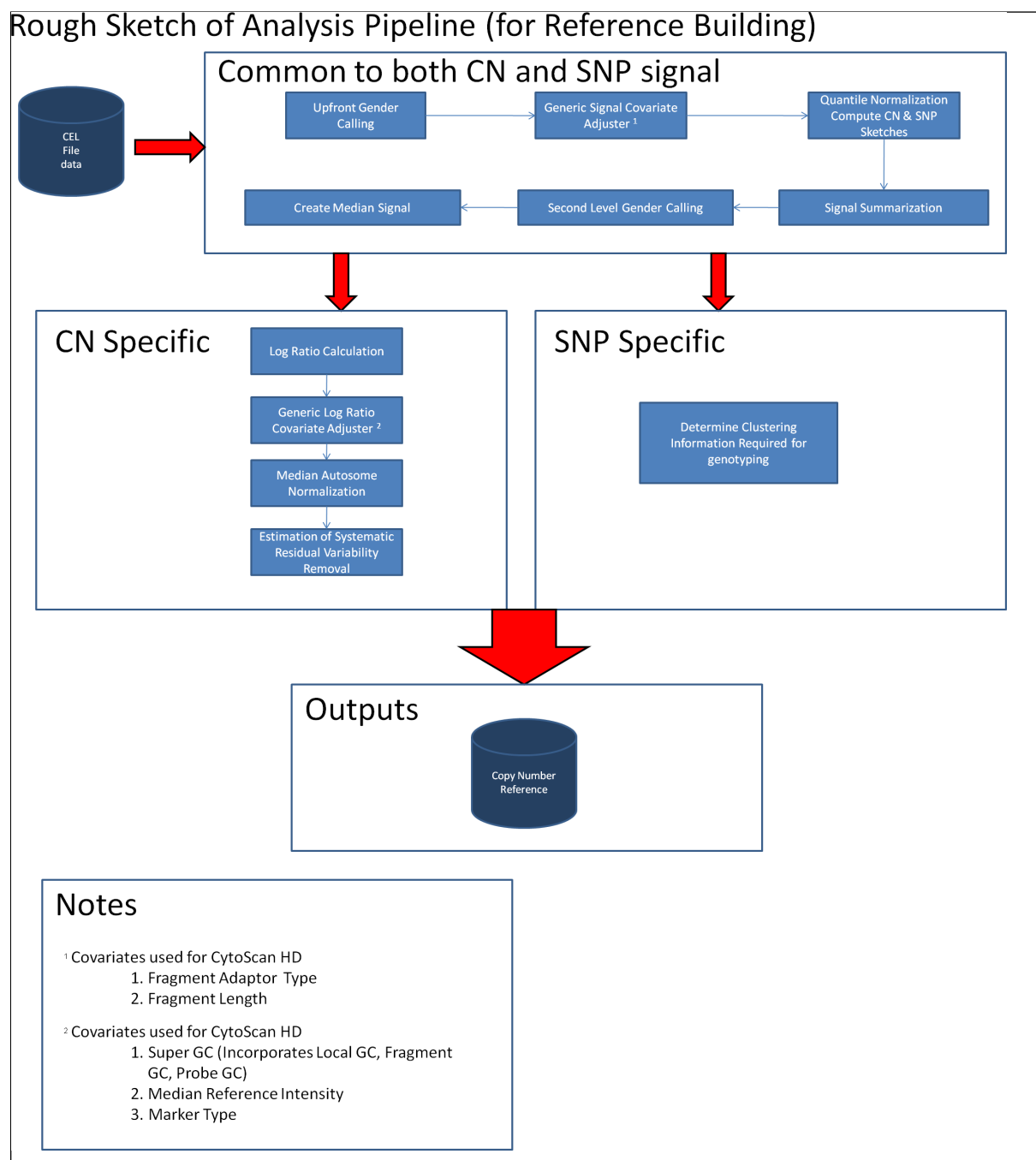
The software automatically *checks the memory available for the analysis using the Analysis Manager* (page 53).

You can track the progress of the analysis in the Analysis Manager. For more details, see *Tracking the Progress of an Analysis*, page 54.

## Creating a Reference File

This section explains how to create a reference file which is required to perform single sample analysis in ChAS. The software analyzes a sample file by comparing it to a reference file. You can use the reference file provided with ChAS, or you can create a reference file using your own sample data.

Figure 3.11 shows an overview of the analyses involved in creating a reference file for the CytoScan™ HD Array.



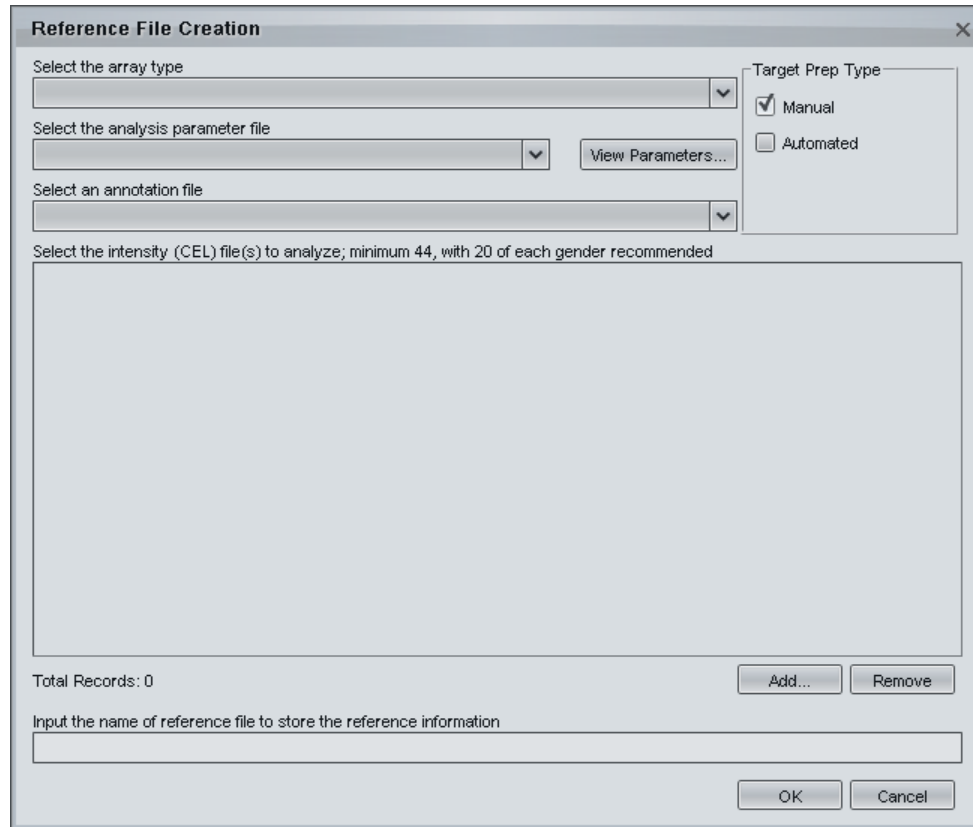
**Figure 3.11 Overview of creating a reference file**

To create a reference file:

 **Note:** When creating a reference file, it is recommended that you use a minimum of 44 CEL files, with at least 20 files from male and 20 from female samples.

1. Select **Analysis > Create Reference** on the menu bar.

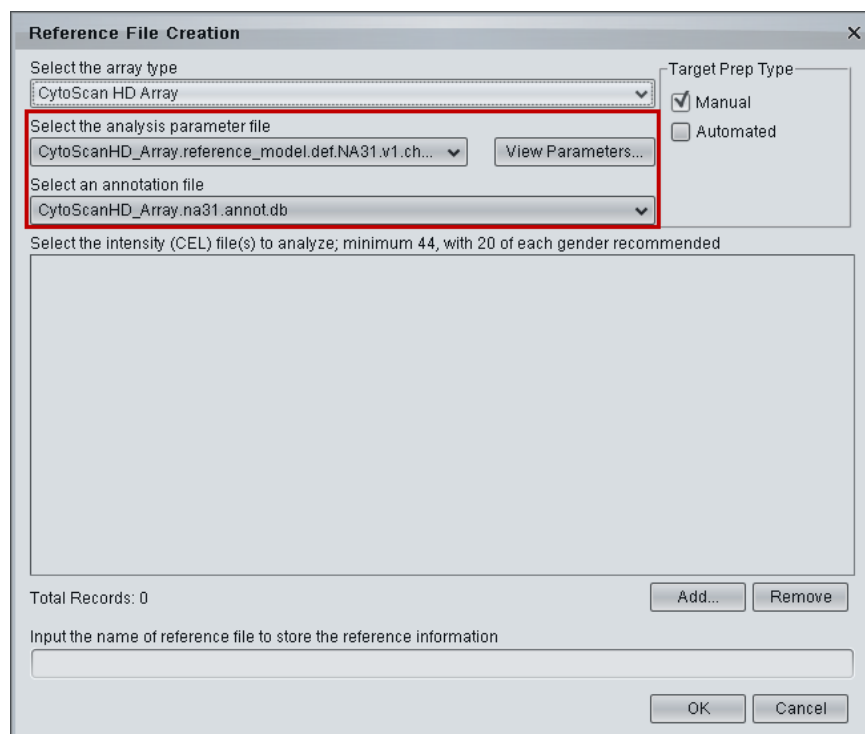
The Reference File Creation dialog box opens.



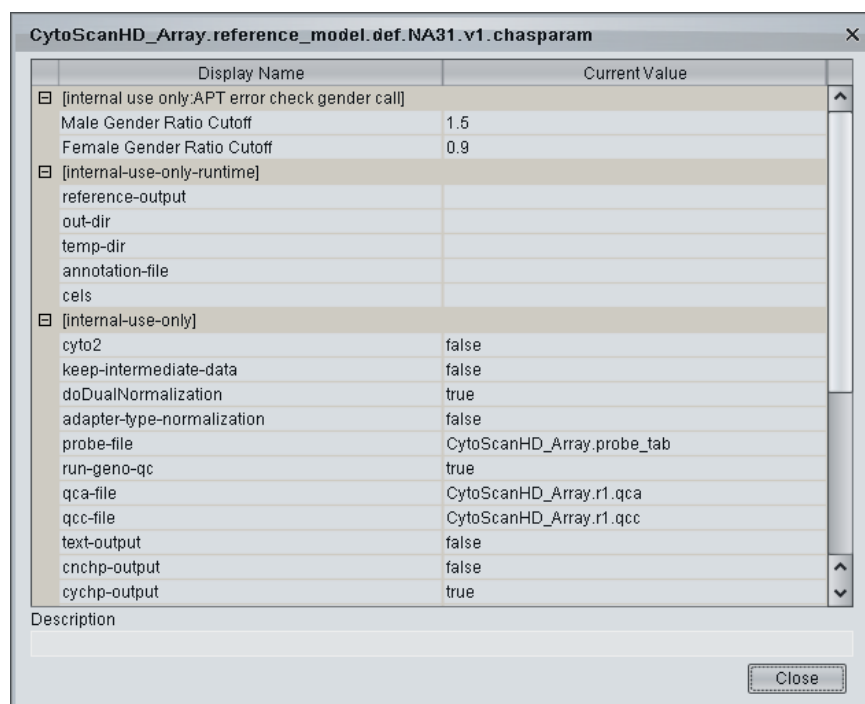
**Figure 3.12 Reference File Creation dialog box**

2. Select an array type from the array type drop-down list.

The analysis parameter file and annotation files are automatically selected based on the selected array type. You can click **View Parameters** to display all the parameters in the parameter file.

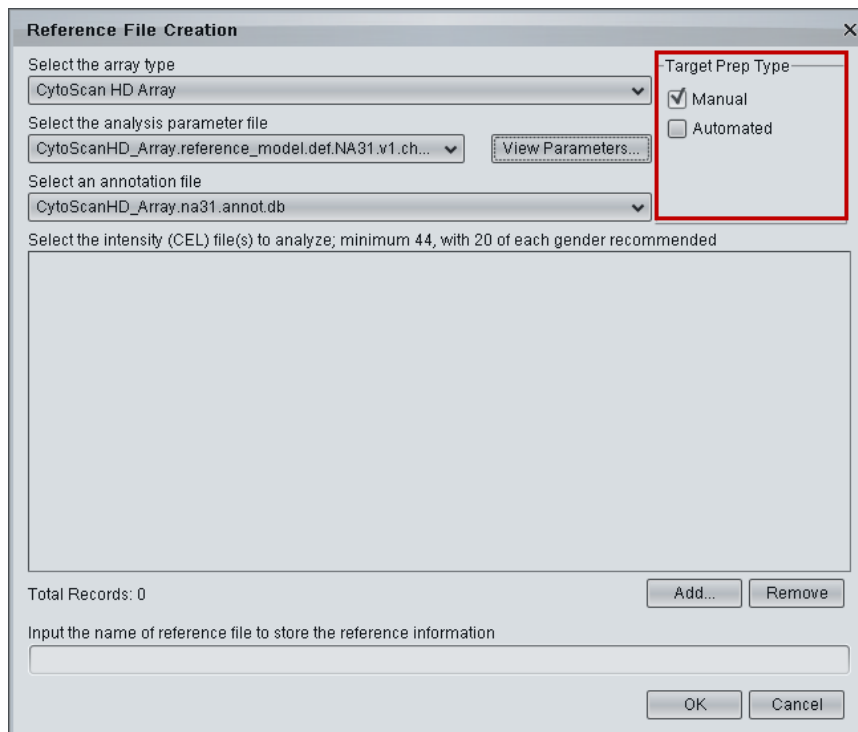


**Figure 3.13 Drop-down lists automatically filtered by array type**



**Figure 3.14 Reference file parameters**

- Put a check mark next to the target preparation method(s) that was used to generate the CEL file data that will provide the input to the Reference Creation: **Manual** or **Automated** (using a robot).

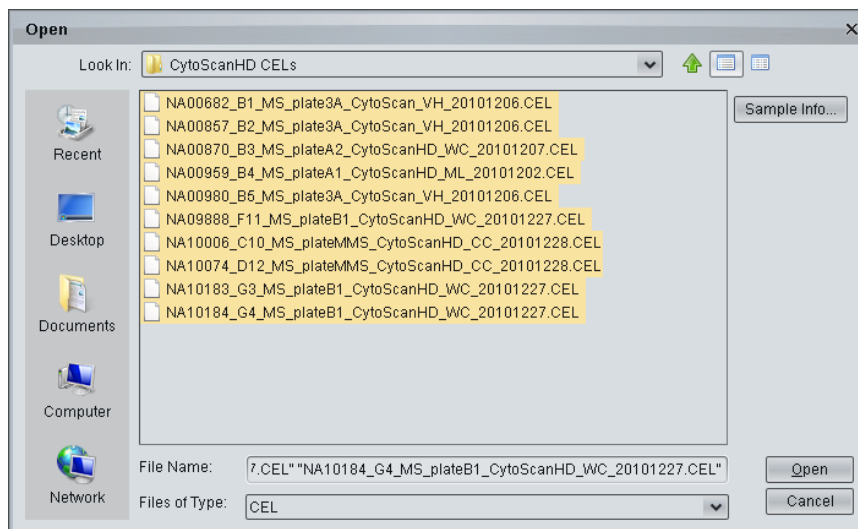


**Figure 3.15 Select the type of target preparation**

4. To select the CEL files for analysis:

a. Click **Add**.


The Open dialog box appears.



**Figure 3.16 Open dialog box**

b. Navigate to the folder with the CEL files.

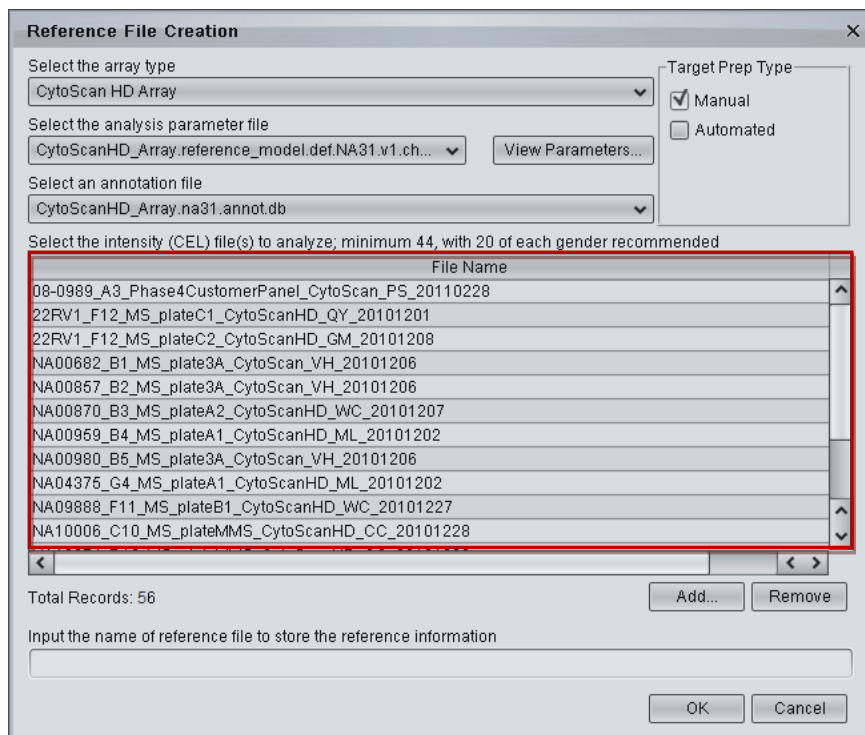
c. Select the CEL files that will be used to create the reference file.

 **Note:** Affymetrix recommends that you use a minimum of 44 CEL files, with at least 20 of each gender to create a reference file.

d. Click **Open**.

The files appear in the CEL file list. To remove a file from the list, select the file and click **Remove**.

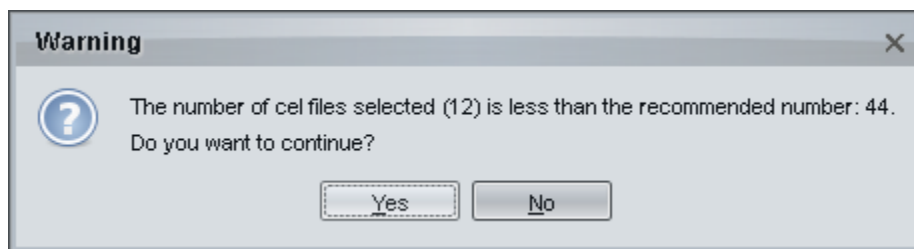
The total number of selected files is displayed below the box.



**Figure 3.17 CEL files selected for creating a reference file**

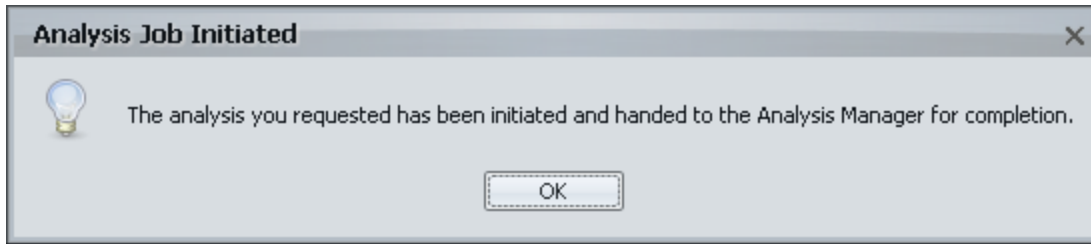
5. Enter a name for the reference file.
6. Click **OK** to begin the analysis.

A warning appears if you selected less than 44 CEL files for creating the reference file. Otherwise, the analysis job begins.



**Figure 3.18 Warning message**





**Figure 3.19 Analysis job started**

7. Click **OK**.

The software automatically [checks the memory available for the analysis using the Analysis Manager](#) (page 53).

You can track the progress of the analysis in the Analysis Manager. For more details see page 54.

## Using the Analysis Manager

The Analysis Manager tracks ongoing analysis tasks for ChAS and alerts you when the computer does not have sufficient memory to perform an analysis. It also delivers the results of analyses and restarts the Browser if it was shut down to free up memory for the analysis.

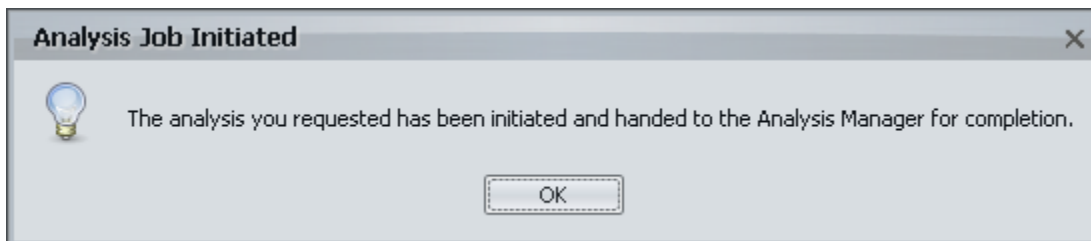
This section covers:

- [Starting an Analysis with the Analysis Manager](#), below
- [Tracking the Progress of an Analysis](#) (page 54)
- [Analysis Results Dialog Box](#) (page 56 )

### Starting an Analysis with the Analysis Manager

1. Click **OK** in the Analysis dialog box.

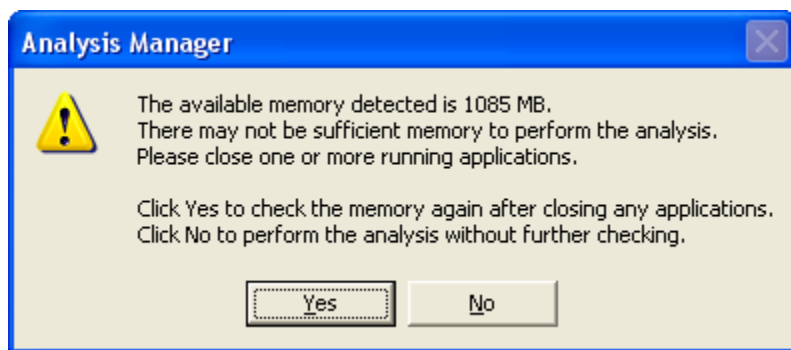
The following notice appears.



**Figure 3.20 Analysis job message**

2. Click **OK** in the Analysis Job Initiated dialog box.

If the Analysis Manager does not detect at least 1200 MB of free memory, the following message appears.



**Figure 3.21 Prompt to close applications**



This message prompts you to close the ChAS window or other applications to provide more memory for the analysis.

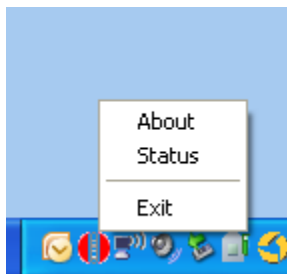
3. Close the ChAS window and click **Yes** to see if enough memory is available.

Click **No** to proceed with the analysis without closing any applications.

You can track the progress in the Analysis Manager window (see below).

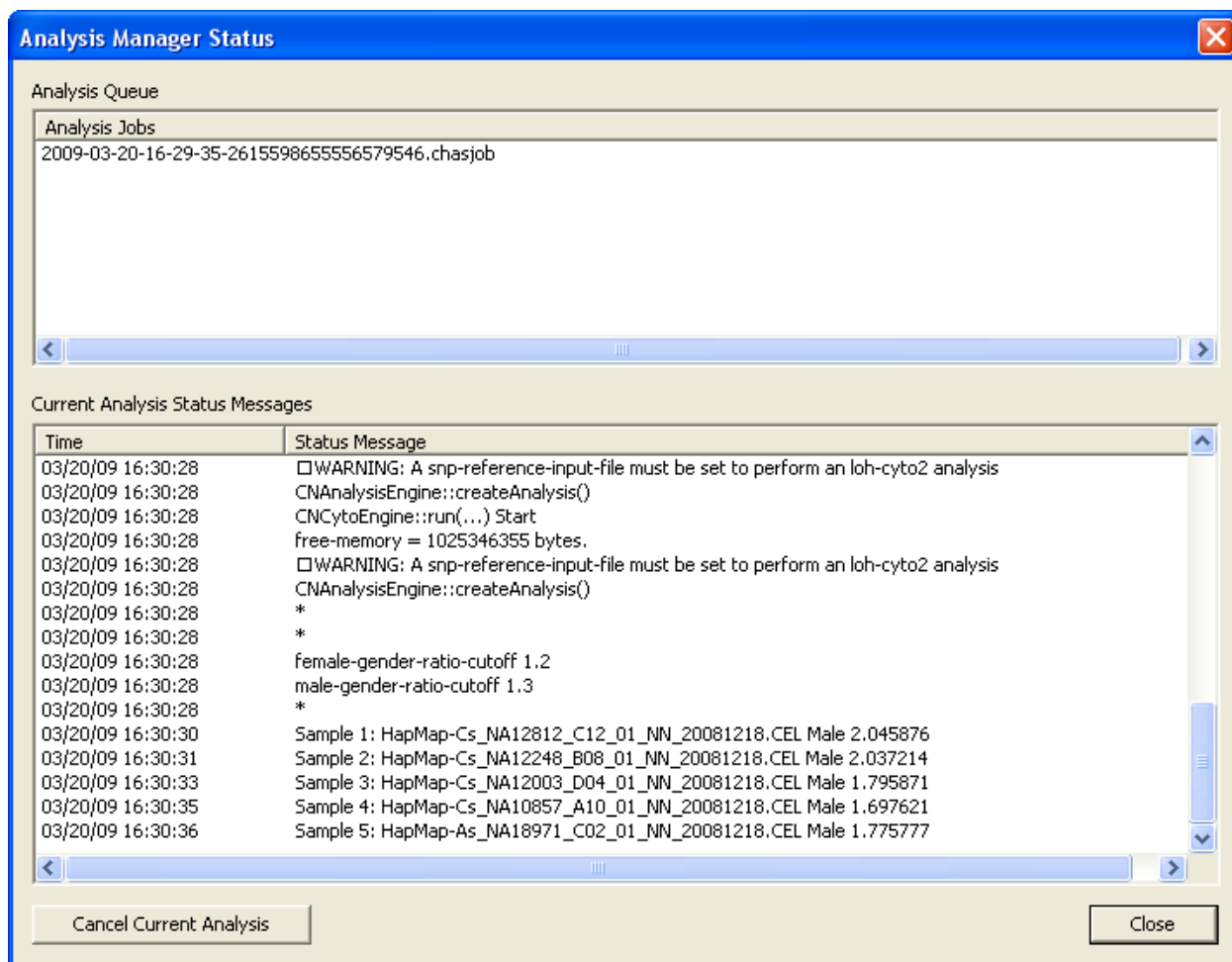
### **Tracking the Progress of an Analysis**

1. To open the Analysis Manager, double-click the Analysis Manager icon  at the far right of the Windows task bar. Alternatively, right-click the  icon and select Status on the shortcut menu that appears..



**Figure 3.22 Analysis Manager in the task tray**

The Analysis Manager Status window appears.



**Figure 3.23 Analysis Manager with analysis running**

The Analysis Manager Status window has the following components:

- Analysis Queue – Lists jobs that are in progress or waiting for analysis
- Current Analysis Status Messages – Provides status messages, with:
  - Time – Time of the message being performed.
  - Status Message – Notice about particular step in the analysis
- Cancel Current Analysis – Aborts the analysis that is currently running
- Close – Closes the Analysis Manager Window



**Important: Closing the Analysis Manager Status window does not stop the current analysis in progress or shut down the Analysis Manager.**

**To cancel the current analysis:**

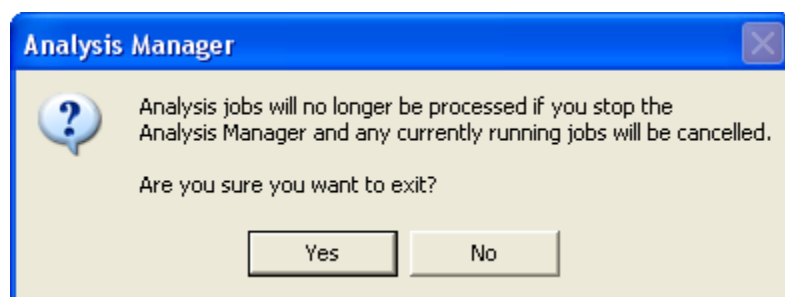
- Click **Cancel Current Analysis**.

If the analysis is already running, it may take a few seconds before the analysis job is cancelled.

**To exit the Analysis Manager and cancel the current analysis:**

1. Right-click the Analysis Manager icon  and select **Exit**.

The following notice appears:




**Figure 3.24 Analysis Manager message**

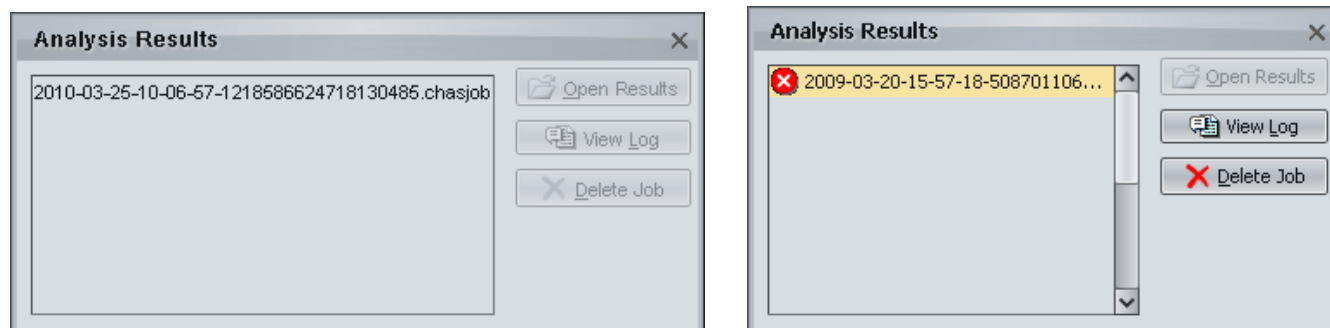
2. Click **Yes** to exit.

Any ongoing analysis stops.

The Analysis Results dialog box opens to display the stopped runs.

### **Analysis Results Dialog Box**

The Analysis Results dialog box displays a list of the finished jobs. Failed or canceled jobs are marked with the  symbol.



**Figure 3.25 Completed job (left), failed or canceled job (right)**

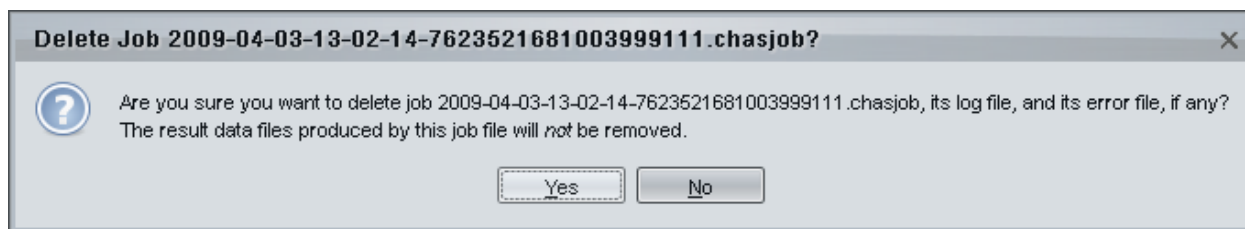
The buttons allow you to perform the following tasks:

- **Open Results** — Opens the folder with results files in the Open dialog box. See [Loading Files](#) (page 67)
- **View Log** — Displays log file with analysis status messages.
- **Delete Job** — Deletes the selected job from the Analysis Results list.

**To delete a job from the list:**

1. Select the job in the list.
2. Click **Delete**.

The following notice appears.



**Figure 3.26 Delete Job message**

3. Click **Yes** to delete the job.

 **Note:** This deletes the job, log file, and error file only. It does not delete the results data files.

## Using the Analysis Parameters Functions

The analysis parameters functions enable you to:

- Create a new analysis parameter file
- Export analysis parameter files
- Import analysis parameter files.

The configuration file parameters are described in [Analysis Parameters](#) (page 233).

### Creating a New Analysis Parameters File (Customizing)

You can create a new analysis parameters file with different parameters.

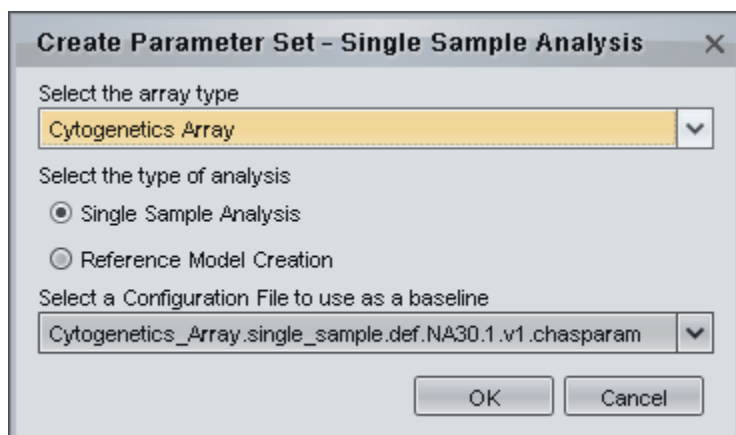
 **Note:** The CytoScan™ HD Array does not have user-modifiable analysis parameters.

1. Select a parameter file to use as a starting point.
2. Edit the parameter file.
3. Save the file with a new name

**To create a new parameter file (customizing):**

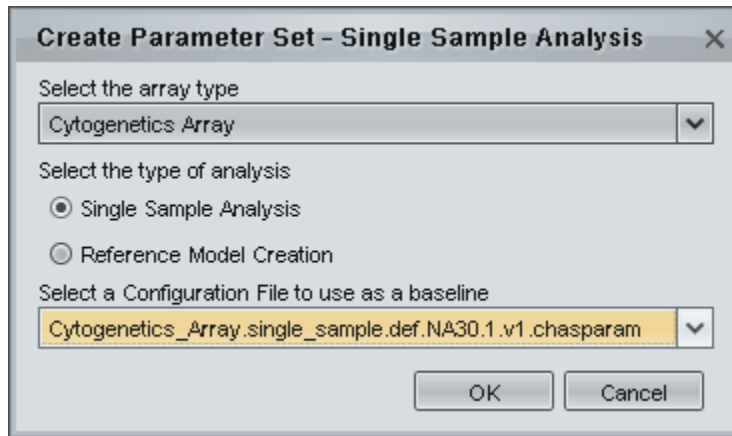
1. Select **Analysis > Customize Analysis Parameters** on the menu bar.

The Create Parameter Set dialog box opens.



**Figure 3.27 Create Parameter Set dialog box**

2. Select the probe array type from the drop-down list.
3. Select the type of analysis you want to create the parameter file for: Single Sample Analysis or Reference Model Creation.
4. Select the configuration file to modify or use as a starting point.



**Figure 3.28 Select a configuration file**

4. Click **OK**.

The Create Parameter Set dialog box for the selected analysis type opens. You can view a brief description of an editable parameter at the bottom of the dialog box by clicking on the parameter name.

The values are also described in [Analysis Parameters](#) (page 233).



**Important:** The analysis algorithm uses different terms to identify the parameters. You may see these terms in various error messages. The internal names are listed in parenthesis at the end of the description and in the material below.

Display Name	Current Value
<input type="checkbox"/> Allele Peaks	
AP Plot Frequency	50
AP Minimum Markers	250
AP Bandwidth Factor	0.45
<input type="checkbox"/> Copy Number Parameters:HMM Parameters	
Diagonal Weight	0.995
MAPD Weight	0.22
Min Segment Size	5
<input type="checkbox"/> Copy Number Parameters:HMM Parameters:Priors	
Mean	-1.63,-0.58,0,0.41,0.65,0.9
Standard Deviation	0.3,0.3,0.3,0.3,0.3,0.3
<input type="checkbox"/> GC Correction	
GC Correction	true
<input type="checkbox"/> LOH	
LOH Minimum Information	100.0
LOH Critical Value	8.0
<input type="checkbox"/> Probe Level Background Correction:Marker Level Normalization	
Median Autosome Median Normalization	true
<input type="checkbox"/> Smooth Signal Graph Output	
Sigma Span	50
<input type="checkbox"/> Weighted Log2 Ratio	
Median Smooth Marker Count	5

Description  
Each Weighted Log2 Ratio value is constructed by taking a window of size "Weighted log2 ratio Median Smooth Marker Count" around each marker and calculating the median log2 ratio. (Internal name=median-smooth-marker-count)

Reset Selected to Baseline Save As Cancel

**Figure 3.29 Editing parameter values**

5. To edit a parameter value:  
Click a numerical value and edit the numbers.  
Change a true/false value and make a selection from the dropdown list.
6. To restore a parameter to its original setting, select the parameter and click **Reset Selected to Baseline**.
7. To save the new parameter file, click **Save As**.  
The Create Parameter Set dialog box appears.

Create Parameter ...0.1.v1.chasparam

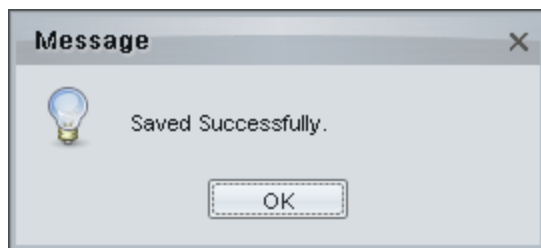
Input Parameter File Name

Custom Parameter Set\_KSA

OK Cancel

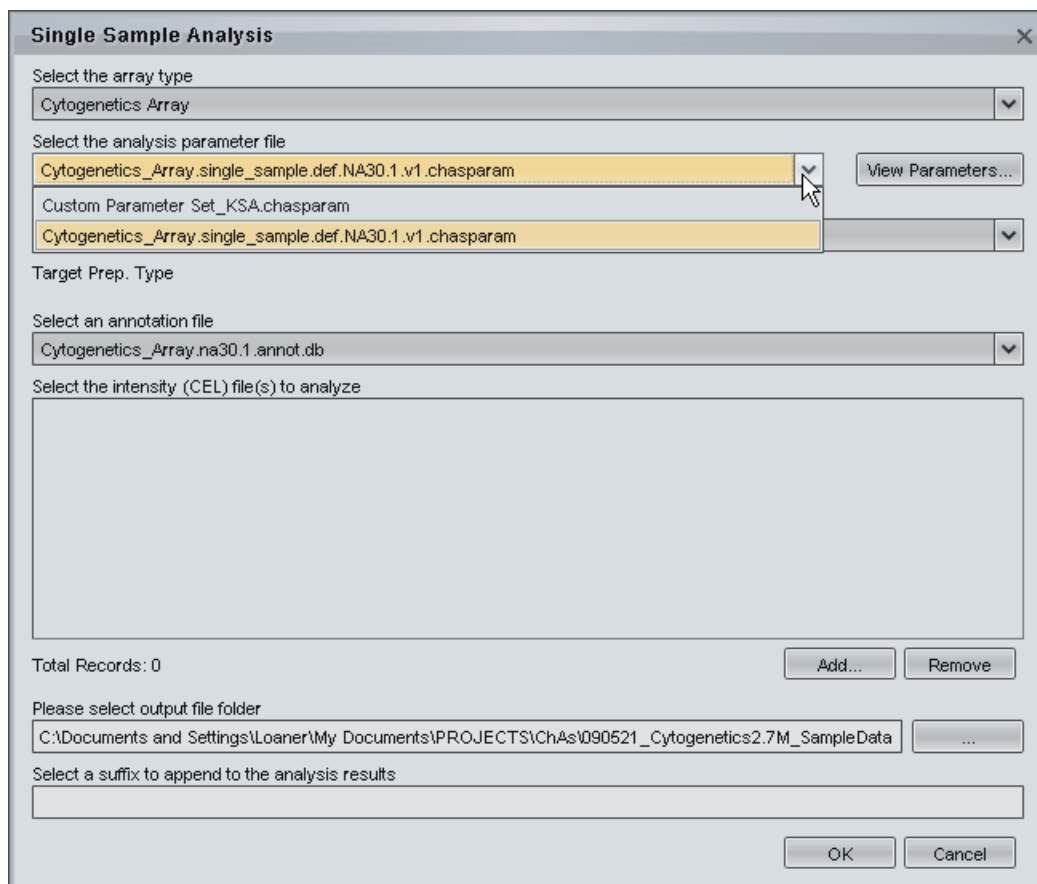
**Figure 3.30 Create Parameter Set File dialog box**

8. Enter a name for the parameter file and click **OK**.  
A message notifies you that the new file has been saved.



**Figure 3.31 Confirmation message**

The new parameter file is now available when setting up an analysis.



**Figure 3.32 New parameter file in the parameter file list**

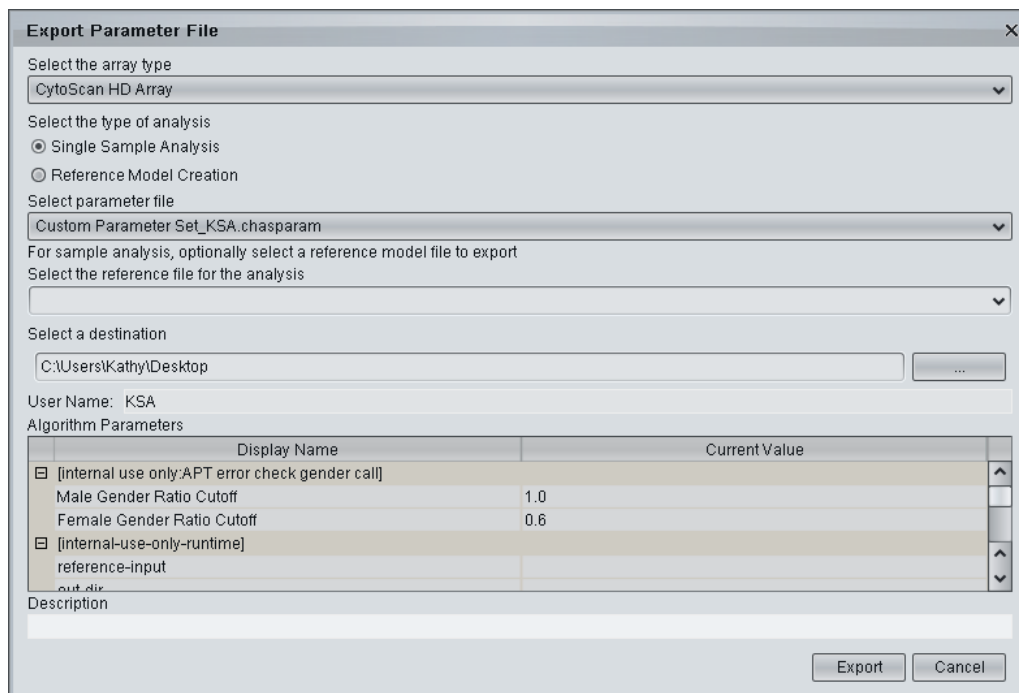
### Exporting an Analysis Parameter File

You can export analysis parameter files to other systems or users.

1. Select **Analysis > Export Analysis Parameters** on the menu bar.

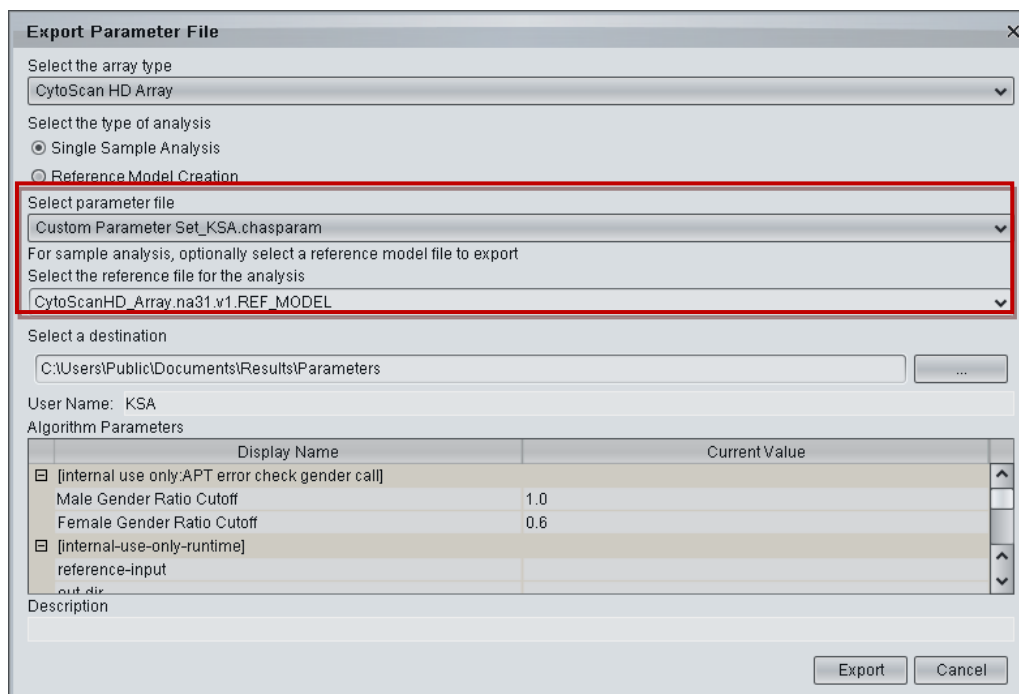
The Export Parameter File dialog box opens.





**Figure 3.33 Export Parameter File dialog box**

2. Make a selection from the array type drop-down list.
3. Choose the type of analysis: Single Sample Analysis or Reference Model Creation
4. Select the parameter file that you want to export. For Single Sample Analysis, you can also select a reference file to export (optional).



**Figure 3.34 Parameter file and reference file selected for export**

5. Select a destination folder.

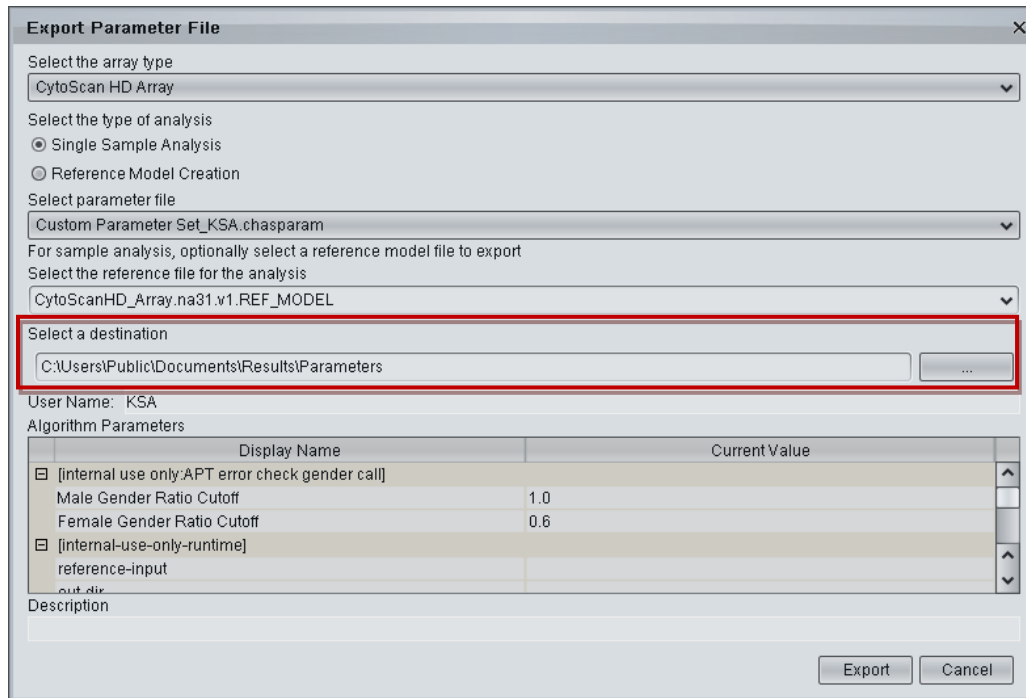




Figure 3.35 Destination folder for the exported files

Enter the path to the destination folder. Alternatively:

- Click the **Browse** button  .
- In the dialog box that appears, select the destination folder. You can also create a new folder (click the  button).
- Click **Save**.

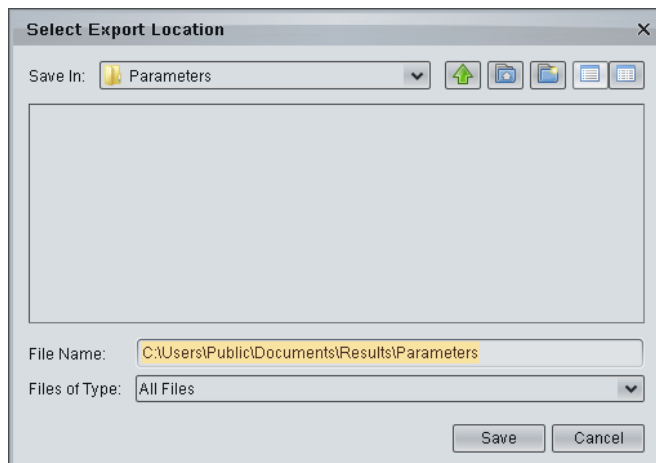
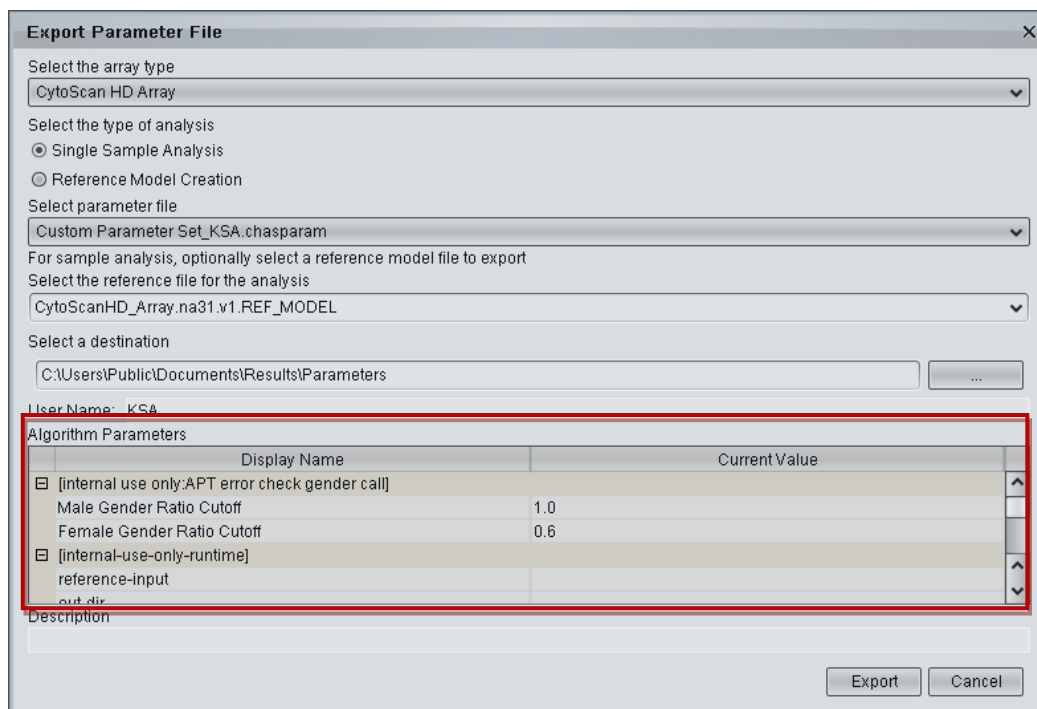


Figure 3.36 Select Export Location dialog box

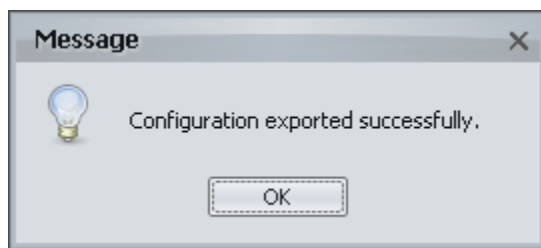
7. Review the selected algorithm parameters.



**Figure 3.37 Algorithm parameters**

- Click **Export**.

The system notifies you of a successful export of the parameter file and reference model file (optional).



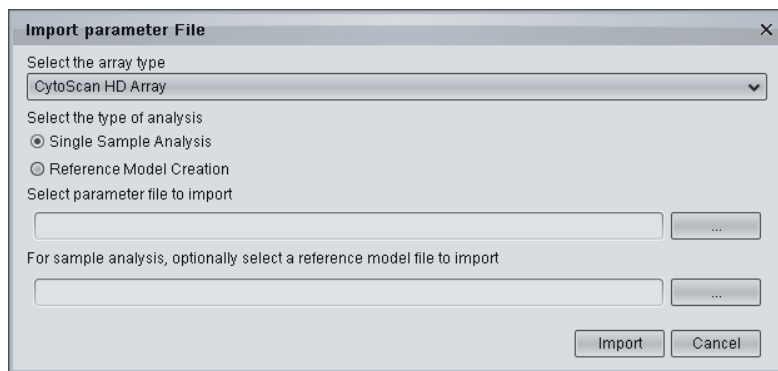
**Figure 3.38 Export confirmation message**

### Importing Analysis Parameter files

You can import analysis parameter files from other systems or users.

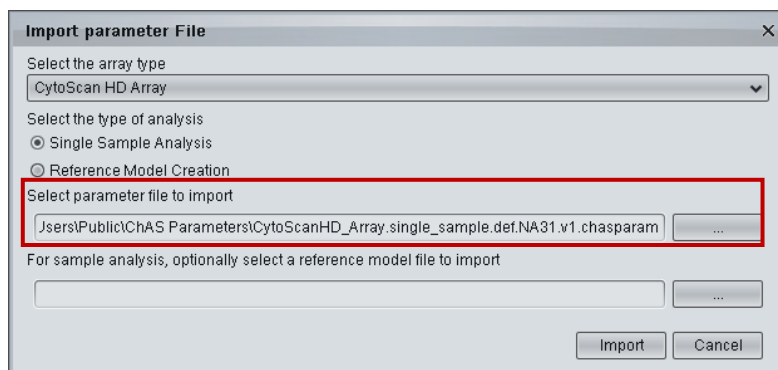
- Select **Analysis > Import Analysis Parameters** on the menu bar.

The Import Parameter File dialog box opens.



**Figure 3.39 Import Configuration File dialog box**

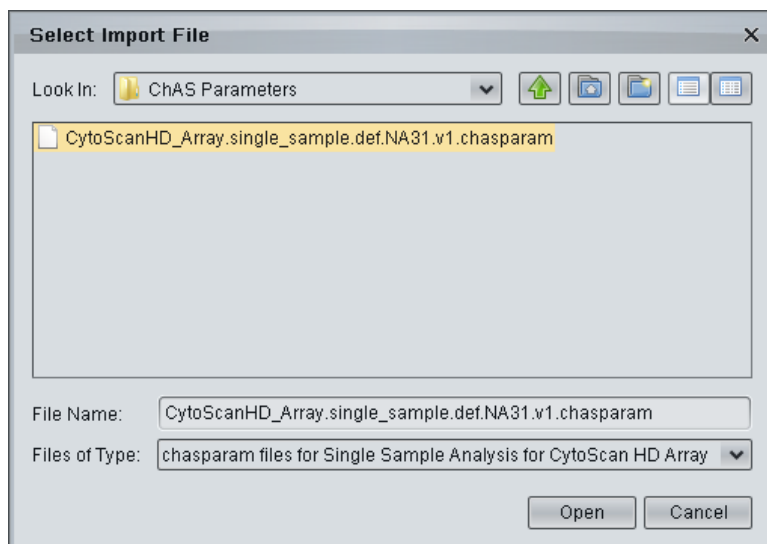
2. Make a selection from the array type drop-down list
3. Choose the analysis type: Single Sample Analysis or Reference Model Creation
4. Select the parameter file that you want to import.



**Figure 3.40 Parameter file to import**

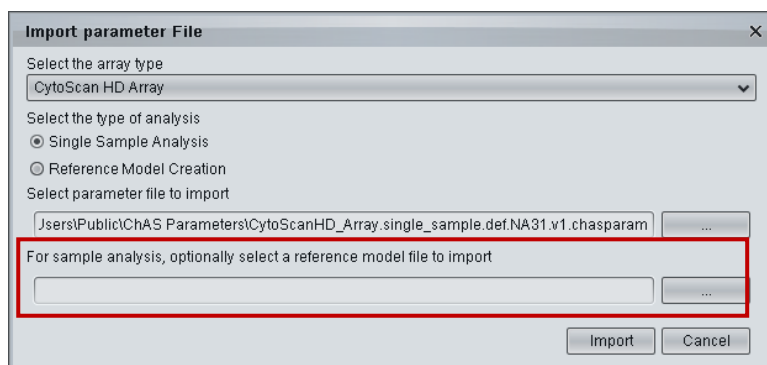
Enter the file path and file name. Alternatively:

- a. Click the Browse button .
- b. In the dialog box that appears, select the parameter file, and click **Open**.




**Figure 3.41 Select Import File dialog box**

6. Select a reference file for single sample analysis to import (optional).

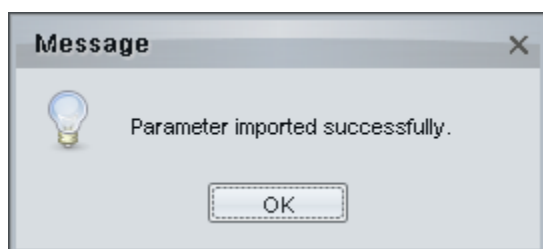


**Figure 3.42 Specify a Reference Model file to import (optional)**

Enter the path and file name in Import Parameter File dialog box. Alternatively:

- a. Click the **Browse** button .
  - b. In the dialog box that appears, select the reference file, and click **Open**.
6. Click **Import**.

The system notifies you of a successful import of the parameter file and reference model file (optional).



**Figure 3.43 Import confirmation message**

## Chapter 4: Loading Data

ChAS can display data from:

- CytoScan™ HD Array CYCHP files or Cytogenetics Whole-Genome 2.7M Array CYCHP files, generated in ChAS
- Genome-Wide Human SNP Array 6.0 CNCHP files, generated in Genotyping Console (GTC)

 **Note:** When referring to steps that apply to both Cytogenetics CYCHP and SNP 6.0 CNCHP data files, the CHP files are described as CxCHP files.

The following sections include:

- [Introduction to Loading Data](#) (below)
- [Loading Files](#) (page 67)
- [Segment Smoothing and Segment Joining \(Optional\)](#) (page 71)
- [Setting QC Parameters](#) (page 76)

### Introduction to Loading Data

The same steps are used to load results (CxCHP files) from a CytoScan™ HD Array, Cytogenetics Whole-Genome 2.7M Array, or Genome-Wide Human SNP Array 6.0.

**When loading CYCHP files into ChAS for viewing, the software:**

1. Selects the run-length encoded segments in the CYCHP file to display as segments.
2. Applies any smoothing or joining that would alter the length and other properties of segments.

 **Important:** In a new user profile, smoothing and joining are turned on by default. The smoothing and joining settings are specific for the array type CYCHP file (for more details on smoothing and joining, see page 71).

3. Examines the data for contiguous segments (segments with no intervening data) with the same copy number. Contiguous segments with the same copy number are displayed as one segment.
4. Displays the segments and graph data:
  - Segment Data
    - Copy Number Gain/Loss
    - Mosaicism (for Cytogenetics Whole-Genome 2.7M CYCHP files only)
    - Loss of heterozygosity (LOH)
  - Graph Data
    - Copy Number State
    - Log2Ratio
    - Weighted Log2Ratio
    - LOH
    - Smooth Signal
    - Allele Peaks or Allele Difference (depending on array type)
    - Genotype Calls (for CytoScan™ HD CYCHP files only)

**When loading CNCHP files into ChAS for viewing, the software:**

1. Performs segment detection by analyzing the CN and LOH graph data in the CNCHP file.



**Note:** When running the Segment Reporting Tool in GTC on SNP 6 data, the software sets the end coordinate such that the segment ends *at the base position of the last marker in the segment*. When loading SNP 6 data into ChAS, the segment detection sets the end coordinate for a segment such that the segment ends *one base after the last marker in the segment*. This may result in a discrepancy between the end position for segments when comparing data analyzed in both GTC and ChAS.



**ChAS uses the median of the aberrant markers' CNStates as the recalculated CNState of the smoothed (and/or joined) segment.**

2. Applies any smoothing or joining that would alter the length and other properties of segments.  
Smoothing is similar to the process applied when running the Segment Reporting Tool in GTC.



**Important:** For CNCHP files from the SNP 6.0 Array, smoothing but not joining is turned on by default in a new user profile.

3. Examines the data for contiguous segments (segments with no intervening “normal” data or aberrant data of the other type, gain, or loss) with the same copy number. Contiguous segments with the same copy number are displayed as one segment.
4. Displays the segments and graph data:

- Segment data
  - Copy Number Gain/Loss
  - Loss of Heterozygosity (LOH)



**Note:** The expected copy number state on the X chromosome of normal males is not constant over its entire length. This is due to the structure of the sex chromosomes. See [Copy Number Segments on the X and Y Chromosomes](#) (page 40) for more information.

- Graph Data
  - Copy Number State
  - Log2 Ratio
  - Allele Difference
  - SmoothSignal
  - LOH

## Loading Files


Loading CxCHP data for viewing in ChAS involves the following steps:


1. **Optional:** Before loading: Select Segment Smoothing and Segment Joining parameters for processing the CN Gain and Loss Segment data.



**Important:** Smoothing is turned on by default in a new user profile. Joining is also on by default for the CytoScan™ HD Array and Cytogenetics Whole-Genome 2.7M Array, but not the Genome-Wide Human SNP Array 6.0.

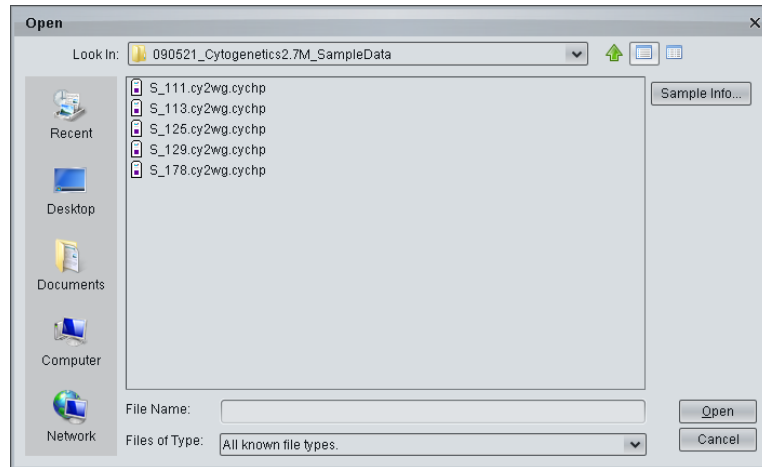
2. Select CYCHP files from CytoScan™ HD Arrays or Cytogenetics Whole-Genome 2.7M Arrays (generated by analyzing CEL files in ChAS) and CNCHP files from Genome-Wide Human SNP Arrays 6.0 (generated by analyzing CEL files GTC).

You can also select region information files in AED and BED format for loading. Use the Open dialog box (click the  button) to load CxCHP data files, Affymetrix Extensible Data (AED), or Browser Extensible Data (BED) annotation files. The AED and BED files that you open will be automatically loaded when a new session is started with the same user profile.

 **Note:** You may want to edit smoothing and joining parameters. This can be done before or after loading the CxCHP data. See [Segment Smoothing and Segment Joining](#) (page 71) for more information.

1. Select **File > Open** on the menu bar. Alternatively, click the **File Open** button .

The Open dialog box appears.




**Figure 4.1** Open dialog box

 **Note:** You can also open the Open dialog box from the Analysis Results dialog box. See [Analysis Results Dialog Box](#) (page 56).

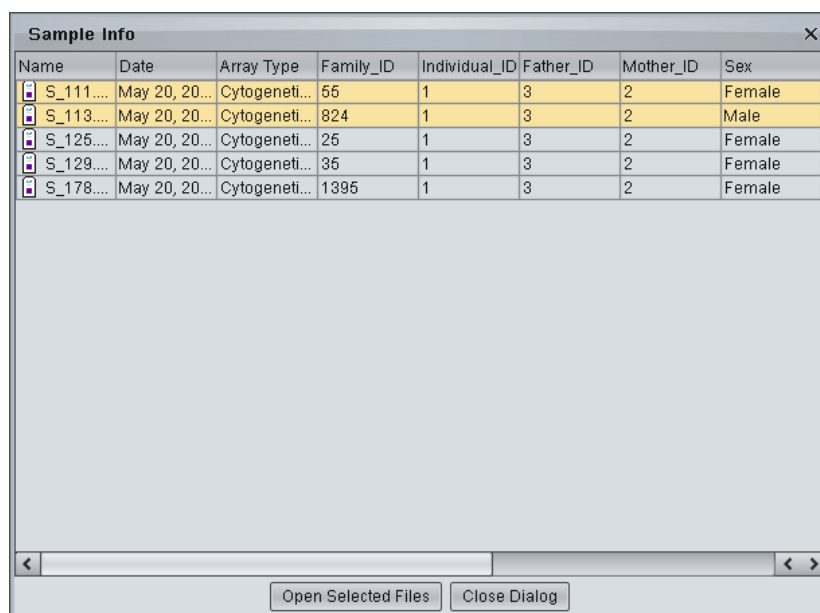
 **Note:** CYCHP or CNCHP files and Region Information files can be in different locations.

2. To view information about results, select one or more files and click **Sample Info**.

The Sample Info dialog box opens (Figure 4.2).

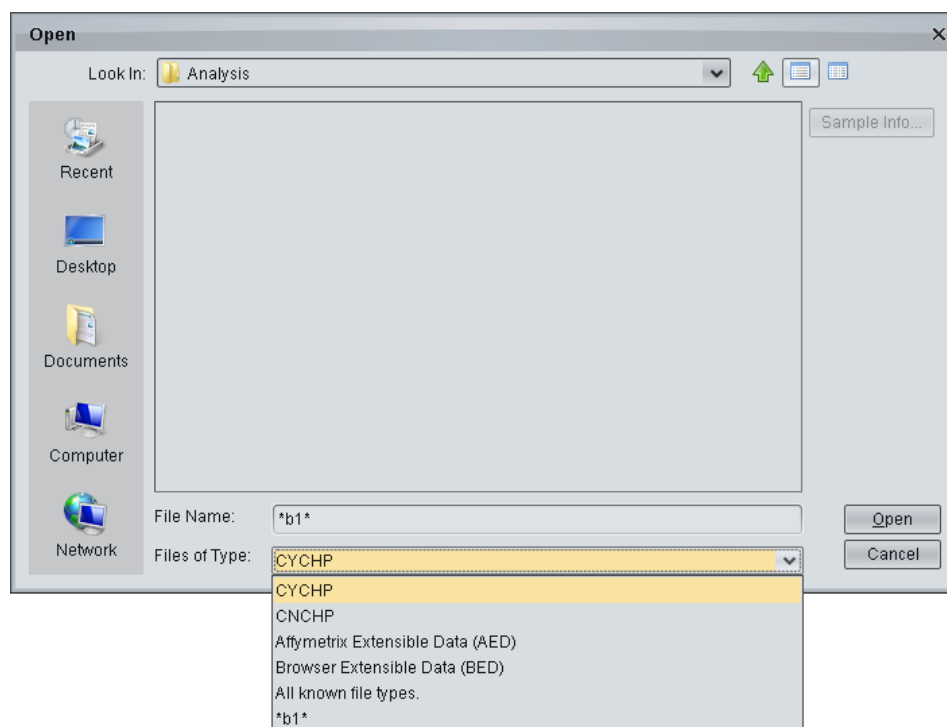
 **Note:** If the CYCHP and ARR files are located in the same folder, the Sample Info dialog box shows information about both the sample and the results.



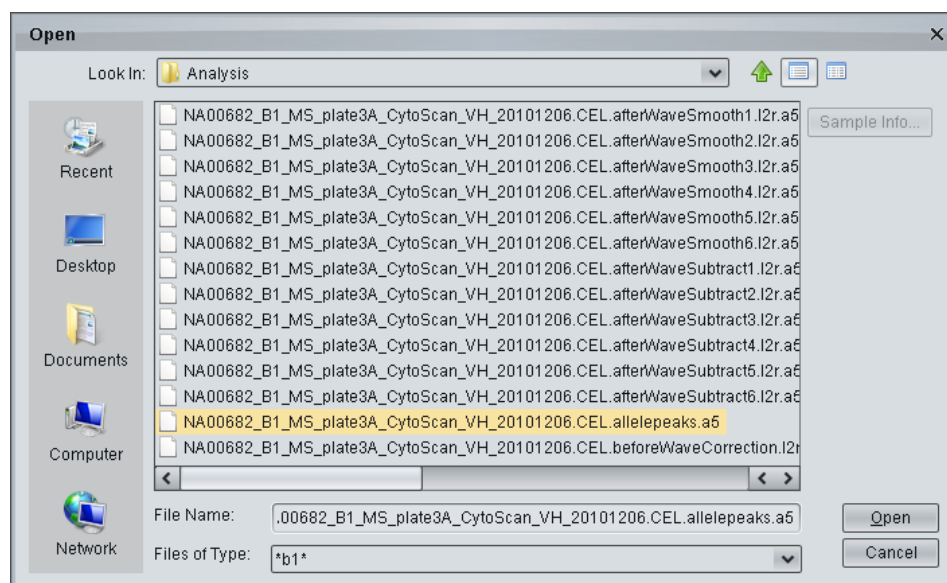


**Figure 4.2 Sample Info dialog box**

3. To load files from the Sample Info dialog box, select the files and click Open Selected Files.  
 Alternatively, close the dialog box and navigate to the folder with the files that you want to load. To find particular files, a search function is available in the Open dialog box. To use the search function:
  - a. Enter a text string with an asterisk (\*) before and after the search term in the File Name field (Figure 4.3).
  - b. Select a file type from the drop-down list.
  - c. Click **Open**.
 Files with names that include the search term are displayed in the Open dialog box.



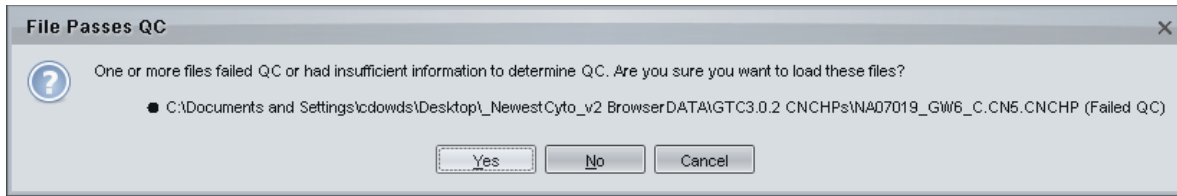
**Figure 4.3 Enter a file name search term (\*text string\*) and select a file type**



**Figure 4.4 Open dialog box shows files with names that include the search term**

4. Select the files (you can use Shift click or CTRL-click to select multiple files)
5. Click **Open**.

If any of the files fail the QC checks, a warning notice appears.

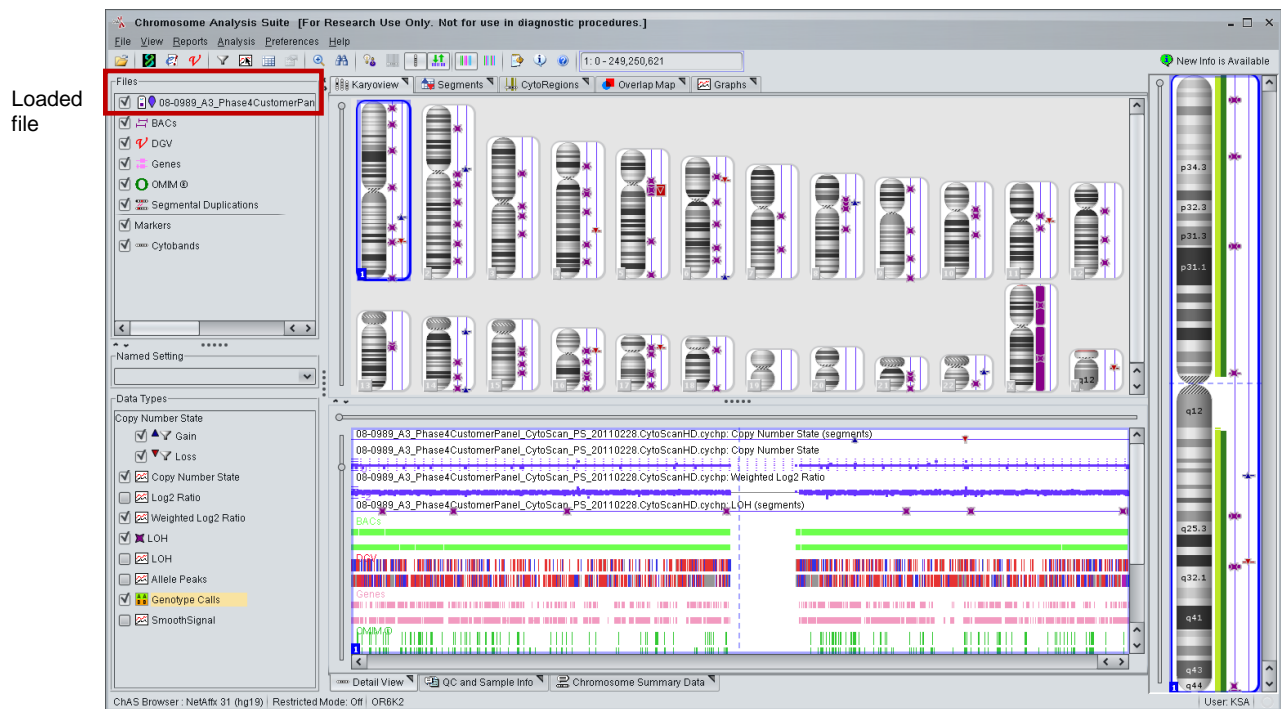


**Figure 4.5 Warning Notice for QC failure**

You can click **Yes** to load the files anyway.

The loaded files appear in the Files list.

 **Note:** During loading, the software checks the data for segments with the same CN state and no intervening data. Such segments will be normalized and displayed as one segment.




**Figure 4.6 Files list shows the loaded data**

## Segment Smoothing and Segment Joining (Optional)

“Smoothing” and “joining” are non-destructive processes that affect the display of Copy Number segments. Smoothing and joining are performed on the Copy Number State data during the loading process, based on settings that are specified before loading. Any data filtering is applied after smoothing and joining.

 **Important:** Smoothing and joining are turned on by default in a new user profile.

Smoothing and joining are specified per array type. The processes do not affect the marker data in the CNCHP or CYCHP file. If these settings are turned off, the Copy Number segment data is displayed without smoothing or joining.

 **Note:** Smoothing and joining affect only data loaded from CNCHP and CYCHP files.

Segments which have been smoothed and/or joined are indicated by a blue check mark in the Smoothed/Joined column of the Segments table. The segment ID name indicates whether smoothing and/or joining has occurred. A red "X" indicates no smoothing or joining has been applied. See [About Smoothing](#) (page 74) and [About Joining](#) (page 75) for more information.

NA	BACs	OMIM ID	Segmental D...	Smoothed/Joined	Segment ID	Sample UUID
RP11-77D9...	607100 Nep...	chr2:880477...		✓	joined109164116	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
RP11-60D16...		chr2:949243...		✓	joined9988282	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTC-774N1...		chrX:622643...		✓	smoothedjoined58117748	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-3126K...	120252 Corn...	chr3:314694...		✗	seg484	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
RP11-423A2...	602600 Myo...	chr6:482242...		✗	seg486	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
RP11-933C6...		chr13:18774...		✗	seg488	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-3015G...		chr11:82193...		✗	seg490	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-2182L2...		chr15:97616...		✗	seg492	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-2042F...	610860 Glyc...	chr5:133319...		✗	seg494	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
RP11-344H1...	605551 QT...			✗	seg498	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-2168F...		chr11:31997...		✗	seg500	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
RP11-787M...				✗	seg502	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99
CTD-2128B...	191045 Card...	chr2:199657...		✗	seg504	uuid:00002ec4-41a3-4e31-1bb1-0071a6002c99

**Figure 4.7 Example Segments table with joined and smooth/joined segments**

**To set smoothing and joining parameters:**

1. Select **Preferences > Edit User Configuration** on the menu bar.
2. In the User Configuration dialog box that appears, click the **Segment Data** tab.

**User Configuration**

Segment Data | QC Thresholds | Color Rules | Files

For CxCHP files of each array type, edit the configuration of Smoothing and Joining to create Copy Number Segments. Smoothing and Joining processing happens at file loading and at configuration saving (always before any CN Gain or Loss Segment Filtering).

☒ Use default segment data rules configuration [Reset customized to defaults](#)

Enable/disable and configure Copy Number State data processing for generating Copy Number Segments

**Smoothing** [merges only contiguous aberrations of the same type (Gain, or Loss)]

☒ Smooth Gain or Loss CNState runs to the most common marker state value

**Smoothing maximum jump limit** [adjacent data points farther apart in CNState will not be smoothed together]

☐ Limit smoothing of CNState data to not smooth aberrant segments more distant than this number of CNStates

**Joining** [merges aberrations of the same type (Gain, or Loss) which are separated by some normal state data]

☐ Join Gain or Loss CNState runs separated by no more than this number of markers of normal state data

☐ Join Gain or Loss CNState runs interrupted by normal state data which are separated from each other by no more than this distance measured in kbp

**Joining maximum jump limit** [aberrant normal-data-flanking data which is farther apart in CNState will not be joined together]

☐ Limit the joining of CNState data (which flanks normal state data) to not join aberrant segments more distant than this number of CNStates

**Figure 4.8 User Configuration dialog box, Segment Data tab**

3. Select an array type from the drop down list.

The following options are available:

#### Use default segment data rules configuration

For the CytoScan™ HD Array, applies the default smoothing and joining rules:

- Smooth Gain or Loss CNState runs to the most common marker value, then generate segments
- Join any “split” CNState runs separated by no more than 50 normal-state markers
- Join Gain or Loss CNState runs interrupted by normal state data which are separated from each other by no more than 200 kbp

For the Cytogenetics Whole-Genome 2.7M Array, applies the default smoothing and joining rules:

- Smooth Gain or Loss CNState runs to the most common marker value, then generate segments
- Join any “split” CNState runs separated by no more than 50 normal-state markers

For SNP 6 arrays, applies the default smoothing rule:

- Smooth Gain or Loss CNState runs to the most common marker value, then generate segments

#### Smooth Gain or Loss CNState runs to the most common marker value

Smoothing to the most common marker state value is only applied to contiguous CNState runs of the same type (gain or loss).

#### Limit smoothing of CNState data to not smooth aberrant segments more distant than this number of CNStates

If this option is chosen, CNState runs which are farther apart than the “smoothing maximum jump limit” will not be smoothed. For example, if the smoothing maximum jump limit is set at 1, then adjacent segments with CNState 3 and 5 will not be smoothed.

#### Join Gain or Loss CNState runs separated by no more than this number of markers of normal state data

If this option is chosen, only Gain or Loss CNState Runs which are separated by less than a threshold number of markers of normal state data will be joined. For example, if the marker threshold is set at 50, then CNState runs separated by more than 50 markers of normal state data will not be joined.

#### Join Gain or Loss CNState runs interrupted by normal state data which are separated from each other by no more than this distance measured in kbp

Choose this option to limit the distance between two CNState runs separated by normal state data that can be joined. For example, if the distance set at 20000 kbp, CNState runs separated by normal state data spanning more than 20000 kbp from the end of the first aberrant segment to the start of the next aberrant segment will not be joined.

#### Limit the joining of CNState data (which flanks normal state data) to not join aberrant segments more distant than this number of CNStates

If this option is chosen, CNState runs that flank normal state data and are farther apart from each other than a user-specified CN threshold will not be joined. For example, if the CNState joining maximum jump limit is set at 1, then CNState runs of 3 and 5 that flank normal state data will not be joined.



**Important: If multiple smoothing and/or joining check boxes are selected, all criteria must be met to smooth and/or join the segments.**

After smoothing and joining, the marker count, mean marker distance and confidence values will be recalculated. For more information, see:

- [About Smoothing](#) (below)
- [About Joining](#) (page 75)

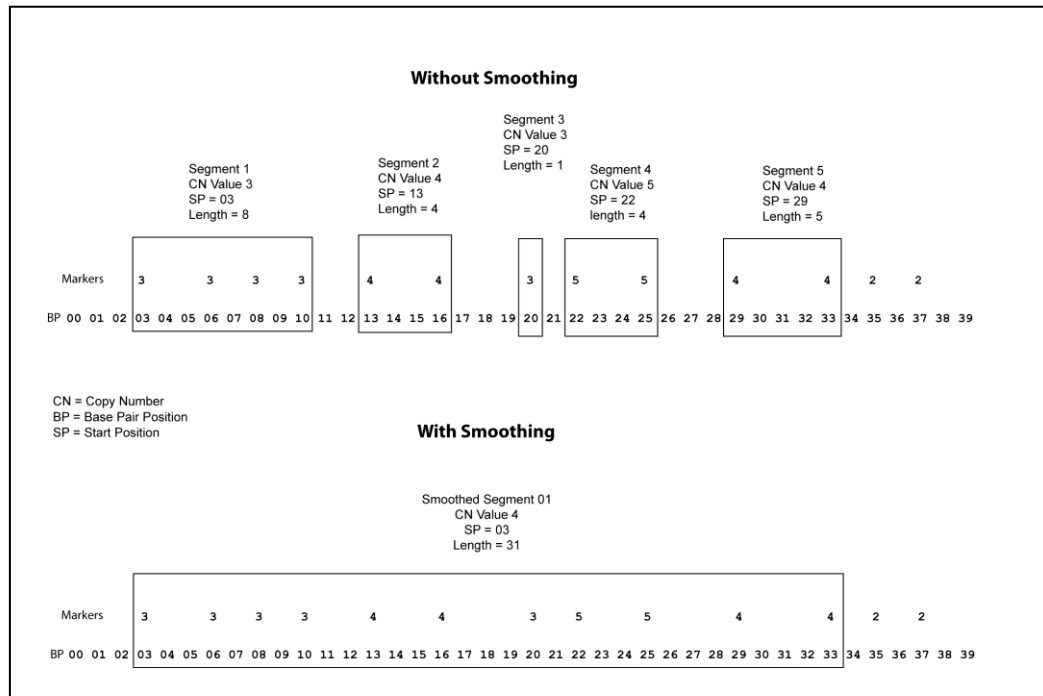
## About Smoothing



**Note:** The examples provided are for a case where the expected copy number is 2. Similar calculations take place for the X and Y chromosomes where the expected copy number may be 0, 1 or 2, depending on gender and whether the segment is located within or outside of the PAR region.

If you have a contiguous set of segments with gain values (for instance, of CN State values of three and four), with no markers of copy number 2 or lower, **without smoothing** they will be treated as a series of individual gain segments. The same rules apply to a set of segments with loss values of 0 or 1.

If you have a contiguous set of markers with gain values of three and four, with no intervening markers of copy number 2 or lower, **with smoothing** they will be consolidated into a single gain segment.



**Figure 4.9 Simplified schematic representation of smoothing**

If you have a contiguous set of markers with loss values of zero and one, with no intervening markers of copy number 2 or higher, after smoothing, they will be consolidated into a single loss segment.

The smoothing process is the same as the process automatically performed by the Segment Reporting Tool in GTC. Different methods are used to assign the CN state value and perform the confidence calculations, as described below. See the *GTC User Manual* for more information.

### Copy Number State for Smoothed Segments

The median CNState of all the markers in the Segment is assigned as the Copy Number State value for the new smoothed segment. The median will thus always be either an integer or a half integer (like 3.5).

For all the half-integer cases:

- Gains are rounded up to the next full integer (3.5 goes to 4)
- Losses are rounded down to the next full integer (0.5 goes to 0, 1.5 goes to 1).

### Confidence Score Calculations for Smoothed Segments

Smoothed segments must have their Confidence scores recalculated after smoothing.

 **Note: Genome-Wide Human SNP Array 6.0 CNCHP file segments do not contain segment confidence scores.**

Recalculations shall use confidence values and marker sums from all component segments in the resulting Smoothed Segment.

Because smoothing stays either above or below normal copy number, the confidence of the smoothed over segment should still be used in the average of the combined segment. The formula used is shown below.

For smoothed segments A, B, and C, if:

$k$  = marker count for segment

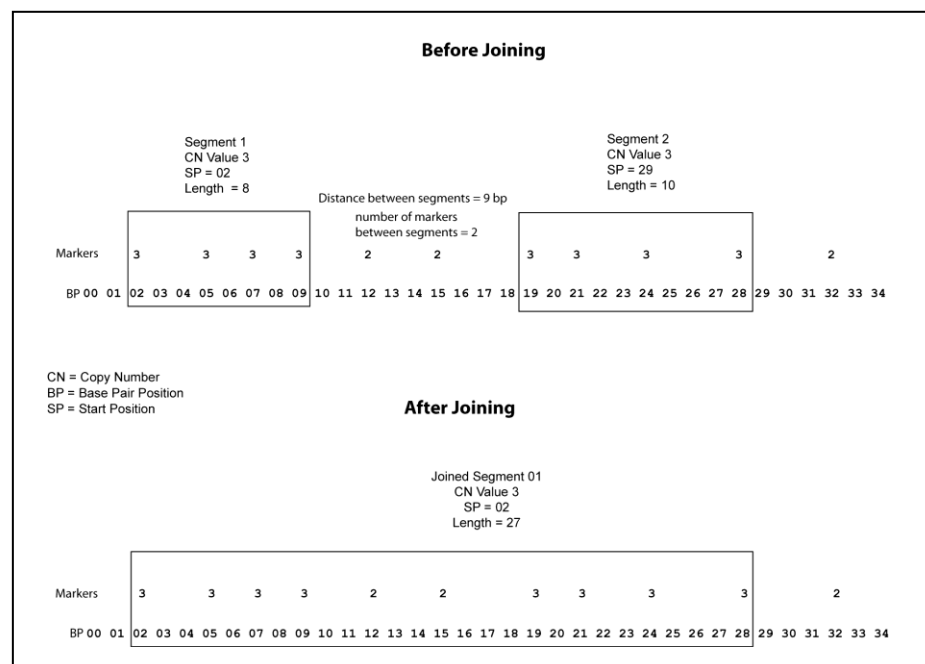
$m$  = mean confidence score for segment

then:

$$\text{Smoothed Segment Confidence} = \frac{k(A) * m(A) + k(B) * m(B) + k(C) * m(C)}{k(A) + k(B) + k(C)}$$

## About Joining

The joining options allow you to join segments with the same type (gain or loss) aberrant CNState that are separated by no more than a specified number of normal-state markers or by no more than a specified distance of normal-state data.



**Figure 4.10 Simplified schematic representation of joining**

The equivalence of CNState of the segments to be joined could have happened as a result of smoothing, or been from “raw” unsmoothed segments with the same CNState.

## Confidence Score Calculations for Joined Segments

 **Note: Genome-Wide Human SNP Array 6.0 CNCHP file segments do not have confidence scores recalculated after smoothing, since the scores are not stored in the CNCHP files.**

Joining is a much different operation than smoothing. Joining IGNORES EXPECTED NORMAL CN STATE DATA for recalculation of the joined segment's CN State and for its Confidence score. In other words, joining imputes copy numbers and the imputed values can have a much different biological interpretation than the observed values. Copy number values that are altered (the expected normal values being smoothed over) due to joining are treated as missing when computing confidence, that is, not included in any confidence score.

Recalculation of Joined segment confidence uses confidence values and marker sums from all aberrant component segments in the resulting Joined Segment, but does NOT factor in the normal data point(s) for their copy number state or confidence.

If segments D, E and F are joined over the normal/expected copy number segment E, then E is implicitly deemed to be unreliable and imputed. For joined segments D, E, and F:

$k$  = marker count for segment

$m$  = mean confidence score for segment

then:

$$\text{Joined Segment Confidence} = \frac{k(D) * m(D) + k(F) * m(F)}{k(D) + k(F)}$$

The data from segment E are lost and do not count toward confidence (or copy number state value) of the joined segment.

## Setting QC Parameters


ChAS checks the CNCHP and CYCHP files for certain QC values. The software notifies you if the QC parameters do not meet the thresholds.

You can adjust the QC threshold values or select different QC metrics.



**Note: Affymetrix does not recommend selecting different QC metrics.**

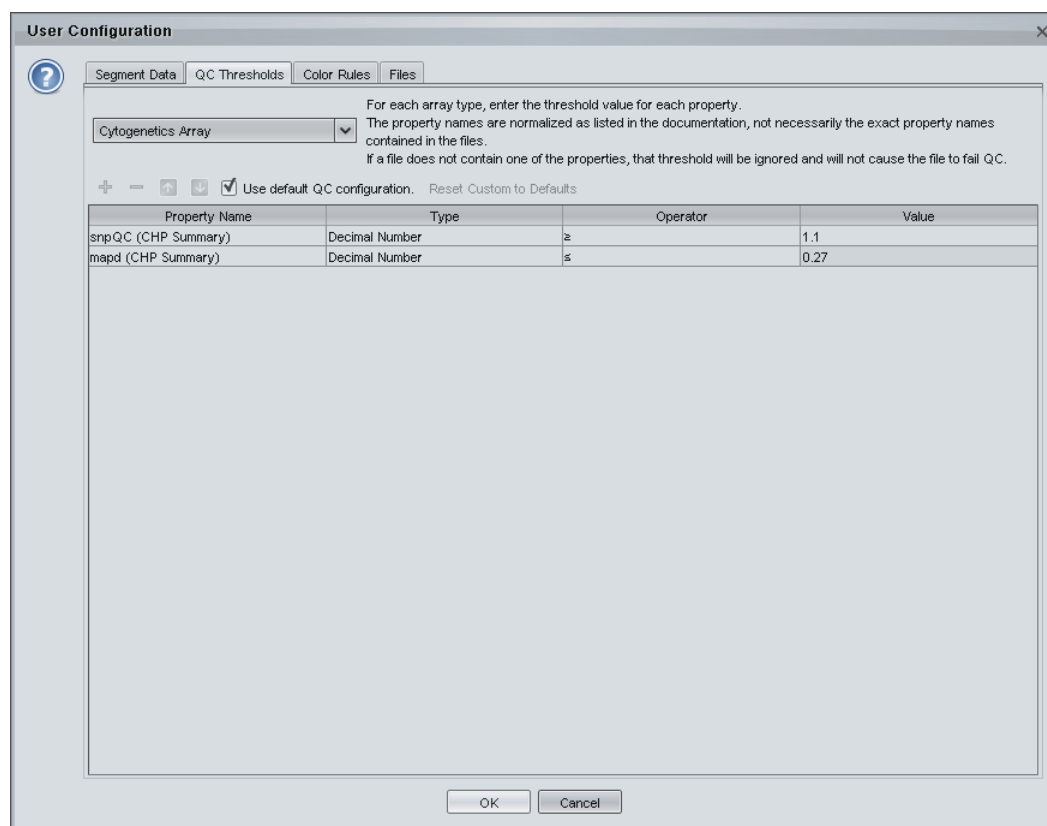
### To view the QC thresholds:

1. Select **Preferences > Edit Configuration** on the menu bar. Alternatively, click the  toolbar button.  
The Configuration dialog box opens.
2. Click the **QC Thresholds** tab.
3. Select an array type from the drop-down list.




**Note: QC parameters are specified per array type.**





**Figure 4.11 Cytogenetics Array QC Thresholds, default settings**

The QC Thresholds table has four columns:

- Property Name**
- snpQC – A QC metric for SNP probes that is derived from polymorphic (SNP) probes
  - mapd – A QC metric for all probes used to determine copy number that is derived from both polymorphic (SNP) and non-polymorphic (CN) probes
-  **Note:** The property names are from the header information of the CxCYP file.

**Type** Value or algorithm used for that type of QC.

**Operator** The type of comparison performed.


**Value** Value assigned to the threshold.

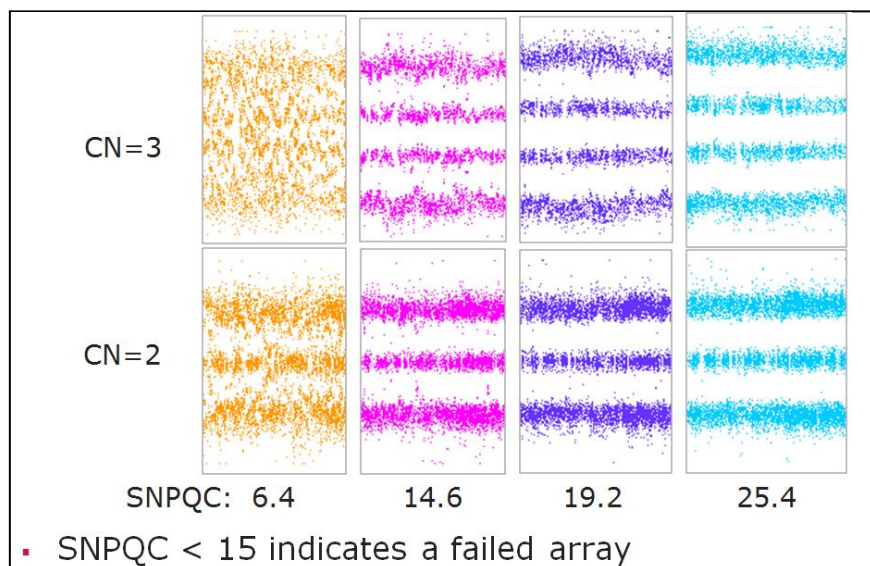
The default QC thresholds are:

Array Type	QC Parameter		
	mapd	snpQC	Waviness SD
CytoScan™ HD Array	≤ 0.25	≥ 15.0	≤ 0.12
Cytogenetics Whole-Genome 2.7M Array	≤ 0.27	≥ 1.1	–

Genome-Wide Human Array SNP 6.0	$\leq 0.35$	–	–
---------------------------------	-------------	---	---

 **Note:** Different snpQC algorithms are used for the CytoScan™ HD Array and Whole-Genome Cytogenetics 2.7M Array.

 **Note:** The waviness SD metric is applicable to blood and cell line data. The waviness SD metric is not intended for alternative sample types such as solid tumor or FFPE samples in which the results may vary as a result of the biological complexity.






**Figure 4.12 SNPQC Metric vs. Track Quality**

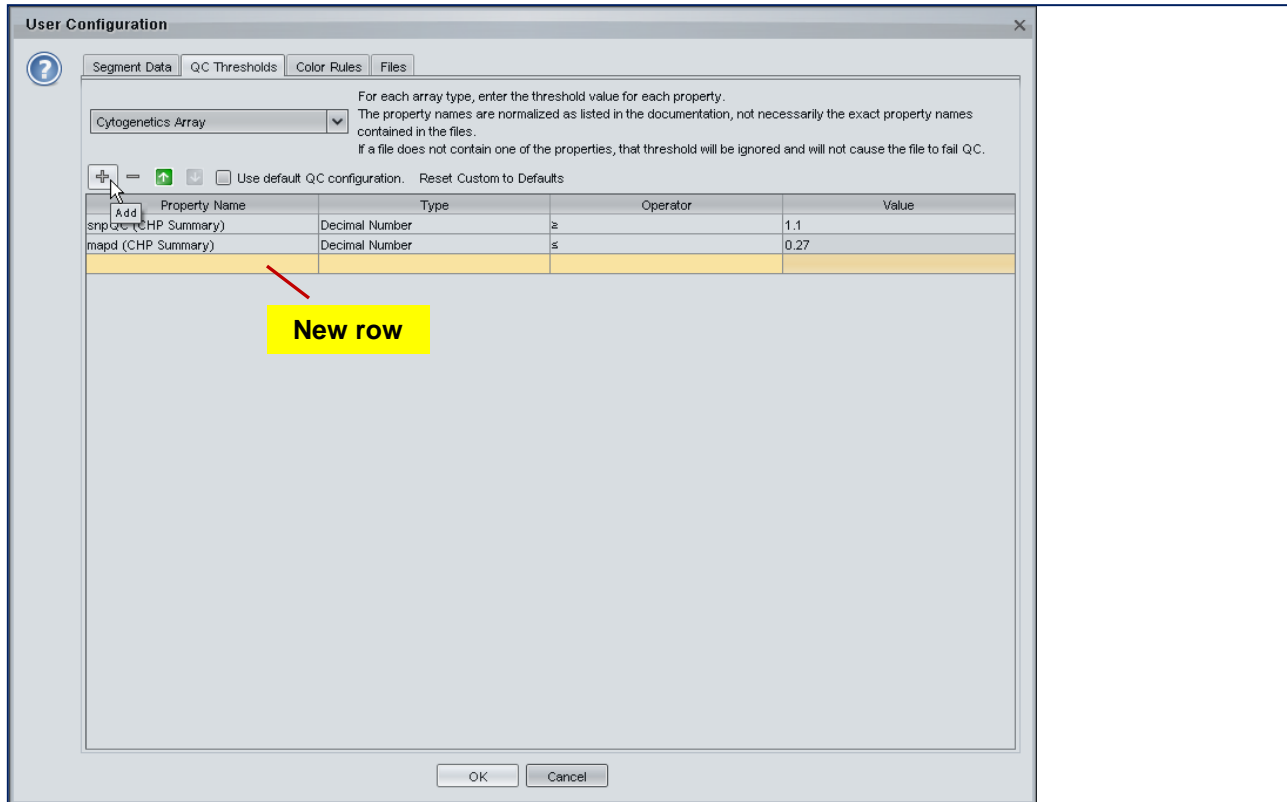
SNPQC is one of the CytoScan™ within-array QC metrics which provides insight into the overall level of data quality from a SNP perspective. When evaluating the SNPQC values, the key consideration is to ensure that the threshold is exceeded. The quality of the SNP allele data is compromised, from an interpretation perspective, when the SNPQC values are below the recommended acceptance threshold as illustrated by the two leftmost graphs representing the two and three copy allelic state. When the SNPQC value is below 15, as illustrated by the data represented in the two graphs at the left, the noise within the array is higher than expected which compromises the overall data quality and clarity of results. However, when the SNPQC value is above 15, the consideration is whether the SNPQC value is above or below the threshold value and not the absolute magnitude.

As long as the SNPQC value exceeds the threshold, there is a retention in the data quality as illustrated by the graphs to the right which demonstrate clear allelic data across a broad range of SNPQC values that exceed the recommended threshold. The threshold was determined from thousands of arrays processed across multiple reagent lots, operators, and sample aberration types. SNPQC is one of the metrics used to assess array quality and should be helpful in determining which experimental data sets are of satisfactory quality to continue with subsequent interpretation.

#### To add a QC property or edit a QC threshold:

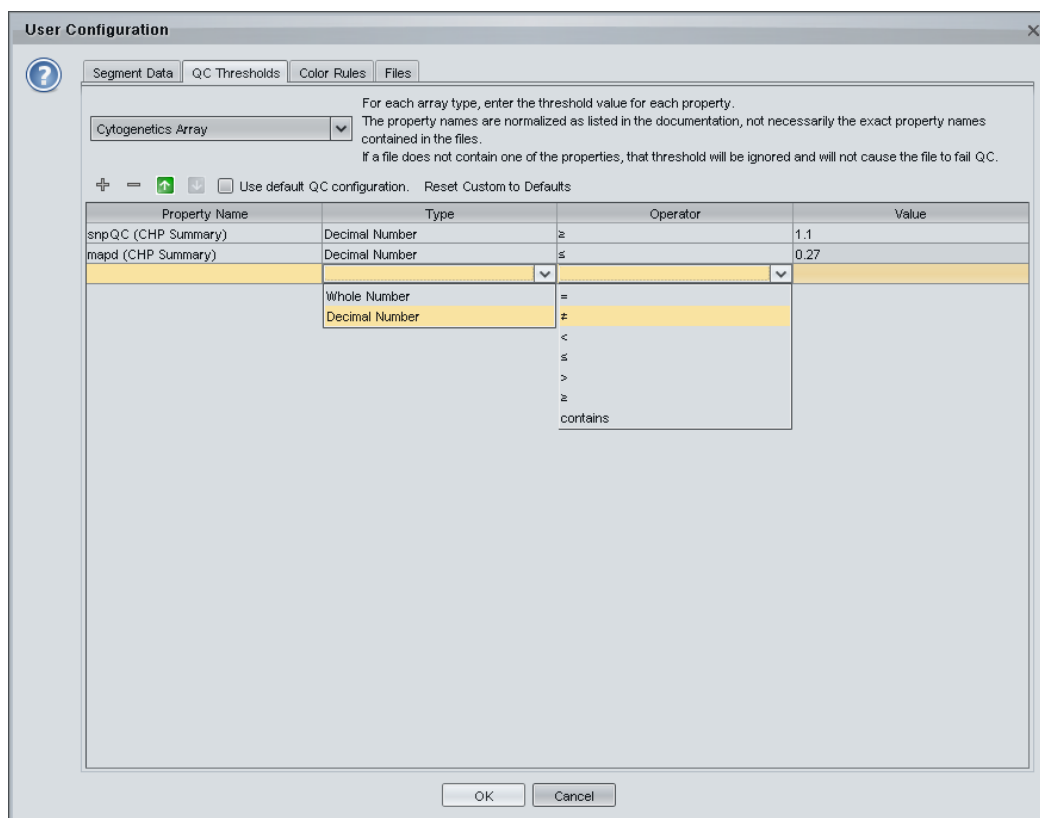
1. Open the Configuration dialog box (click the  toolbar button).
2. Select an array type from the drop-down list.
3. In the QC Thresholds tab, remove the check mark next to “Use default QC configuration”.
4. Click the **Add** button .

- A new row appears in the table. To delete a property row, select the row and click the **Remove**  button.
- Click the Property Name cell and enter the new QC property name.



**Figure 4.13 Adding a new row to the QC Thresholds table**

- Click the Type cell and select Decimal Number or Whole Number.
- Click the Operator cell and select an operator from the drop-down list.
- Double-click the Value cell and enter the threshold value for the QC property.



**Figure 4.14 QC Selecting Type and Operator values for the QC property**

9. To specify another QC property, repeat step 1 to step 7.
10. Click **OK** to apply the new QC thresholds.

## Chapter 5: Viewing Data in ChAS

The loaded CYCHP or CNCHP data can be displayed in:

- Graphic Displays (see [Displaying Data in Graphic Views](#), page 89).
- Tables (see [Displaying Data in Table Views](#), page 177).

After the data is loaded, you can:

- Filter the segments by Segment Parameters to hide segments that do not meet your requirements for significance.

See [Filtering Segments](#) (page 130).

- Select a region information file for use as a CytoRegion file and:
  - Perform differential filtering for segments in CytoRegions and in the rest of the genome.
  - Display only segments that appear in CytoRegions using Restricted Mode.

See [Using CytoRegions](#) (page 134).

- Select a region information file for use as an Overlap Map and use the Overlap filter to conceal segments that appear in the Overlap Map items. This functionality may be helpful for tracking or filtering out benign copy number change regions.

See [Using the Overlap Map](#) (page 144).

- Add selected features of the genome to new or existing Region (AED) files, and edit annotation data on existing annotations.

See [Creating and Editing AED Files](#) (page 152).

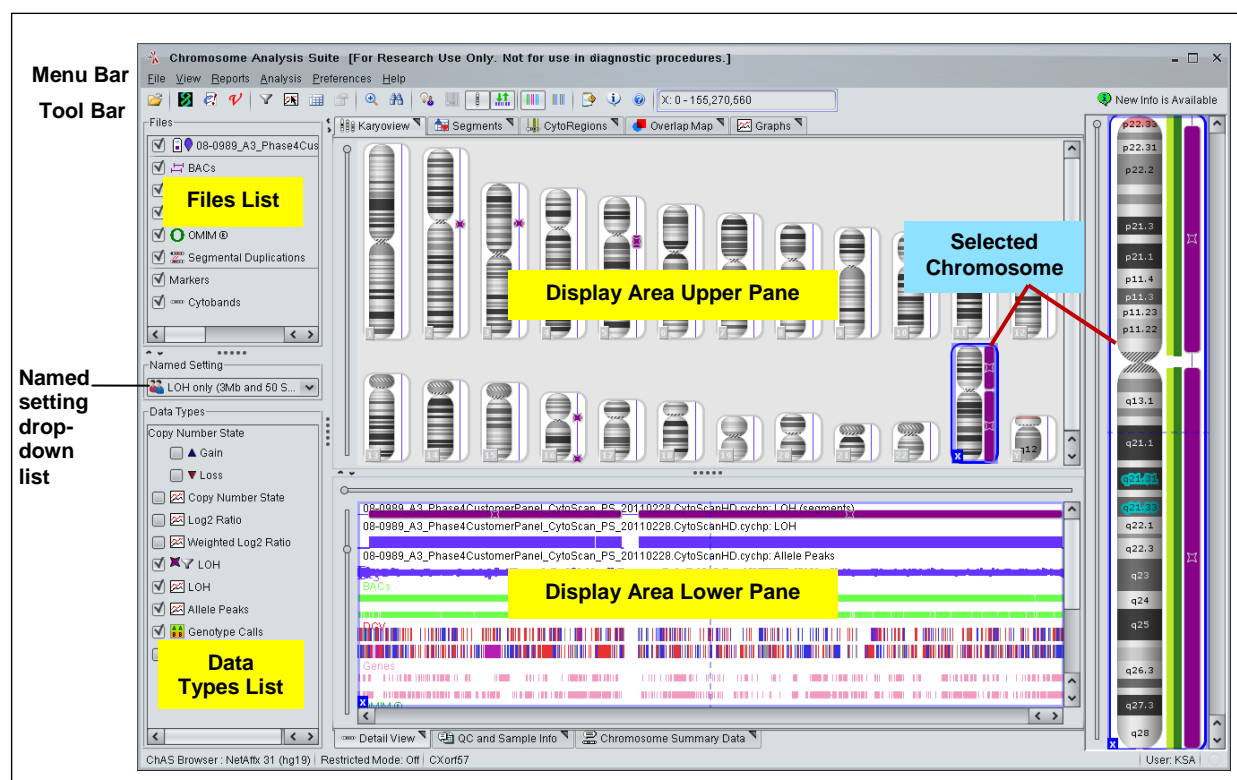
- Prepare reports on your findings by exporting graphics and table data in PDF and other formats.

See [Reporting Results](#) (page 206).

- Save setups of ChAS for different tasks in user profiles and named settings.

See [User Profiles and Named Settings](#) (page 225).

## Overview of ChAS Window Components



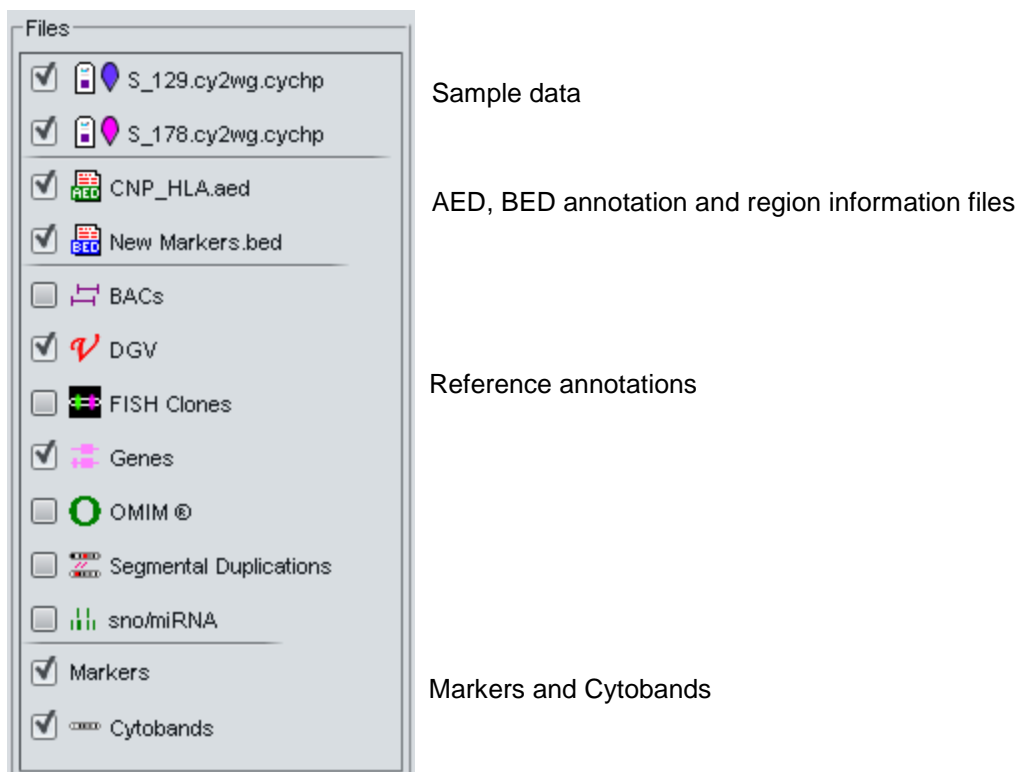
**Figure 5.1 ChAS window**

The ChAS window components include:

- **Menu Bar:** Access to the functions of the software.
- **Tool Bar:** Quick access to commonly used functions.
- **Files List** (page 83): Displays data and annotation files that can be displayed.
- **Data Types List** (page 84): Displays the type of data available in the files.
- **Named Settings** (page 85): Displays a list of the previously saved display settings for ChAS.
- **Status Bar** (page 85): Displays information on the status of the software, the ChAS Browser NetAffx Genomic Annotation file version, the hg version, and information about the annotation or probe that the mouse pointer is nearest to in the Detail View **Display Area** (page 86): Displays the following data in graphical and table formats:
  - CYCHP or CNCHP graph data
  - Detected Segments
  - Regions
  - Reference Annotations

## Files List


The Files list displays the different sources of data and annotations that are loaded in the Chromosome Analysis Suite. Files are grouped by type in the Files list.




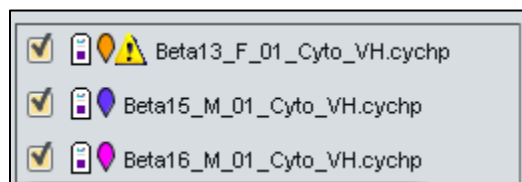
**Figure 5.2 Files list**

Displays the files grouped by type:

- Sample Data



Colored nibs  display the color used for the data lanes for that sample in the Karyoview, Selected Chromosome View, and Detail View.

If a loaded file has a QC parameter that is out of range, an alert symbol  appears next to the file name.



**Figure 5.3 QC alert for Sample Data files**

- Region Information Files

Icons indicate the file type (AED or BED) and whether the loaded files have been selected as a CytoRegion file  or Overlap Map  file.

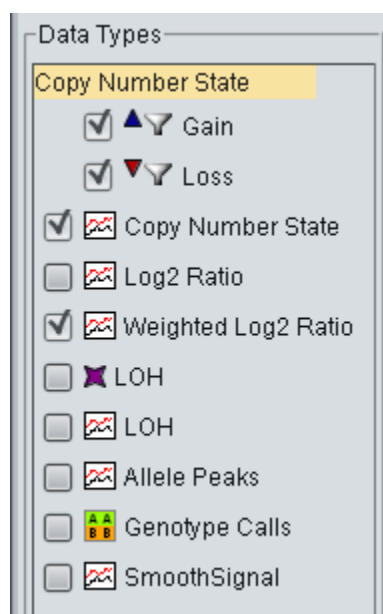
- Reference Annotations (loaded during software start-up)  
Icons indicate the annotation type.
- Cytobands (Separated from other reference annotations because cytobands cannot be moved in the displays)

See [Selecting Data for Display](#) (page 107) for information on using the Files list to select loaded data and region information files, and reference annotation for display.

You can export the feature information in these files to a new AED region information file (see [Exporting Information in AED or BED Format](#), page 174).

## Data Types List

The Data Types list shows the types of data available for display in the Karyoview, Selected Chromosome View, and Detail View. The available data types may vary, depending upon the type of sample data available. See [Introduction to Loading Data](#) (page 66).





**Figure 5.4 Data Types list**

The Data Types list allows you to select from Segments data and graph data.

The Segments data is displayed graphically in:

- Karyoview
- Selected Chromosome View
- Detail View

If filtering is applied to a segment type, a funnel icon  appears next to the segment symbol in the list.

Graph data, indicated by the  icon, is displayed only in the Detail View. See [Detail View](#) (page 98) for more information.

Unselected data is also concealed from the different tables.

See [Selecting Data Types for Display](#) (page 109) for information on using the Data Types list to select different data types for display.



# Named Settings

The Named Settings drop-down list enables you to apply a previously created setting for ChAS. The settings may include things like:

- Segment Filter and Overlap Map Filter settings
- Types of data to be displayed
- Restricted Mode on/off status

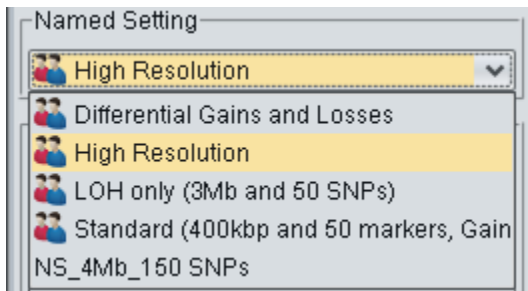


Figure 5.5 Named Settings

 **Note:** Affymetrix-supplied Named Settings (indicated by the  icon) should not be deleted from the system because they are shared by all user profiles.

See [User Profiles and Named Settings](#) (page 225) for information on creating and using Named Settings.

# Status Bar

The status bar displays information on:

- NetAffx Genomic Annotation database and its hg version
- Restricted Mode status
- Cursor Position
- User Profile ID

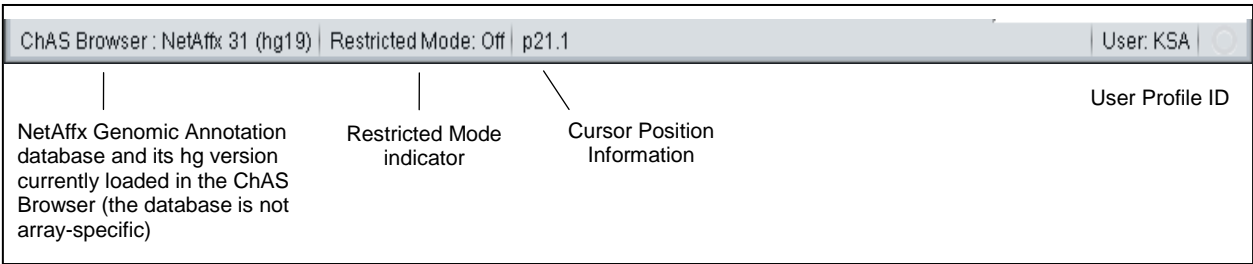
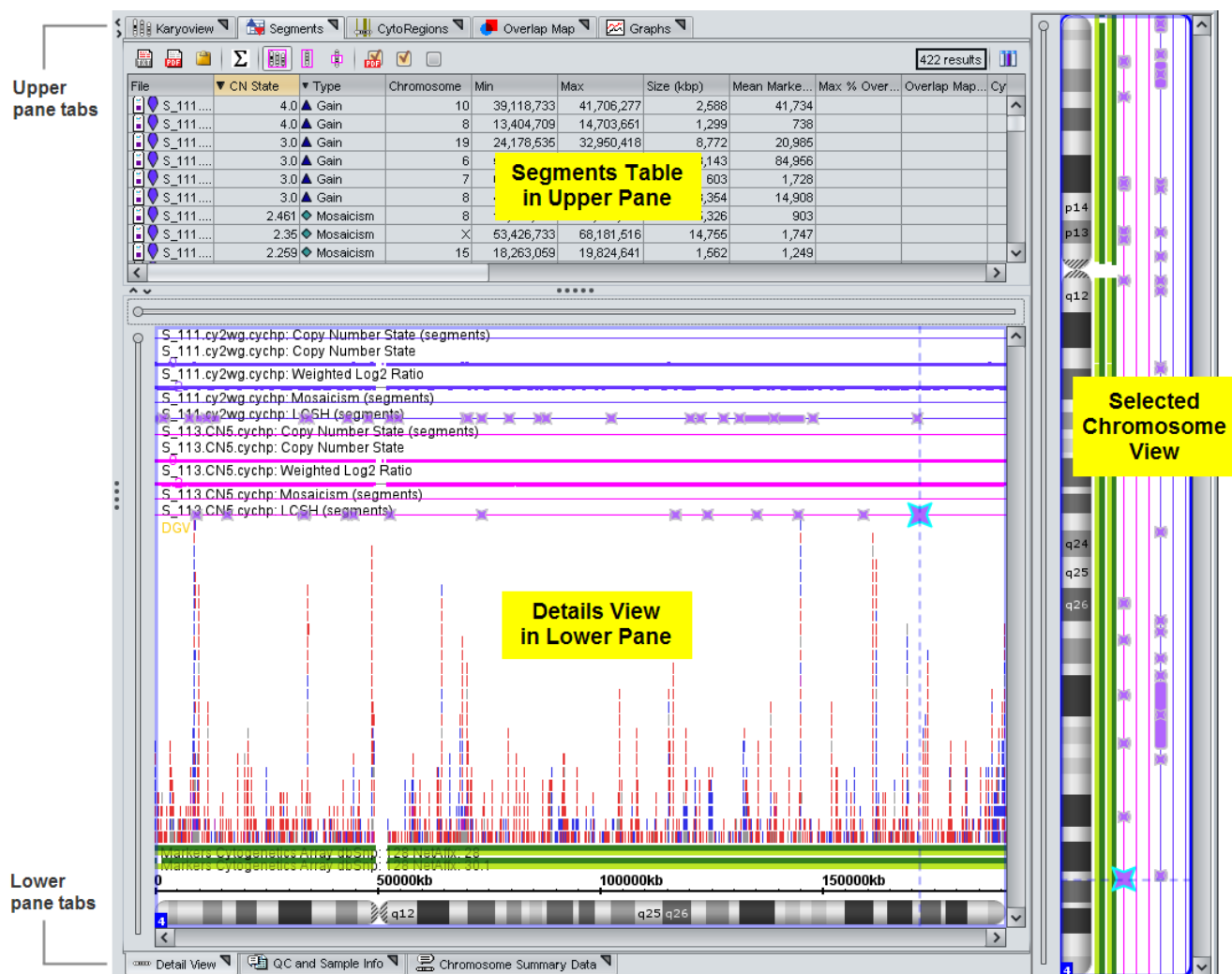


Figure 5.6 Status Bar

## Display Area



The Display Area is divided into three panes:

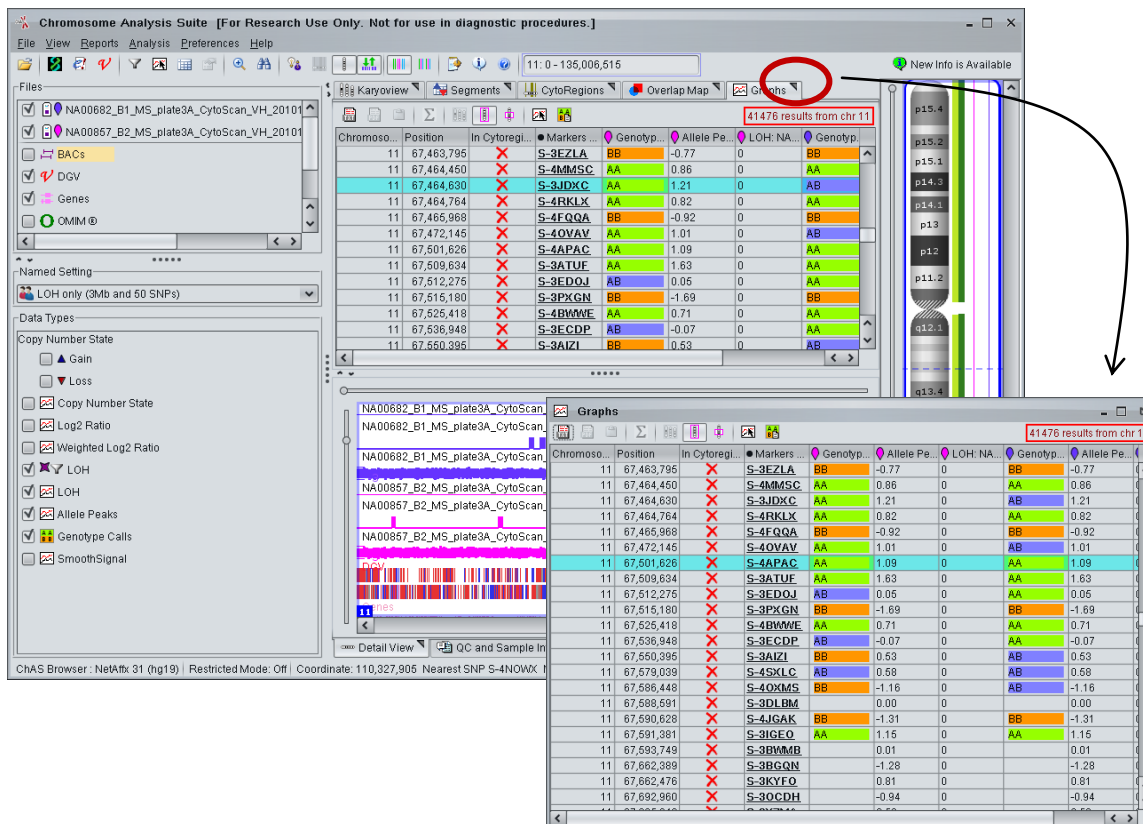
- [Upper Panes](#) (page 87)
- [Lower Panes](#) (page 87)
- [Selected Chromosome View](#) (page 87)



**Figure 5.7 Display Area showing Segments Table and Detail View**

The tabs in the upper and lower panes display different types of data, in both graphical and table formats. Data from the same sample files is displayed in all three panes.

You can display a pane in a separate window by clicking the  icon on the tab. To close the window and return the information to the tab panel, click the  icon in the window.



**Figure 5.8 Graphs pane opened in new window**

## Upper Panes

The Upper pane displays data in graphics and table formats:

- [Karyoview](#) (page 90): Displays selected segment types for selected sample files for all chromosomes.
- [Segments Table](#) (page 181): Displays a list of the detected segments in the selected sample files.
- [CytoRegions Table](#) (page 138): Displays a list of the Regions in the AED/BED file selected as the CytoRegion file. Includes information on detected segments which lie in CytoRegions.
- [Overlap Map Table](#) (page 148): Displays a list of the Regions in the AED/BED/Reference Annotation file selected as the Overlap Map. Includes information on detected segments that are overlapped by Overlap Map Items.
- [Graphs Table](#) (page 191): Displays marker data for the loaded and selected CxCHP files.

## Selected Chromosome View

The Chromosome View displays detected segments in selected sample files for the chromosome selected in the Karyoview.

See [Selected Chromosome View](#) (page 95) for more information.

## Lower Panes

The Lower Pane displays:


- [Detail View](#) (page 98): Displays the selected section of the chromosome displayed in Selected Chromosome View, with
  - Detected segments and graph data in selected CxCHP files

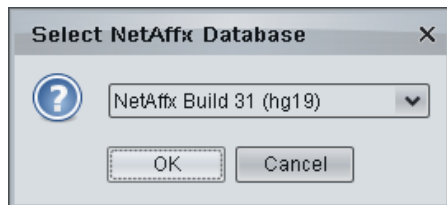
- AED/BED file regions and annotations.
- Reference annotation files
- [QC and Sample Info Tab](#) (page 193): Displays information about loaded Sample and Region (AED) files and other settings.
- [Chromosome Summary Data](#) (page 198): Summarizes particular data across each chromosome in the loaded sample data files (for example, proportion of each chromosome found to be in the state of LOH).

## Changing the ChAS Browser NetAffx Genomics Annotation File Version

You can change the NetAffx Genomic Annotations file loaded in the ChAS browser.

1. Select **File > ChAS browser NetAffx Genomic Annotation file version** on the menu bar
2. In the dialog box that appears, select a NetAffx Build from the drop-down list and click **OK**.

 **Note:** If there are loaded CxCHP files with hg version different from the selected NetAffx database, a message informs you that these data files will be closed before the NetAffx database is loaded.



**Figure 5.9 Select a NetAffx Genomics Annotation file version**

## Chapter 6:     **Displaying Data in Graphic Views**

ChAS provides different graphic views for the detected segment and other data. These views enable you to:

- Get an overview of the detected segments
- Compare segments between samples
- Drill down to examine areas of interest in more detail
- View the graph and marker information used to generate the detected segments
- Take advantage of reference annotations and external web sites
- Create your own Affymetrix Extensible Data (AED) files with regions of interest and annotations

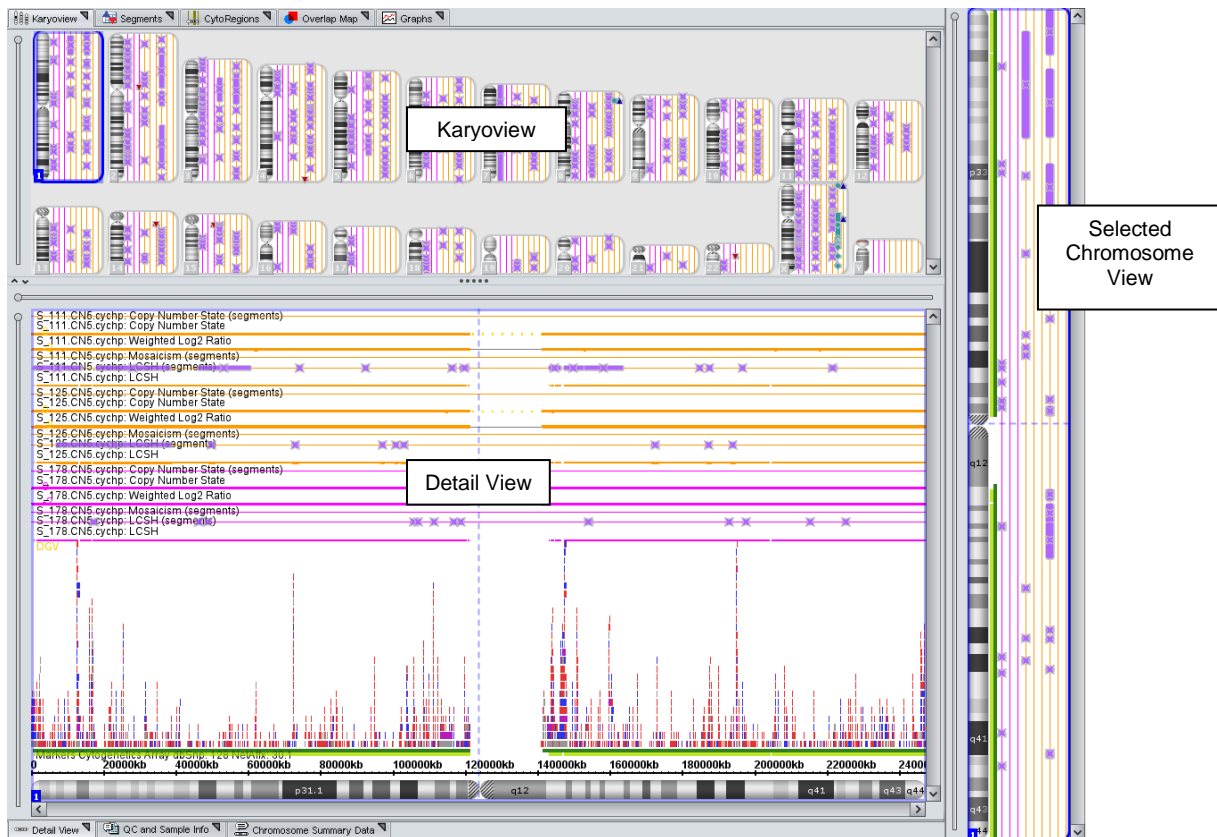
This chapter discusses:

- [The Graphic Display Views](#), below
- [Controlling the Display of Data](#) (page 105)
- [Learning More About Features](#) (page 122)

### **The Graphic Display Views**

The data can be displayed in the following graphic views:

- [Karyoview tab in upper pane](#) (page 90)
- [Selected Chromosome View](#) (page 95)
- [Detail View in Lower Pane](#) (page 98)



**Figure 6.1 Graphic display views**

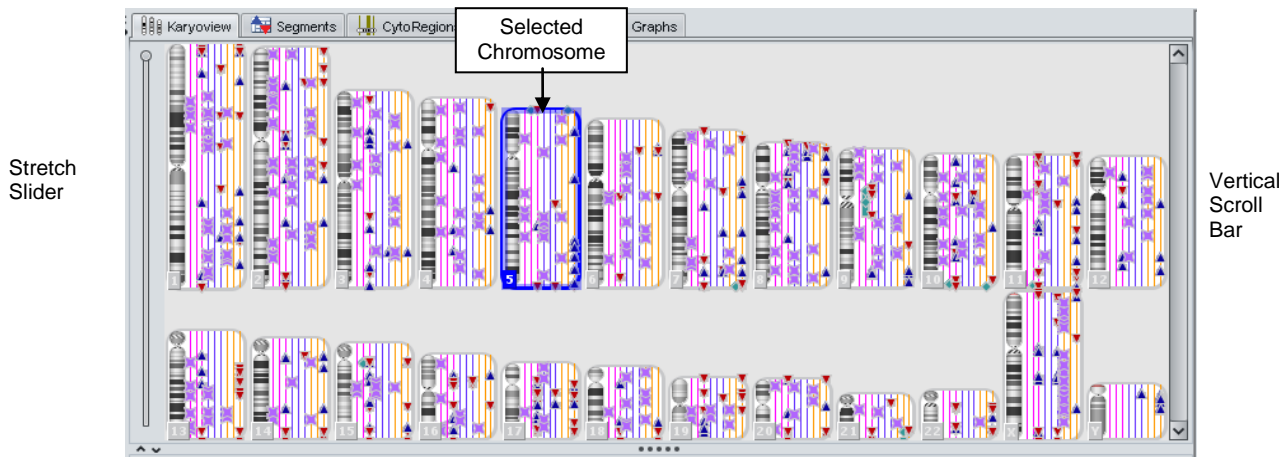
Data from the same sample files is displayed in all the views at different scales.

If an item in any of the views is selected, the icon for that item is enlarged or highlighted in the views.

## Karyoview

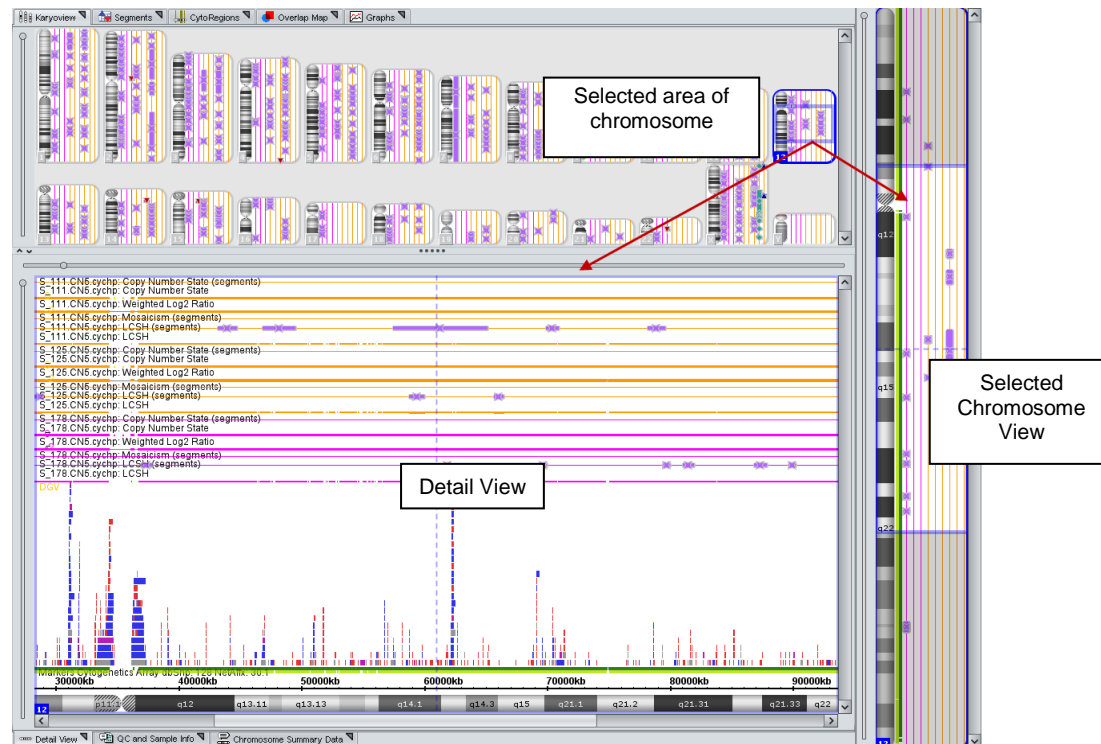
The Karyoview displays a genome-wide view of the detected segments and other data. In the Karyoview:

- Click a chromosome in the Karyoview to select it.
- Press Ctrl + Left/Right Arrow keys to move between chromosomes.
- To jump to chromosome 1, press Ctrl+Home
- To jump to chromosome Y (last chromosome in the Karyoview), press the Ctrl+End.



**Figure 6.2 Karyoview controls**

Using the mouse, click and drag on a selected chromosome to select an area for display in the Detail View. This area is highlighted in the Selected Chromosome View.



**Figure 6.3 Karyoview controls**

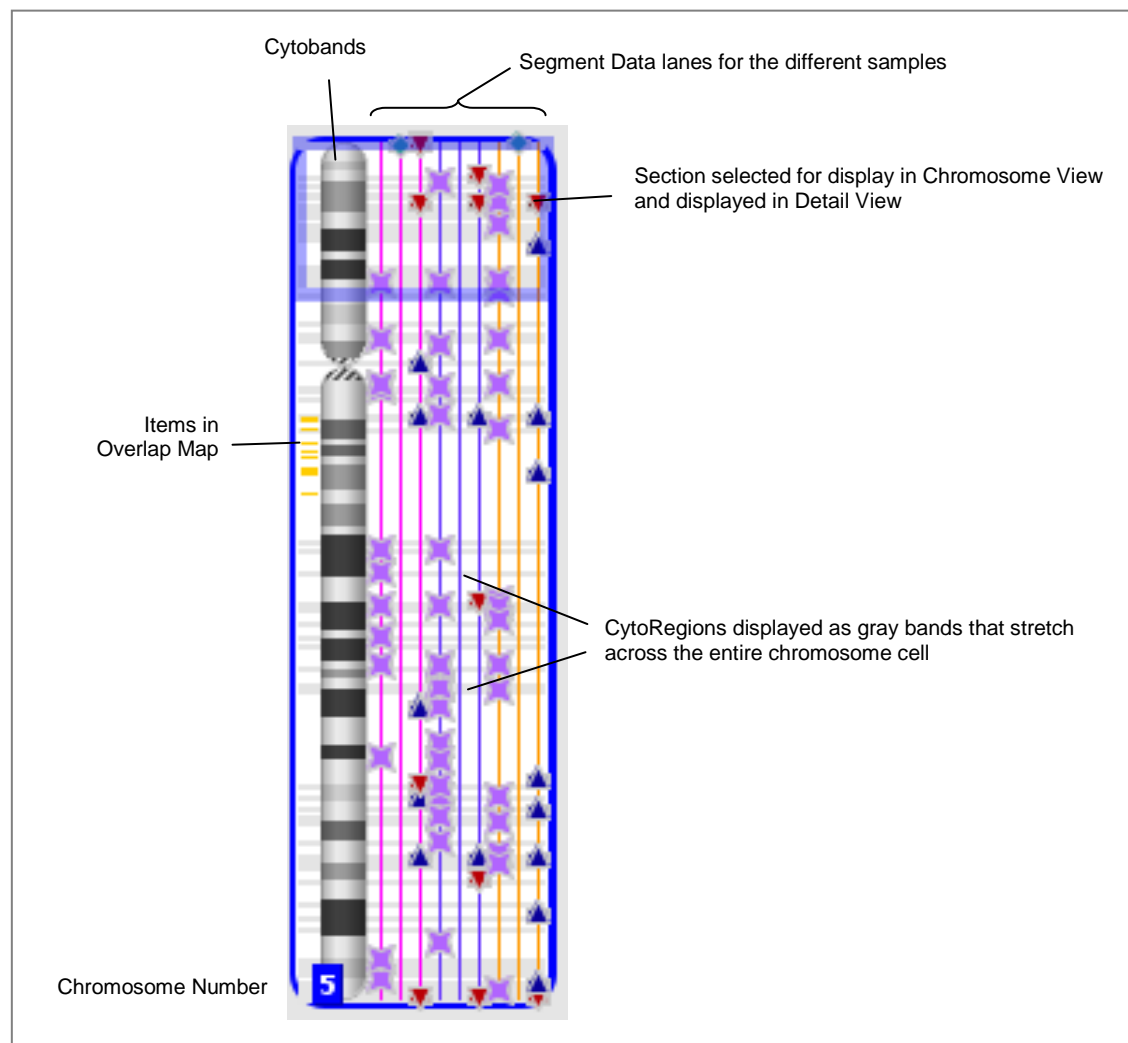
You can use the Stretch Slider and Vertical Scroll Bar to zoom in on a section of the Karyoview. You can also use the mouse wheel as shown below:

- Alt + mouse wheel stretches the display
- Mouse wheel scrolls up and down

The following information is displayed for each chromosome:

- Chromosome number

- Cytobands
- Segment Data, with separate lanes for each sample file and each displayed segment type. Each sample is assigned a unique color in the display that is used for the lane.  
You can:
  - [Change the grouping](#) of samples and segment types (page 110).
  - Change the position of [samples](#) (page 107) and [data types](#) (page 109) in a group.
- CytoRegions in selected CytoRegion File
- Items in selected Overlap Map file



**Figure 6.4 Detail of Karyoview**

Click on a chromosome in the Karyoview to select it for display in the Selected Chromosome View and the Detail View. The selected Chromosome is highlighted in the Karyoview.

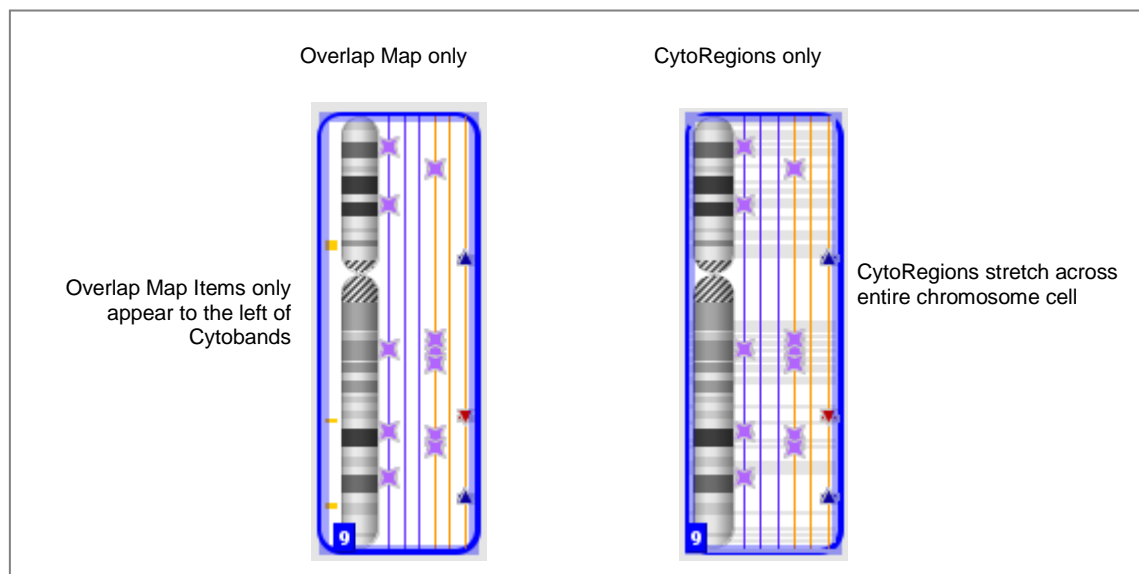
The Stretch Slider and the vertical scroll bar controls the vertical stretch of the Display area.

At higher magnifications, more details of the cytobands are displayed in the Karyoview. You can see cytoband labels if the display has room.



The portion of the chromosome selected in the Selected Chromosome View and displayed in the Detail View is highlighted in the Karyoview.

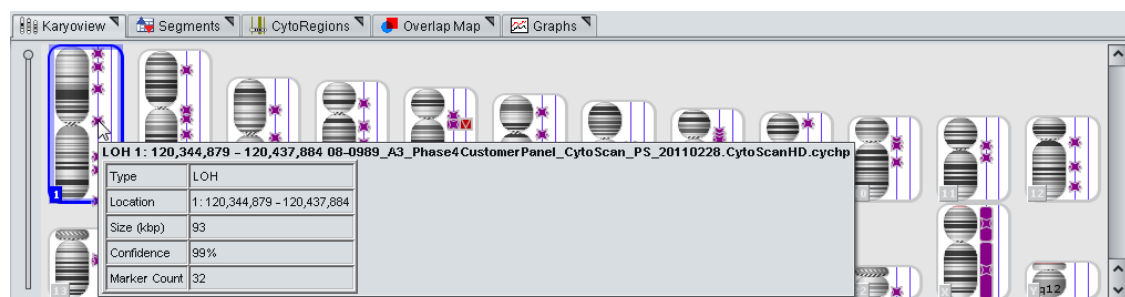
If you have selected a CytoRegions file, the cytoregions are displayed as gray bands that stretch across the entire chromosome cell, from right to left of the cytobands.



**Figure 6.5 Display of Overlap Map items and CytoRegions in Karyoview**

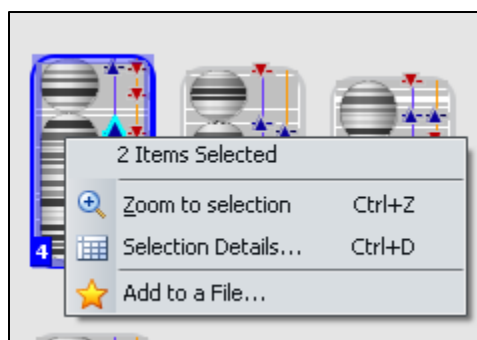
If you selected an Overlap Map file, the overlap map items are displayed as small rectangles to the left of the cytobands. They are displayed in a different color from the CytoRegions bands.

You can mouse over a feature in the Karyoview, Selected Chromosome View, or Detail View to display a pop-up with information about the feature.



**Figure 6.6 Karyoview pop-up**

You can right-click a feature in the Karyoview or Chromosome Display to open a shortcut menu of options.

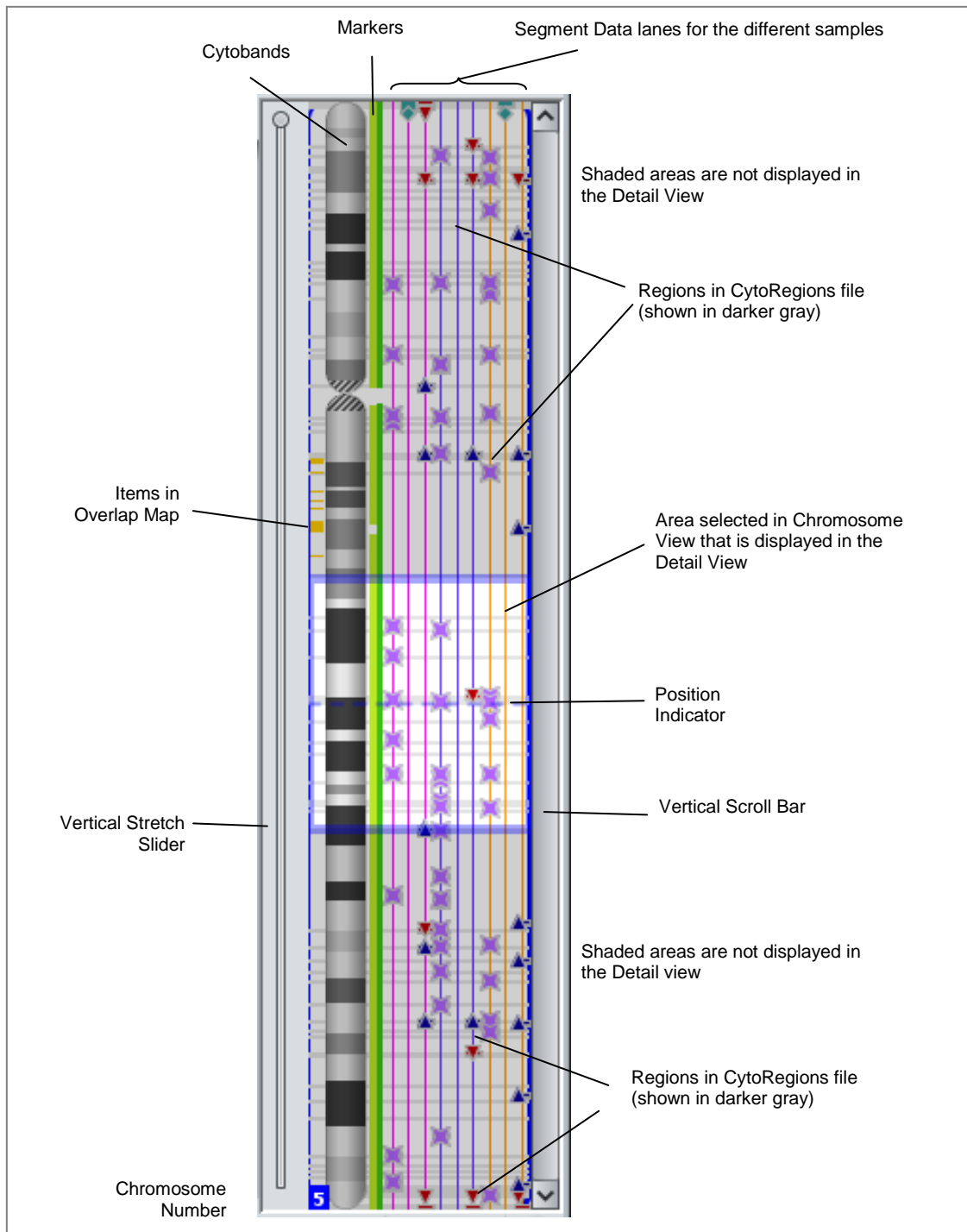


**Figure 6.7 Karyoview shortcut menu**

See [Learning More About Features](#) (page 122) for more information.

## Selected Chromosome View

The Selected Chromosome View is similar to the Karyoview, but it displays a single selected chromosome at higher magnification. Click and drag in the Chromosome View to select an area for display in the Detail View.

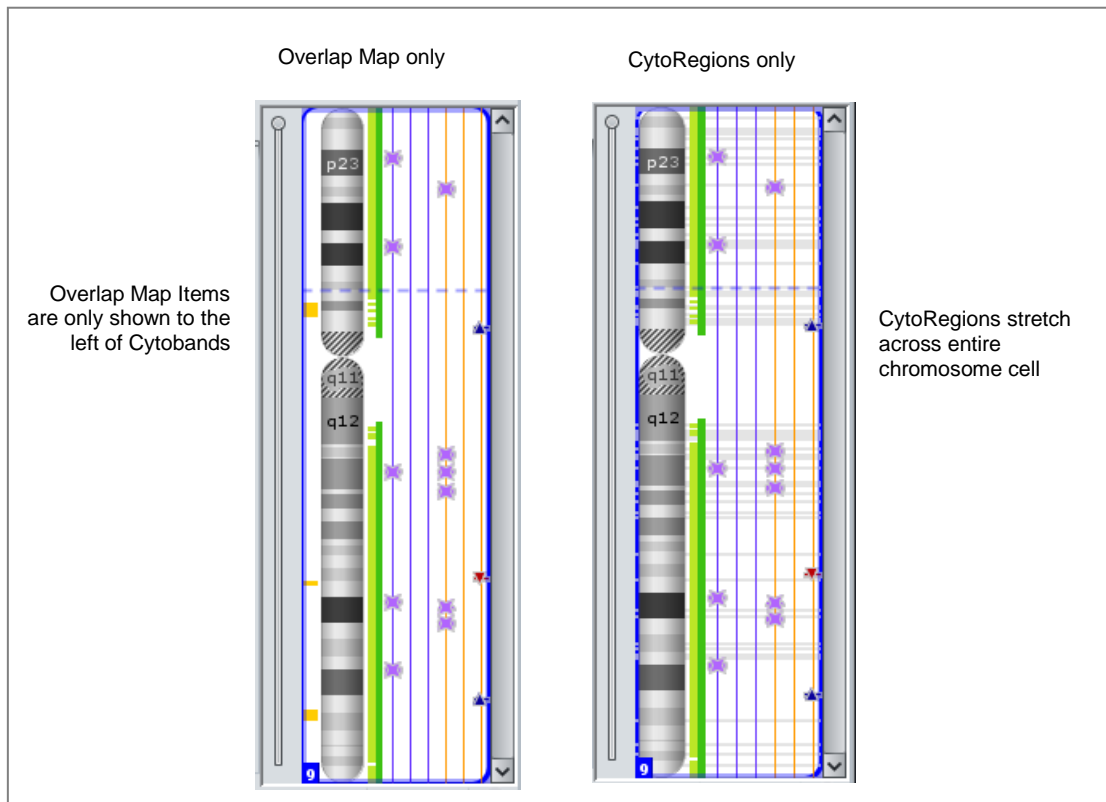


**Figure 6.8 Selected Chromosome View**

You can use the Stretch Slider and Vertical Scroll Bar to zoom in on a section of the Selected Chromosome View. You can also use the mouse wheel as shown below:

- Alt + mouse wheel stretches the display

If you have selected a CytoRegions file, the cytoregions are displayed as gray bands that stretch across the entire chromosome cell, from right to left of the cytobands.

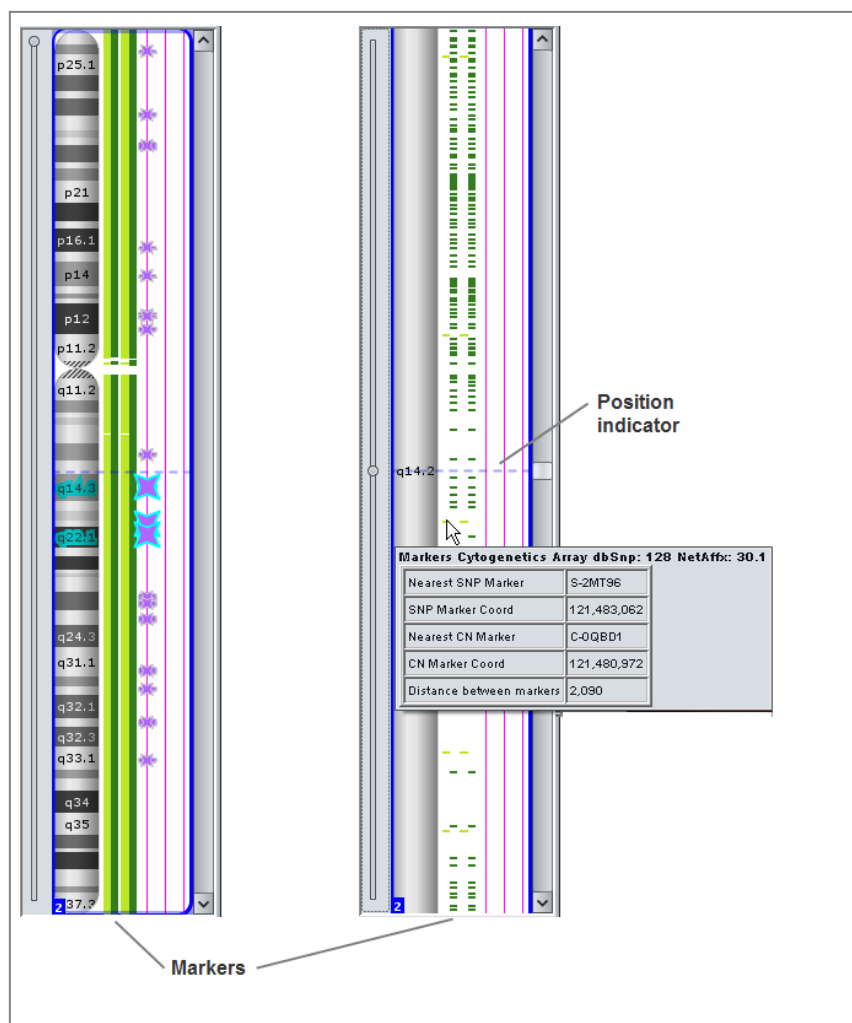


**Figure 6.9 Display of Overlap Map items and CytoRegions in Selected Chromosome View**

If you have selected an Overlap Map file, the overlap map items are displayed as small rectangles to the left of the cytobands. They are displayed in a different color from the CytoRegions bands.

The Position Indicator is a dashed horizontal blue line. Click in the Selected Chromosome View to set the position of the indicator. The position is highlighted in the graphs table and used as the center point when zooming.

The marker positions are displayed to the right of the cytobands. When zoomed out, they appear as green ribbons. When zoomed in, the markers and their positions can be seen.



**Figure 6.10 Selected Chromosome View, markers in zoomed-out view (left) and zoomed-in view (right)**

SNP markers are displayed in the light green band nearest the cytobands. The SNP marker/probe names in the CytoScan™ HD Array and the Cytogenetics Whole-Genome 2.7M Array start with the letter 'S'.

Copy Number markers are displayed in the dark green band nearest the detected segments. The non-polymorphic copy number probe names on the CytoScan™ HD Array and the Cytogenetics Whole-Genome 2.7M Array start with the letter 'C'.

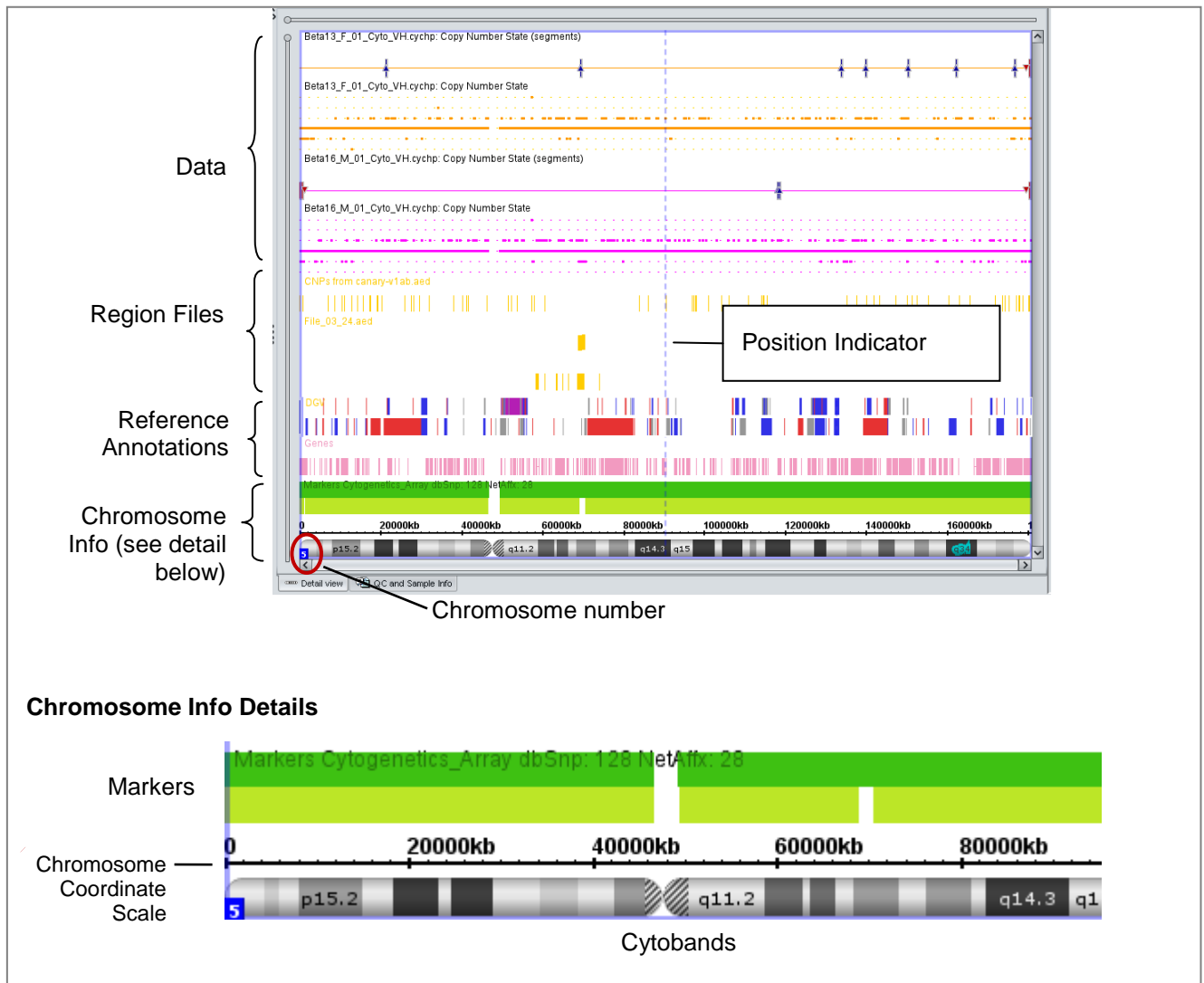
You can mouse over a marker to learn more about it.

Segments selected in any view are highlighted in the Selected Chromosome View.

For information on the other features of the Selected Chromosome View, see [Karyoview](#), page 90.

## Detail View

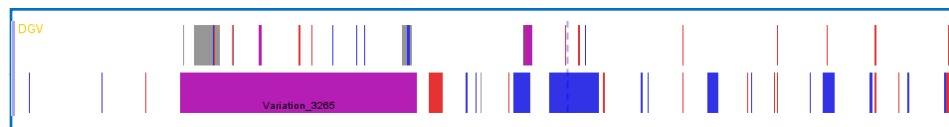
The Detail View enables you to look in detail at the detected segments, marker data, regions, and reference annotations for the loaded files.



**Figure 6.11 Detail View**

## Annotation Color Codes

In the Detail View, DGV annotations are color-coded to indicate association with gain or loss.



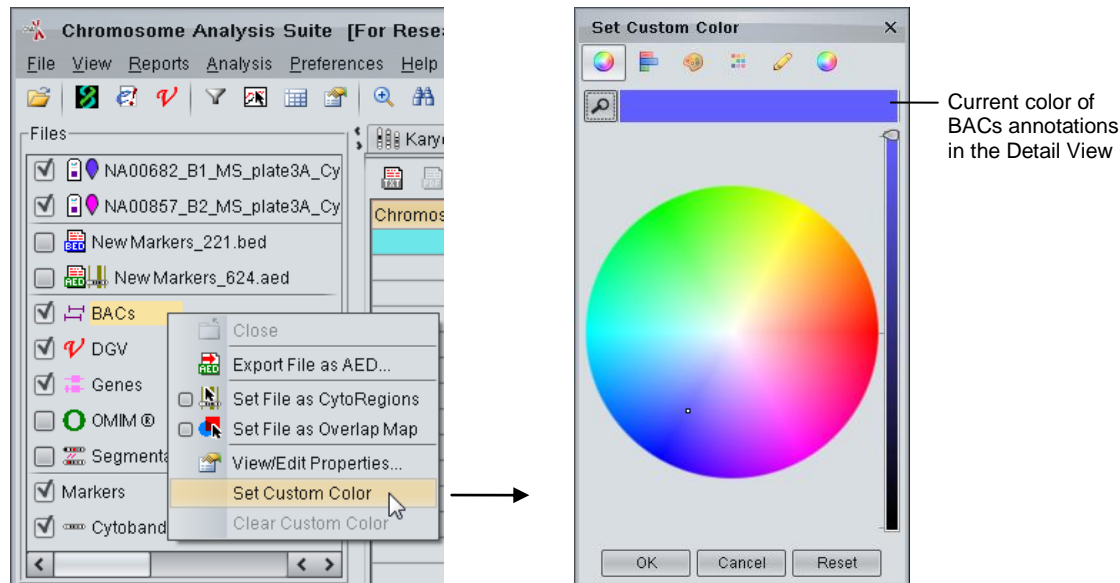
- Purple – Gain and loss are associated with the region
- Red – Only loss is associated with the region
- Blue – Only gain is associated with the region

- Gray – Copy number variation is associated with the region, but information regarding the number of times gains or losses were observed is not present in the annotation record in the DGV database

The color codes for DGV annotations cannot be modified. However user-selected colors can be applied to the other types of annotations.

#### To change an annotation color:

1. Right-click an annotation type in the Files windowpane and select **Set Custom Color** on the shortcut menu (Figure 6.12).



**Figure 6.12 Open the color palette**

2. Specify a color for the selected annotation type using the color controls in the color palette, and click **OK**.  
The new color is applied to the annotations in the Details View.
3. To return to the default annotation color, right-click the annotation in the Files windowpane, and select **Clear Custom Color** on the shortcut menu.


#### Data in the Detail View

The Detail View displays the following kind of data for CytoScan™ HD or Cytogenetics Array data (CYCHP):

**Table 6.1 Data for CytoScan™ HD Array (CYCHP)**

Data Types	Definition
<b>Detected Segment Types</b>	
<b>Gain</b>	Amplifications or duplications
<b>Loss</b>	Hemizygous or homozygous deletions
<b>LOH</b>	Loss of Heterozygosity
<b>Probe array data (displayed as graph data)</b>	

<b>Copy Number State</b>	HMM-derived integer Copy Number State
<b>Log2 Ratio</b>	Per marker log2 ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.
<b>Weighted Log2 Ratio</b>	Contains the Log2 ratios processed through a Bayes wavelet shrinkage estimator. These processed values are input to the CNState algorithm HMM.
<b>LOH</b>	Loss of Heterozygosity
<b>Allele Peaks</b>	Filtered and smoothed values for individual markers. Nonparametric estimation is used to understand possible regional peak structure towards which the data is smoothed. The amount of filtration and smoothing is dynamically adapted based on sample quality. Allele difference is computed based on differencing A signal and B signal, then standardizing based on reference file information.
<b>Genotype Calls</b>	SNP genotype calls (single sample, BRLMM-P-plus algorithm)
<b>Smooth Signal</b>	Smoothed Calibrated Copy Number Estimate

 **Note:** Starting with the NetAffx annotation version NA30.2, there is a subset of ~55,000 SNP probes which are used for allelic information analysis but which are not used for Copy Number analysis, for the Cytogenetics Array. For these SNP probes, LOH and Allele Peaks data will be displayed, but these SNP probes will not have Log2 Ratio, Weighted Log2 Ratio, SmoothSignal, or Copy Number State data displayed, nor will they be used for ascertainment of Mosaicism. The calculation of Segment data for all the various Segment types takes this change into account. All non-polymorphic (copy number) and the vast majority of SNP probes will NOT be affected by this change, and will continue to display all previous data points from the Cytogenetics Array CYCHP files.

**Table 6.2 Data for Cytogenetics Whole-Genome 2.7M Array (CYCHP)**

<b>Data Types</b>	<b>Definition</b>
<b>Detected Segment Types</b>	
<b>Gain</b>	Amplifications or duplications
<b>Loss</b>	Hemizygous or homozygous deletions
<b>Mosaicism</b>	Regions of Non-Integer Copy Number Gain or Loss (between CN State 1 and 3).
<b>LOH</b>	Loss of Heterozygosity
<b>Probe array data (displayed as graph data)</b>	
<b>Copy Number State</b>	HMM-derived integer Copy Number State
<b>Log2 Ratio</b>	Per marker log2 ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.
<b>Weighted Log2 Ratio</b>	Running median of a genomic window of log2 ratios that filters outliers.



<b>LOH</b>	Loss of Heterozygosity
<b>Allele Peaks</b>	Filtered and smoothed Standardized Centered Allelic Ratio (SCAR) values for individual markers. Nonparametric estimation is used to understand possible regional peak structure towards which the data is smoothed. The amount of filtration and smoothing is dynamically adapted based on sample quality.
<b>Smooth Signal</b>	Smoothed Calibrated Copy Number Estimate

The Detail View displays the following kind of data for Genome-Wide SNP Array 6.0 Array data (CNCHP):

**Table 6.3. Data for Genome-Wide SNP Array 6.0 (CNCHP)**

<b>Data Types</b>	<b>Definition</b>
<b>Detected Segment Types</b>	
<b>Gain</b>	Amplifications or duplication
<b>Loss</b>	Hemizygous or homozygous deletions
<b>LOH</b>	Loss of Heterozygosity
<b>Probe array data (displayed as graph data)</b>	
<b>Copy Number State</b>	HMM-derived integer Copy Number State
<b>Log2 Ratio</b>	Per marker log2 ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.
<b>LOH</b>	Loss of Heterozygosity
<b>Allele Difference</b>	Difference of A signal and B signal, each standardized with respect to their median values in the reference.
<b>Smooth Signal</b>	Smoothed Calibrated Copy Number Estimate
<b>LOH</b>	Loss of Heterozygosity

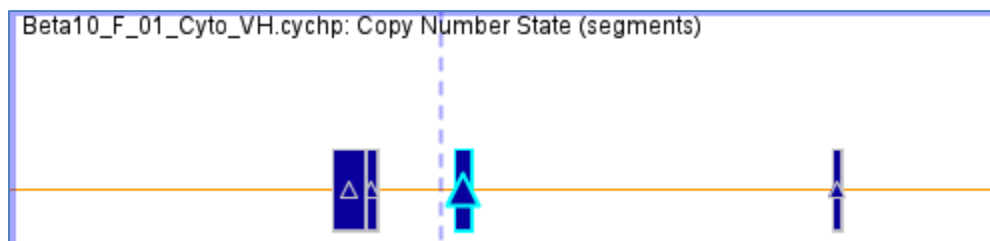
See Changing Graph Appearance for more information about controlling the display of graph data.

In addition, the Detail View displays:

- **Regions:** Features in the various region files loaded into ChAS, including CytoRegions and Overlap Map items.
- **Annotations:** Indicate the known or suspected locations of features, such as mRNAs, exons, FISH/BAC clones, structural variants, and so forth.  
You can expand or contract the annotations. See [Expanding and Contracting Annotations](#) (page 115).
- **Chromosome info, with:**
  - Coordinate scale

- Marker position information
- Chromosome number
- Cytoband information

Selected segments are displayed with enlarged icons; selected regions or annotations are outlined and highlighted.



**Figure 6.13 Selected Segment**

### Navigating the Detail View



**Figure 6.14 Detail View navigation controls**

The Detail View has the following navigation controls:

<b>Chromosome Coordinates scale</b>	Shows the position along the genome.
<b>Zoom Slider</b>	Controls the horizontal zoom and the area of the chromosome displayed.
<b>Stretch Slider</b>	Controls the vertical stretch of the Display area.
<b>Scroll bars</b>	Used to select the area displayed after zooming or stretching the vertical or horizontal scale.

<b>Position Indicator</b>	Dashed vertical blue line. Click in the view to set the position of the indicator
	The position that is highlighted in the graphs table.
	The position that is used as the center point when zooming.

You can use the Stretch Slider and Vertical Scroll Bar to zoom in on a section of the Detail. You can also use the mouse wheel as shown below:

- Alt + mouse wheel stretches the display
- Ctrl + mouse wheel zooms in on the horizontal scale
- Mouse wheel scrolls up and down

### Selecting a Chromosome Section for Display

Data from the same sample files is displayed in all three views, displayed at different scales.

You can select a particular chromosome, or a section of the chromosome, for detailed study using:

- [Karyoview and Selected Chromosome View](#), page 103
- [Coordinate Range Box](#), page 104
- [Zoom to a Selected Item](#), page 105
- [Zoom Slider and Horizontal Scroll Bars](#), page 102

You can also double-click on an item in a table to zoom to the region of the chromosome where that item is located.

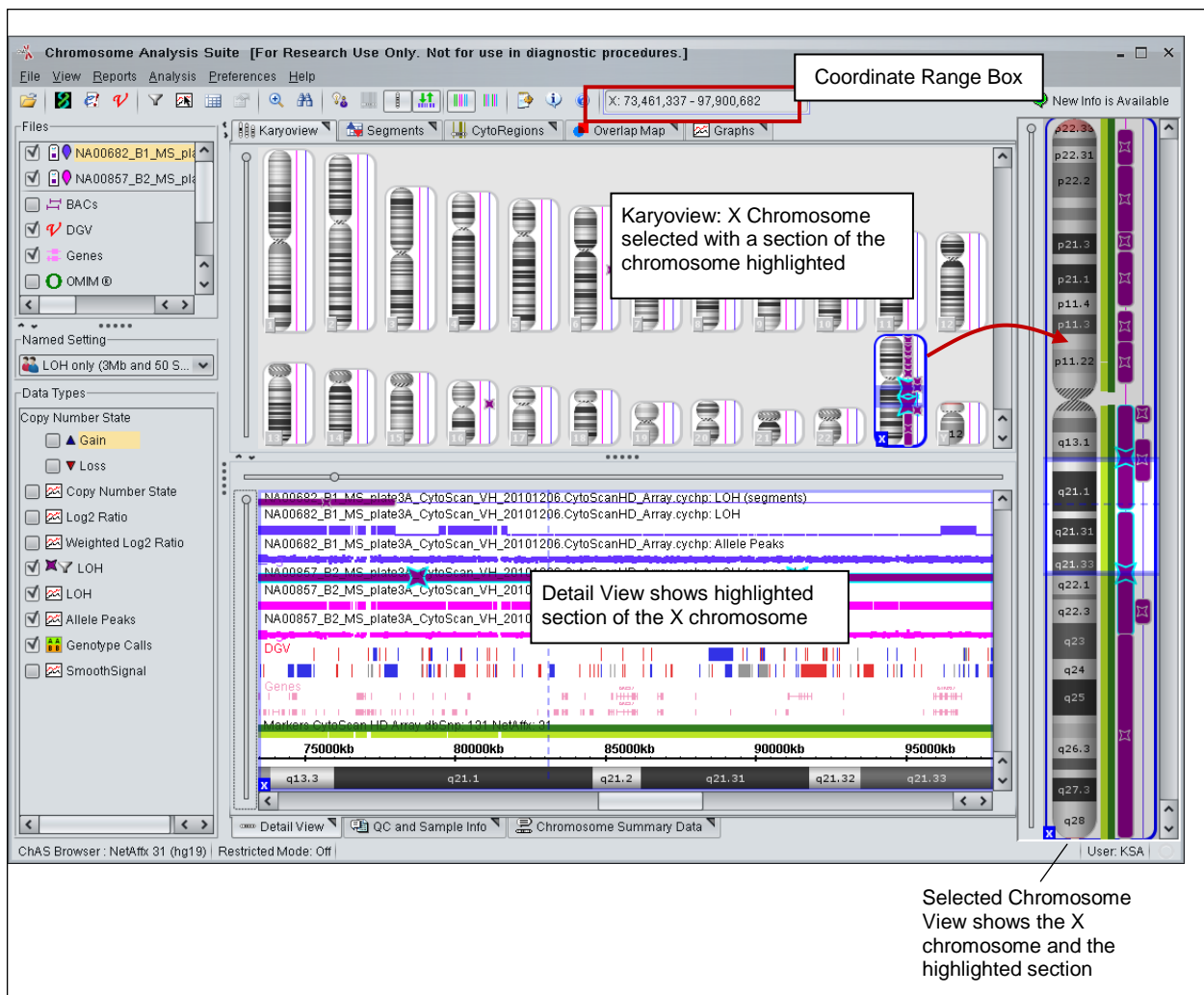
### Karyoview and Selected Chromosome View

#### To select a chromosome for detailed examination:

- Click a chromosome in the Karyoview.
- The chromosome is displayed in the Selected Chromosome View and the Detail view.

#### To examine a section of the chromosome:

- Click and drag on the section in the Karyoview or the Selected Chromosome View.
- The selected section is displayed in the Detail View.



**Figure 6.15 Areas displayed in Karyoview, Selected Chromosome View, and Detail View**

### Coordinate Range Box

The Coordinate Range box is located in the ChAS main toolbar. It shows the selected chromosome and the start and stop positions displayed in the Detail View. You can enter coordinates in the box to update the Detail View.



**Figure 6.16 Coordinate Range Box in Main toolbar**

**To go to a specific coordinate or coordinate range:**

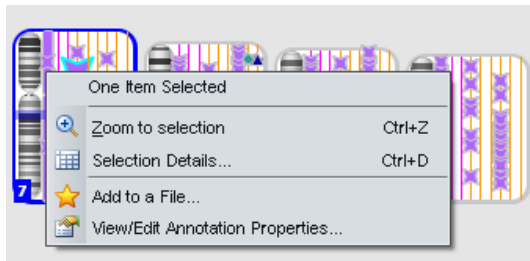
- Enter the desired location in any one of these formats then press the <Enter> key:
  - “**chromosome number : start - end**”: sets the view to the given start and end coordinates on the given chromosome.

- “**start : end**” or “**start - end**”: sets the view to the given start and end coordinates of the current chromosome.

### Zoom to a Selected Item

There are several ways to zoom in on a feature.

- Click a segment in the Segment table, CytoRegions table, or Overlap Map
- Double-click a feature in the Karyoview or selected chromosome view
- Double-click an annotation in the Detail view
- In the Karyoview, Selected Chromosome View, or Details View, you can use the “Zoom to selection” option in the feature right-click menu to go to the start and stop coordinates of the selected feature.



**Figure 6.17 Right-click menu options for a feature**

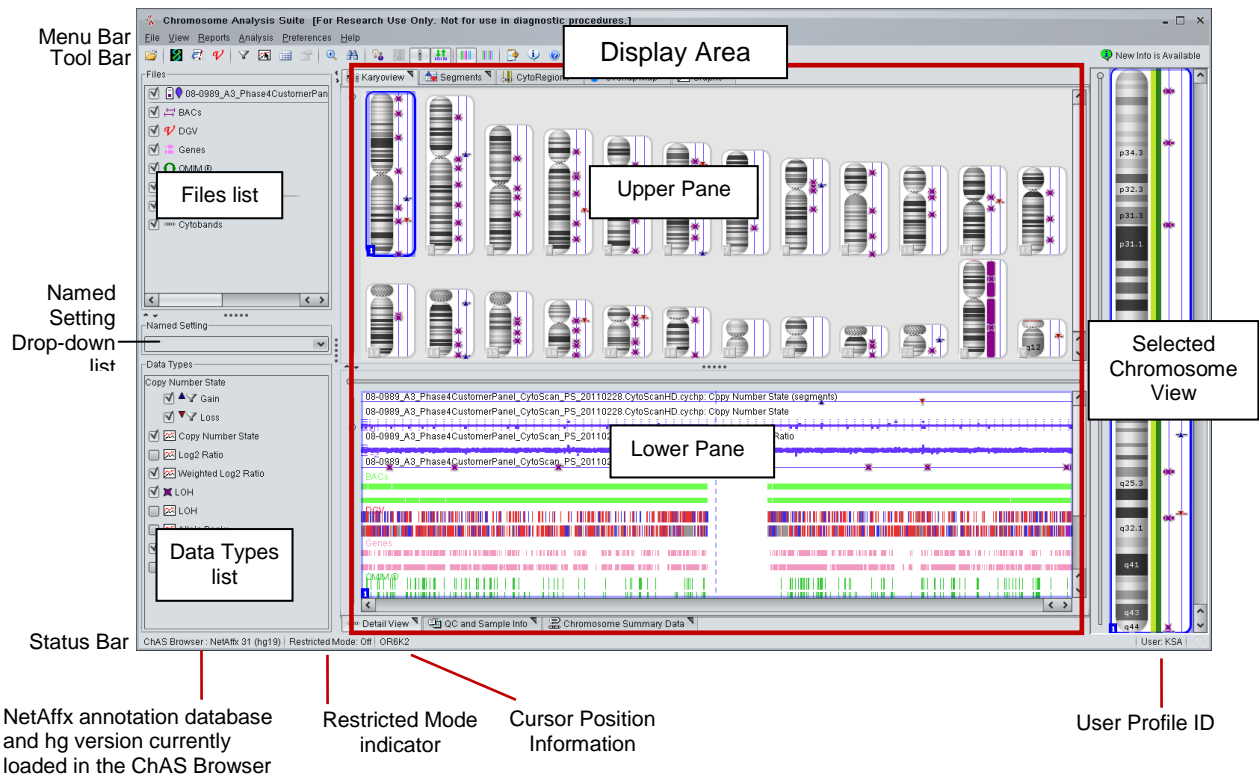
- Use the [Zoom slider and Horizontal Scroll Bars](#) to control the Detail View.

## Controlling the Display of Data

Chromosome Analysis Suite provides controls for:

- [Selecting Data for Display](#) (page 107)
- [Selecting Data Types for Display](#) (page 109)
- [Changing the Grouping of Samples and Data Types](#) (page 110)
- [Selecting Light or Dark Schemes for Display](#) (page 115)
- [Expanding and Contracting Annotations](#) (page 115)
- [Changing Graph Appearance](#) (page 116)

Later chapters explain other options for filtering data, how to specify certain regions for extra attention or ignoring, and how to create Region files with region information and annotations.



**Figure 6.18 ChAS with data loaded**

## Selecting Data for Display

The Files list allows you to select sample data, region files, and reference annotations for display in the graphic view.

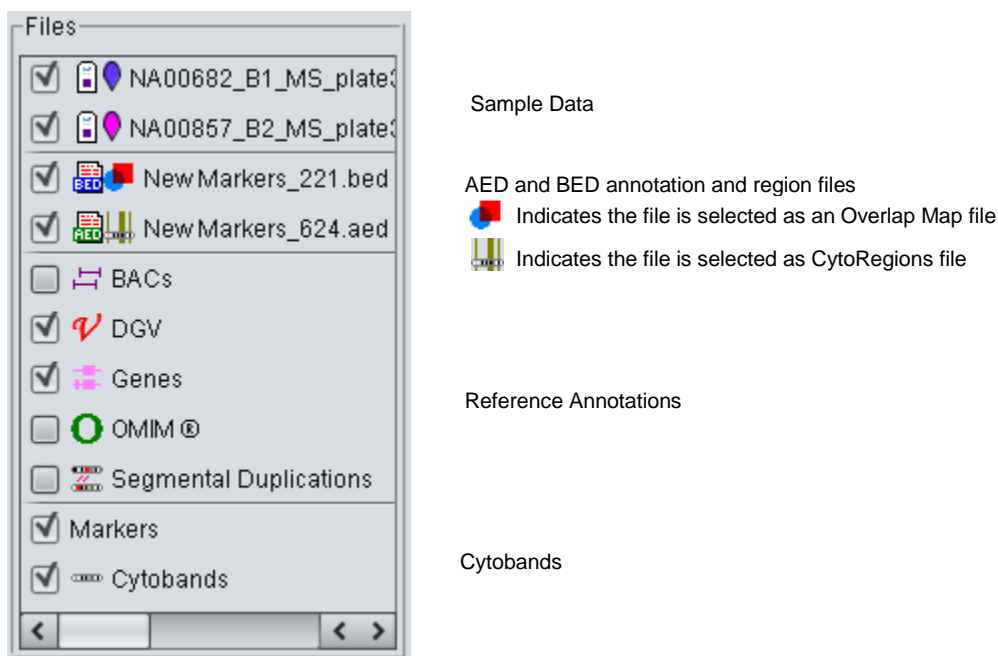








Figure 6.19 Files list

The Files list displays the files grouped by:

- **Sample Data**  
Colored nibs  display the color used for the data lanes for that sample in the Karyoview, Selected Chromosome View, and Detail View.  
If a loaded file has a QC parameter that is out of range, an alert symbol  appears next to the file name.
- **Region Data Files**  
Icons indicate the file type (  AED or  BED) and whether the loaded files have been selected as a CytoRegions file  or Overlap Map file .
- **Reference Annotations:** Loaded during software installation and startup. Only displayed in Detail View.
- **Cytobands:** Separated from other reference annotations because they cannot be moved in the displays, and because they are also displayed in the Karyoview and Selected Chromosome View.

### To select and deselect files for display:

- Click in the checkbox next to the file name.

The order in the Files list determines the order of display of the lanes in the Karyoview, Chromosome view, and Detail View.

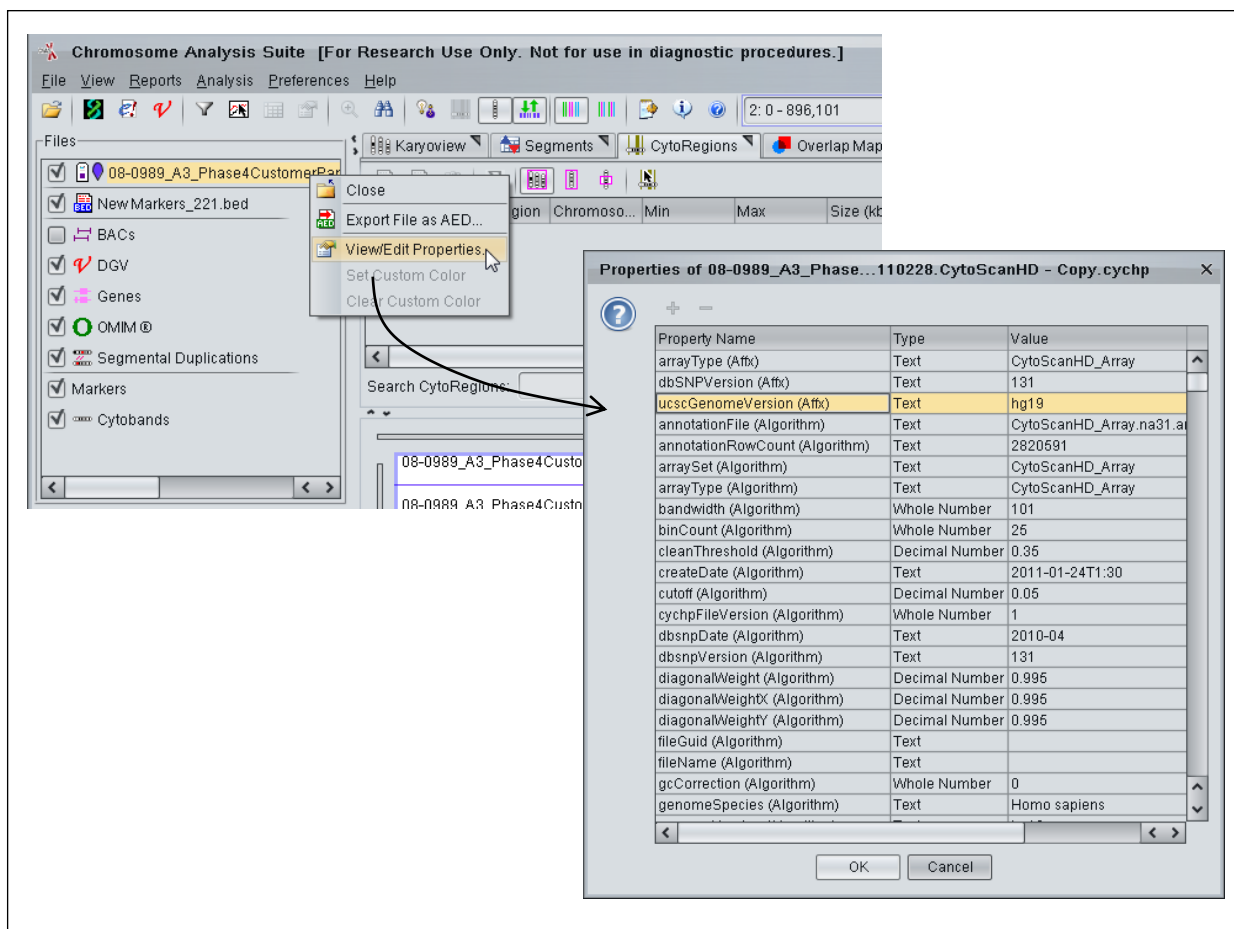
### To change the order of the Sample lanes or reference annotations:

- In the Files list, click a file name and drag it to a new position.

## To view data properties:

1. Right-click a file and select **View/Edit Properties** on the shortcut menu.

The Properties dialog box appears.



## 6.20 View data properties

### To close a file:

1. Right-click on the file you wish to close.
2. Select **Close** from the menu.

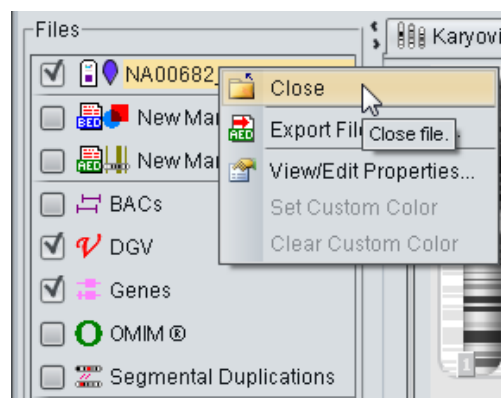



Figure 6.21 Closing a file

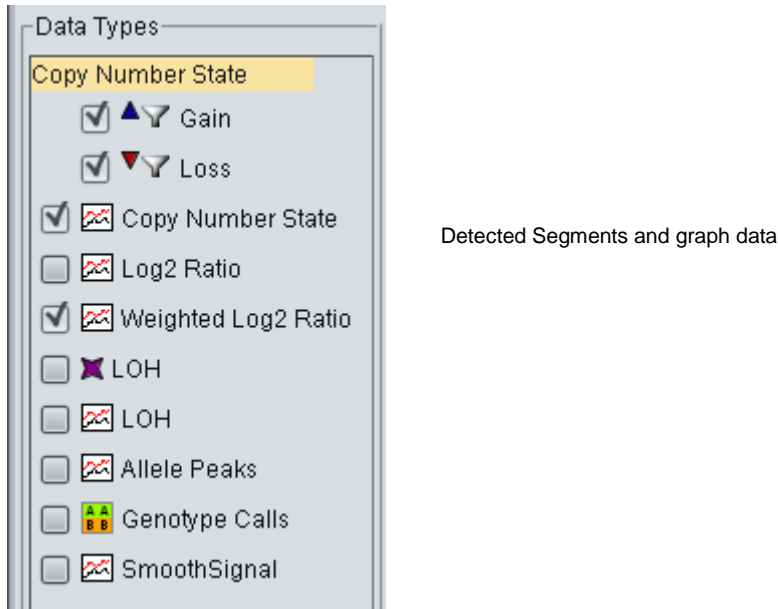


The file is removed from the Files list and the data is no longer displayed.

 **Note:** If you change the smoothing or joining parameters, the new rules are applied to the original, raw segments.

### Selecting Data Types for Display

The Data Types list shows the data types that can be displayed in ChAS.




**Figure 6.22 Data Types list**

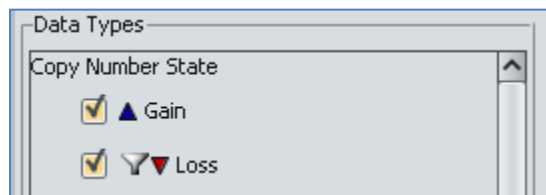
It displays a list of data that can be displayed in the Karyoview, Selected Chromosome View, Detail View, and tables. The available data types may vary, depending upon the type of sample data available.

It allows you to select from Segments data and CN/LOH discrete graph data.

The Segments data is displayed in:

- Karyoview
- Selected Chromosome View
- Detail View

If segment parameter filters have been applied to a segment type, a funnel symbol  appears next to the segment type name.



**Figure 6.23 Funnel symbol for Loss segments**

The Graph data is displayed only in the Detail View.

**To select and deselect data types for display:**

- Click in the checkbox next to the Data Type name.

**To change the order of the data types:**

- In the Data Types list, click a file name and drag it to a new position.

The selections made here can be saved with a Named Setting (see [Creating and Using Named Settings](#), page 228).

**To turn the symbols used for segments on or off:**

- From the View menu, select or deselect Segment Symbols.

**Changing the Grouping of Samples and Data Types**

A unique color is assigned to each sample and used for the lanes in the Karyoview, Selected Chromosome View, and Detail View.

Each segment type is assigned to its own lane and has its own symbol.

You can group the lanes by:

- [Sample](#), page 111
- [Segment and Graph Type](#), page 113

This allows you to do different types of comparisons between samples and segment types.

**To change the grouping:**

- From the View menu, select **Group by Sample** or **Group by Type**; or  
In the Toolbar, click the **Group by Sample**  or **Group by Type**  button.

## Lanes Grouped by Sample

When the lanes are grouped by sample, the different segment types for each sample are kept together in the Karyoview and Selected Chromosome View (see Figure 6.24) and in the Detail View (see Figure 6.25)

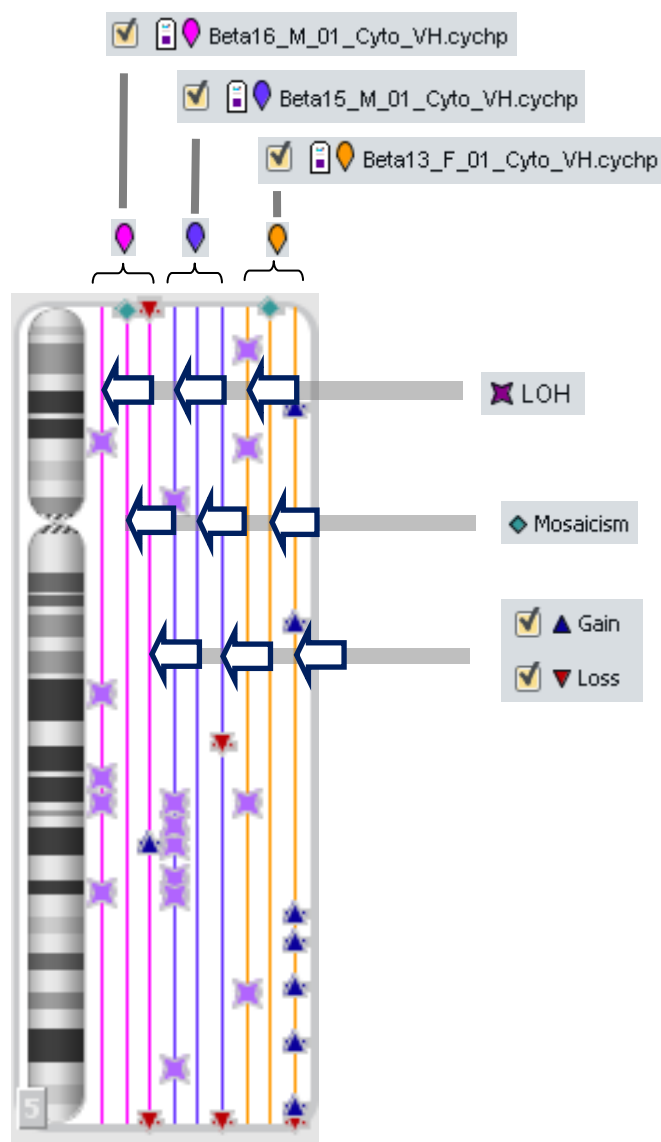


Figure 6.24 Lanes grouped by sample in Karyoview

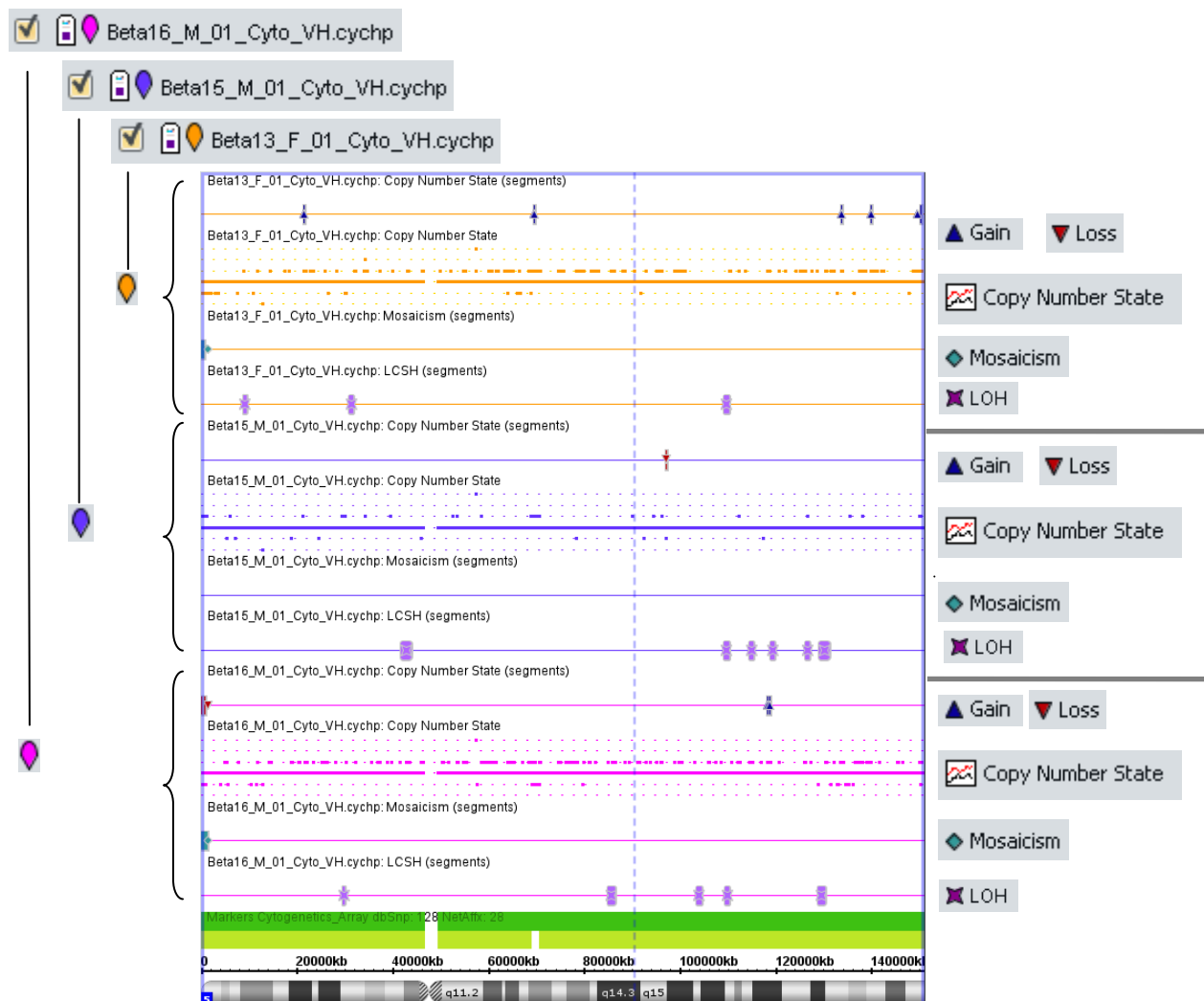


Figure 6.25 Lanes grouped by Sample in Detail view

### Tracks Grouped by Data Type

When the lanes are grouped by data type, the lanes for different samples are kept together for each segment or graph type in the Karyoview and Selected Chromosome View (see Figure 6.26) and in the Detail View (see Figure 6.27).

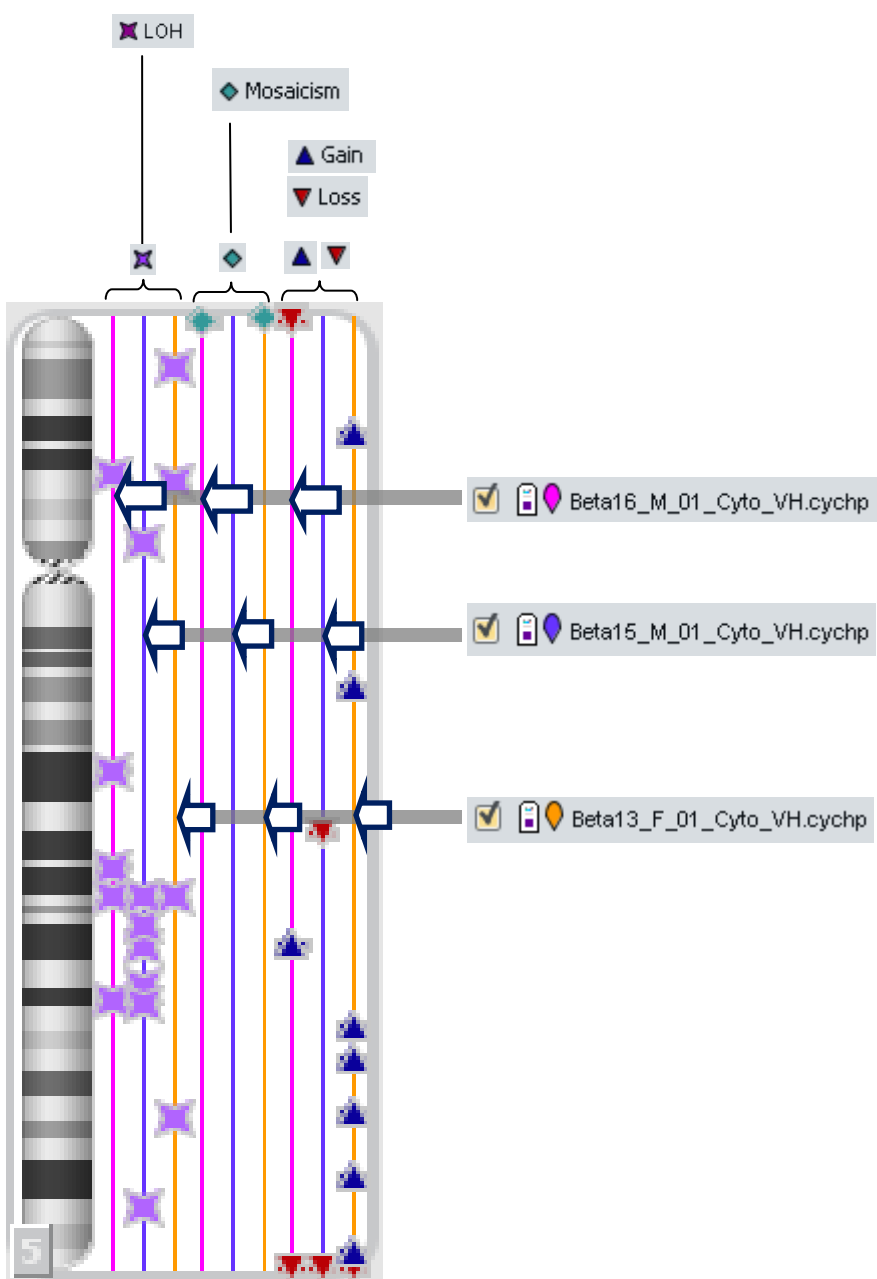
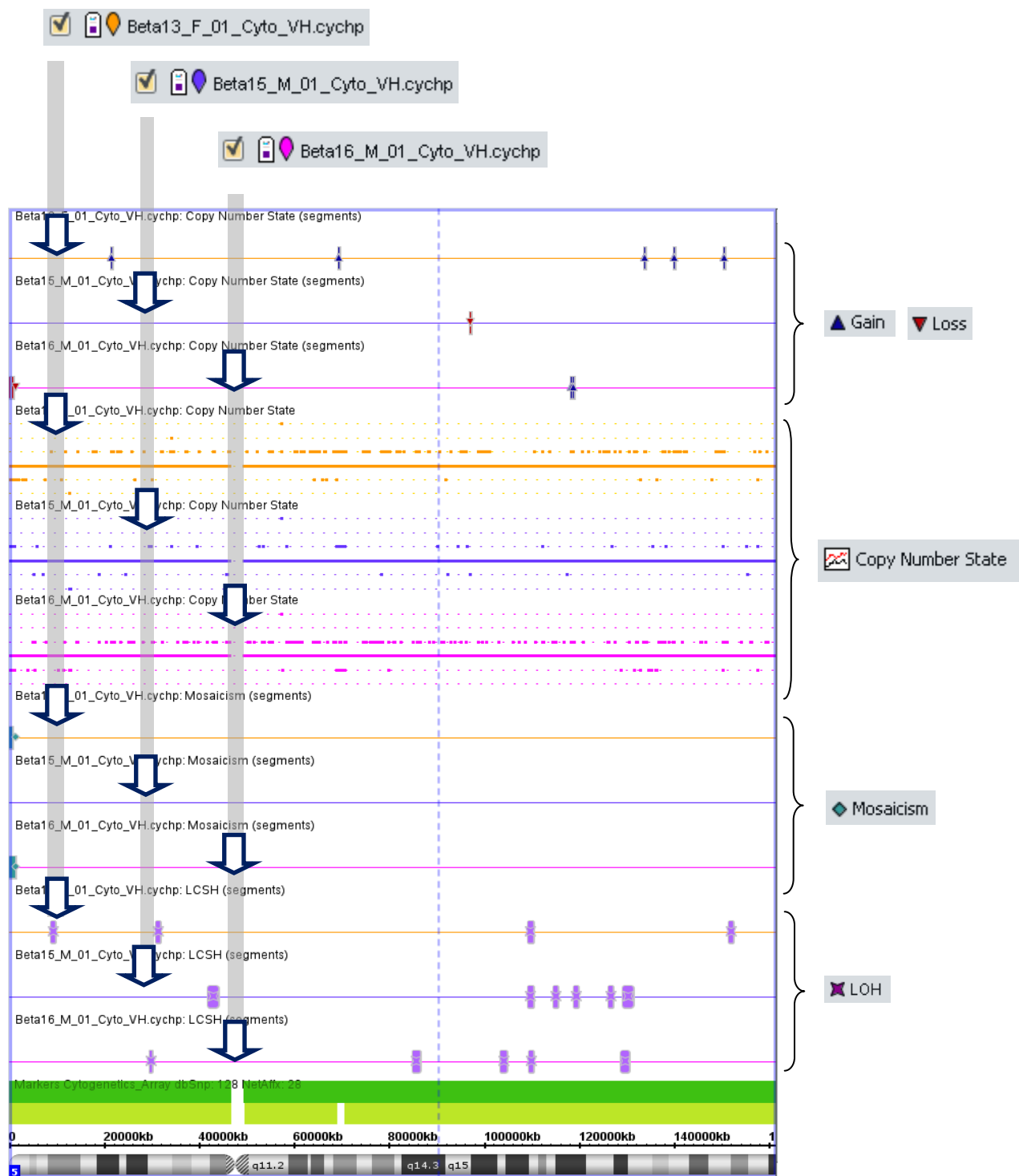


Figure 6.26 Lanes grouped by data type in Karyoview



**Figure 6.27 Lanes grouped by Data Type in Detail View**


You can change the order of samples and Data types in the views by clicking and dragging in the Files and Data Types list.

## Selecting Light or Dark Schemes for Display

You can choose to display the graphics using a light background or a dark background (called Dark Scheme).

To select Dark Scheme for display:

- From the View menu, select **Use Dark Scheme**; or

Click the **Dark Scheme**  button in the Main Tool Bar.

The Dark/Light scheme is used.

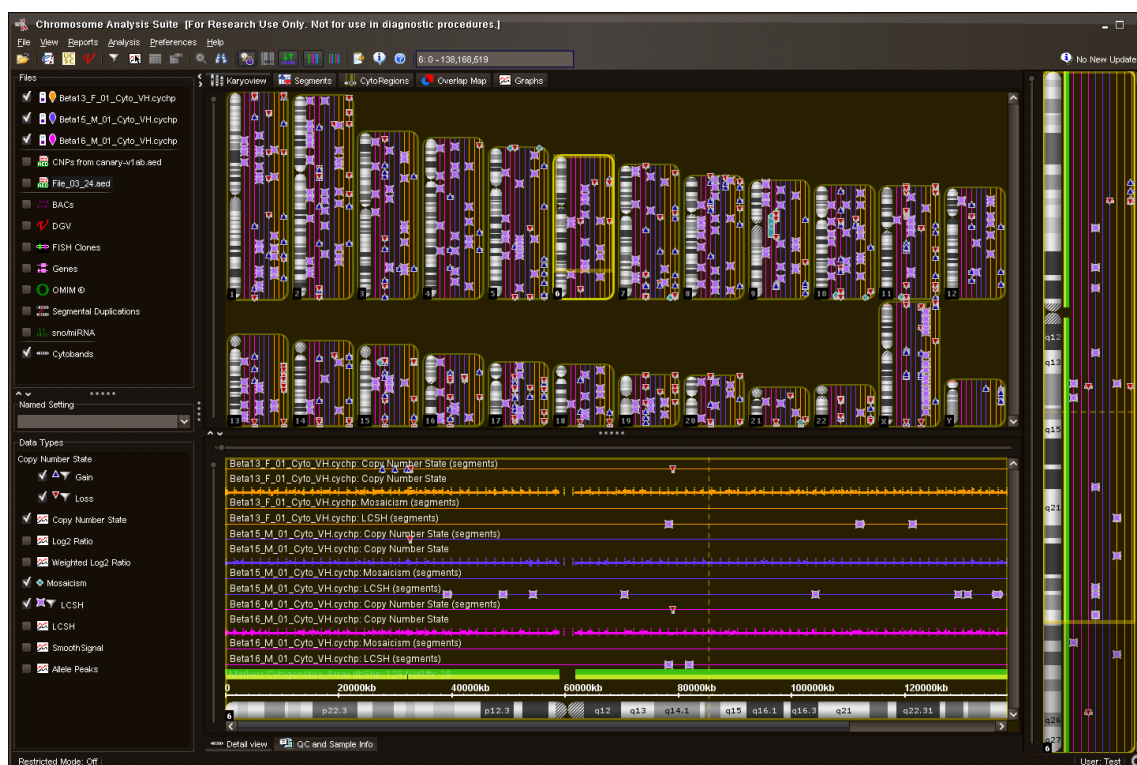


Figure 6.28 ChAS with Dark Scheme selected

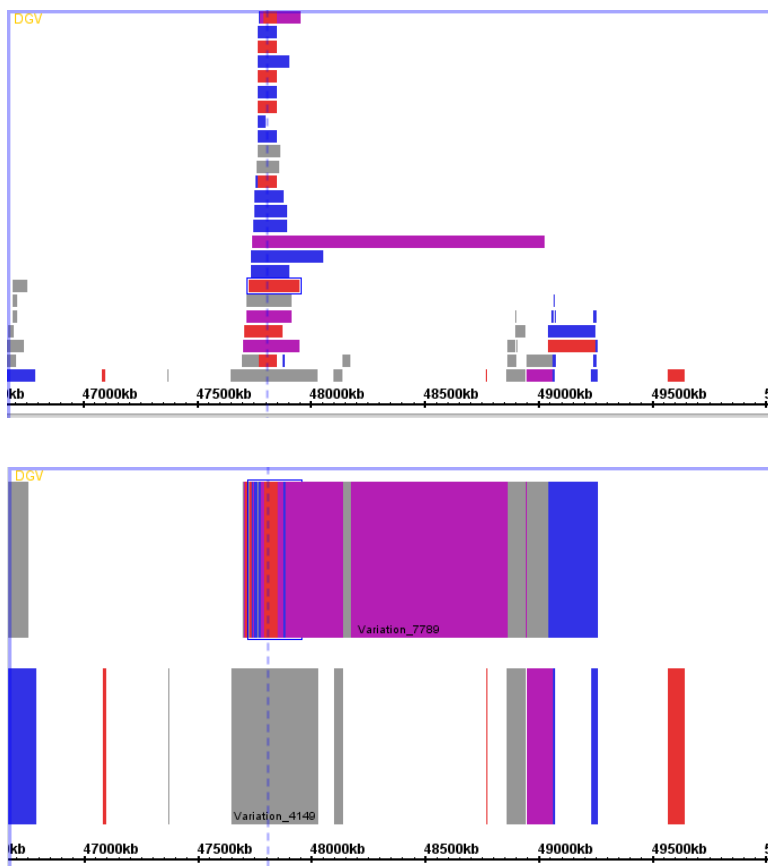
## Expanding and Contracting Annotations

For tracks containing multiple rows of annotations, collapsing tracks consolidates all rows within a track into two rows. Any annotations after the first one will be placed in the second row.

When there are multiple annotations of one type at the same coordinate, the separate annotations will be shown on separate rows.


Collapsing tracks is useful if you don't need to see all the details. However, be aware that in collapsed tracks larger annotations may obscure smaller ones; annotations with introns may be obscured by annotations that don't show the intron.

 **Note:** The maximum number of tracks that can be displayed for any reference annotation is 25.



**Figure 6.29 DGV annotations, collapsed (top) and expanded (bottom)**

**To toggle between collapsed and expanded display of annotations:**

- From the View menu, select **Expand/Collapse Annotations**  
or
- Click the **Expand/Collapse**  button on the main toolbar.

### **Changing Graph Appearance**

You can modify many properties of the graphs in the Detail View. ChAS provides options for:

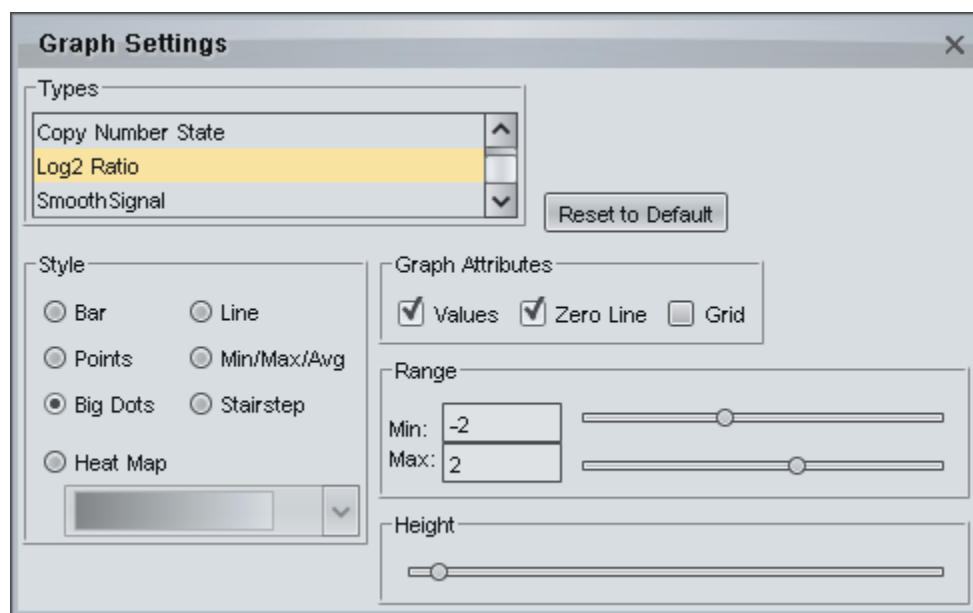
- [Selecting Different Graph Styles](#) (page 117)
- [Changing Graph Attributes](#) (page 120)
- [Changing Scale](#) (page 121)

Settings and adjustments that are specific for graphs can be made using the **Graph Settings** dialog box.

**To open the Graph Settings dialog box:**

- From the View Menu, select **Graph Settings**; or  
Click on the Graph Settings button in the Graphs Tab toolbar.  
The Graph Settings dialog box opens.





**Figure 6.30 Graph Settings dialog box**

The Types box displays the graph data types being displayed in the Detail View.

**To change the settings for a graph type:**

1. Click on a data type to change the settings for that type.
2. Make changes to the graph settings by typing in new values or by operating sliders in the Graph Adjuster panel. For details, see:
  - [Selecting Different Graph Style](#) (below)
  - [Changing Graph Attributes](#) (page 120)
  - [Changing Scale](#) (page 121)

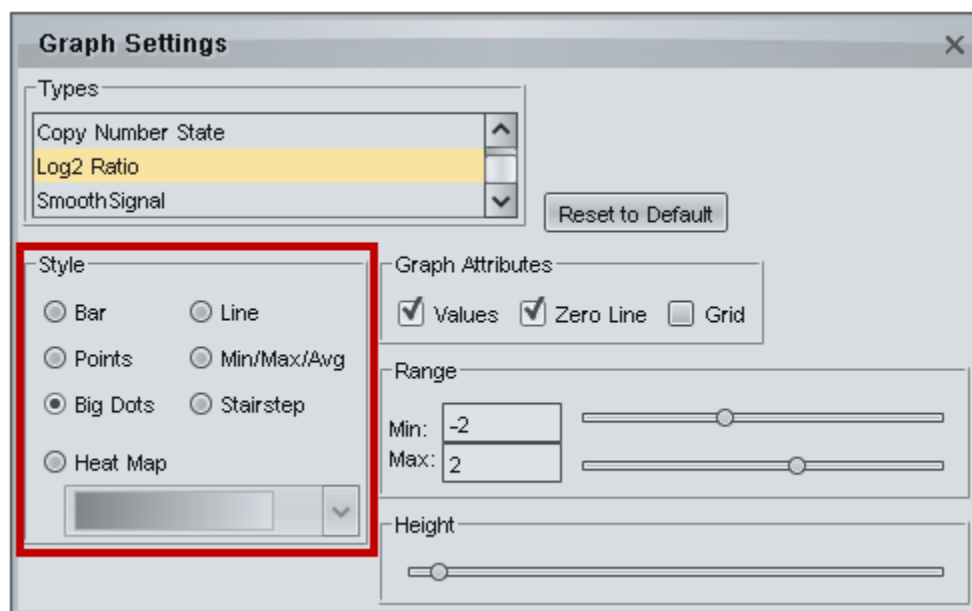
Any changes you make to the values in the **Graph Settings dialog box** will apply to all currently selected graph types.

**Selecting Different Graph Styles**

Graphs can be shown in various representational styles. The type of graph that is most appropriate depends on the type of question being asked about the data. For example, when comparing trends and patterns, it is very useful to use the line graph display method. The user is encouraged to experiment with the different display types to find out which method works best for specific purposes and at specific zoom in magnifications.

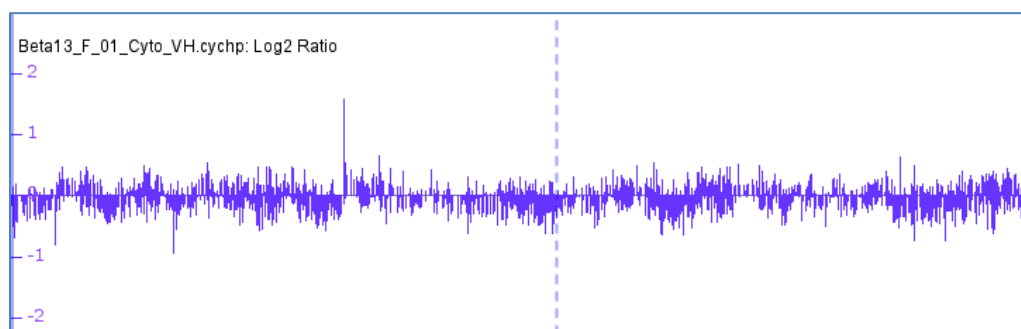
**To change the graph style:**

- In the **Style** section choose one of the options.



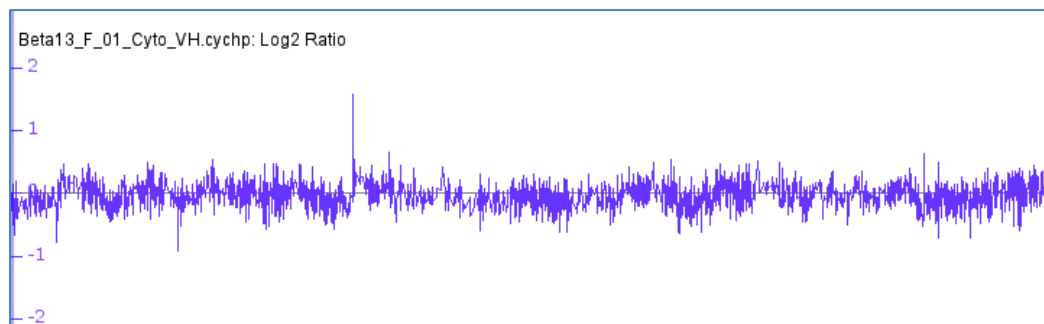
**Figure 6.31 Style options available for the Detail View**

- **Bar** – Individual values are shown as vertical bars that are one base wide for position graphs.



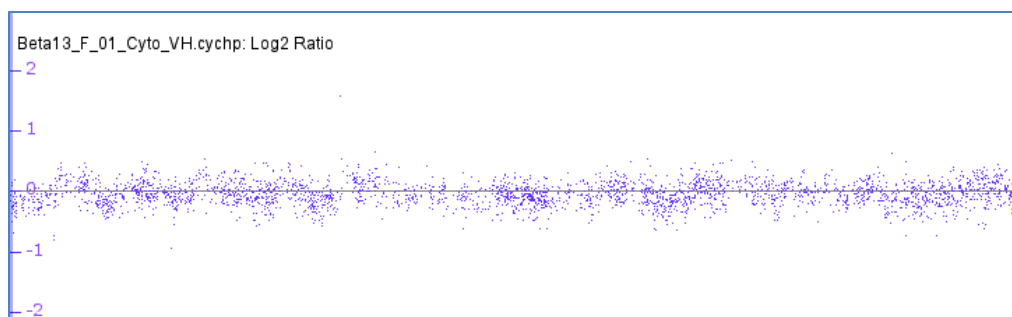
**Figure 6.32 Bar**

- **Line** – Subsequent values are linked with a line. Even if the input file was not sorted, the values will be connected in order along the genomic coordinate axis.



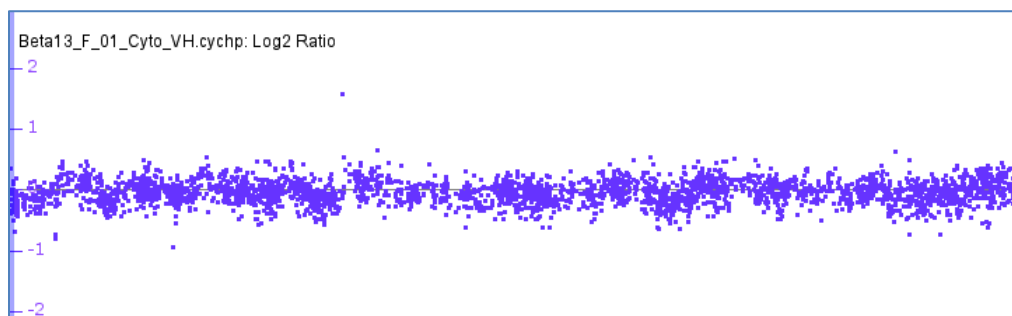
**Figure 6.33 Line**

- **Points** – Shows a single dot for each data value.



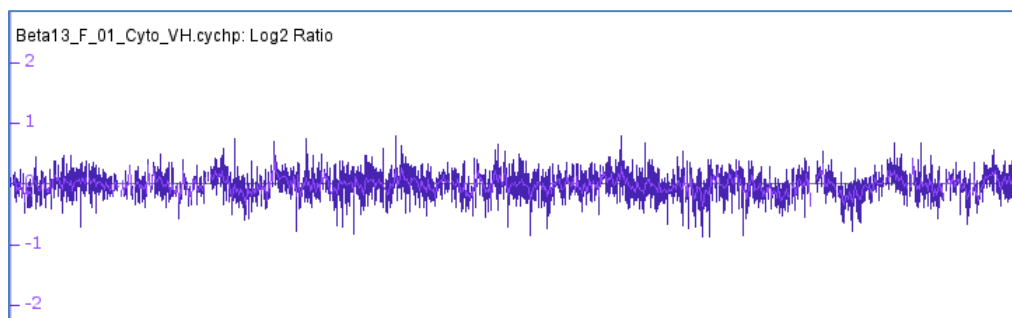
**Figure 6.34 Points**

- **Big dots** – Shows a single big dot for each data value.



**Figure 6.35 Big dots**

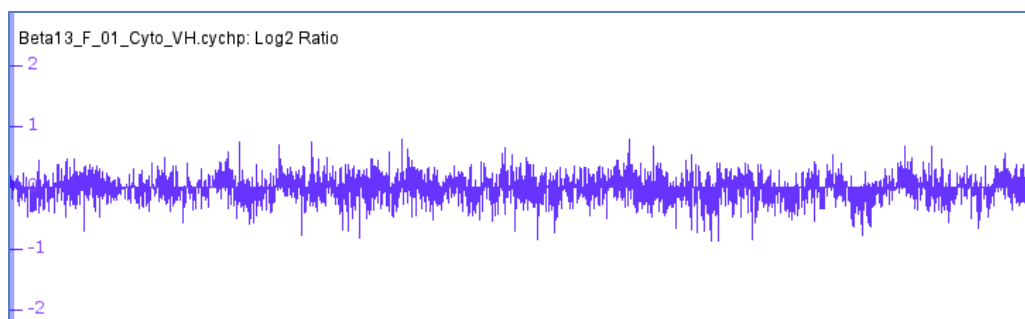
- **Min/Max/Avg** – This style is especially useful for showing very densely populated graphs with data points for large numbers of positions.



**Figure 6.36 Min/Max/Average**

When Detail View is zoomed all the way in, the display is equivalent to the **Line** style. When zooming out, ChAS starts to summarize values. When the scale of the display reaches the point where individual x-values are associated with multiple score values, ChAS picks the maximum and minimum values and draws a vertical bar between them. In addition, ChAS draws lines through the average of all the data points represented at each x value.

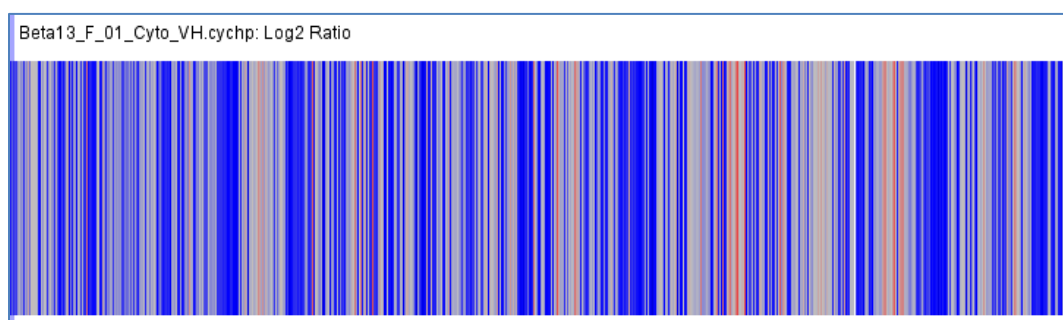
- **Stairstep** – Similar to the bar graph style, except that bar widths along the horizontal axis are stair-stepped.



**Figure 6.37 Stairstep**

For example, if position 100 has a value of 50 and position 200 has a value of 75 and there are no values in between, then ChAS will draw a bar of height 50 that starts at position 100 and stops at position 200. Then, at position 200, ChAS will draw a new bar of height 75 that terminates at the next location with a value.

- **Heat map** – Instead of showing relative intensity via the height of the line at each pixel or coordinate as in most other graph styles, a heat map shows expression levels via color or brightness of the line at each pixel or coordinate. This graph style is useful if you want areas of unusual values to jump out at you. If a graph does not render or is hard to see, adjust the visible bounds of the graph until features are readily visible. Several heat map color schemes are available to choose from.



**Figure 6.38 Heat map**

There are now two Red/Gray/Blue heat maps.

One is designed to look good for copy number data scaled from 0 to 4 (or 1 to 3), with 2 = normal, and the other is designed to look good for copy number data scaled from 0 to 5 with 2 = normal.

The user must be careful to test that the heat map scaling is appropriate for the data.

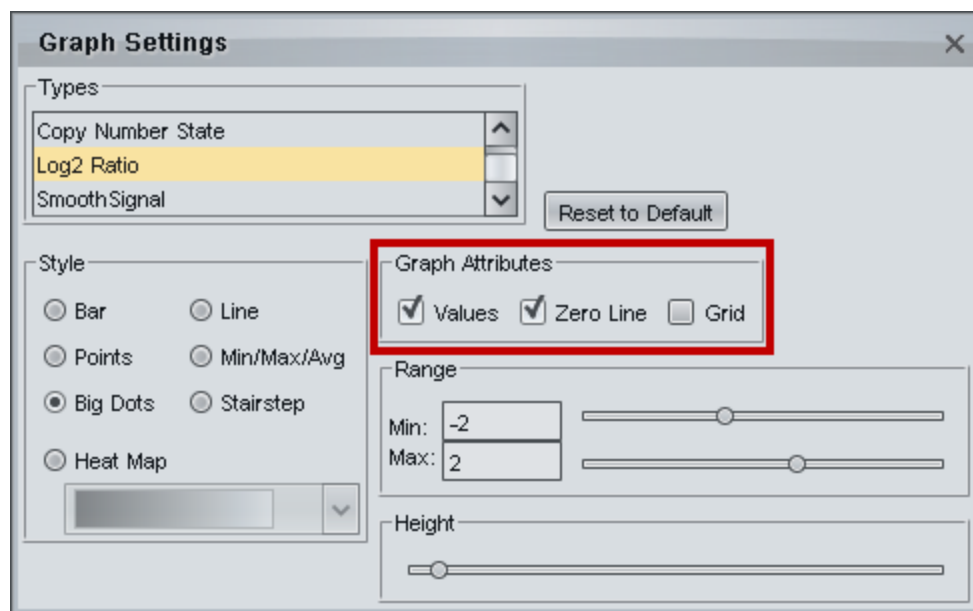
### Changing Graph Attributes

You can display graphs with:

- Value scale
- Zero Line
- Grid

### To change the graph attributes:

- Select the attributes you want to display in the Detail View.



**Figure 6.39 Graph attributes**

### Changing Scale

Changing the visible bounds involves changing the scale of the graph by setting the maximum and minimum values to be displayed.

To set these visible bounds, use the **Range** section of the **Graph Properties** dialog.



### To set specific minimum and maximum values:

- Use the sliders, or type in values to the boxes.

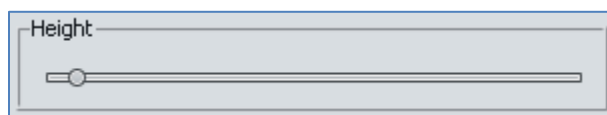
These values will be applied to each selected graph. You are free to set maximum and minimum values that cover a range smaller or larger than the actual range of your data.



**Note:** The algorithm detects CN state up to 5 for Cytogenetics Whole-Genome 2.7M Array CYCHP files or up to 4 for Genome-Wide Human SNP Array 6.0 CNCHP files and CytoScan™ HD Array CYCHP files. In the Graph Settings, the copy number state range can be set to a maximum of 10, however, this does not change the maximum CN detected by the algorithm.

### To change the vertical height of a graph:

- Use the Height Slider to stretch all the graph type in the vertical direction.



The graph height slider is used to increase or decrease the size of a given graph type.

The size is specified in a relative manner. The final graph size will depend on the number of other graphs and annotations being displayed.

## Learning More About Features

You can use the following tools to learn more about features in the different views:

- [Pop-ups](#) (below)
- [Right-click Menu Options](#) (page 123)
- [Selection Details Table](#) (page 125)
- [Going to External Websites](#) (page 128)

### Pop-ups

You can mouse over a feature in any of the views to display a popup box with information on the feature. The information provided depends on the type of data that the mouse arrow is on.

Pop-ups are available for:

- Cytobands
- Detected segments
- Graph data
- Marker position indicators
- Reference annotations



**Note:** you should expand the reference annotations before selecting one to avoid selecting multiple annotations. See [Expanding and Contracting Annotations](#) (page 115).

- Displayed Region files, including
  - Overlap Regions
  - CytoRegions

The information displayed differs depending upon the type of feature selected, as shown in the samples below.

For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see [Appendix E](#), page 253).

You can learn more about the terms used in the pop ups in [Selection Details Table](#) (page 125).

#### To turn pop-ups on or off:

- From the View menu, select **Mouse-over Pop-ups**.

The information can include custom properties created by a user (see [Viewing and Editing Annotations](#), page 159 for more information).

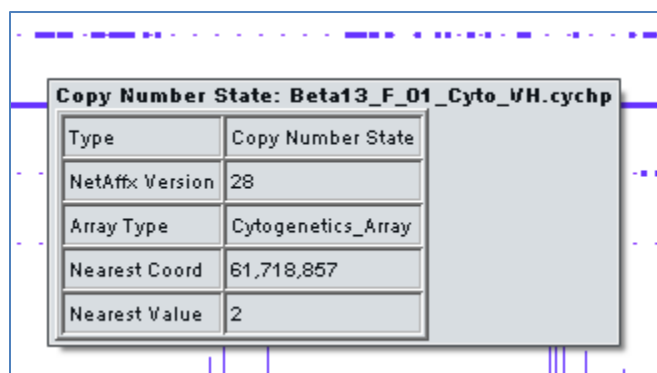


Figure 6.40 Pop-up for CN State Graph

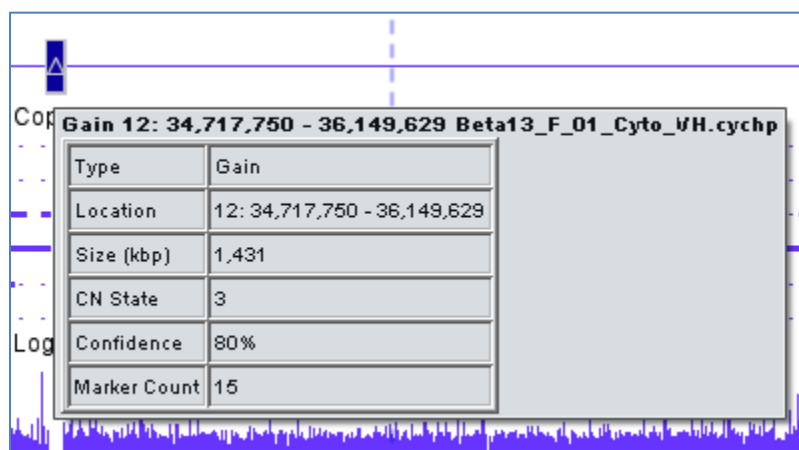


Figure 6.41 Pop-up for a Gain Segment

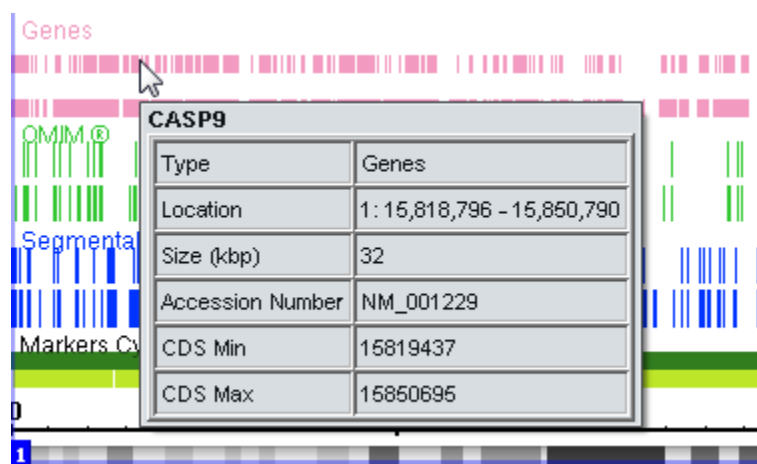


Figure 6.42 Pop-up for a Gene

### Right-click Menu Options

You can right-click on any of the following types of features to open a menu with options for learning more about the feature:


- Detected segments

- Reference annotations, including cytobands

 **Note:** You should expand the reference annotations before selecting one to avoid selecting multiple annotations. See [Expanding and Contracting Annotations](#) (page 115).

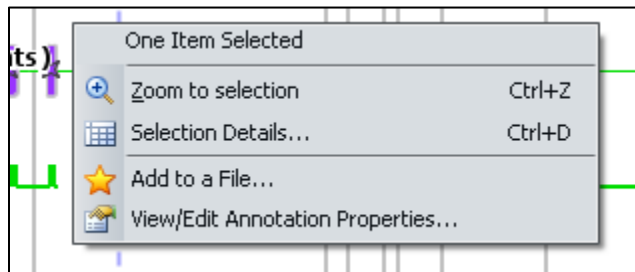
- Displayed Region files, including:
  - Overlap Regions
  - CytoRegions

Not all options are available for the different feature types.

 **Note:** In the Detail View you can select multiple items of different types. The available options will differ, depending upon the number and types you have selected.

Some menu options are common to the different types of features:

- Number of items selected (if more than one item is selected, the options available may differ, depending upon the type and number of items).
- Zoom to selection: See [Zoom to a Selected Item](#) (page 105).  
There are several ways to zoom in on a feature, see page 105).
- Selection Details: Opens the Selection Details box with information about the feature. See [Selection Details](#) (page 125).
- Add to a file: Add the selected segment, annotation, or region to a region file. See [Adding Regions to an Existing AED File](#) (page 156).
- View/Edit Annotation Properties: displays the Annotations Properties dialog box for the selected feature. You may or may not be able to edit the properties. See [Viewing and Editing Annotations](#) (page 159).

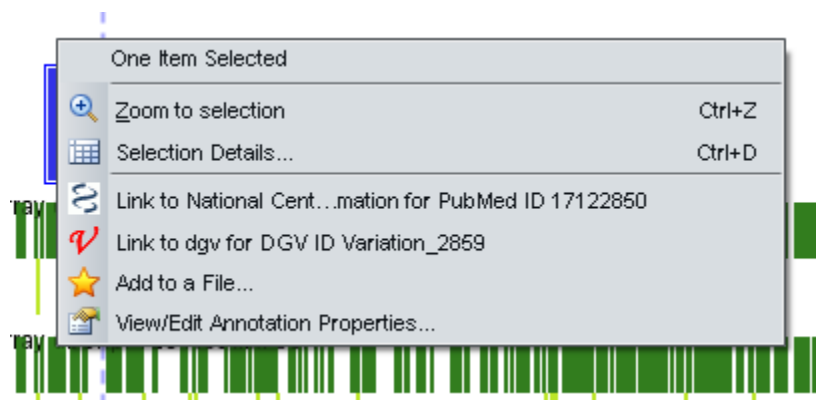


**Figure 6.43 Segment right-click menu**

Special options for annotations include:

- Link to remote web site for more information.

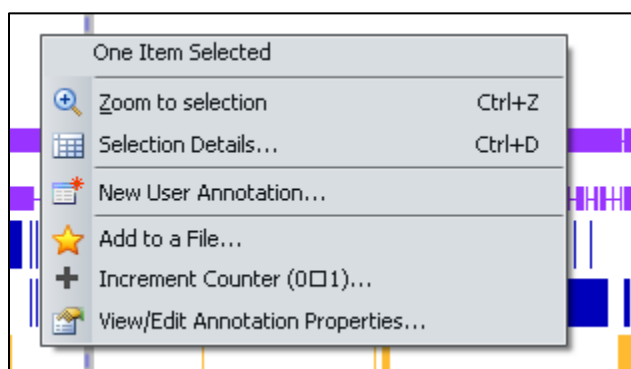




**Figure 6.44 Annotation right-click menu**

Custom functions for regions include:

- New User Annotation: Opens the New User Annotation dialog box, which allows you to add user annotation to a region. See [New User Annotations](#) (page 167).
- Increment counter: Increments the counter in the annotation properties, allowing you to track the number of times a feature has been seen. For more information about the annotation properties, see
- [Viewing and Editing Annotations](#) (page 159).



**Figure 6.45 Region right-click menu**

## Selection Details Table

The Selection Details table displays information available for items selected in the graphic display views (Karyoview, Chromosome View, and Detail View). It is accessed by right-clicking on an item in one of the views and selecting **Selection Details**.

The information is presented in two tables:

- Upper table shows one item per row – Provides summation feature and PDF or tab-separated text file export capabilities.
- Lower table shows one item per column – Provides ability to export the table to a tab-separated text file.

To reorder the columns in the upper table, drag a column header left or right. The corresponding row in the lower table is automatically moved to the new location in the table.

Selection Details

19 results

Annotation	Chromosome	Min	Max	Size (kbp)	Type	DGV ID	Gain	Loss	Frequency	Method
Variation_32	6	81,341,647	81,350,014	8	DGV	Variation_32	0	5		Agilent Cust
Variation_65	6	66,316,761	66,318,755	2	DGV	Variation_65		1		Sequence tr
Variation_75	6	77,496,501	77,510,013	14	DGV	Variation_75	0	17		Agilent 185k
Variation_36	6	83,023,799	83,037,377	14	DGV	Variation_36	1	0		Paired End
Variation_32	6	78,933,954	78,949,535	16	DGV	Variation_32	0	3		Agilent Cust
Variation_23	6	103,842,682	103,870,485	28	DGV	Variation_23	0	1		Paired End
Variation_10	6	104,097,165	104,099,987	3	DGV	Variation_10		1		Illumina Hum

\*\*\*\*\*

	Variation_51944	Variation_44272	Variation_32820	Variation_6500
Min	6	6	6	6
Max	67,075,448	81,682,707	81,341,647	66,316,7
Size (kbp)	67,105,020	81,687,381	81,350,014	66,318,7
Type	DGV	DGV	DGV	DGV
DGV ID	Variation_29774	Variation_51944	Variation_44272	Variation_32820
Gain	1	0	0	0
Loss	0	201	1	5
Frequency				
Method	Illumina HumanHap550 Bead...	Illumina HumanHap550 V1 B...	Illumina DNA sequencing	Agilent Custom CGH Arrays
Total Gain/L...	1	201	1	5
Reference	Jakobsson et al. (2008)	Shaikh et al. (2009)	Bentley et al. (2008)	Perry et al. (2008)
Sample	485 control samples (Human...	2,026 healthy controls	1 male Yoruba individual fro...	30 control samples (HapMap)
Var Type	CopyNumber	CopyNumber	CopyNumber	CopyNumber
PubMed ID	18288195	19592680	18987734	18304495

Click an arrow to hide or show one of the tables. The display choice persists the next time ChAS is started.

Upper table:  
One item per row

Lower table:  
One item per column

Figure 6.46 Example Selection Details table

The information displayed in the Selection Details table includes:

## Common

**Annotation** Identifier for the item.

**Chromosome** Chromosome on which the item is located.

**Min** Zero-based index position of the first base pair in the sequence.

**Max** Zero-based index position of the last base pair in the sequence, plus one. Adding one ensures that the length of any (hypothetical) segment containing a single marker would be one, and ensures that the coordinates match the coordinate system used in BED files.

For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see [Appendix E](#), page 253).

**Size (kbp)** Size of the item.

**Type** Type of segment (Gain, Loss, LOH, Mosaicism) or annotation.

**Strand** The sequence strand of the item.

## Segments

<b>CN state</b>	Copy number state.  The expected copy number state on the X chromosome in males is not constant over its entire length. This is due to the structure of the sex chromosomes. See <a href="#">Copy Number Segments on the X and Y Chromosomes</a> (page 40) for more information.
<b>Mean marker Distance</b>	Mean Marker Distance: the length of the segment divided by the number of markers in the segment.
<b>confidence</b>	Confidence score determined during segment detection, not a p-value, but a measure of confidence that a Segment is truly aberrant (non-normal data). For CN Segments, confidence is determined on a marker by marker basis by evaluating the concordance of the log2ratio at each marker with the copy number state assigned by the HMM. The average confidence score of markers in Gain and Loss segments determines the confidence score of that segment.

## Genes

<b>Accession Number</b>	Unique identifier assigned to the sequence in GenBank.
<b>CDS Min</b>	Minimum position of the coding sequence (BED-style coordinates).
<b>CDS Max</b>	Maximum position of the coding sequence (BED-style coordinates).

## FISH Clones

<b>Labs</b>	Lab where clone was produced.
<b>accNames</b>	Accession associated with the clone.
<b>accCount</b>	Number of accessions associated with the clone.
<b>beCount</b>	Number of BAC end sequences associated with this clone.
<b>placeCount</b>	Number of times FISH'd.
<b>bandEnds</b>	End FISH band.
<b>stsCount</b>	Number of STS markers associated with this clone.
<b>stsNames</b>	Names of STS markers.
<b>placeType</b>	How clone was placed on the sequence assembly.
<b>bandStarts</b>	Start FISH band.
<b>beNames</b>	Accession IDs of BAC ends
<b>DGV</b>	Database of Genomic Variants
<b>PubMed ID</b>	Link to PubMed.
<b>Gain</b>	Number of gains in the sample population.
<b>Loss</b>	Number of losses in the sample population.

<b>Total Gain/Loss</b>	Number of total Gains and Losses observed in the sampled population.
<b>Method</b>	Method used to identify the variant.
<b>Sample</b>	Sample(s) used to identify the variant.
<b>Reference</b>	PubMed ID for paper.
<b>VarType</b>	Variation type.

## OMIM®

<b>OMIM Morbid Map ID</b>	ID number for those OMIM entries in OMIM's MorbidMap
---------------------------	--

## Segmental Duplications

<b>Score</b>	Score based on the raw BLAST alignment score. The score for segmental duplications is set to zero in NetAffx annotation 31 and higher.
<b>FracMatch</b>	The fraction of matching bases.
<b>FracMatchIndel</b>	The fraction of matching bases with Indels.



**Note:** Affymetrix does not generate or verify the information for genes, FISH clones, Segmental Duplications, sno/miRNAs, or DGV annotations.



**Note:** Segmental Duplication and Sno/miRNA annotations do not have any unique terms; but sno/miRNA annotations use the “type” field to indicate subtypes like “cdBOX” and “HAcaBOX”.

Some information may not be displayed, depending upon the feature type. The information can include custom properties created by a user (see [Viewing and Editing Annotations](#), page 159 for more information).

You can export data from the table using the standard table export tools (see [Exporting Table Data](#), page 212).

You can perform multi-column sorts (see [Sorting by Columns](#), page 178).

## Going to External Websites

You can view select an area in the Detail View for viewing at one of the following public sites:

- UCSC
- Ensembl
- Toronto DGV

To view the selected area at a public site:

1. In the Detail View, zoom and scroll to the area of interest.

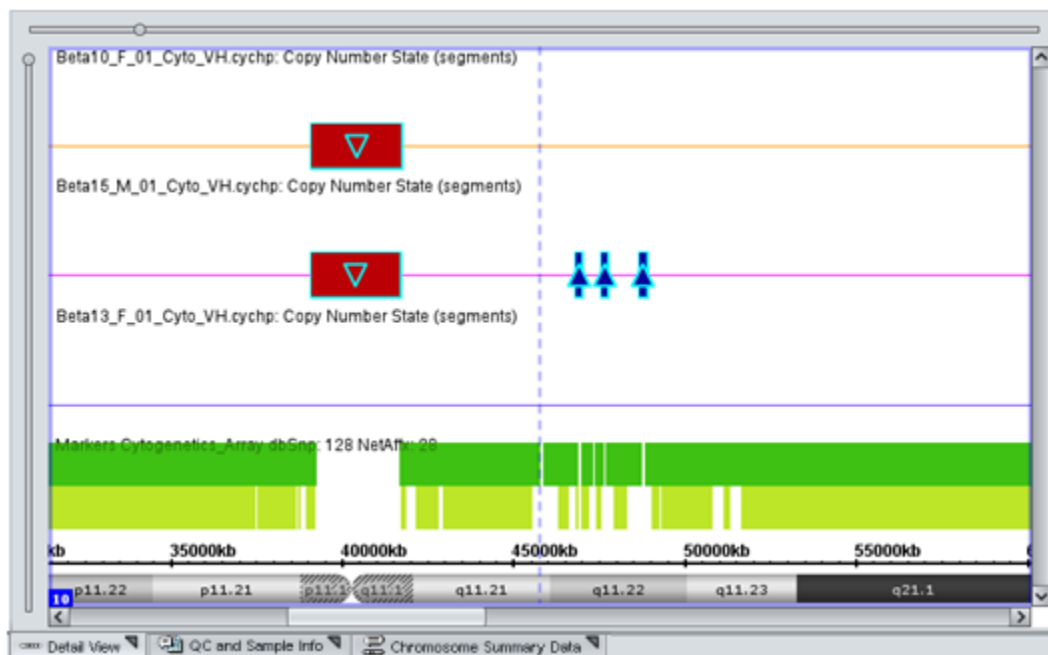





Figure 6.47 Selected area in Detail View

2. From the View menu, select **View Region at [site name]**. Alternatively, click a toolbar button:

-  UCSC
-  Ensembl
-  Toronto DGV

A browser opens, displaying the selected area of the chromosome.

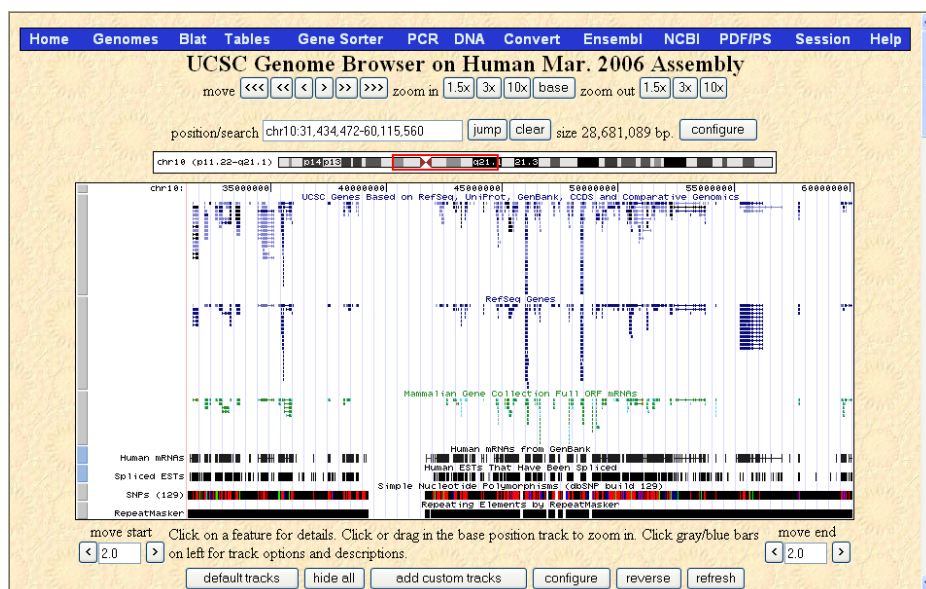


Figure 6.48 Selected area displayed in UCSC Genome Browser

## Chapter 7: Filtering Segments

ChAS enables you to filter the detected segments using different segment parameters, concealing segments that do not meet requirements for significance for:

- Marker Count
- Length
- Confidence Value

You can apply these filters to different segment types, using different parameters for each type. The filtering is done on the fly, with changes to the parameters reflected in the different views as they are made.

 **Note:** SNP6 CNCHP file segments cannot be filtered by confidence scores because the scores are not stored in the CNCHP files.

A segment must pass all filter requirements for the segment type to be displayed.


You can apply different filter values for areas inside CytoRegions and areas outside the CytoRegions (Genome-wide). See [Using Filters with CytoRegions](#) (page 141).


The Overlap Map filter is described in [Using the Overlap Map](#) (page 144).

Filter settings are saved when a Named Setting is created and can be reapplied. See [User Profiles and Named Settings](#) (page 225).

### Applying Segment Parameter Filters

To open the Segment Filters dialog box, do one of the following:

- Select **View > Segment Filters** on the menu bar
- Click the **Segment Filters**  toolbar button
- Right-click on a segment type in the Data Types list and select **Filters...** from the right-click menu.

 **Note:** If you use the right-click menu option, only the filter settings for the selected segment type are displayed.

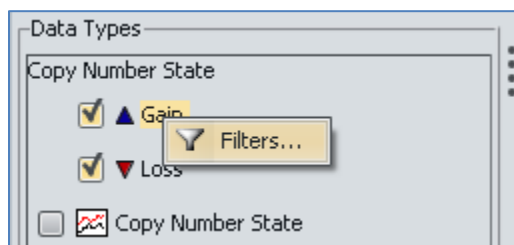



Figure 7.1 Right-click menu

The Segments Filters dialog box opens.

Different  
Segment types

Genome/CytoRegions tabs  
(displayed only when  
CytoRegions file is selected)

**Figure 7.2 Segment Filters dialog box**

 **Note:** the Overlap Map filtering parameter is set using the same dialog box. The Overlap Map function is described in [Using the Overlap Map](#), page 144.

You can filter using the following Segment Parameters:

- |                     |   |
|---------------------|---|
| <b>Marker Count</b> | The number of markers the segment encompasses from start to finish. A segment must have at least as many markers as you specify to be displayed.<br><br>Each marker represents a probe which represents a sequence along the genome at a particular spot. Markers are probe sequences of DNA, each sized from 12-50 base pairs long, depending on the type of array data. The 12-50 bp sequence is unique to that one spot on the genome it represents. |
| <b>Size</b>         | Based on the start and end markers of a segment. Because each segment represents a single place in the genome, you can measure from start to end, in DNA base pairs, and by filtering, demand a segment be at least that long to be visualized.   |

## Confidence

The Confidence Score for a segment is generated during analysis and recorded in the CYCHP file. It is an indicator of the likelihood that the segment represents a real change in that region of the genome and is a measure of how likely the data fits the assigned state for that marker.

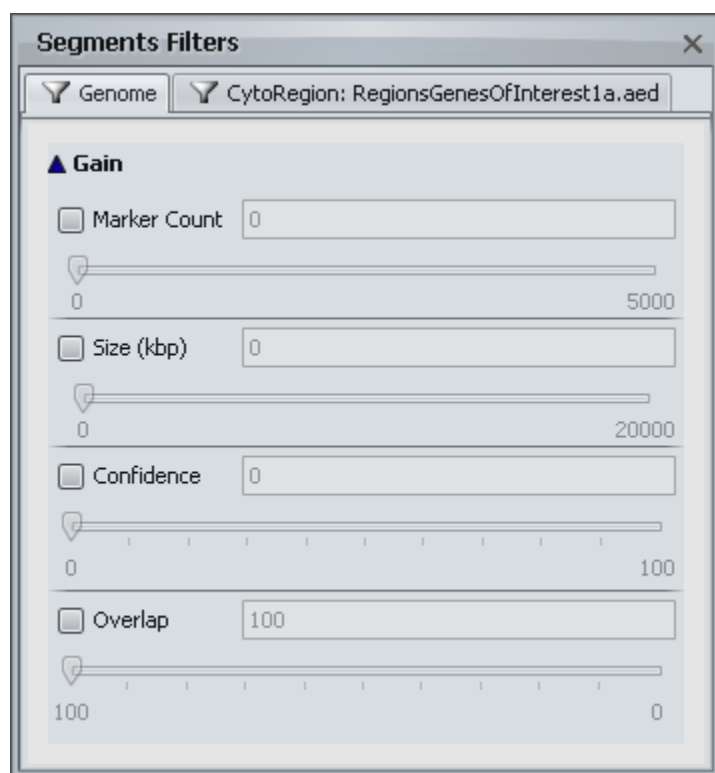
For CN Segments, confidence is determined on a marker by marker basis by evaluating the concordance of the log2ratio at each marker with the copy number state assigned by the HMM. The average confidence score of markers in Gain and Loss segments determines the confidence score of that segment.

Segments which are produced by smoothing or joining are assigned a confidence score based on weighted confidence scores in the original segments. This recalculated confidence is the value which is compared against the Confidence filter value.

See [Segment Smoothing and Segment Joining \(Optional\)](#) (page 71).



**Note:** SNP6 CNCHP file segments cannot be filtered by confidence scores because the scores are not stored in the CNCHP files.



**Figure 7.3 Gain Segment filter settings**

### To use the Segment Parameter filters:

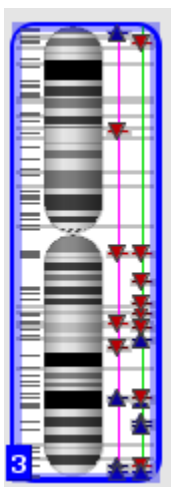
1. Open the Segment Filters dialog box.
2. Put a check mark next to the parameters that will be used as filters.
3. Set the value for the parameters with the slider or enter a value in the box.

Moving sliders from left to right progressively removes more and more segments.

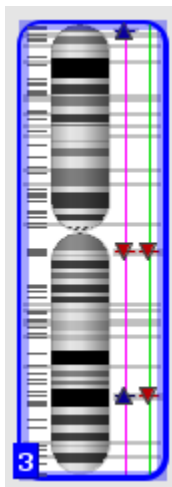
The filtered results are seen instantly in all tables and graphs.



No Filters Applied



Segment parameter filters  
Applied



**Figure 7.4 Filtering results**

See [Using the Overlap Map](#) (page 144) for information about using the Overlap setting.

See [Using Filters with CytoRegions](#) (page 141) for information on using different filtering settings in CytoRegions.

## Chapter 8: Using CytoRegions

The CytoRegions feature enables you to define parts of the genome that are of special interest to you. To use CytoRegions, you need to select a file with position information for regions of the genome as the CytoRegions file.

You can select a Region information file in AED or BED format. You can use a pre-existing Region file, or create a new one in AED format in ChAS and add regions to it by selecting segments, annotations, or regions in other loaded files.

You can add annotations to regions to help you track the information. See [Creating and Editing AED Files](#), page 152.



**Note:** The CytoRegions feature is intended for use with up to a few thousand regions. Larger numbers of regions can be used, but will impact performance. A reference annotation file, such as BAC or Genes, is not recommended for use as a CytoRegions file due to the large number of reference annotations.

After selecting a CytoRegions file, you can:

- Use the Restricted Mode to display only Segments and graph data that appear in those regions. Annotations are not hidden by CytoRegions or by the application of Restricted Mode.
- Use differential filtering options for these regions and for the rest of the genome.

This chapter describes:

- [Selecting a CytoRegions Information File](#) (page 134)
- [Viewing CytoRegions](#) (page 136)
- [Using Filters with CytoRegions](#) (page 141)
- [Using Restricted Mode](#) (page 142)

### Selecting a CytoRegions Information File

Select the CytoRegions file from the available region information files. See [Loading Files](#) (page 67). The software automatically checks the hg version of an AED or BED file before loading (see Figure 8.1 for an example BED file). The file will not be loaded if the hg version does not match what is loaded in the ChAS Browser. If an hg version is not found for the AED or BED file, a warning message is displayed.

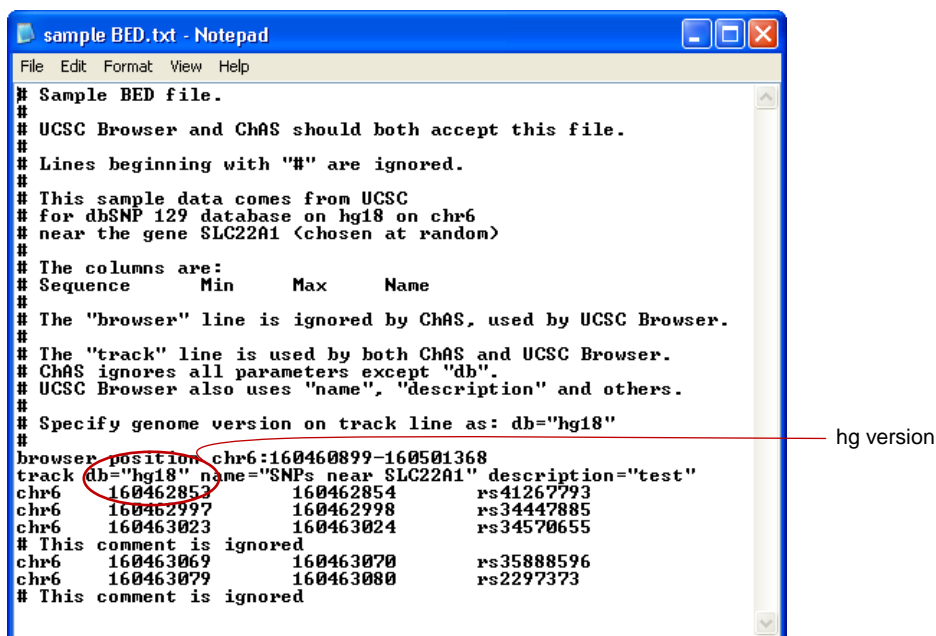


Figure 8.1 Example BED file

Do either of the following to select a CytoRegions file:

- In the files list, right-click a file and select **Set File as CytoRegions** on the shortcut menu.

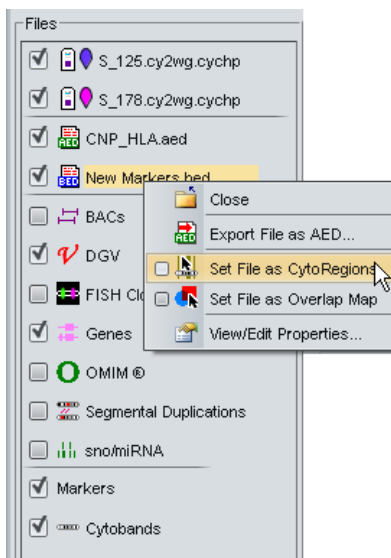


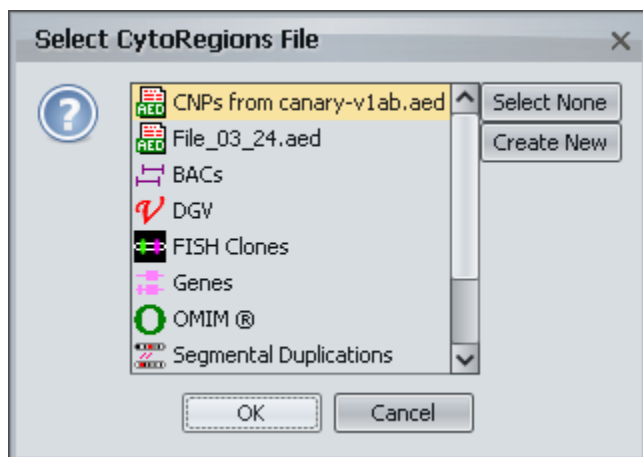
Figure 8.2 Select a CytoRegions file from the file list

OR

1. Select **View > CytoRegions** on the menu bar.

Alternatively, in the CytoRegions Tab of the Upper display, click the **Set CytoRegions** button .



The Select CytoRegions dialog box opens.



**Figure 8.3 Select CytoRegions File dialog box**

2. Select a regions information file to use for cytoregions and click **OK**.

The **Create New** feature is described in [Creating an AED File of Annotations](#) (page 152).


-  **Note:** If a region file was selected for the CytoRegion file before shutting off the software, that file is automatically loaded as the CytoRegion file when the software is opened the next time with the same user profile.
-  **Note:** To clear a CytoRegions file map from ChAS, choose “Select None” in the dialog box (Figure 8.2). Alternatively, right-click the file in the Files list and select “Set File as CytoRegions” on the shortcut menu to toggle the check mark off.

After selecting a CytoRegions file, you can:

- [View the CytoRegions in the Karyoview, Selected Chromosome Display, and Detail View](#) (page 137).
- [View CytoRegion information in the CytoRegions Tab](#) (page 138).
- [Use Segment filters with CytoRegions](#) (page 141).
- [Use Restricted Mode](#) (page 142).

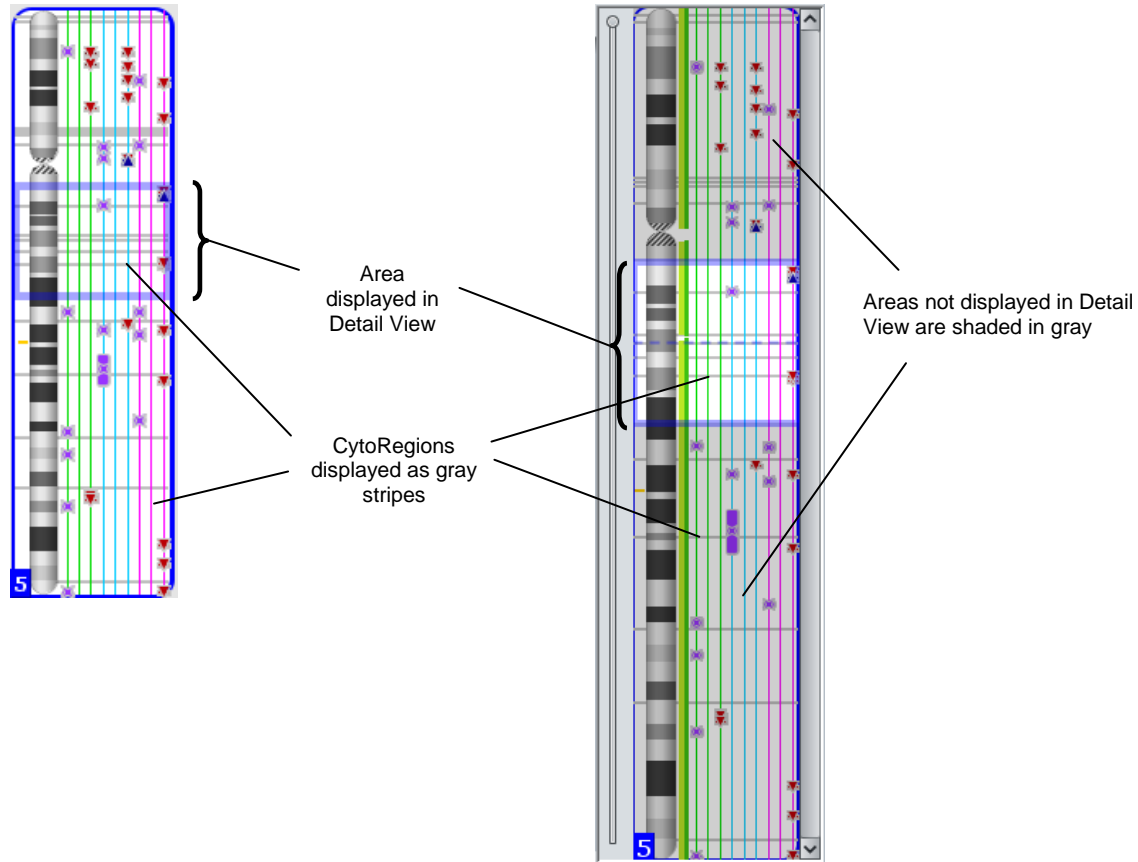
## Viewing CytoRegions

CytoRegions can be displayed in:

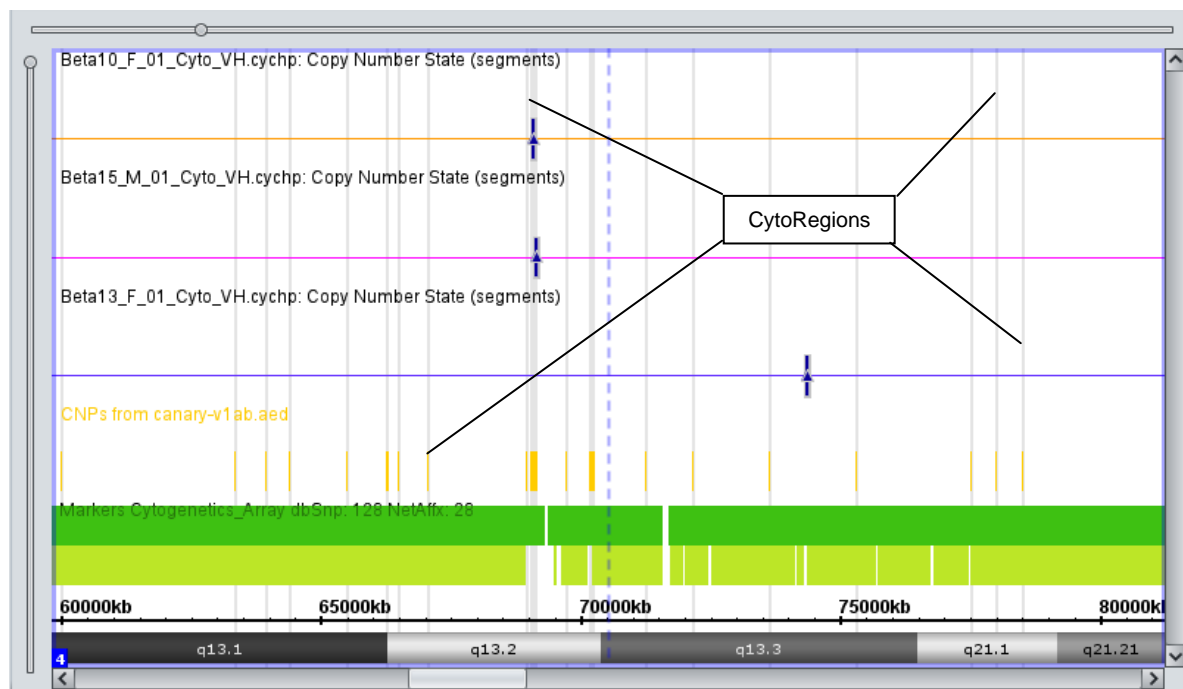
- [The ChAS graphic views](#) (below)
  - [CytoRegions Table](#) (page 138)
-  **Note:** CytoRegions that share genomic coordinates with a detected segment are listed in the “CytoRegions” column of the Segments table. See [Segments Table](#) (page 182).

## CytoRegions in the Graphic Views

Regions specified in the CytoRegions file are displayed as dark gray stripes in the Karyoview, Chromosome Display, and Detail View.



**Figure 8.4 Karyoview and Selected Chromosome View with CytoRegions Displayed**

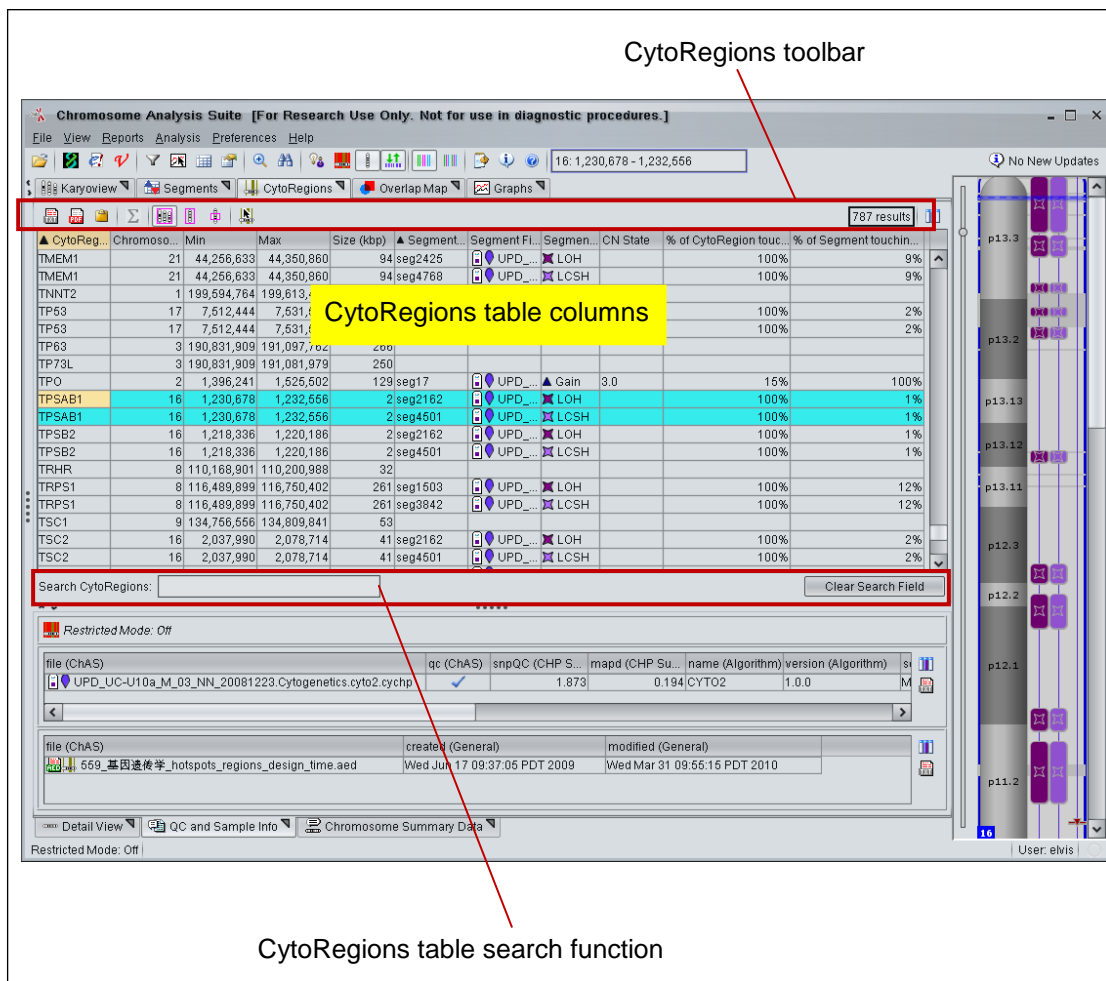


**Figure 8.5 Detail View with CytoRegions displayed**

### **CytoRegions Table**

The CytoRegions table shows the intersections between regions in the designated CytoRegions file and the segments in the Segments table. Every region in the CytoRegions file will be listed on at least one line of the CytoRegions table, even if it does not intersect any segments. For those regions which intersect one or more segments, there will be one table row for each intersection. Depending on the columns which have been used to sort the CytoRegions table, these rows may or may not be near each other. A segment that intersects more than one region in the CytoRegions file appears multiple times in the CytoRegions table, one row for each intersection.

In the CytoRegions table, the “CN State” value corresponds to the state of copy-number and mosaicism segments that intersect the cytoregion. There is no CN State value for other types of segments that do not correspond to a copy-number call.



**Figure 8.6 CytoRegions Table**

### Highlighting Regions in the CytoRegions Table and Details View

If the Details View displays the CytoRegions file (the CytoRegions file is check marked in the Files list), you can conveniently find and view items.

- Click a row in the CytoRegions table to select the corresponding annotation from the CytoRegions file. All of the lines for that region are highlighted in the table. The Details View zooms to the currently selected region.

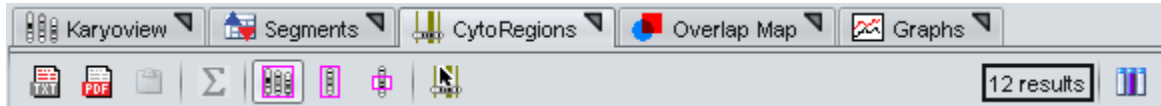


**Note:** If the Details View does not automatically zoom to the selected region, confirm that the Auto-zoom option is selected (select “View > Auto-zoom to table selection” on the menu bar.)


- In the Detail View, click a region or select multiple regions of the CytoRegion file to highlight all of the corresponding rows in the CytoRegions table. The CytoRegions table automatically scrolls to show at least one of the highlighted rows.
- To quickly find a particular Segment in the CytoRegions table, first double-click that segment in any of the views or in the Segments table (the current region will be set to that segment), then press the toolbar button in the CytoRegions table to show only the cytoregions in the current region.

## CytoRegions Toolbar

The Toolbar provides quick access to table functions. The standard functions are described in [The Standard Toolbar](#) (page 178).



The Toolbar has one specialized button.

	Select CytoRegions file (see <a href="#">Selecting a CytoRegions Information File</a> , page 134).
---	--

## CytoRegions Table Columns

The table can display the following columns:

<b>CytoRegion Type</b>	Type of file or element from which the CytoRegion is derived. Default User Annotations are annotations derived from AED or BED file.
<b>CytoRegion</b>	Identifier for region.
<b>Chromosome</b>	Chromosome in which the region is located.
<b>Min</b>	Starting position of the region.
<b>Max</b>	Ending position of the region.
<b>Size (kbp)</b>	Size of the region.
<b>Segment ID</b>	Unique ID assigned to the detected segment by ChAS.
<b>Segment File</b>	Sample File that the segment was detected in.
<b>Segment type</b>	Type of segment: <ul style="list-style-type: none"><li>• CN loss or gain</li><li>• Mosaicism</li><li>• LOH</li></ul>
<b>Segment Min</b>	Starting position of segment.
<b>Segment Max</b>	Ending position of segment.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>Segment Size (kbp)</b>	Size of the segment.
<b>CN State</b>	Copy Number State (Displayed for Gain, Loss, and Mosaicism segment types).
<b>Shared Size</b>	Size of the contact between segment and cytoregion.



**% of CytoRegion touching Segment**

How much of the CytoRegion is contacted by the Segment.

**% of Segment touching CytoRegion**

How much of the detected segment is contacted by the CytoRegion.

You can:

- Use the common table functions to control the display of data in the table (see [Common Table Operations](#), page 178)
- Export data in various formats (see [Exporting Table Data](#), page 212).
- Search CytoRegions to display only regions of interest (see below).

## Searching CytoRegions

The Search CytoRegions feature allows you to search the “CytoRegion” column for text strings that match a search string. You can use \* as a wild card.

### To search cytoregions:

1. Enter the search string in the Search CytoRegions box at the bottom of the CytoRegions table.
2. Press **Enter**.

The table displays only annotations that match the search string.

3. To restore the table, click **Clear Search Field**.

## Using Filters with CytoRegions

If you have CytoRegions information loaded into ChAS, you have the option to apply different segment filtering parameters to the parts of the genome that are defined as CytoRegions and the parts of the genome that are not within these regions (Genome). For segment filtering to function as described below, the CytoRegion segment filters should always be set identical to or less stringent than the counterpart Genome segment filter settings.

The filter settings for CytoRegions and for the rest of the genome interact in different ways, depending upon whether:

- A cytoregion information file is selected
- Restricted Mode is selected

If a cytoregion information file is selected, but Restricted Mode is not selected:

- Segments wholly within a cytoregion are filtered using the CytoRegions filter settings.
- Segments that don't touch a cytoregion are filtered using the Whole Genome settings.
- A segment that touches both Genome and CytoRegions must pass both the CytoRegions filter thresholds and Whole Genome filter thresholds, otherwise it will not be shown.

If Restricted Mode is selected:

- Segments in or straddling cytoregion boundaries are filtered using the CytoRegions filter settings.

These rules also apply when a region information file has been selected as an Overlap Map and the Overlap filters are used.


## Using Restricted Mode

The Restricted mode allows you to view detected segments and graph data only in regions you have defined in advanced in the CytoRegions file.

 **Note:** Restricted mode is not available unless a region information file or one of the Reference Annotations is selected for the CytoRegions function.

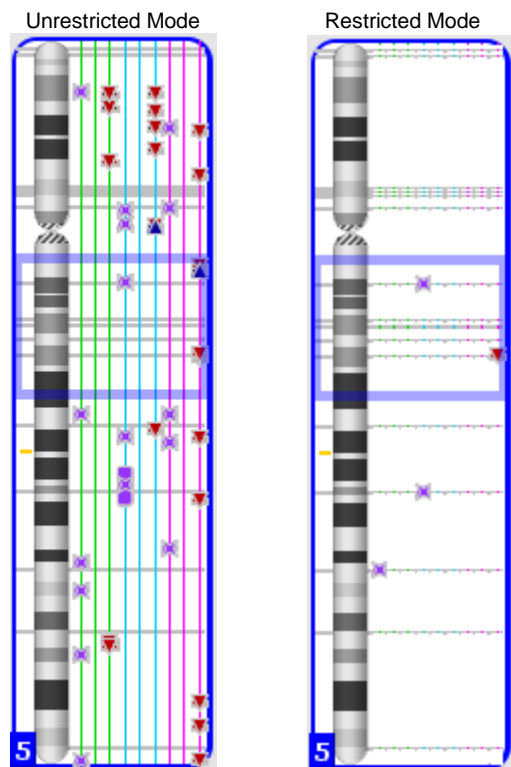
**To select/deselect Restricted Mode:**

- From the View menu, select **Restrict to CytoRegion**; or

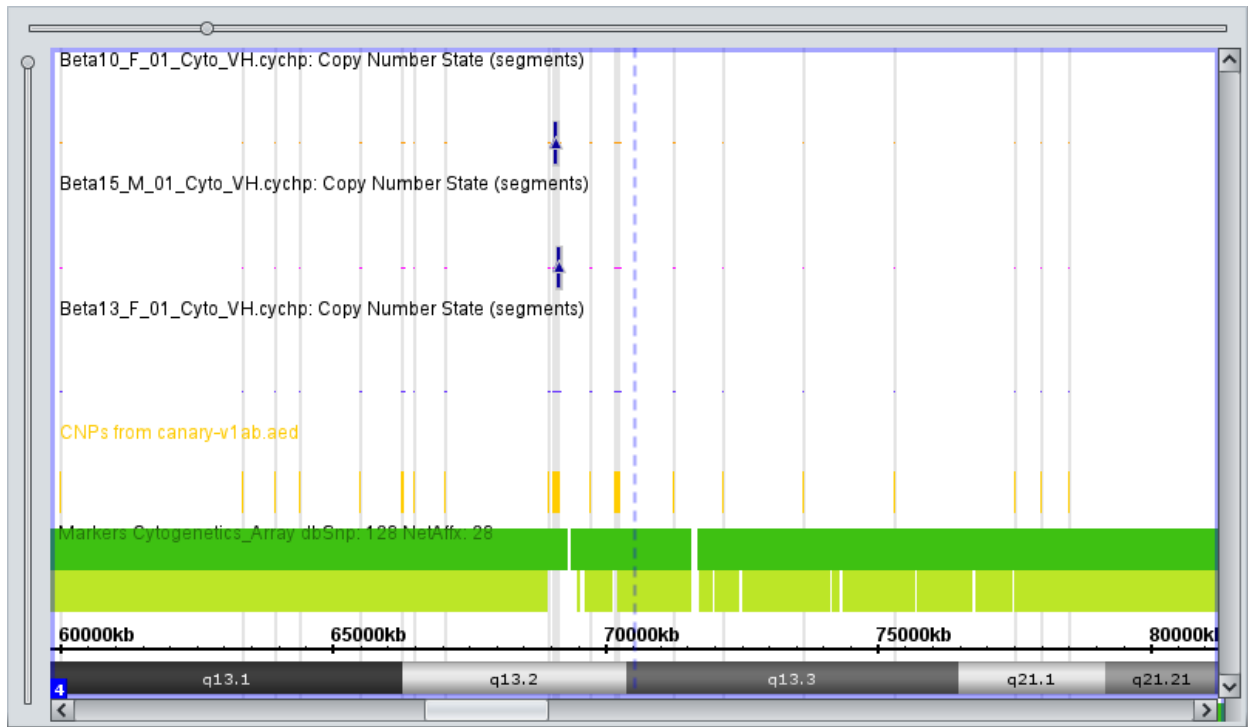
Click the Restricted Mode button  in the main toolbar.

Segments and graph data that are not in the defined cytoregions are concealed in the display.

Segments in or straddling cytoregions boundaries are filtered using the CytoRegions filter settings.

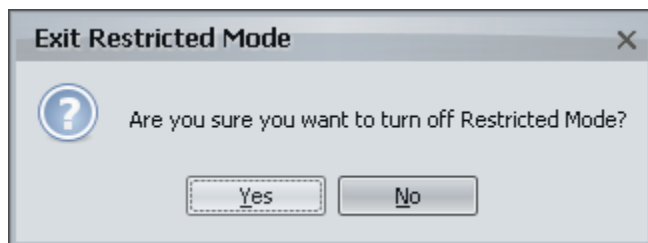


**Figure 8.7 Karyoview in unrestricted and restricted mode**



**Figure 8.8 Detail View in Restricted Mode**

When you deselect Restricted Mode, the following dialog box appears:



**Figure 8.9 Exit Restricted Mode notice**

Click **Yes** to exit Restricted Mode

## Chapter 9: Using the Overlap Map

The Overlap Map allows you to show or hide segments in areas of the genome that you aren't interested in, for example, in regions of Benign Copy Number Polymorphism, by:

- Specifying regions of the genome in an Overlap Map File.
- Optionally filtering out detected segments that are overlapped by these regions.

You can specify the percentage of the segment (between 1 and 100%) that must be overlapped by the Overlap Map items before being filtered from display.

The Overlap Map filter operates separately from the CytoRegions features, but you can apply different overlap map filtering parameters to CytoRegions and to areas outside of CytoRegions.

### To use the overlap filter:

1. [Select a region information file for the Overlap Map](#) (below).
2. [Set Percent Overlap value](#) for each segment type you wish to filter in the Segment Filters dialog box (page 150).

You can view the Overlap Map regions in:

- [The ChAS graphic views](#) (page 147)
- [The Overlap Map tab](#) (page 148)

## Selecting the Overlap Map File

You can use the following file types for the Overlap Map:

- Region information files in AED and BED format.
- Position information from Reference annotation files.

The software automatically checks the hg version of an AED or BED file before loading (see Figure 8.1 for an example BED file). The file will not be loaded if the hg version does not match what is loaded in the ChAS Browser. If an hg version is not found for the AED or BED file, a warning message is displayed.

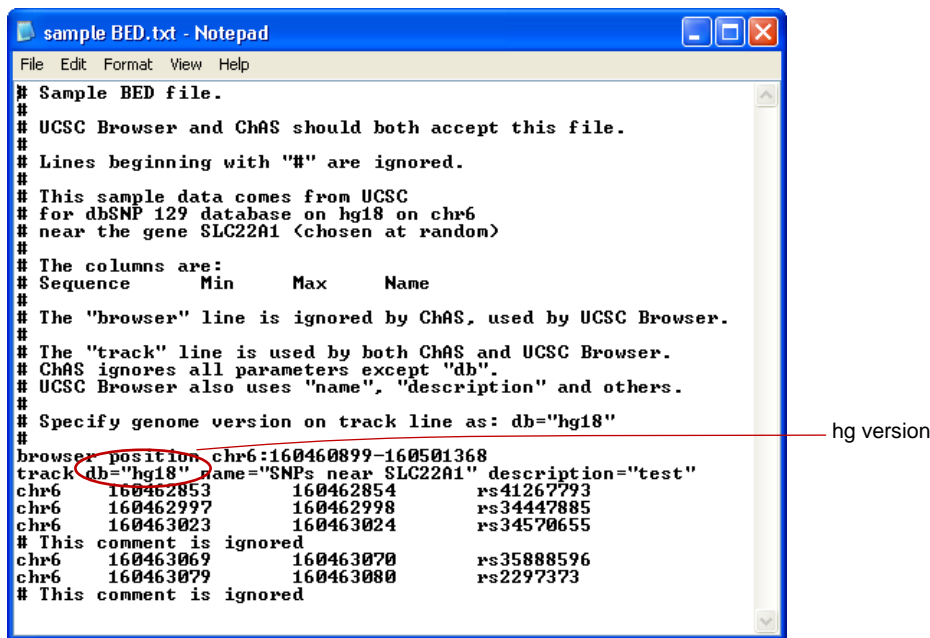


Figure 9.1 Example BED file

Do either of the following to select an overlap map:

- In the files list, right-click the file and select **Set File as Overlap Map** on the shortcut menu.

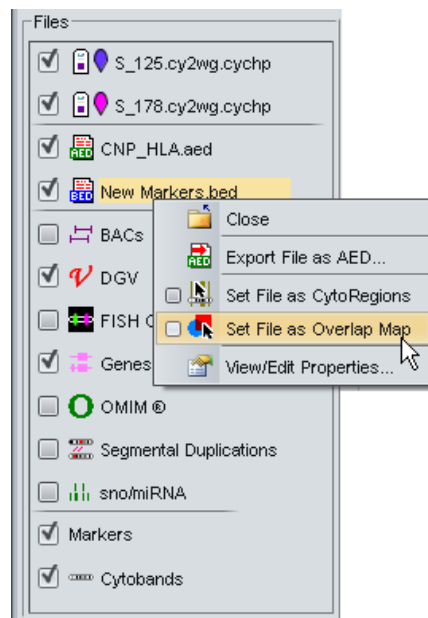


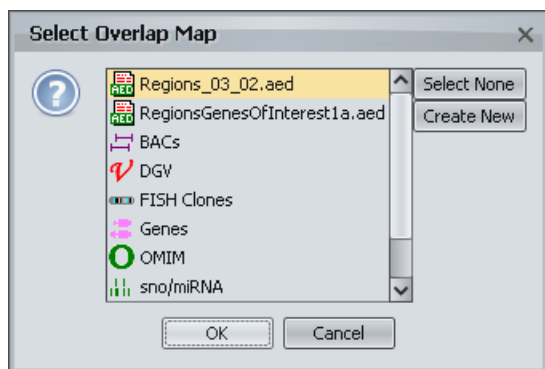
Figure 9.2 Select an Overlap Map from the files list

OR

1. **Select View > Overlap Map** on the menu bar.

Alternatively, in the Overlap Map tab in the upper display, click the **Select Overlap Map**  toolbar button.

The Select Overlap Map dialog box appears.



**Figure 9.3 Select Overlap Map dialog box**

The dialog box displays a list of the region and annotation files you can select for an overlap map.

2. Select the file you wish to use and click **OK**.

The region file is loaded and the regions are displayed with overlap information in the Overlap Map tab.


The Overlap Map icon  appears next to the selected file in the File List. 

 **Note:** To clear an Overlap map, click “Select None” in the dialog box (Figure 10.2). Alternatively, right-click the file in the files list, and select “Set File as Overlap Map” on the shortcut menu.

## Viewing Overlap Regions

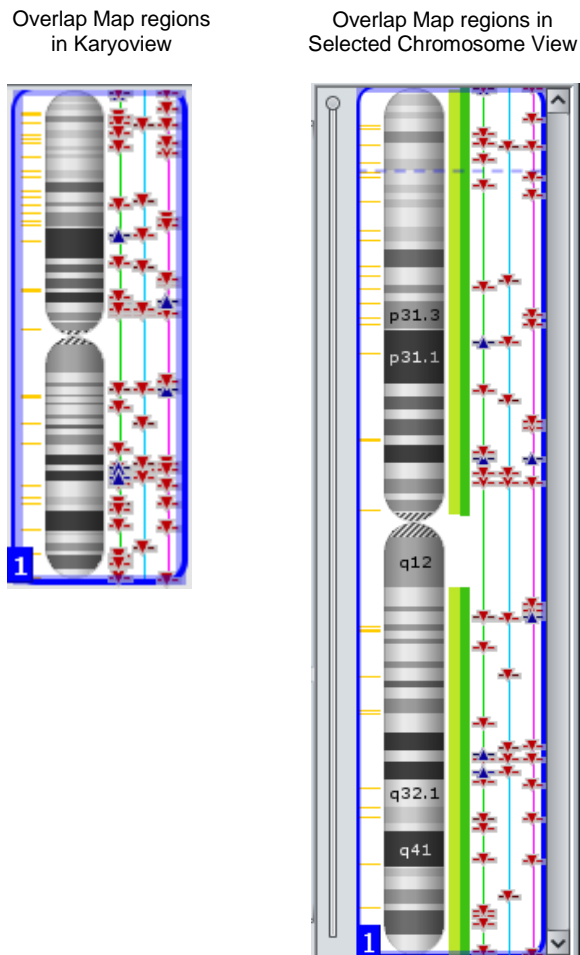
You can view the Overlap Map regions:

- [In the graphic display views](#) (below)
- [In the Overlap Map Tab](#) (page 148)

 **Note:** Overlap Map items that are covered by a detected segment are listed in a column of the Segments table. See [Segments](#) (page 182).

## Viewing Overlap Map Regions in the Graphic Displays

In the Karyoview and Selected Chromosome View, regions specified in the Overlap Map file are displayed as short rectangles to the immediate left of the cytobands.

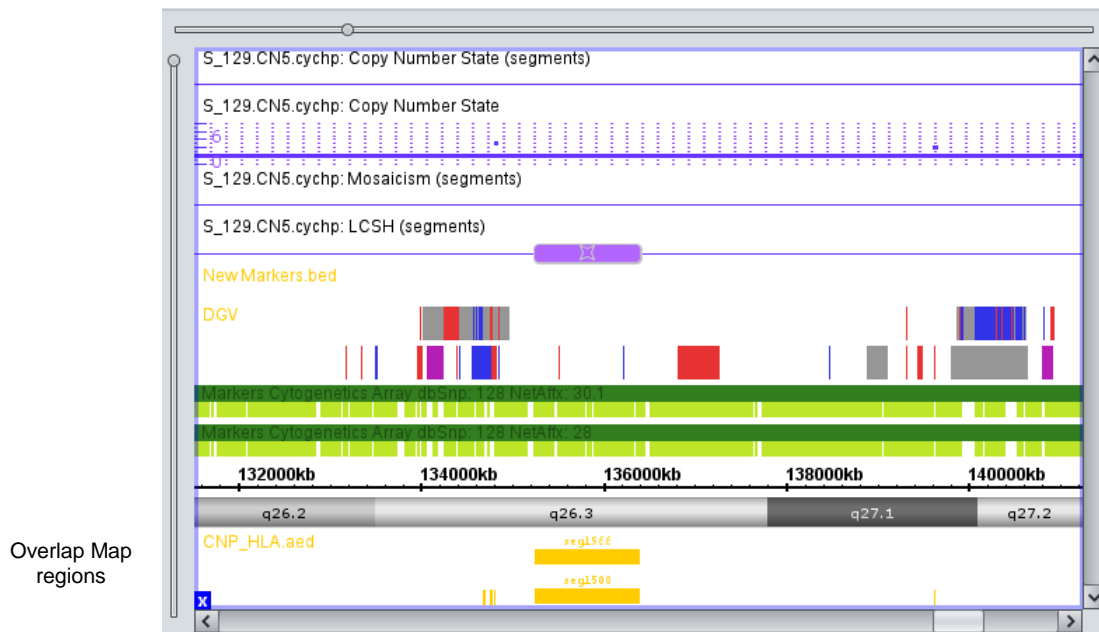


**Figure 9.4 Karyoview and Selected Chromosome View with Overlap Map Regions Displayed**

In the Detail View, the overlap map regions are displayed as rectangles below the cytobands. The default color is yellow, but you may select different colors for displaying regions in a region information file.

You can change the colors used to display items by using:

- Specifying color when adding the item to an AED file (see [Entering General Information](#) (page 163))
- Using the Color Rules (see [AED/BED Color Rules](#) (page 169))



**Figure 9.5 Detail View with Overlap Regions displayed**

### Viewing the Overlap Map Table

The Overlap Map table displays a list of overlapping items from the overlap map file and the Segments table. Each region in the Overlap Map file will be listed on at least one row of the table, even if it does not overlap any segments. For those regions which overlap one or more segments, there will be one for each overlap. Depending on the columns which have been used to sort the Overlap Map table, these rows may or may not be near each other. A segment that overlaps more than one region in the Overlap Map file will appear multiple times in the Overlap Map table, one row for each overlap.


### Highlighting Overlap Regions in the Overlap Map Table and Details View

If the Details View displays the Overlap Map file (the Overlap Map file is check marked in the Files list), you can conveniently find and view items.

- Click a row in the Overlap Map table to select the corresponding annotation from the Overlap Map file. All of the rows for that region are highlighted in the table. The Details View zooms to the currently selected region.



**Note:** If the Details View does not automatically zoom to the selected region, confirm that the Auto-zoom option is selected (select “View > Auto-zoom to table selection” on the menu bar.)

- In the Detail View, click a region or select multiple regions from the Overlap Map file to highlight all of the corresponding rows in the Overlap Map table. The Overlap Map table automatically scrolls to show at least one of the highlighted rows.
- To quickly find a particular Segment in the Overlap Map table, first double-click that segment in any of the views or in the Segments table (the current region will be set to that segment), then press the  toolbar button in the Overlap Map table to show only the overlaps in the current region.



Overlap Map Item	Chromosome	Min	Max	Size (kbp)	Segment ID	Segment File	Segment Type	Segment Min
Variation_31239	5	69,000,764	69,578,673	577	seg2209	Beta13_F_01_Cyto_VH.cychp	▲ Gain	69,502,982
Variation_31241	5	69,876,869	70,578,999	702	seg2217	Beta13_F_01_Cyto_VH.cychp	▲ Gain	70,232,999
Variation_31240	5	69,680,605	69,875,037	294	seg2211	Beta13_F_01_Cyto_VH.cychp	▲ Gain	69,802,877
Variation_0283	5	68,604,421	70,408,242	1,803	seg2211	Beta13_F_01_Cyto_VH.cychp	▲ Gain	69,802,877
Variation_0283	5	68,604,421	70,408,242	1,803	seg2217	Beta13_F_01_Cyto_VH.cychp	▲ Gain	70,232,999
Variation_0283	5	68,604,421	70,408,242	1,803	seg2209	Beta13_F_01_Cyto_VH.cychp	▲ Gain	69,502,982
Variation_0283	5	68,604,421	70,408,242	1,803	seg2195	Beta13_F_01_Cyto_VH.cychp	▲ Gain	68,766,136
RPL35	9	126,659,978	126,664,061	4	seg3898	Beta13_F_01_Cyto_VH.cychp	▲ Gain	126,686,233
RP11-348K2	9	126,439,465	126,643,093	203	seg3898	Beta13_F_01_Cyto_VH.cychp	▲ Gain	126,686,233
RP11-62A6	9	126,439,614	126,592,579	152	seg3898	Beta13_F_01_Cyto_VH.cychp	▲ Gain	126,686,233
RP11-174P6	9	126,048,268	126,895,604	847	seg3898	Beta13_F_01_Cyto_VH.cychp	▲ Gain	126,686,233
Variation_7787	X	40,635,684	42,901,712	2,266	seg7580	Beta13_F_01_Cyto_VH.cychp	▲ Gain	41,099,363
Variation_7787	X	40,635,684	42,901,712	2,266	seg7578	Beta13_F_01_Cyto_VH.cychp	▲ Gain	41,022,335

**Figure 9.6 Overlap Map Table**

## Overlap Map Toolbar

The Toolbar provides quick access to table functions. The standard functions are described in [The Standard Toolbar](#), page 178.



The Toolbar has one specialized button.

	Select Overlap Map file (see <a href="#">Selecting the Overlap Map</a> (page 144)).
--	---

## Overlap Map Table Columns

The table can display the following columns:


<b>Map Item Type</b>	Source of the position information (CN Gain or Loss segment, reference annotation, etc.)
<b>Overlap Map Item</b>	Identifier for the item.
<b>Chromosome</b>	Chromosome in which the item is located.
<b>Min</b>	Starting position of the item.
<b>Max</b>	Ending position of the item.
<b>Size (kbp)</b>	Size of the item.
<b>Segment ID</b>	Unique ID assigned to the detected segment by ChAS.
<b>Segment File</b>	Sample File that the segment was detected in.
<b>Segment Type</b>	Type of segment: <ul style="list-style-type: none"> <li>• CN loss or gain</li> <li>• Mosaicism</li> <li>• LOH</li> </ul>
<b>Segment Min</b>	Starting position of segment.

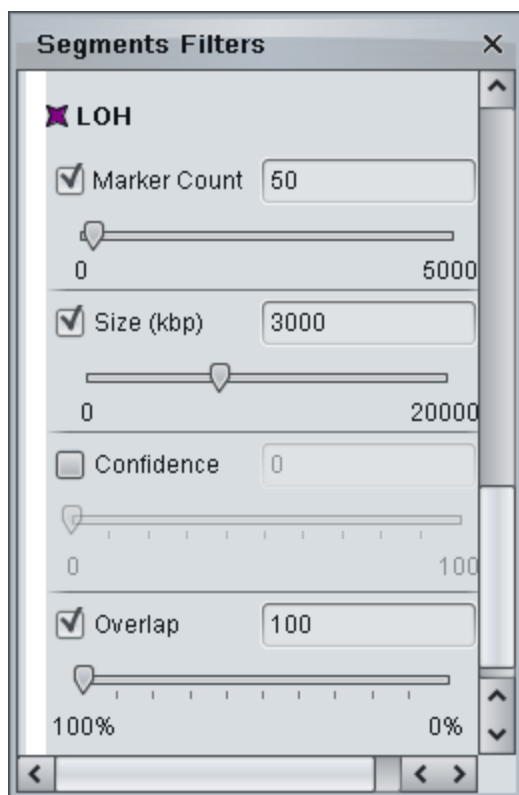
<b>Segment Max</b>	Ending position of segment.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>Segment Size</b>	Size of the segment.
<b>CN State</b>	Copy Number State (Not displayed for other segment types).
<b>% Overlap</b>	How much of the detected segment is covered by the Overlap Map item. A Segment which has 20% of its length somehow encompassed within the boundaries of an Overlap Map item has an Overlap value of 20%.  This percentage value is used to filter segments out of the displays and tables when filtering segments by “Overlap” in the filter slider dialogs.
<b>% Coverage</b>	How much of the Overlap Map item is covered by the Segment.
<b>Shared Size</b>	Size of the overlap between segment and Overlap Map item. Coverage values are not presently used in filtration of Segments from the displays or tables.

## Setting the Overlap Filter

After selecting an Overlap Map Region file, you have the option to set Overlap filters for the different segment types.

### To use the Overlap filters:

1. Select a region information or Reference Annotation file for the Overlap Map (see [Selecting the Overlap Map File](#) on 144).
2. Open the Segment Filters dialog box (click the  button).  
See [Applying Segment Parameter Filters](#) (page 130).



**Figure 9.7 LOH filter settings, Overlap selected.**

3. Select the Overlap check box(es) for the segment types you want to use to filter.
4. Set the value for the parameters with the slider or enter a value in the box.

Moving the slider farther to the right (or entering smaller values in the box) will cause more and more of the Overlapped segments to be removed from the display.

The detected segments must share at least the specified percentage of their length with the Overlap Map region to be filtered out and hidden from display. A Segment which has 20% of its length somehow encompassed within the boundaries of an Overlap Map item has an Overlap value of 20%.

The minimum value of the setting is 1%.

The results of filtering are seen instantly in all tables and graphs.

See [Using Filters with CytoRegions](#) (page 141) for information about the interactions of the Overlap Map filter with the CytoRegions.

## Chapter 10: Creating and Editing AED Files

You can create AED files that contain:

- User-selected annotations. The annotations in an AED annotation file can be edited in ChAS.
- Position information on regions of the chromosome, as well as additional annotation information on the regions. AED region information files can be used for CytoRegions or Overlap Map functions.
- Third party reference annotations converted to AED file format

User-generated AED region information files can be used for CytoRegions or Overlap Map functions. You can add a region to a new or previously created AED file by selecting the following feature types:

- Detected Segments
- Reference Annotations
- Regions in previously loaded files

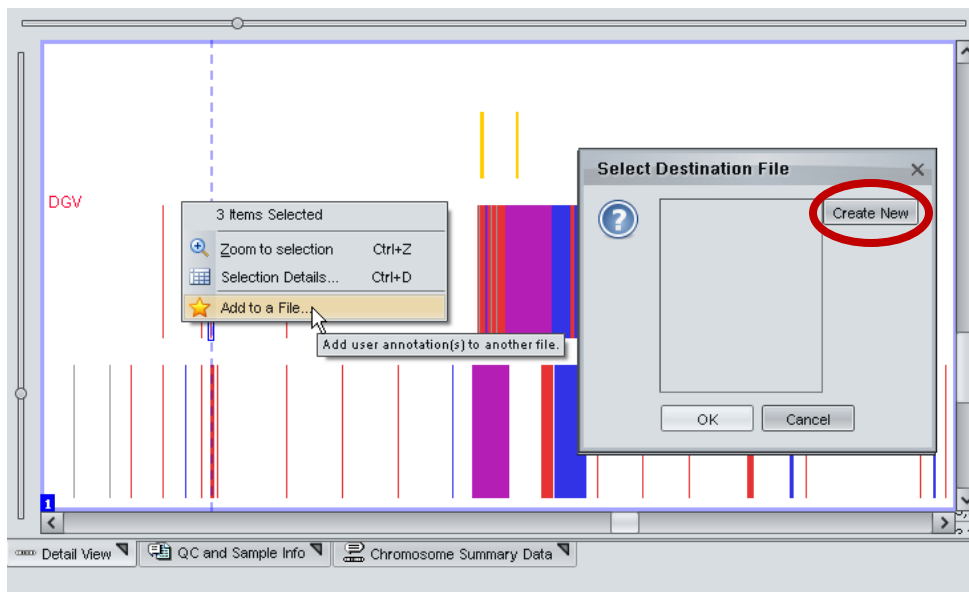
This chapter discusses:

- [Creating an AED File of Annotations](#) (below)
- [Creating an AED File of Regions](#) (page 154)
- [Adding Regions to an Existing AED File](#) (page 156)
- [Deleting Regions from an AED File](#) (page 158)
- [Viewing and Editing Annotations](#) (page 162)
- [Viewing or Editing AED File Properties](#) (page 159)
- [AED/BED Color Rules](#) (page 169)
- [Exporting Information in AED or BED Format](#) (page 174)

The AED file format is described in [AED File Format](#) (page 237).

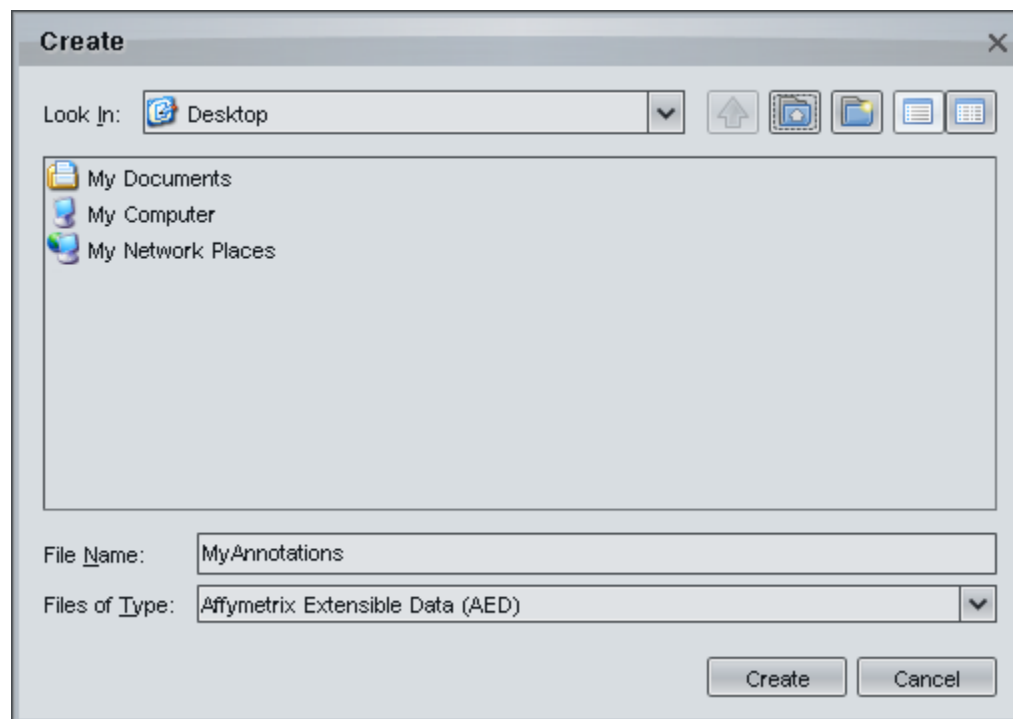
### Creating an AED File of Annotations

1. In the Detail View, select the non-AED annotation(s) for the AED file:
  - Right-click an annotation and select **Add to a File** on the shortcut menu
  - OR
  - Draw a box around multiple annotations, right-click the selection, and select **Add to a File** on the shortcut menu
2. In the dialog box that appears, click **Create New**.



**Figure 10.1 Detail View, create a new AED annotation file**

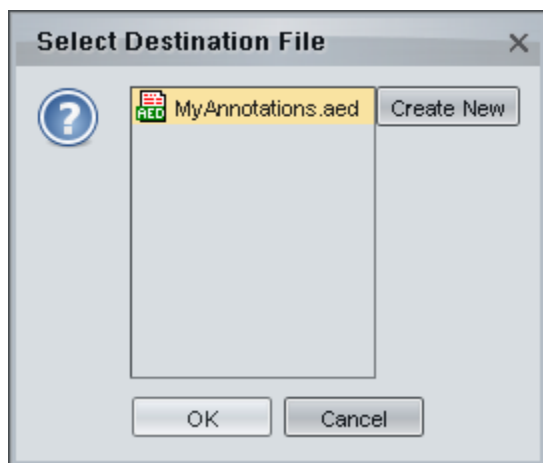
3. In the Create dialog box that appears next, select a folder and enter a file name.



**Figure 10.2 Create dialog box**

3. Click **Create**.

The Select Destination File dialog box shows the name of the new AED file.



**Figure 10.3 Select Destination File dialog box**


4. Click **OK**.

The Details View shows the new annotation (AED).

 **Note:** The AED file is automatically assigned the same genome assembly version (i.e., “hg18”, “hg19”, etc.) as the currently loaded NetAffx annotations.

### **Adding Annotations to an AED File**

1. In the Detail View, select the annotation(s) that you want to add to an AED file:
  - Right-click an annotation and select **Add to a File** on the shortcut menu
  - OR
  - Draw a box around multiple annotations, right-click the selection, and select **Add to a File** on the shortcut menu
2. In the Select Destination File dialog box (Figure 11.3), select an AED file and click **OK**.

 **Note:** Adding annotations to an AED file does not modify the genome assembly version. If the AED file does not specify a genome assembly version, none is automatically assigned to the AED file when annotations are added.

## **Creating an AED File of Regions**

You may wish to put new features from a set of samples in a new Regions file.

 **Note:** You can also use the Export feature to export data in existing files to an AED file. See [Exporting Information in AED or BED Format](#) (page 174).

**To create a new CytoRegions file in AED file format:**

1. From the View menu, choose **Select CytoRegions** file or **Select Overlap Map**.  
The appropriate Select File dialog box opens.

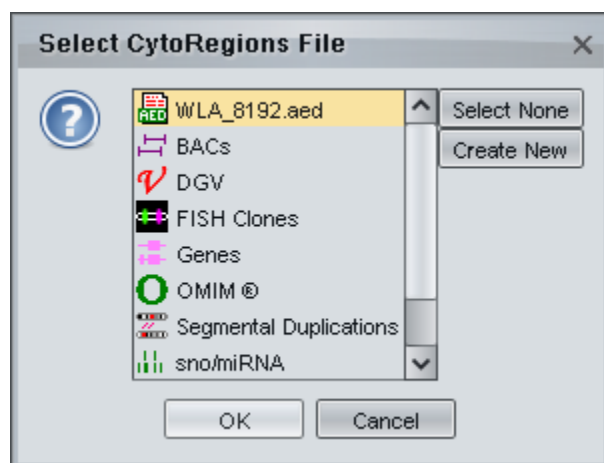


Figure 10.4 Select CytoRegions file dialog box

 **Note:** You can also create a new AED file when adding a region to an AED file.

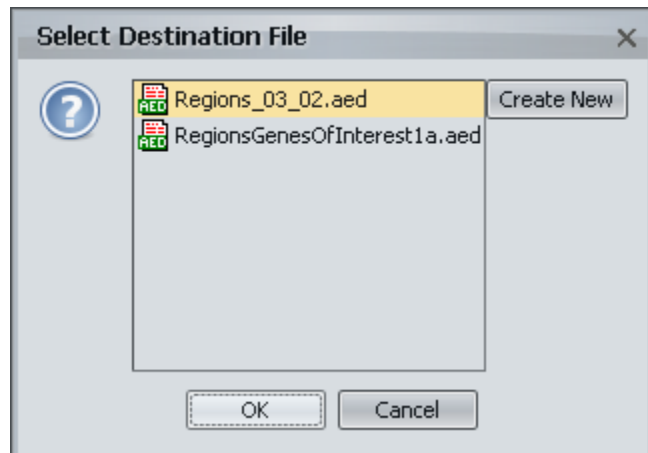
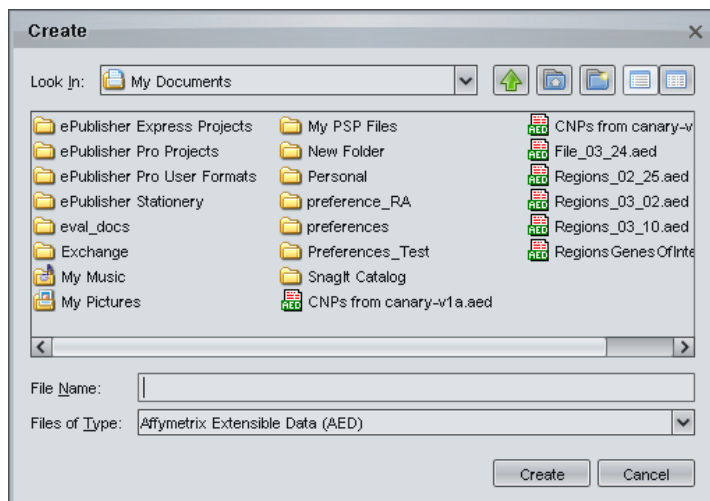


Figure 10.5 Select Destination File dialog box

2. Click **Create New** in the Select File dialog box.  
The Create dialog box opens.

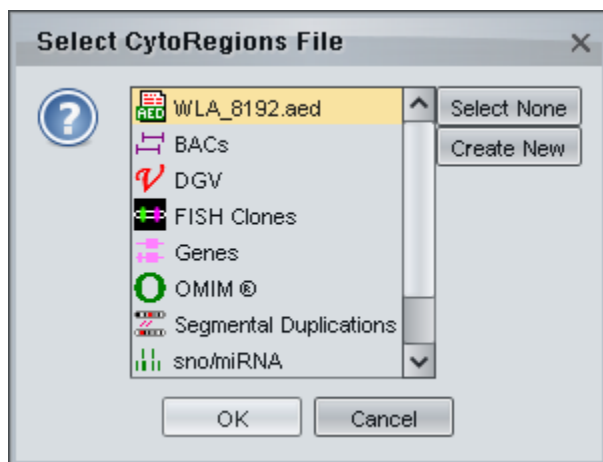


**Figure 10.6 Save dialog box**

3. Use the dialog box controls to browse to a folder for the AED file.
4. Enter a file name.
5. Click **Create** in the Create dialog box.

The Select File dialog box appears with the new file selected.

**Note:** The AED file is automatically assigned the same genome assembly version (i.e. “hg18”, “hg19”, etc.) as the currently loaded NetAffx annotations.



**Figure 10.7 Select Destination File**

You can select the new file or add regions to it, depending upon what function you were performing initially.

**Note:** To open an AED file, click the  button or select **File > Open** on the menu bar.


### **Adding Regions to an Existing AED File**

You can add a new region to an existing region information file in AED format by selecting the following features in the ChAS graphic views:

- Detected Segments



- Reference Annotations
- Regions in previously loaded files.

 **Note:** You can edit the color of DGV annotations that have been added to an AED file by creating a color rule (see page 170). Alternatively, you can edit the color of a particular DGV annotation added to an AED file, in the Detail View (see page 163).

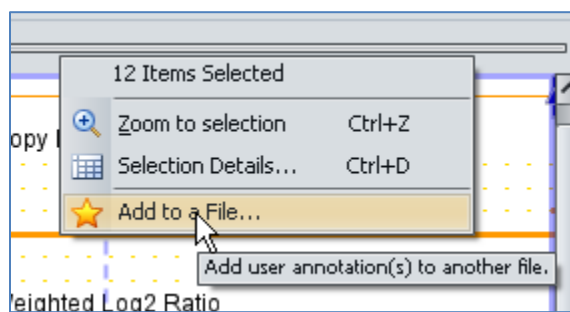
**To add a section to a new region (AED) file:**

1. Right-click on any of the following feature types in the graphic displays and select **Add to a File** from the right-click menu:

- Segment
- Reference Annotation (including Cytobands)

 **Note:** You should expand the reference annotations before selecting one to add to an AED file to avoid selecting multiple annotations. [Expanding and Contracting Annotations](#), page 115.

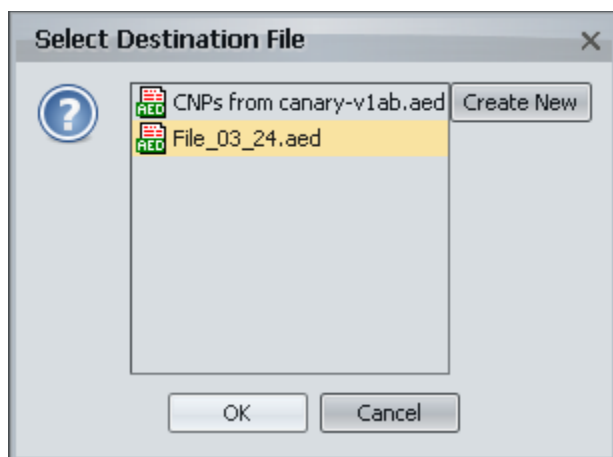
- Region



**Figure 10.8 Right-click menu**

The Select Destination File dialog box opens.

 **Note:** Some options may not be available, depending upon the number of type of items you have selected.

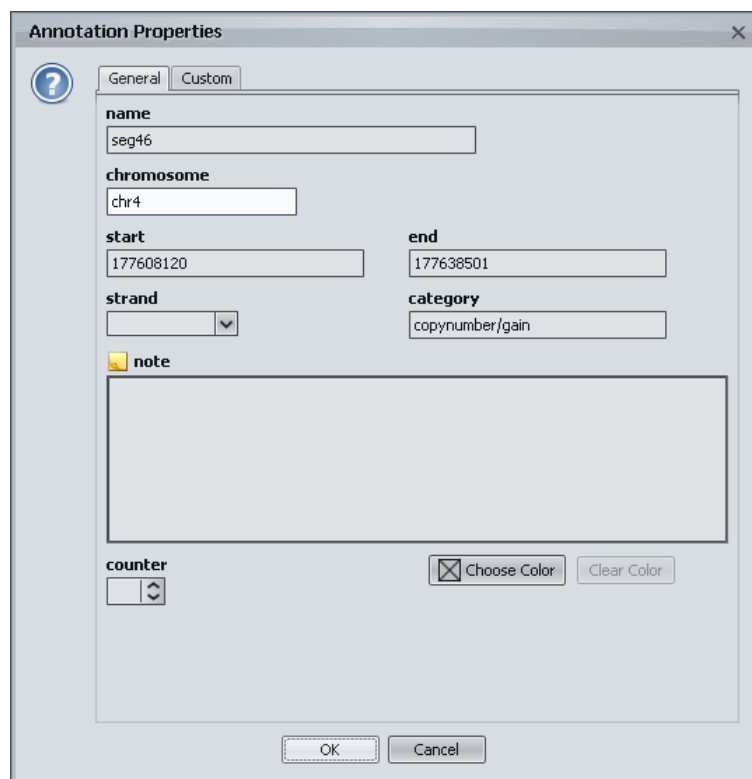


**Figure 10.9 Select Destination File dialog box**

The Select Destination File dialog box displays a list of the currently existing AED files to which you may add the segment.

2. Select the region file you wish to use and click **OK**.

The Annotation Properties dialog box opens if you have selected a single item. If you have selected multiple items, the dialog box will not open.



**Figure 10.10 Annotation Properties dialog box**

See [Viewing and Editing Annotations](#) (page 159) for information on editing the annotations.

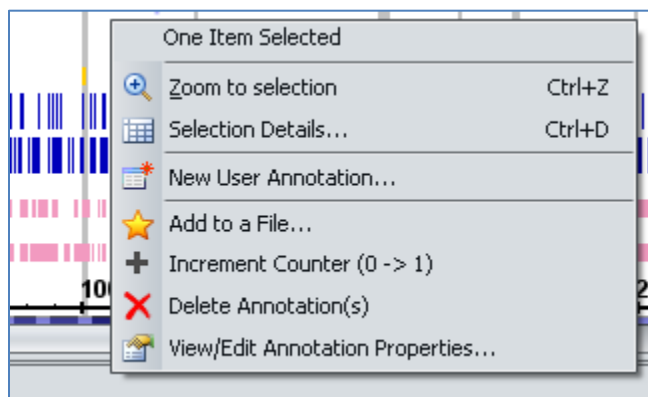
3. After editing the annotation properties, click **OK** in the Annotation Properties dialog box.

The section is saved in the Region file.

## **Deleting Regions from an AED File**

**To delete a region:**

1. Right-click on a region in a region file and select **Delete Annotation(s)**.



**Figure 10.11 Right-click menu for region**

The region is deleted from the Region (AED) file.

## Viewing or Editing AED File Properties

In ChAS, you can:

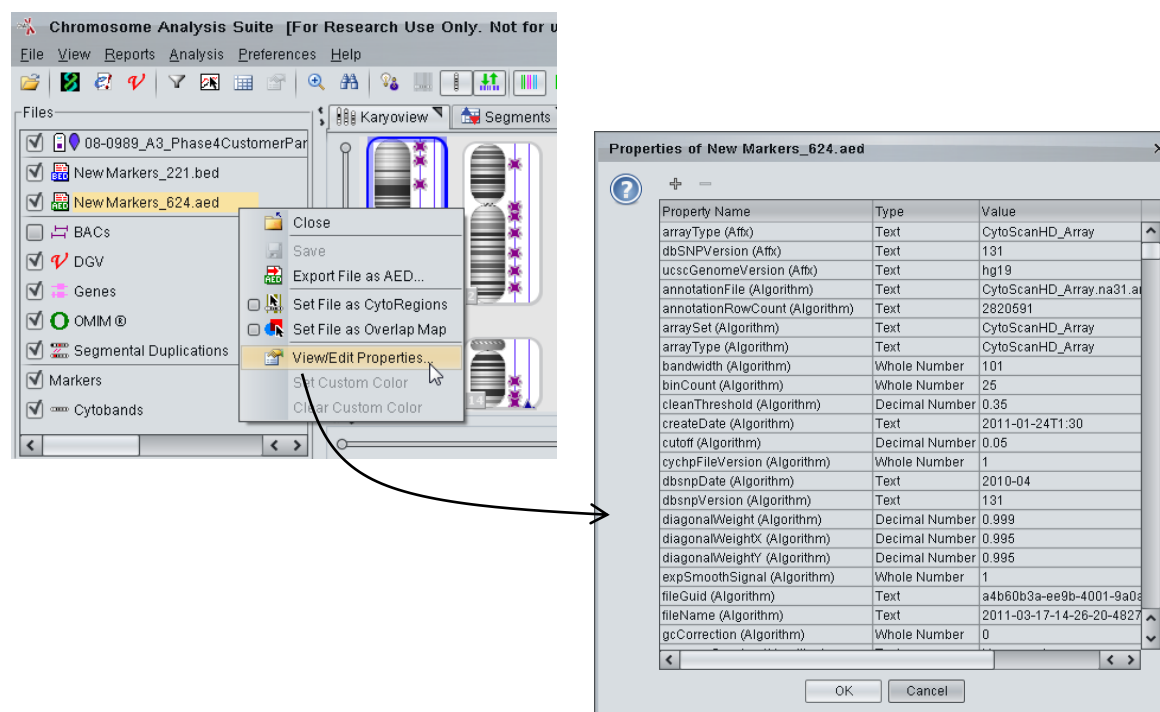
- View AED or BED file properties
- Edit AED file properties (modify, add, or delete properties)

 **Note:** BED files cannot be edited in ChAS. For more information on editing BED files, see page 159.

**To view file properties:**

1. Right-click the file and select **View/Edit Properties** on the shortcut menu.


The Properties dialog box appears.



**Figure 10.12 View AED or BED file properties**

## Viewing the Genome Assembly Version

The assigned genome assembly version of a loaded file can be viewed in the Properties box (Figure 10.12). The property, if it has been set for the file, is shown as “ucscGenomeVersion{Affx}”. An AED file that is created within ChAS is automatically assigned the same genome assembly version as the loaded NetAffx annotations (for example, “hg19”). If you add annotations to an existing AED file, its genome assembly version will not be modified; and if no version is specified for the AED file, no version will be assigned to it.


 **Note:** When you save a file as AED or BED, the current value of the genome assembly version property, if present, will be saved in the file. If two or more files are merged into an AED or BED file, the current value of the genome assembly version, if present in at least one of the files, will be saved in the merged file.

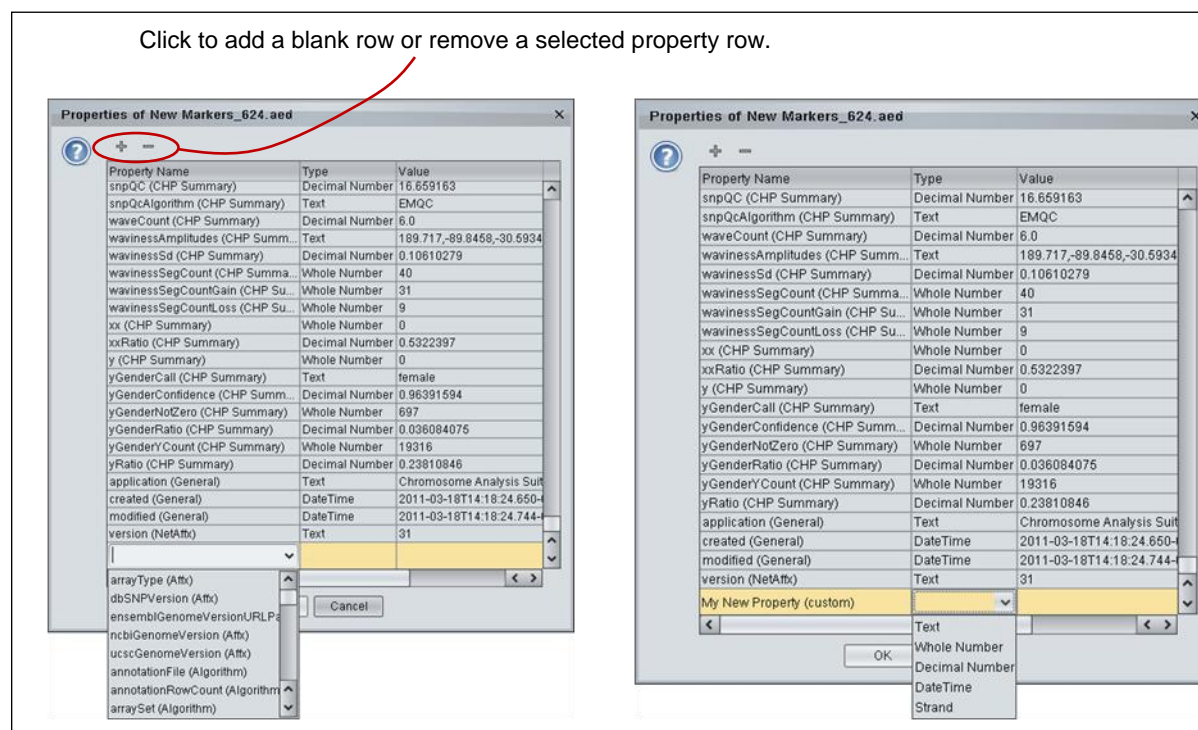
If an AED file does not include a genome assembly version, you can manually set it. To do this, in the Properties dialog box:

1. Add a new property row (see Figure 10.13).
2. Select the Property Name “ucscGenomeVersion(Affx)” from the drop-down list.
3. Select “Text” under the Type drop-down list.
4. In the “Value” column, enter the genome assembly version (for example, “hg19”).

 **Note:** You can manually set the genome version of an AED file by editing the “ucscGenomeVersion(Affx)” property. For more details on editing a property value, see page 160.Editing AED File Properties

#### To add a property:


1. In the Properties dialog box, click the  symbol to create a blank row.
2. In the “Property Name” cell, enter a property name or make a selection from the Property Name drop-down list.



**Figure 10.13 Specify a property name and type**

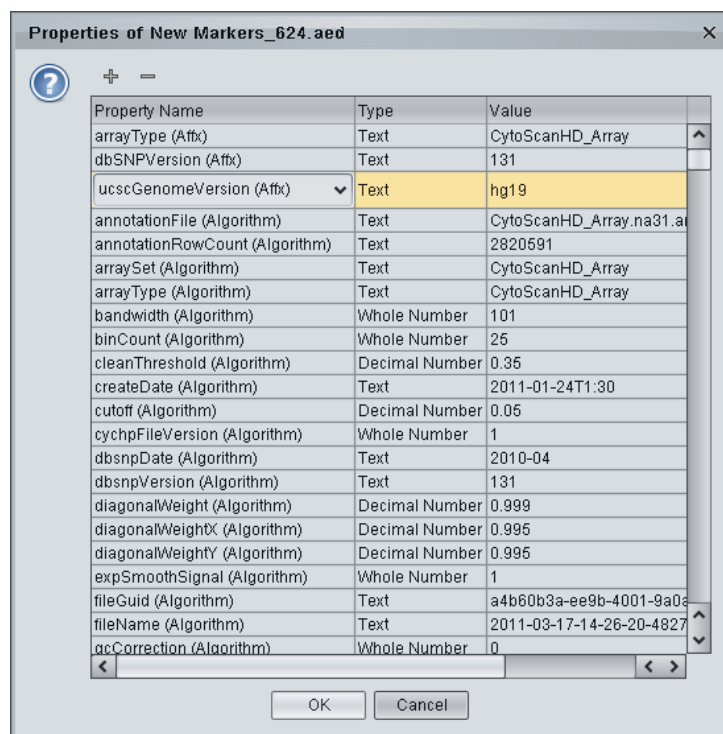
3. Click the “Type” cell and make a selection from the drop-down list.
4. Click the “Value” cell and enter the value.

#### To remove a property:

1. In the Properties dialog box, select the row that you want to delete.
2. Click the  icon.

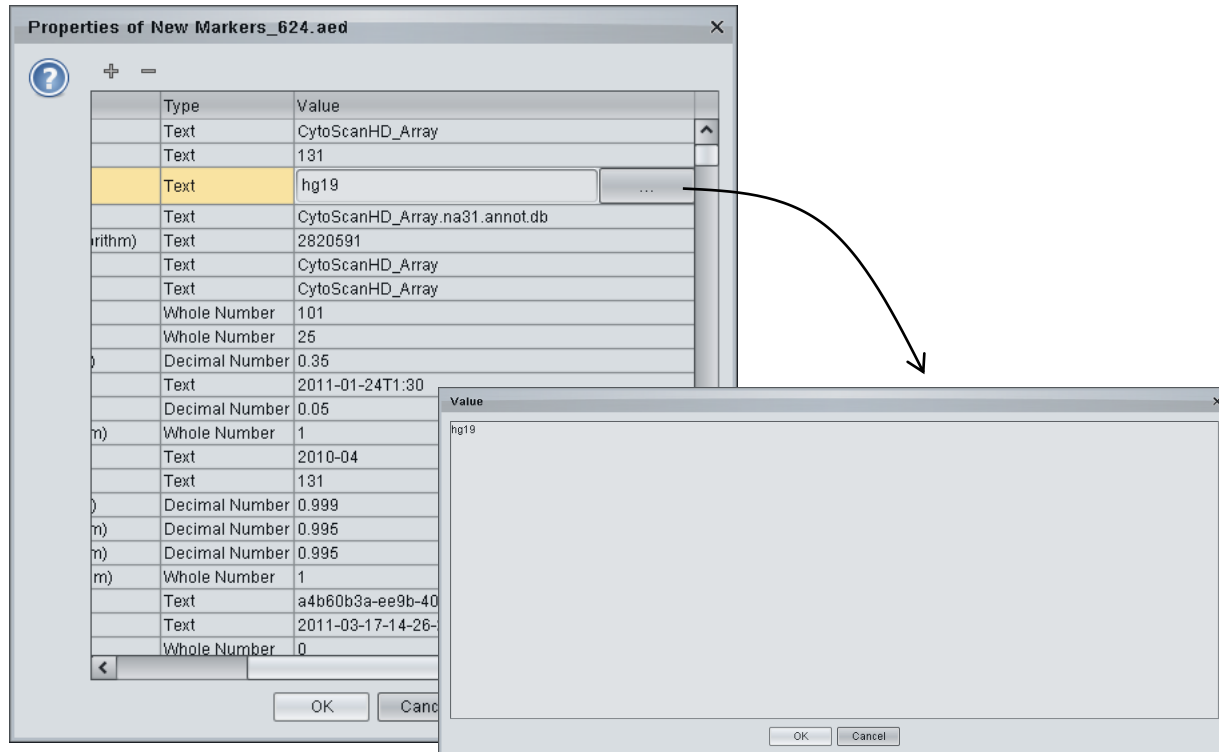
#### To change a property value:

1. Select the row in the Properties dialog box.



**Figure 10.14 Select a property to edit**

2. Double-click the "Value" cell, then click the Browse button that appears.



**Figure 10.15 Set a new property value**

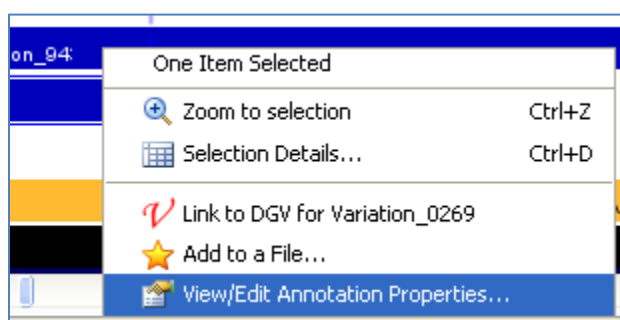
3. Enter a new value in the Value dialog box and click **OK**.

## Viewing and Editing Annotations

The Annotation Properties box opens:

- When adding a region to a Region (AED) file
- When you select View/Edit Annotation Properties from the right-click menu for the following types of features:
  - Detected Segments (view annotations only)
  - Reference Annotations (view annotations only)
  - Regions (view and edit annotations)

 **Note:** You can only edit annotations in AED files.



**Figure 10.16** Right-click menu for Reference Annotation in the Detail View

 **Note:** The View/Edit Annotation Properties menu option is not available if you have more than one feature selected in the Detail View.

The Annotation Properties dialog box has two tabs:

- [General](#) (below)
- [Additional](#) (page 164)

You can also create new user annotations if you select an element. This feature allows you to create a region that is not based on a segment or reference annotation. See [New User Annotations](#) (page 167).

## Entering General Information

The screenshot shows the 'Annotation Properties' dialog box with the 'General' tab selected. The fields are as follows:

- name**: seg1586
- chromosome**: chr3
- min**: 79551711
- max**: 81612828
- strand**: (dropdown menu)
- category**: lcsb
- note**: (large text area)
- reference**: (text area)
- counter**: (spin box)
- color**: (Choose Color button, Clear Color button)

At the bottom are 'OK' and 'Cancel' buttons.

Figure 10.17 Annotation Properties dialog box

### To edit general annotations:

1. In the General Tab, enter, edit, or select:

<b>name</b>	ID assigned to the region in the Region (AED) file.
<b>chromosome</b>	Cannot be edited in Annotation Properties box. See <a href="#">New User Annotations</a> , page 167.
<b>start</b>	Starting Position for the Region (can be edited)
<b>end</b>	Ending position for the region (can be edited).
<b>strand</b>	The Sequence Strand of the item.
<b>category</b>	Information on the source of the region. If the region was added by selecting a segment, the segment type is saved.
<b>note</b>	Information and comments about the region.
<b>counter</b>	Allows you to track the number of times something has been seen.

2. To change the color used for an individual annotation in the Detail View:

- A. Click **Choose Color**.

The Choose Color dialog box opens.



**Figure 10.18 Choose Color dialog box**


- B. Choose a color in the Swatches tab, or click the HSB or RGB tab to define a color.
- C. Click **OK** in the Choose Color dialog box.



**Important:** The color set in the Annotation Property dialog overrides the colors specified by a Color Rule created in the User Configuration dialog box. For more details on creating a Color Rule, see page 169.

### **Adding Custom Properties**

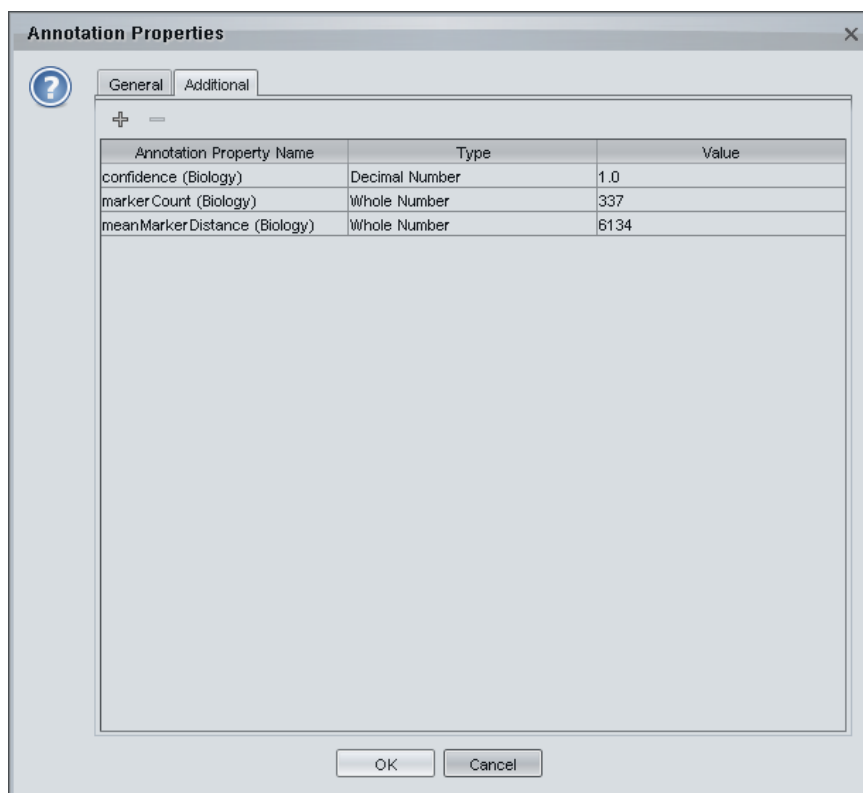
In the Additional Tab you can enter new annotation information for an AED annotation. The information will be displayed in the:

- Tooltip when you position the mouse arrow  over an annotation in the Details View
- Selection Details dialog box

#### **To add custom properties:**

1. In the Annotation Properties dialog box, click the **Additional** tab.





**Figure 10.19 Additional Annotation Properties tab**

The tab has three columns:

**Annotation Property Name** Name assigned to the property.


**Type** Type assigned to the property.

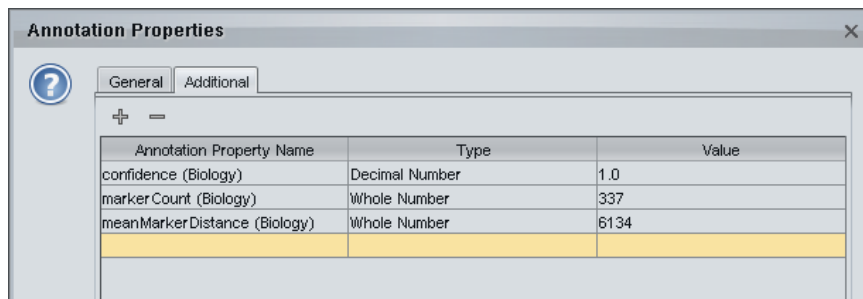
**Value** Value assigned to the property.

Different default properties may already be assigned to the annotations added from reference annotations.

2. Click the **Add**  button at the top of the table.

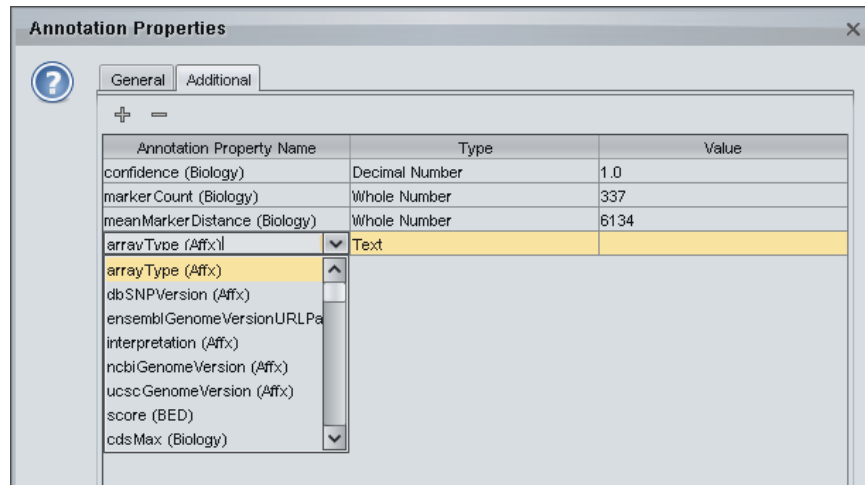
A new row appears in the table.

You can delete a property by selecting a row and clicking the **Remove**  button.



**Figure 10.20 New property row added**

- Click in the row under the Property Name column and enter a name for the property. For more details, see [Appendix C](#), page 246.

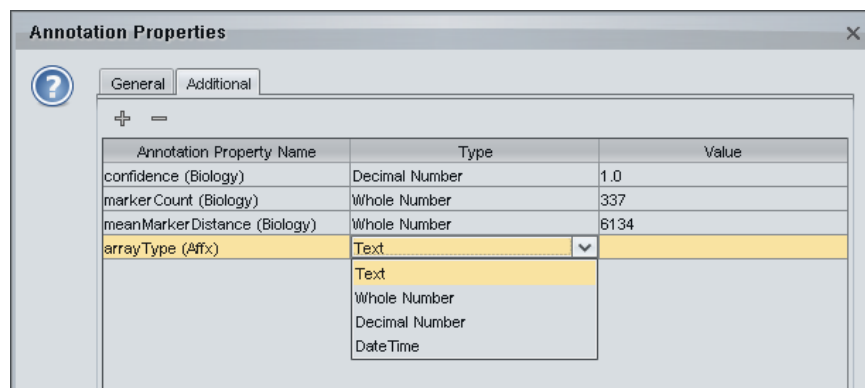


**Figure 10.21 Selecting property name**

- Click in the row under Type and select a property type from the drop-down list. For more details, see [Appendix C](#), page 246.

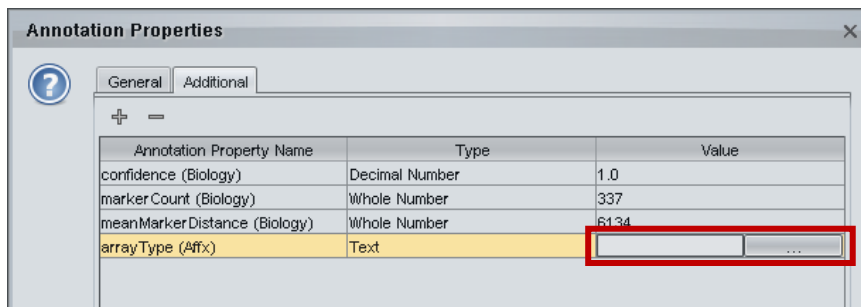
You can choose from:

- Text
- Whole Number
- Decimal Number
- DateTime



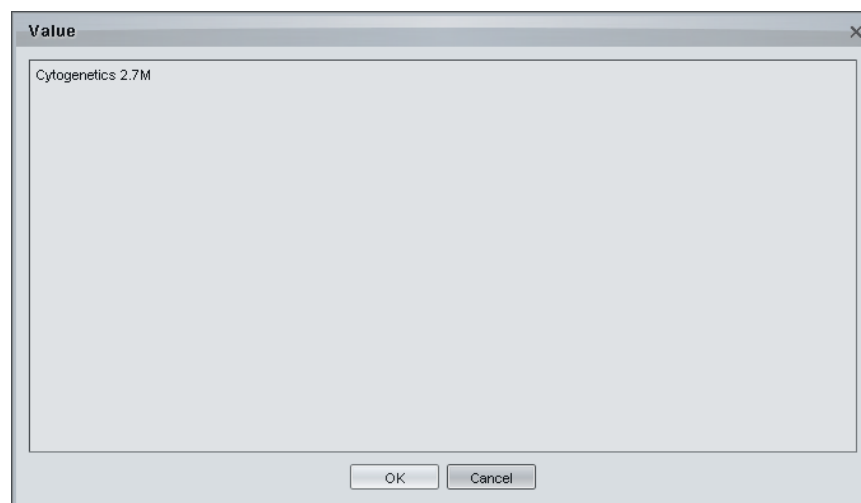
**Figure 10.22 Selecting property type**

- Click in the row under Value and enter the value.



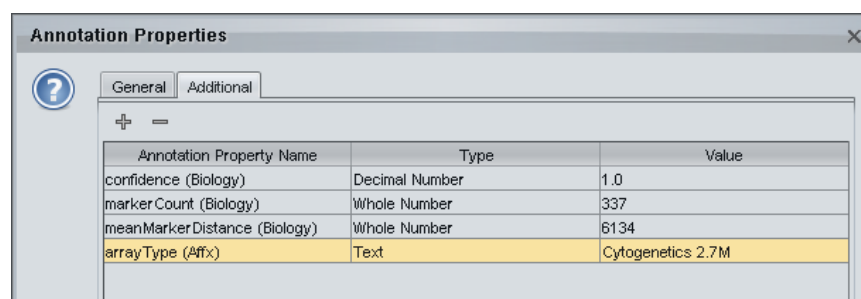
**Figure 10.23 Enter a value for the property type**

Alternatively, click the Browse button, and in the Value dialog box that appears, enter the property value and click **OK**.



**Figure 10.24 Value dialog box**

The property entry is completed.



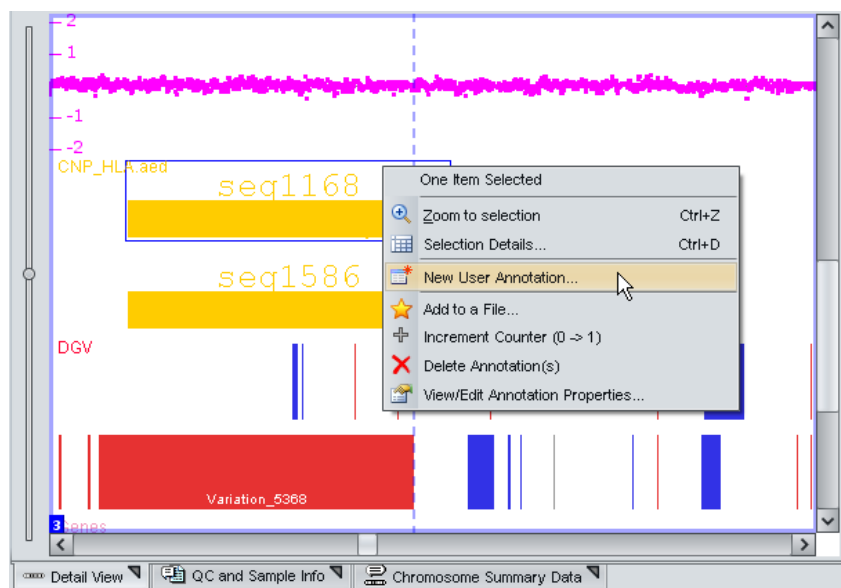
**Figure 10.25 Property entry completed**

6. Click **OK** in the Annotation Properties dialog box.

## New User Annotations

You can create a new annotation from an AED annotation.

1. In the Details view, right-click an AED annotation and select New User Annotation on the shortcut menu.



**Figure 10.26 Detail View**

2. In the dialog box that appears, enter the annotation information. For more details, see [Entering General Information](#), page 163 and [Adding Custom Properties](#), page 164.

**Figure 10.27 Enter information for the new annotation**

3. Click **OK**.  
The new user annotation is created and saved in the AED file



**Note:** The default New User Annotation information includes only the chromosome number. It does not include any information or properties associated with the AED annotation.

## AED/BED Color Rules

The Color Rules feature enables you to set display colors for annotations in AED files by various annotation properties, depending upon the original source of the region (detected segment, reference annotation, etc.). You can color annotations using the properties of the annotations, including:

- name
- category
- markerCount
- confidence

Using the color rules, you can assign a different color to annotations with different properties, making it easier to track the different types of data. For example, you could assign different colors to different categories:

- Mosaic
- Gain
- Loss

You can also perform comparisons for numerical values, coloring only values above or below a certain level.


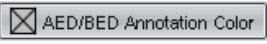
By default, ChAS displays the regions in an AED or BED file in a single color. There are several ways to edit annotation color:

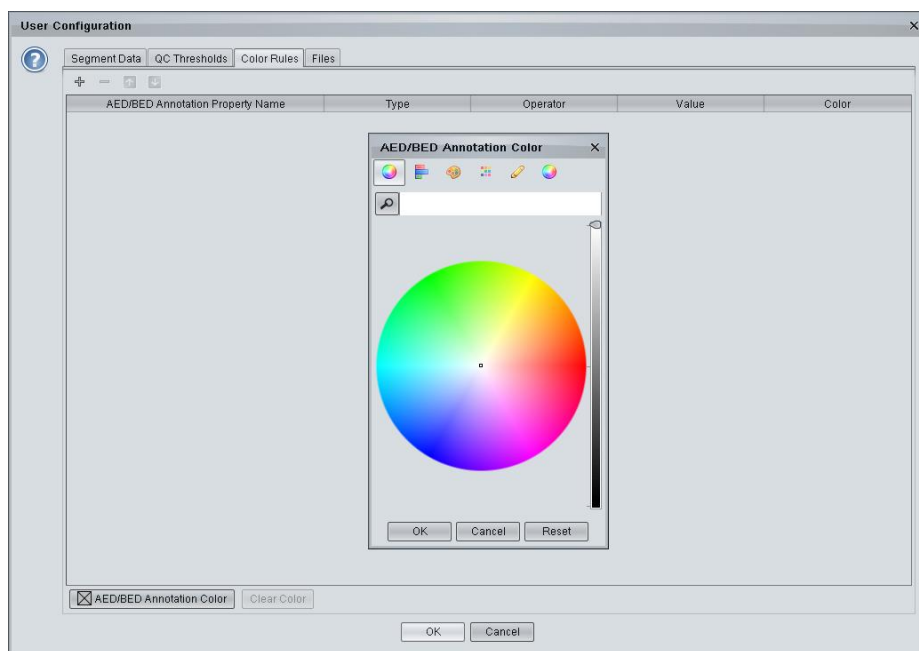
- Select a new color for all AED or BED files (see below)
- Create a color rule that specifies a color for annotations which meet user-specified requirements (page 170)
- In the Detail View, choose a color for a particular annotation in the Annotation Property dialog box (page 159).



**Important:** An AED annotation color set in the Annotation Properties dialog box (accessed in the Detail View) takes precedence over a color rule and the default AED/BED file color. A color rule can overwrite the AED or BED file default color.


**To select a default color for loaded AED or BED files:**

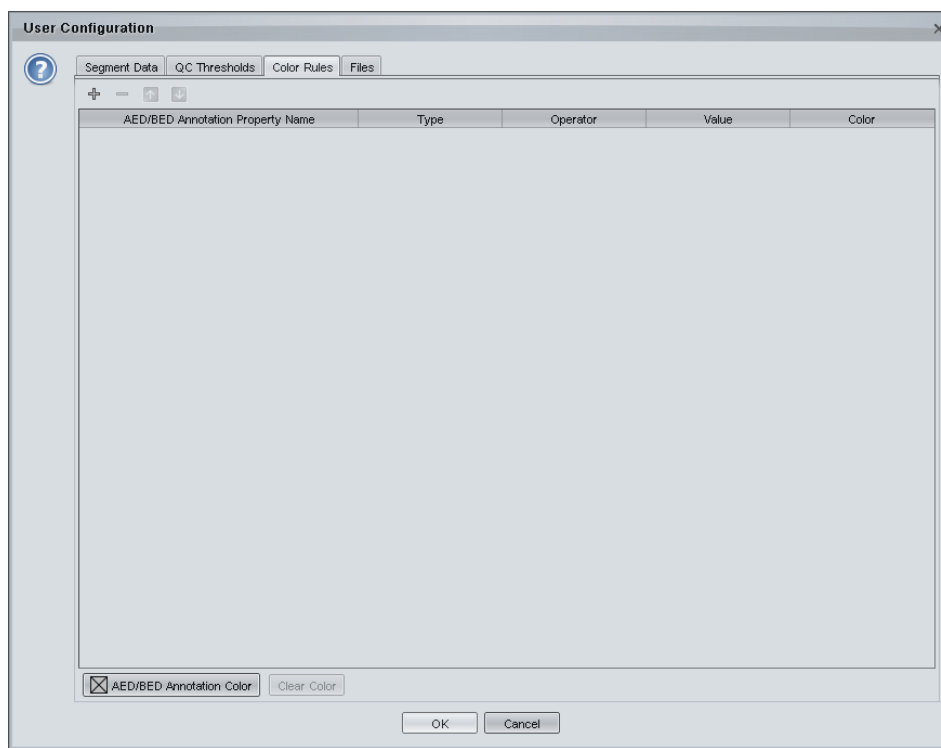
1. Open the User Configuration dialog box (click the  button or select **Preferences > Edit User Configuration** on the menu bar).
2. In the Color Rules tab, click the  button.
3. In the dialog box that appears, choose a color swatch or use the color controls to specify a color.
4. Click **OK** in the AED/BED Annotation Color dialog box.



**Figure 10.28 AED/BED Annotation Color dialog box**

**To create a color rule:**

1. Open the User Configuration dialog box (click the  button or select **Preferences > Edit User Configuration** on the menu bar).
2. Click the **Color Rules** tab.




**Figure 10.29 User Configuration dialog box, Color Rules tab**

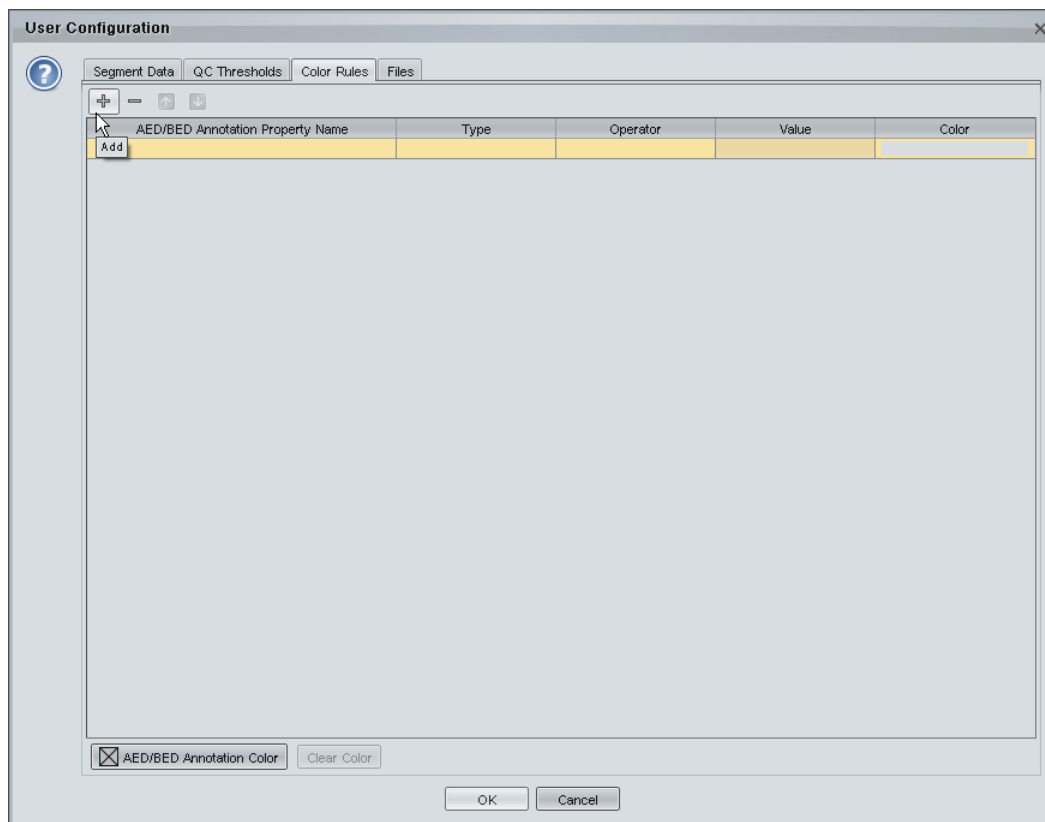
The tab has five columns:

<b>AED/BED Annotation Property Name</b>	Name assigned to the property.
<b>Type</b>	Type assigned to the property (for example, text).
<b>Operator</b>	Type of comparison with value performed .
<b>Value</b>	Value assigned to the property.
<b>Color</b>	Color assigned to the region property and value.

2. Click the **Add**  button at the top of the table.

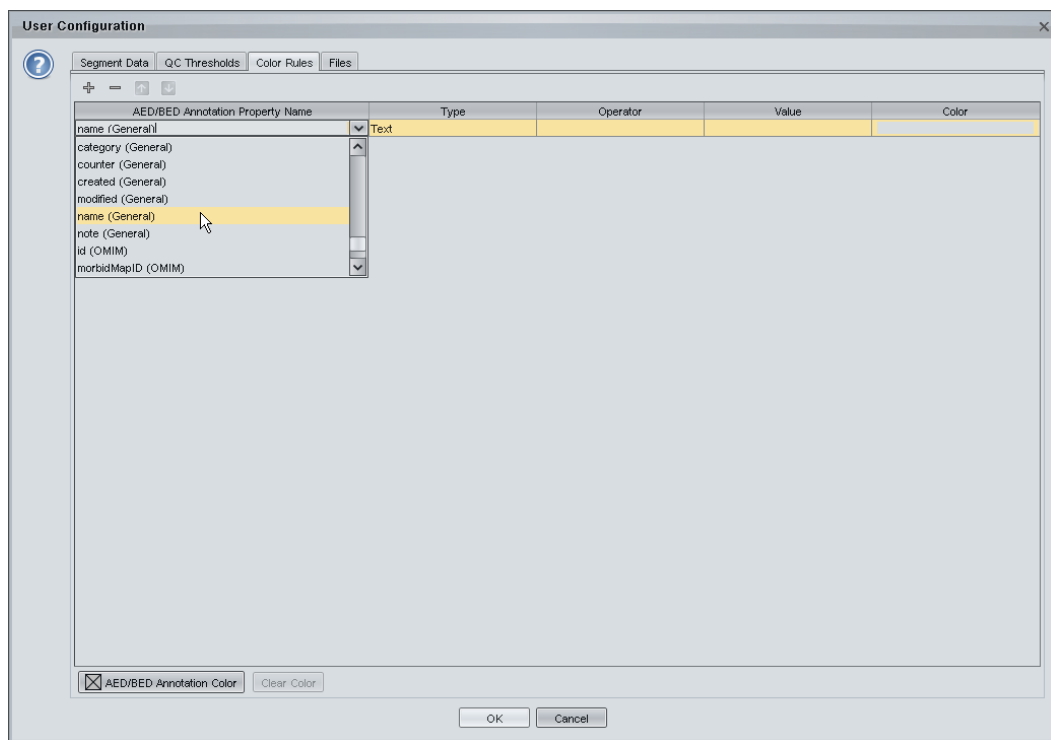
A new row appears in the table.

You can delete a property by selecting a row and clicking the **Remove**  button.



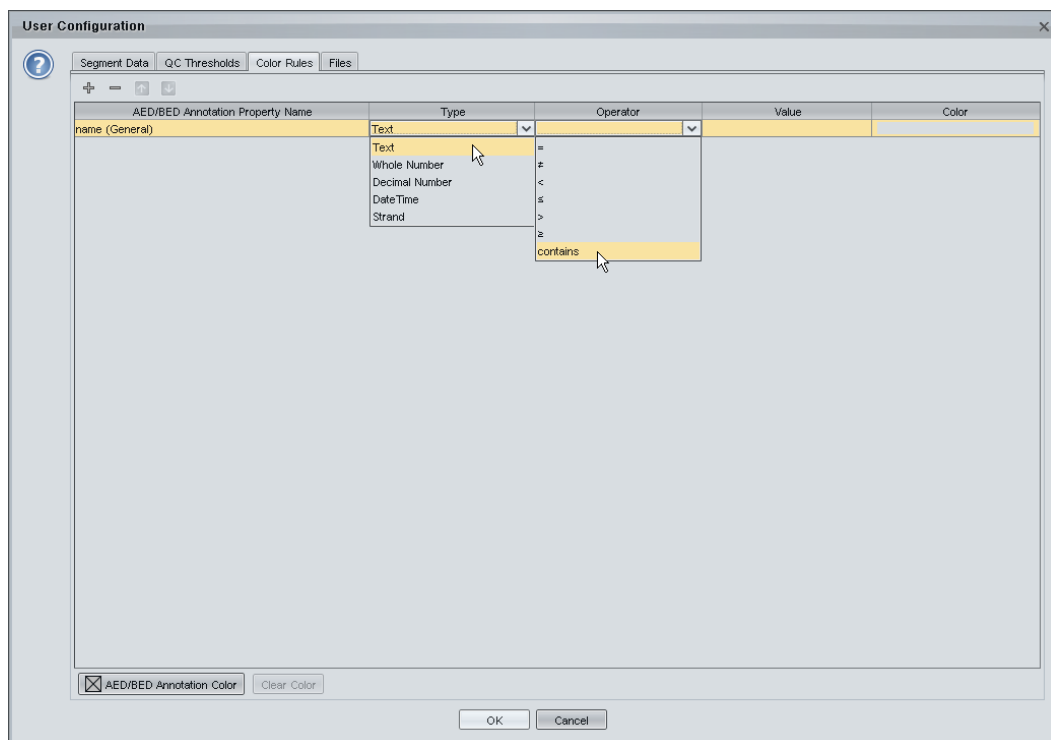
**Figure 10.30 New row added to the Color Rules table**

3. Click in the row under the Property Name column and enter a name for the property or select a property from the drop-down list. For more details, see [Appendix C](#), page 246.



**Figure 10.31 Enter or select a property name**

4. Click in the row under Type and select a property type from the drop-down list.
5. Click in the row under Operator and select an operator from the drop-down list.



**Figure 10.32 Selecting property type and operator for the comparison**



6. Click in the row under Value and enter a value for the property.
7. Click in the row under Color.

The Pick a Color dialog box opens.

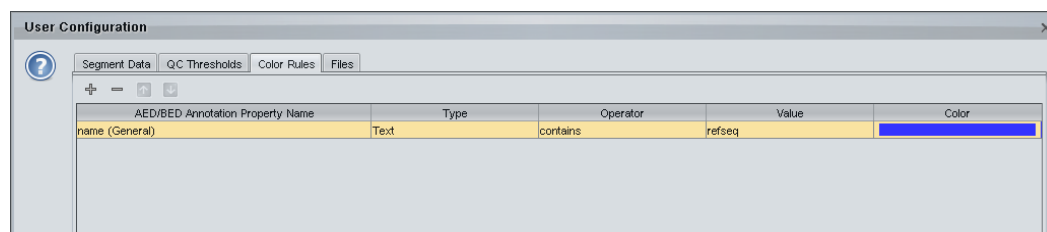


**Figure 10.33 Pick a Color dialog box**

The dialog box provides several options for selecting a color.

8. Select a color and click **OK** in the Pick a Color dialog box.

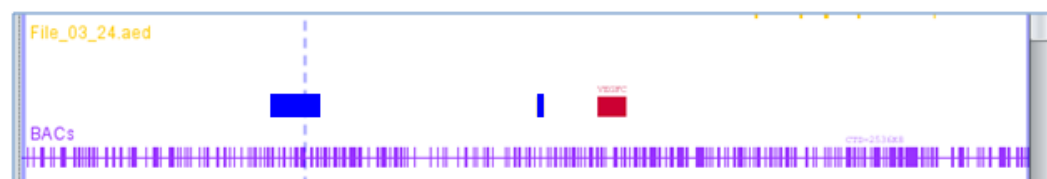
The property entry is completed.



**Figure 10.34 Property entry completed**

Click **OK** in the Color Rules tab dialog box.


Regions that satisfy the property comparison are displayed in the selected color in all views.



**Figure 10.35 Regions displayed in color**

## Exporting Information in AED or BED Format

You can export position data for the different features to AED or BED file format. The exported BED file contains only the names and locations of the detected segments. The exported AED file contains additional information, such as header information, feature ID, and hg version (which is the same as the NetAffx Genomic Annotations Database file loaded in the Browser at export).

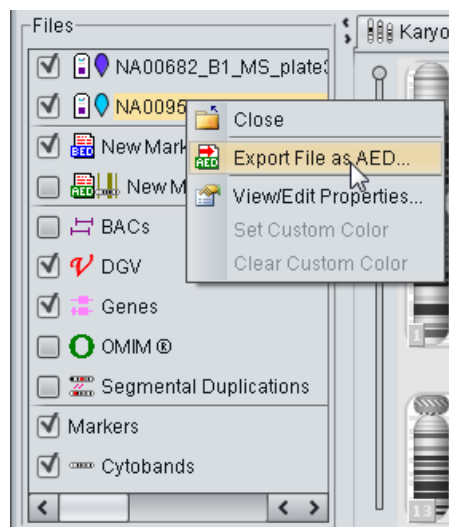
 **Note:** AED or BED files created in ChAS 1.0, 1.0.1, or 1.1 do not automatically include the hg version.

Position Data	Export to AED File	Export to BED File
Detected Segments for CYCHP/CNCHP files	Regions, names, and properties	Regions and names
Annotation Features in Reference Annotation files	Regions, names, and properties	Regions and names

**To export position data as an AED or BED file:**

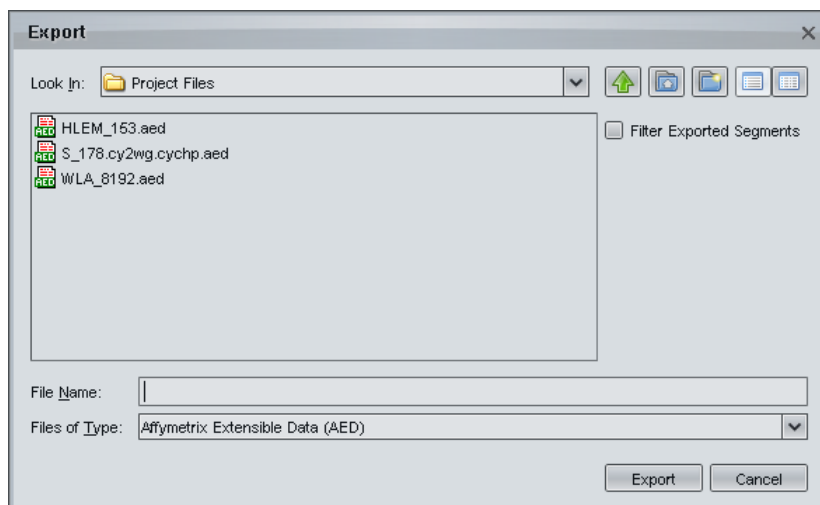
1. From the File menu, select **Export as AED...**

Alternatively, right-click the file in the Files list and select **Export File as AED** on the menu.



**Figure 10.36** Right-click menu

The Export dialog box opens.



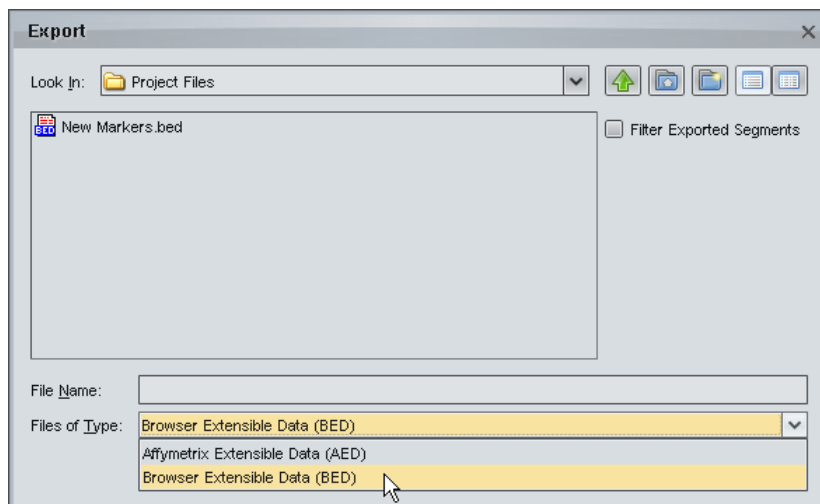
**Figure 10.37 Export dialog box, AED file format selected**

2. Choose the “Filter Exported Segments” option to restrict the export to the contents of the Segments table. If filters have been applied to the data, only the retained segments will be exported. Graph data and Chromosome Summary data will not be exported.

 **Note:** If this option is not selected, all segments which were loaded with the CxCHP file will be exported along with header information, regardless of whether filters are applied. The export includes all segment data, regardless of check mark status (ON or OFF) in the Files windowpane.

3. Select a folder location for the file using the navigation tools.
4. To export to AED file format, enter a name for the file.

To export to BED file format, enter a name for the file, and select “Browser Extensible Data (BED)” from the Files of Type drop-down list.



**Figure 10.38 Export dialog box, BED file format selected**

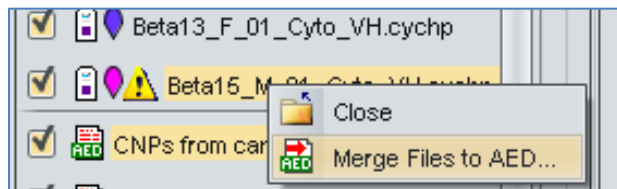
5. (Optional) Select the Filter Exported Segments dialog box to export only segments that meet filter criteria.
6. Click **Export** in the Export dialog box.

The AED file is exported and can be loaded as a region information file or sent to another user.

You can also merge feature position data from multiple different files.

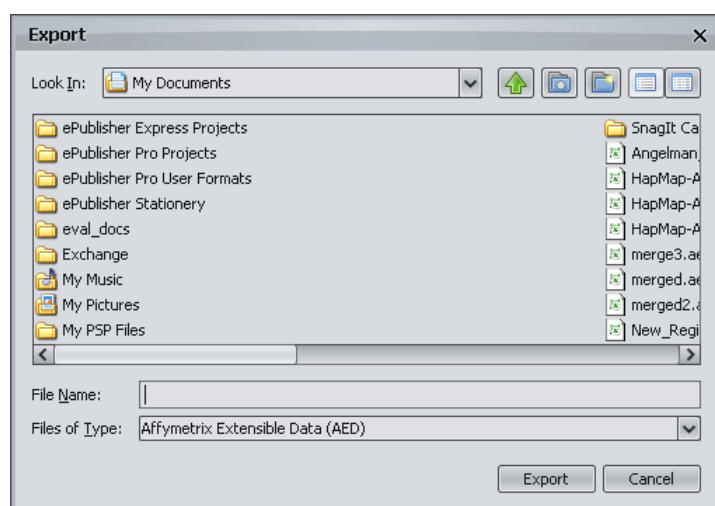
**To merge and export feature position information for multiple files:**

1. Select the files in the File List by clicking on them while pressing the CTRL key.
2. Right-click the selected files and choose **Merge Files to AED** on the menu.



**Figure 10.39** Files selected for merging AED outputs

The Export dialog box opens.



**Figure 10.40** Export dialog box

3. Use the navigation controls to select a folder for the merged AED file and enter a file name for the file.
4. Click **Export**.

The file with the merged AED region information is created and can be used as a region information file in ChAS.

 **Important:** After two AED files are merged, the original metadata in the header is not retained.

 **When two or more files are merged into AED/BED format, the current value of the genome assembly version property, if present in at least one of the files, will be saved in the merged file.**

## Chapter 11: Displaying Data in Table Views

The data in the CYCHP files can also be displayed and exported in tabular format, as well as the graphic representation in the Karyoview, Selected Chromosome View, and Detail View.

These tables include:

- [Selection Details](#) (page 125): Displays information about selected feature(s) in the graphic views.
- [CytoRegions Table](#) (page 140): Displays information about regions in the CytoRegions file.
- [Overlap Map Table](#) (page 149): Displays information about regions in the Overlap Map file.
- [Segments](#) (page 182): Displays information about detected segments in the sample files.
- [Graphs Table](#) (page 192): Displays graph data from the sample files.
- [Searching Results](#) (page 200)
- [Finding Intersections](#) (page 203)

This chapter covers:

- *Common Table Operations* (below)
- *Segments Table* (page 181)

File	CH Stat e	Type	Chro mos ome	Min	Max	Size (dop )	Max n Mark er Distance	Max % Over lap	Over lap Map Item s (% of Segment overlapped)	Cyto Regl one	Inter pret ation	Mark er Count	Conf iden ce	Cyto band Start	Cyto band End	Gen es	DGV	FIS H Clon es	sno/ mRN A	BAC s OM	Segment al Dupl icat ions	Sno othe r Join ed
08.0989_AS_Phase4Custom.ePMap_CfuoScan_P3_20110228.CfuoScanHD.chen.p	1.0	Loss	16	32277220	32872263	896043	4407		Overlap Map Not Set	Cytor egion : Not Set		136	0.8866743	p11.2	p11.2	TP63 TG3 TG3	Varia tion_ B, 4002, Varia tion_ TG3 3078 1, Varia tion_ 3117, Varia tion_ 3763 2, Varia tion_ 3897 8, Varia tion_ 7106, Varia tion_ 4844, Varia tion_ 3880 9, Varia tion_ 3717 9, Varia tion_ 7264 0, Varia tion_ -	N/A	N/A	RP11 - 7763 13, RP11 - 86L1 3, CTD- 2318 O3, CTC- 463H 6, CTD- 2848 N16, CTD- 3048 K26, RP11 - 161w 16, Varia tion_ 12A2 3880 9, CTA- 37w1 2, CTA- 174H 8, RP11 - 62P2 1, CTC-	chr16 -3293 2087, chr16 -3268 -6338, chr16 -3324 1271, chr16 -3387 3078, chr16 -2010 4241, chr7: 6601 6106, chr7: 6811 2444, chr7: 6480 1269, chr2: 9163 8714, chr2: 9041 3788, chr16 -1819 4483, RP11 -4284 1692, chr16 -3307	true

- *Figure 11.14* PDF file, page 3
- Graphs Table (page 191)
- *Chromosome Summary Data* (page 198)
- *Searching Results* (page 200)
- *Finding Intersections* (page 203)

## Common Table Operations








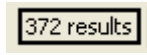

The controls that are common to all tables are described in this section. They include:

- [The Standard Toolbar Controls](#) (below)
- [Sorting by Columns](#) (page 178)
- [Ordering Columns](#) (page 178)
- [Selecting Columns for Display](#) (page 179)

### The Standard Toolbar Controls

The Toolbar provides quick access to table functions.


Some of the toolbar controls are standard to all tables.

At Left:	
	Export as TXT file. See <a href="#">Export Tables as TXT File</a> (page 217 ).
	Export as PDF Report. See <a href="#">Export Table Data in a PDF File</a> (page 212).
	Copy selected cells to clipboard. See <a href="#">Transfer to Clipboard</a> (page 218).
	Calculate the sum of the selected values from a numeric column.
	Display results for entire genome.
	Display results for selected chromosome.
	Display results for portion of chromosome displayed in Detail View.
At Far Right:	
	The number of rows in the table.
	Opens the Select Columns dialog box that enables you to choose the column headers to show or hide.

The Export functions are described in [Exporting Table Data](#) (page 212).

### Sorting by Columns

You can sort a table by a single column's values, or by the values in up to three columns.

 **Note:** You may sort on any column except, for reasons of efficiency, the marker name column in the graphs table.

Sorting on certain columns can cause a noticeable decrease in performance. For example, it is recommended that you do not sort a table using the columns in the Segments table that show the overlapping RefSeq, FISHClones, or other items. The data for these table cells is calculated only on an as-needed basis when it needs

to be displayed. Using such a column for sorting would force the calculation of the data for all such cells. Since the sorting would be alphabetical, it is unlikely to be useful. Similarly, for reasons of efficiency, sorting based on the marker name column in the Graphs table is not allowed.

### To sort a table by a single column:

- Click in the header of the column to sort the table by that column's values.

A triangle appears in the header.

- Triangle pointing upwards  indicates ascending sort order.
- Triangle pointing downwards  indicates descending sort order.

Click the header to toggle between Ascending, Descending, and no sort order.



**Note:** The “Type” column in the Segments table sorts segments based on the order that they appear in the Data Types windowpane, not in ascending/descending alphabetical order.

### To perform a multi-column sort:

- Click in the header of the first column.
- Click in the header of the second column.
- Click in the header of the third column, if desired.

The last selected column has sort priority.

Sort priority is indicated by the size of the triangle in the header.

CytoRegion Type	CytoRegion	Chromosome	Min	Max	▲ Size (kbp)	▼ Segment ID	▼ Segment File	Segment
Default User Annotation	CNP2477	20	20,000,000	20,110,000	60 seg120	▼	Beta10_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP606	3	163,996,361	164,106,575	111 seg684	▼	Beta10_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP2269	17	41,521,619	41,719,991	198 seg6392	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP12506	17	26,040,482	26,043,732	3 seg6312	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP2197	16	72,953,795	73,009,537	55 seg6151	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP2156	16	21,422,575	21,498,841	76 seg5977	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP147	1	194,997,668	195,068,695	71 seg587	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP148	1	195,089,940	195,168,372	78 seg587	▼	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP933	6	32,539,530	32,681,749	142 seg572	▼	Beta15_M_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP901	5	180,311,316	180,350,709	39 seg537	▼	Beta15_M_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP12104	13	69,640,329	69,670,896	30 seg5218	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP1952	13	68,149,981	68,166,243	16 seg5216	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP12099	13	66,945,140	66,954,900	9 seg5214	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP12094	13	65,094,528	65,103,708	9 seg5210	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP1946	13	63,122,789	63,134,693	11 seg5208	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP12093	13	64,198,454	64,209,619	11 seg5208	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss
Default User Annotation	CNP12091	13	63,227,094	63,303,323	76 seg5208	▼	Beta13_F_01_Cyto_VH.cychp	▼ Loss

Figure 11.1 Table sorted by descending order of Segment ID and ascending order of Size

## Ordering Columns

### To change the order of columns in a table:

- Click and drag in the column header to move the column to a new location.

### Selecting Columns for Display

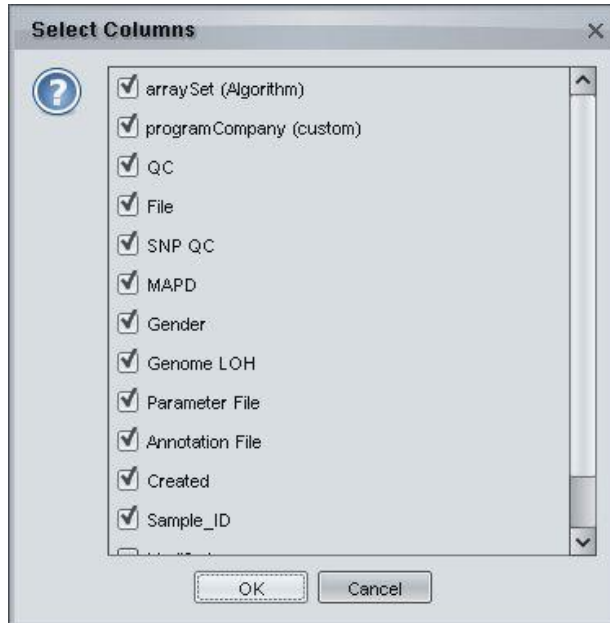
You can select columns using two methods:

- Select Columns dialog box
- From the menu that appears when you right-click on a column header

To select columns to display or hide:

1. Click the Select Columns toolbar button .

The Select Columns dialog box opens. The checkboxes for the columns are listed in the order in which the columns are currently displayed. Columns which are currently hidden in the table appear at the end of the list.



**Figure 11.2 Select Columns dialog box**

 **Note:** The specific items will vary, depending upon the type of table and data being displayed.

2. Select or deselect the checkboxes to display or hide the data and click **OK**.

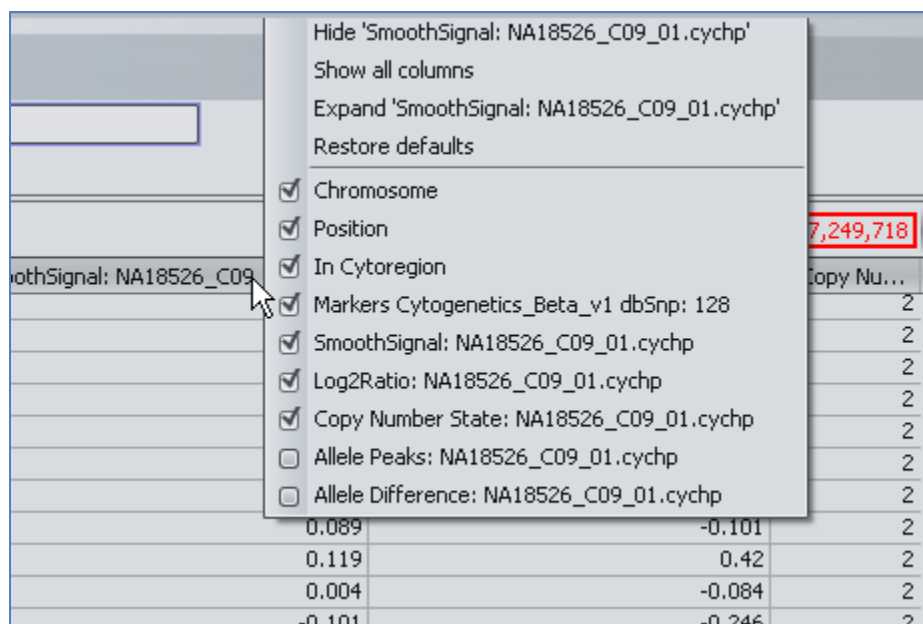
The selected columns are displayed. These choices are remembered between sessions.

### Right-click Menu

You can right-click in a column header to display a menu that will let you:

- Hide the selected column
- Show all columns, including hidden ones
- Expand selected column to display complete heading.
- Restore the default selection of columns.
- Select other columns to hide or show.





**Figure 11.3 Column heading right-click menu**

**To use the heading right-click menu:**

1. Right-click in the heading.

The menu appears.



**Note:** The specific items will vary, depending upon the type of table and data being displayed.

2. Select the desired option from the menu.

## Segments Table

The Segments table displays a list of the detected segments in the loaded sample data files.

Karyoview Segments CytoRegions Overlap Map Graphs											
File	▼ CN State	▼ Type	Chromosome	Min	Max	Size (kbp)	Mean Marke...	Max % Over...	Overlap Map...	CytoRegions	Use In Report
S_111...	4.0	▲ Gain	8	13,405,640	14,703,847	1,298	822				✓
S_111...	3.0	▲ Gain	X	58,186,054	62,269,568	4,084	19,261				✓
S_111...	3.0	▲ Gain	X	1,034,120	1,620,357	586	4,844			PAR1	✓
S_111...	3.0	▲ Gain	X	2,274,261	2,325,565	51	841			PAR1	✓
S_113...	3.0	▲ Gain	X	1,289,655	1,586,449	297	4,946			PAR1	✓
S_111...	2.466	◆ Mosaicism	8	10,389,935	17,167,028	6,777	1,165				✓
S_111...	2.326	◆ Mosaicism	X	50,129,131	92,389,241	42,260	1,912				✓
S_111...	2.228	◆ Mosaicism	X	99,724,529	107,568,304	7,844	1,342				✓
S_111...	2.22	◆ Mosaicism	X	125,958	3,337,170	3,211	1,775			PAR1	✓
S_111...	2.212	◆ Mosaicism	X	108,081,093	109,525,168	1,444	1,342				✓
S_111...	2.203	◆ Mosaicism	X	140,637,962	140,994,866	357	1,230				✓
S_111...	2.203	◆ Mosaicism	X	107,604,073	107,688,354	84	1,204				✓
S_111...	2.203	◆ Mosaicism	X	123,823,789	123,907,942	84	914				✓

**Figure 11.4 Segments Table**

The Segment Tab has:

- Toolbar
- Segments

There are certain columns in the Segments table which dynamically compute intersections of reference annotations with the segments. The data in these columns is computed on an as-needed basis for each cell. You may see text such as “<Working...>” in these cells while the data is being calculated. The results for all cells will be calculated when exporting to PDF or TXT, or copying to the clipboard. Hiding columns that are not needed may improve performance, particularly during export operations. Sorting the table based on columns could be slow.

You can also create a PDF report for selected segments only, using the specialized export controls of the Segments table. See [Export Selections in PDF File](#) (page 187).

**To highlight segments in the views or the table:**

- Double-click in a row of the table to zoom to the segment in the Karyoview, Selected Chromosome and Detail Views.
- Click on a segment in the Karyoview, Selected Chromosome or Detail View to highlight the segment in the Segments table.

**Toolbar**

The Toolbar provides quick access to table functions. The standard functions are described in [The Standard Toolbar](#), page 178.



**Figure 11.5 Segments Table Toolbar**

The Toolbar has three specialized buttons.

	Export Checked Rows to a PDF File
	Check all in column “Use in Report”
	Uncheck all in column “Use in Report”.

These buttons are described in [Export Selections in PDF File](#) (page 187).

The Results box displays the number of detected segments in the displayed sample files.

**Segments Table**

In the Segments table, “NA” means that the information (for example, FISH Clones or sno/mRNA) is not available in the NetAffx database because the information has not yet been mapped. For example, FISH Clones or sno/mRNA files will not appear in the Files list for the NA31 (hg19) ChAS Browser NetAffx Genomic Annotations file.

Annotations which share genomic coordinates with a segment are listed in order of start coordinate value, smallest values first (i.e. from left to right in the Details View). For annotations with the same start coordinate (for example, isoforms of a single gene), the one with the smallest end coordinate is listed before others with larger stop coordinates.

If a column in the Segments table contains more than 10 items, “...” is displayed after the 10<sup>th</sup> item to indicate that some data are not displayed in order to save calculation time. For example, “...” will follow the 10<sup>th</sup> name in the Genes column. However, a complete list of the genes will be included when the information is copied to the system clipboard or exported to reports. For gene isoforms with identical names, only one instance of the gene locus will be listed in the Segment table to reduce duplicate gene names.

The table displays each segment with the following information:

<b>File</b>	File the segment was detected in
<b>CN State</b>	Copy Number State (not displayed for LOH segment types).  The expected copy number state on the X chromosome in normal males is not constant over its entire length. This is due to the structure of the sex chromosomes. See <a href="#">Copy Number Segments on the X and Y Chromosomes</a> (page 40) for more information.
<b>Type</b>	Type of segment, for example, LOH. When sorting by this column, the segments of a particular sample are listed in the same order that they appear in the Data Types windowpane.
<b>Chromosome</b>	Chromosome on which the segment was found.
<b>Min</b>	Start position of segment.
<b>Max</b>	End position of segment.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>Size (kbp)</b>	Size of the segment.
<b>Mean Marker Distance</b>	Length of the segment in base pairs divided by the number of markers in the segment.
<b>Max % Overlap</b>	The highest percentage by which some item(s) in the Overlap Map overlaps the segment. Segments completely overlapped by an Overlap Map item are 100% overlapped. This number is used for Filtering Segments out by "Overlap".
<b>Overlap Map Items (% of Segment overlapped)</b>	Item(s) in the Overlap Map which overlap the segment, followed by the percentage by which the segment is overlapped by that Item.
<b>CytoRegions</b>	Names of the CytoRegions which the segment touches.
<b>Use in Report</b>	Allows manual selection of Segments for export to a Segments Table PDF, rather than all segments in the table.
<b>Marker Count</b>	Number of markers in the segment.
<b>Confidence</b>	Confidence score calculated by algorithm, a measure of the likelihood that a region is different from "normal".
<b>Cytoband Start</b>	Cytoband in which the segment begins.
<b>Cytoband End</b>	Cytoband in which the segment ends.
<b>Genes</b>	List of RefSeq genes from the Genes track that overlap with the segment.
<b>DGV</b>	List of DGV variations that overlap with segment.

<b>FISH Clones</b>	List of FISH clones that overlap with segment.
<b>sno/miRNA</b>	List of sno/miRNA features that overlap with segment.
<b>BACs</b>	List of BACs that overlap with segment.
<b>OMIM</b>	List of OMIM features that overlap with segment.
<b>Segmental Duplications</b>	List of Segmental Duplications that overlap with a segment.
<b>Smoothed/Joined</b>	Indication that segment was created by smoothing or joining two or more segments in the initial segment detection.
<b>Segment ID</b>	File-specific unique identifier assigned to the detected segment.
<b>Sample UUID</b>	Affymetrix array ID number of the array cartridge
<b>Max % Coverage</b>	The highest percentage by which a segment covers some item(s) in the Overlap Map.
<b>Number of Overlap Map Items</b>	Number of Overlap Map items which share genomic coordinates with the segment.
<b>% of Overlaps Map Item covered by Segment</b>	Overlap Map Item and the percentage by which it is covered by the segment

### **Obtaining All Annotations Associated With a Segment**

1. Select a segment in the Segments table, Karyoview, Detail View, or Selected Chromosome View  
The exact length of the selected segment fills the entire width of the Detail View.

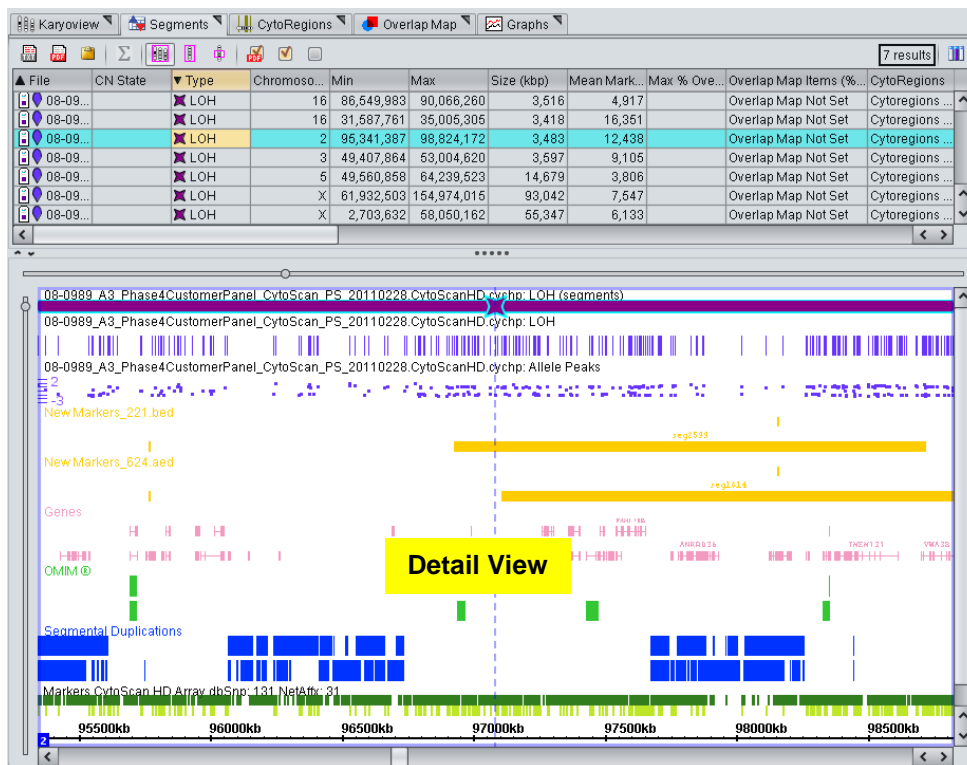

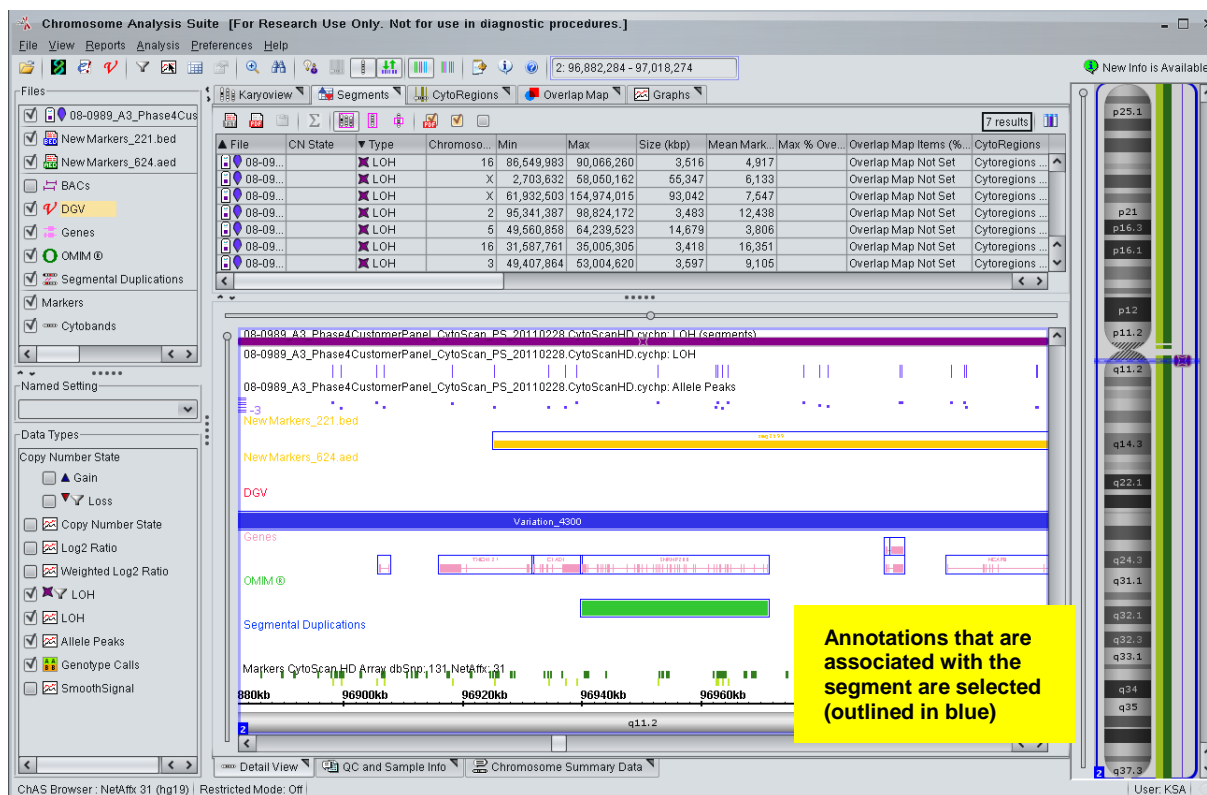



Figure 11.6 Detail View zoomed in on the segment selected in the Segments table

- Click the  toolbar button to expand all annotation tracks.

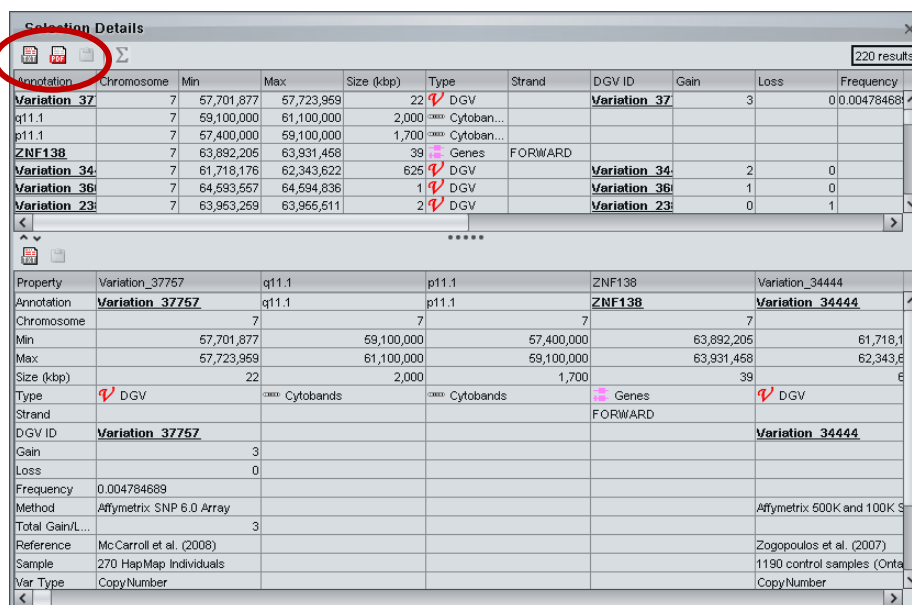


**Figure 11.7 Detail View, annotations have been expanded and selected**

- Using the mouse, draw a box around all of the genes and annotations of interest. When you release the mouse button, a blue box outlines the selected items in the Detail View.
- Click the  toolbar button.

The Selection Details table appears. It includes all of the items selected in the Detail View. For more details on the table, see page 125.

Click a button to export the information



Annotation	Chromosome	Min	Max	Size (kbp)	Type	Strand	DGV ID	Gain	Loss	Frequency
Variation_37	7	57,701,877	57,723,959	22	DGV		Variation_37	3		0.004784689
q11.1	7	59,100,000	61,100,000	2,000	Cytobands					
p11.1	7	57,400,000	59,100,000	1,700	Cytobands					
ZNF138	7	63,892,205	63,931,458	39	Genes	FORWARD				
Variation_34	7	61,718,176	62,343,622	625	DGV		Variation_34	2	0	
Variation_36	7	64,593,557	64,594,836	1	DGV		Variation_36	1	0	
Variation_23	7	63,953,259	63,955,511	2	DGV		Variation_23	0	1	

Property	Variation_37757	q11.1	p11.1	ZNF138	Variation_34444
Annotation	Variation_37757	q11.1	p11.1	ZNF138	Variation_34444
Chromosome	7	7	7	7	7
Min	57,701,877	59,100,000	57,400,000	63,892,205	61,718,176
Max	57,723,959	61,100,000	59,100,000	63,931,458	62,343,622
Size (kbp)	22	2,000	1,700	39	625
Type	DGV	Cytobands	Cytobands	Genes	DGV
Strand				FORWARD	
DGV ID	Variation_37757				Variation_34444
Gain	3				
Loss	0				
Frequency	0.004784689				
Method	Affymetrix SNP 6.0 Array				Affymetrix 600K and 100K S
Total Gain/Loss	3				
Reference	McCarroll et al. (2008)				Zogopoulos et al. (2007)
Sample	270 HapMap Individuals				1190 control samples (Ontario)
Var Type	Copy Number				Copy Number

Figure 11.8 Selection Details table

## Export Selections in PDF File

In the Segments table, you can select segments to include in a PDF report.





**Important:** The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.

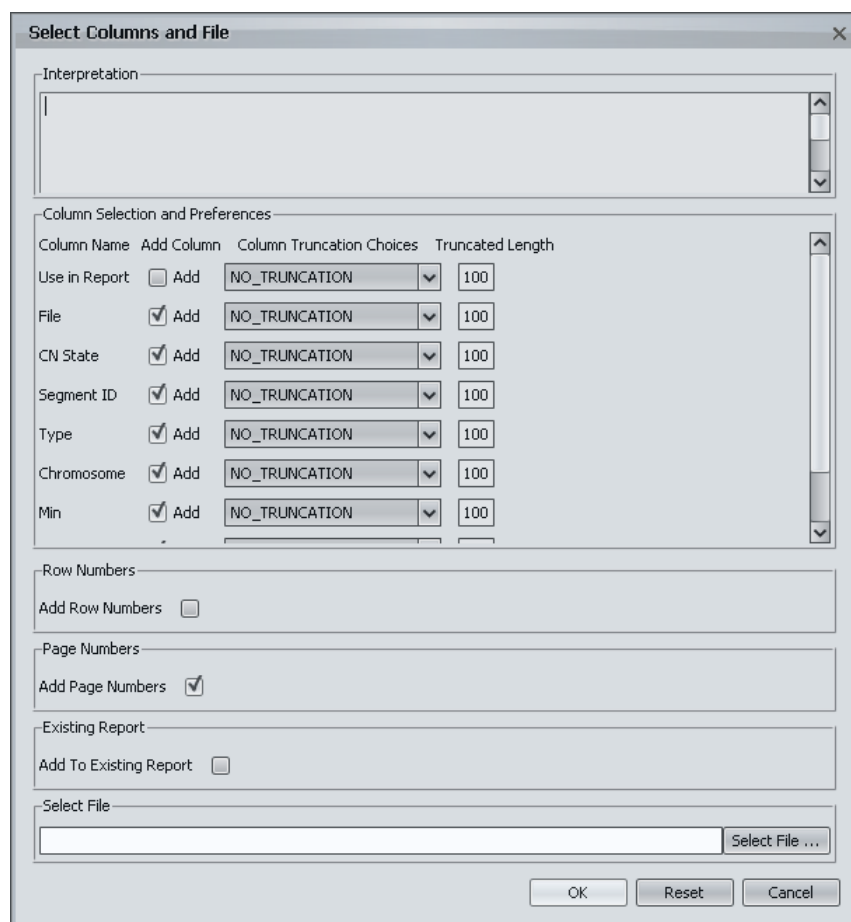


**Note:** You can also use the standard table export functions to export data from the Segment table, but you cannot select segments for export using the [Export Table Data in a PDF File](#) function (page 212).

### To export selected segments in a PDF file:

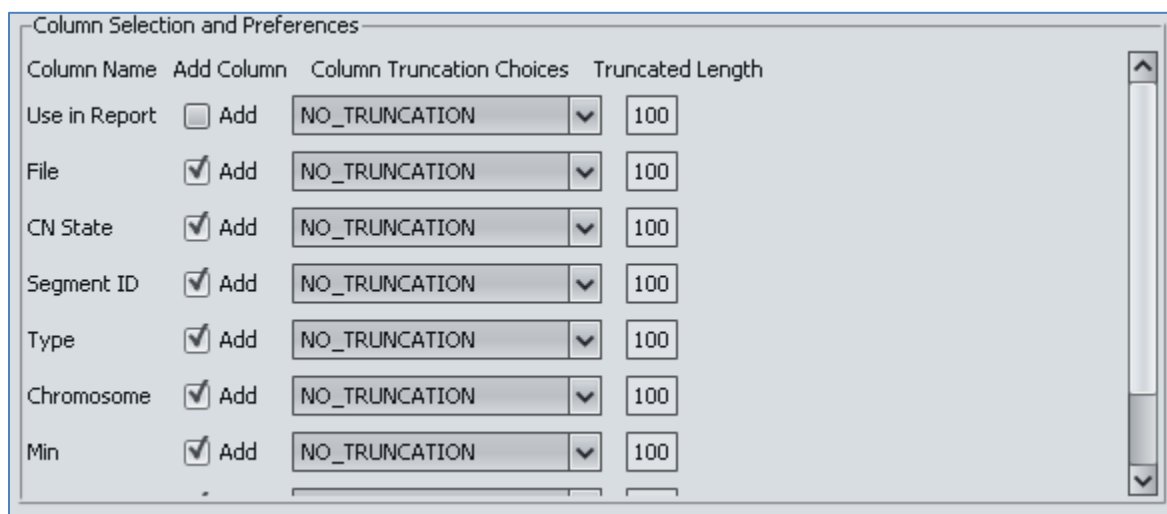
1. Select the columns for display.  
Limiting the columns of information will make the PDF easier to read.
2. Select and de-select the checkboxes in the Use in Report column to select segments for export.  
You can:

- Click the **Check All**  button in the table toolbar to select all segments.
  - Click the **Uncheck All**  button in the table toolbar to deselect all segments.
3. Click the Export Checked Rows button to export the checked segments to a PDF file.  
The Select Columns and File dialog box opens.



**Figure 11.9 Select Columns and File dialog box**

4. Select the columns to be displayed and the truncation rules for the content of the columns.



**Figure 11.10 Column Selection and Preferences**

**Column Name** Header of column in table.



**Add Column** Select checkbox to display column in PDF.

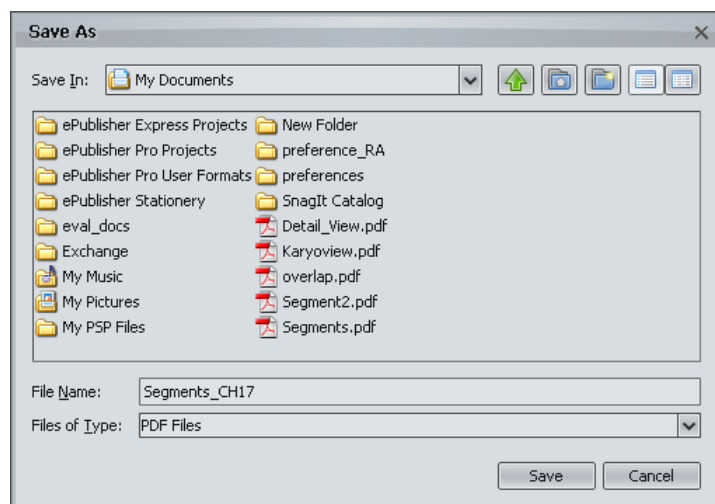
**Column Truncation Choices** Select Truncation Options:

- No\_Truncation: field is exported as is, using wrap-around if necessary.
- Truncate\_Beginning: Truncates content at the beginning of the field, leaving as many characters as specified in Truncated Length box.
- Truncate\_Middle: Truncates content in the middle of the field, leaving characters at the beginning and end, with ellipses (...) to mark the truncated characters.
- Truncate\_End: Truncates content at the end of the field, leaving as many characters as specified in Truncated Length box.

**Truncated length** Number of characters displayed after truncating the data.

5. Select the options for adding row numbers and page numbers, if desired.
6. Select the option for adding to existing report, if desired. See [Combining PDFs into a Single Report](#) (page 219).
7. Click **Select File...**

The Save As dialog box opens.



**Figure 11.11 Save As dialog box**

8. Select a folder location for the file using the navigation tools.
9. Enter a name for the PDF file, or select a file for the information to be appended to.
10. Click **Save** in the Save As dialog box.
11. Click **OK** in the Select Columns and File dialog box.

A PDF file is created with the selected segment data.

## Segments Table

For research use only. Not for use in diagnostic procedures.

### Data Files In View

File Label	File Name
08-0989_A3_Phase4CustomerPanel_CytoScan_PS_20110228.CytoScanHD.cychp	08-0989_A3_Phase4CustomerPanel_CytoScan_PS_20110228.CytoScanHD.cychp

Named Setting - Standard (400kbp and 50 markers, Gains and Losses)

### Genome Filters

Segment Type	Filter Type	Minimum Filter Value
Gain	Segment Length	400000
Gain	Marker Count	50
Loss	Segment Length	400000
Loss	Marker Count	50

Settings for 08-0989\_A3\_Phase4CustomerPanel\_CytoScan\_PS\_20110228.CytoScanHD.cychp

Setting Type	Setting
Smoothing	On
Smoothing Max Jump Limit	Off
Joining	50 Markers, 200 kbp

Segments Table - Page 1

Figure 11.12 PDF file, page 1

Setting Type	Setting
Joining Max Jump Limit	Off
Restricted Mode	Off

### Segments Details

Segments Table - Page 2

Figure 11.13 PDF file, page 2

File	CN State	Type	Chromosome	Min	Max	Size (kb)	Mean Marker Distance	Max % Overlap	Overlap Map Items (% of Segment overlap)	Cyto Regions	Interpretation	Marker Count	Confidence	Cyto band Start	Cyto band End	Genes	DGV	FISH Clones	SNP/mRNA	BACs	OMI M	Segmental Duplication	Sim other deleted
08-0989_A3_Phasu4Cuason-eRe-ne_C-fuSc-ar_P-S_20-1102-28.C-fuSc-sonD-ofen-P	1.0	Loss	16	32277220	32872263	596.043	4407		Overlap Map Not Set	Cytoregion Not Set		136	0.88657434	p11.2	p11.2	TP63TG3B, TP63TG3	Varia-son_4002, Varia-son_30781, Varia-son_3117, Varia-son_37632, Varia-son_38978, Varia-son_7106, Varia-son_4944, Varia-son_38839, Varia-son_37179, Varia-son_72540, Varia-son_...	N/A	N/A	RP11-776J13, RP11-66L13, CTD-231B10, CTD-453H8, CTD-42411, N16, CTD-3048K24, RP11-161M16, CTD-12A21, CTD-37M12, CTD-174H8, RP11-62P21, CTC-...	chr16:33232087, chr16:32665335, chr16:33241271, chr16:33573078, chr16:20104241, chr7:66016106, chr7:65112444, chr7:64601269, chr2:91636714, chr2:90413788, chr16:16194483, chr10:42641692, chr16:3307...	true	

Segments Table - Page 3

Figure 11.14 PDF file, page 3

## Graphs Table

The Graphs table displays the marker data used to create the graphs in the Detail view. Markers that are not used for the graphs currently displayed do not appear in this table. And, as in the Detail View, only markers from a single chromosome are displayed.

The column headings are colored according to the tracks used for the Karyoview, Selected Chromosome View, and Details View.

The Graphs table includes genotype SNP calls for CytoScan™ HD Array results (Figure 11.15).

Karyoview Segments CytoRegions Overlap Map Graphs									
Chromosome	Position	In CytoRegion	Markers	Genotype	Allele Pe...	LOH: NA...	Genotype	Allele Pe...	LOH: NA...
3	98,950,964	X	S-3ZZMMW	BB	-0.86	0	BB	-1.12	0
3	98,951,834	X	S-4OICY	AA		0	AA	0.84	0
3	98,952,443	X	S-3IYYV	BB	-1.44	0	AB	0.19	0
3	98,964,877	X	S-4CHOC	AA	0.73	0	AA	0.67	0
3	98,969,437	X	S-4HXZU	BB	-0.90	0	AB	0.00	0
3	98,971,751	X	S-3UIQV	AA	0.79	0	AA	0.94	0
3	98,976,066	X	S-4PLOX	AB	-0.01	0	AB	-0.04	0
3	98,987,381	X	S-4RZBP		-0.16	0		-0.11	0
3	98,987,607	X	S-3NFFM	AB	-0.18	0	AB	-0.19	0
3	98,988,001	X	S-3BEAX	AB	0.02	0	AB	0.11	0
3	98,988,161	X	S-3EGSQ	BB	-1.44	0	BB	-1.43	0
3	98,990,993	X	S-4NDAG	AB	0.01	0	AB	-0.06	0
3	99,013,299	X	S-4BPQE	BB	-0.98	0	BB	-1.07	0
3	99,013,781	X	S-3XSEU	AA	0.79	0	AA	0.85	0
3	99,013,836	X	S-3KYKV	BB	-0.76	0	BB	-1.12	0
3	99,013,936	X	S-3ASGB		-0.95	0		-0.95	0

Figure 11.15 Example Graphs table with Genotype Calls for CytoScan™ HD results

The table displays each data point in the displayed graphs.

#### To highlight markers in the views or the table:

- Click in a row of the table to place the cursor on the marker in the Chromosome and Detail Views.
- Click on a marker in the Selected Chromosome or Detail View to highlight the marker in the Graphs table.

The Graphs tab has the following elements:

- Toolbar
- Graph Data table

#### Toolbar



**Figure 11.16 Graphs Toolbar**


The toolbar allows you to:

- Export data in TXT format only. See [Export Tables as TXT File](#) (page 217)
- View data for chromosome and selected chromosome region only. You cannot display data from the whole genome in Graphs tab.
- Export genotype results

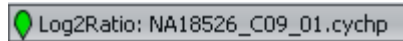
The Graphs Settings button  opens the Graph Settings panel, enabling you to change the style of graph, scale, and other features for the data graphs. See [Changing Graph Appearance](#) (page 116).

#### Graphs Table

The Graphs table displays:

<b>Chromosome</b>	Chromosome the marker is located on.
<b>Position</b>	Position where the marker is located in the chromosome.
<b>In Cytoregion</b>	Whether marker is located in a cytoregion or not: <input checked="" type="checkbox"/> In Cytoregion <input checked="" type="checkbox"/> Not in Cytoregion. See <a href="#">Using CytoRegions</a> (page 134) for more information.
<b>Markers</b>	Marker ID. Right-click to link to NetAffx information about the marker.  <b>Note: For efficiency reasons, it is not possible to sort the table on this column.</b>
<b>Genotype</b>	The SNP genotype call.

## Graph Data types




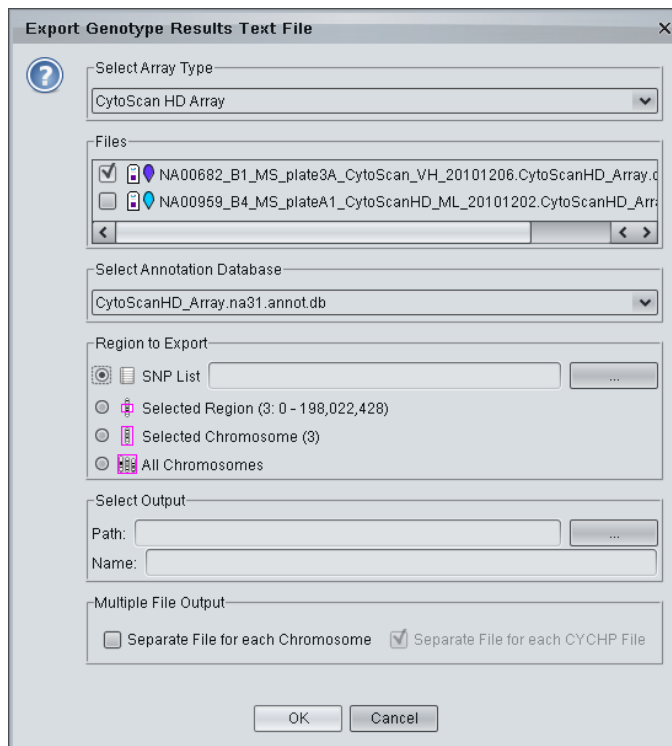
The Column header displays:

- Color Nib
- Data Type
- Name of sample file

The table cell displays the value for the marker.

## To export genotype calls:

1. Click the  toolbar button. Alternatively, select Reports > Export Genotype Results Text File from the menu bar.
2. In the dialog box that appears (Figure 11.17), select the array type, results file(s) (CYCHP), and annotation database to use for the export.




**Figure 11.17 Export Genotype Results**

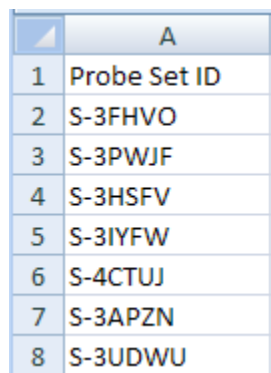
3. Specify the region to export:

**Export Option**

**Exports Genotypes for:**

SNP List	SNPs specified in a user-created SNP list (.txt).
Selected Region	SNPs included in the chromosome region selected in the Karyoview.
Selected Chromosome	SNPs on the chromosome selected in the Karyoview.
All Chromosomes	SNPs on all 24 chromosomes.

 **Note:** The SNP list should include one column header named “Probe Set ID” or “ProbeSet\_ID” and one probe set name per row. The export will not proceed without one of these column headers.



	A
1	Probe Set ID
2	S-3FHVO
3	S-3PWJF
4	S-3HSFV
5	S-3IYFW
6	S-4CTUJ
7	S-3APZN
8	S-3UDWU

**Figure 11.18 Example SNP list**

- Enter the path name or click the Browse button to select a folder for the output.
- Enter a file name prefix. If only one output file is created (see below), this will be the file name. If multiple files are created, a suffix will be added to this string to create the file name. Do not include the file extension here.
- Select a Multiple File Output option which determines if a separate file will be created for each chromosome and/or CYCHP file.

Selected Output Option(s)	Files Created
<b>None</b>	One output file will be exported that contains all chromosome and all CYCHP file data. There will be separate data columns for each CYCHP file in the exported file.
<b>Separate File for each Chromosome</b>	Creates a separate file for each chromosome in the output data. If all chromosomes are selected, 24 files will be created. There will be separate data columns for each CYCHP file in the exported file.
<b>Separate File for each CYCHP File</b>	Creates one text file per CYCHP file. Each file contains genotype calls for all chromosomes.
<b>Separate File for each Chromosome and Separate File for each CYCHP File</b>	Create a separate file for each CYCHP file and for each chromosome. If three CYCHP files are selected and all chromosomes are reported on, this will create 72 files.

- Click **OK**.


```

Genotypes_10624_NA00682_B1_MS_plate3A_CytoScan_VH_20101206.CytoScanHD_Array - Notepad
File Edit Format View Help
# Annotation DB Used: C:\ProgramData\Affymetrix\Chromosome Analysis Suite\Library\CytoScanHD_Array.na31.annot.db
# Array Type Name: CytoScan HD Array
# Array Type Internal Name: CytoScanHD_Array
# Export GUID: 57825efe-111e-4c1f-b246-0ac30e90a955
# Array Annotation Database NetAffx Build: 31
# UCSC Genomic version: hg19
# NCBI Genomic version: 37
# dbSNP Version: 131
# CHP File: C:\Users\Public\Documents\Results\NA00682_B1_MS_plate3A_CytoScan_VH_20101206.CytoScanHD_Array.cychp (NA31)
# Input Chromosome: X
# Input Region: 32467670 to 39977200
# Output Chromosome: All
Probe Set ID    Call Codes    Confidence    Signal A    Signal B    Forward Strand Base Calls    dbSNP RS ID
S-3FHVO AB      2.6645353E-15 1439.5713    1465.9479    AG    rs5972570    X    32467970
S-3PWJF BB      3.164554E-4    1819.0908    3027.3113    CC    rs6631589    X    32474756
S-3HSFV AA      3.6859404E-14 2010.413    633.6292    TT    rs6653872    X    32474793
S-3YFW BB      2.1022117E-10 553.2077    1964.6304    TT    rs58133832    X    32474885
S-4CTUJ BB      1.1914025E-11 587.60876    2515.3096    GG    rs73619093    X    32475588
S-3APZN AA      2.220446E-16    1630.4667    325.15146    CC    rs7886431    X    32477868
S-3UDWU AB      0.0    1046.0934    973.08466    GA    rs228392    X    32487337
S-4TMJH BB      0.0    205.25491    1140.2894    GG    rs1033462    X    32491061
S-4QILH AB      4.6629367E-15 343.2968    343.11258    CT    rs228397    X    32491351
S-3RURA BB      3.2804226E-10 1042.2152    2928.8074    CC    rs16998310    X    32491564
S-3XUYV AB      1.26565425E-14 1885.5472    2282.4426    TC    rs228403    X    32499010
S-4SGGR AA      8.881784E-16    1264.3896    256.46527    CC    rs5927077    X    32499987
S-4EVRJ BB      0.0    203.82622    902.2788    CC    rs7884521    X    32508989
S-4RTKS BB      0.0    215.45128    1334.9247    GG    rs16990410    X    32511411
S-3GSOW AB      6.661338E-16    855.1964    985.78174    AC    rs228314    X    32511992
S-4GEV AA      0.0    1065.4402    196.95076    CC    rs2023557    X    32515117
S-4NKSJ BB      0.0    542.1877    2601.7148    TT    rs41500547    X    32518960
S-3HOWD AB      0.0    1236.4248    1168.713    AG    rs228333    X    32536452
S-3WUFU AB      6.1460135E-9    1957.5522    2141.443    AT    rs228337    X    32539807
S-3HYJZ AB      8.245182E-12    1427.3093    1196.1893    GA    rs228338    X    32540070
S-3ZAGU AA      0.0    2544.556    436.40463    AA    rs73453749    X    32540396
S-3QCZV AB      2.0320567E-10 931.8176    614.83685    CT    rs119301    X    32543241
S-3RENZ NoCall 0.10250127    635.82434    1383.4319    CT    rs16998320    X    32543741
S-3WOT AA      0.0    3115.7256    554.69006    GG    rs170606    X    32544026
S-4JAXT AA      5.4622307E-11 1249.4127    410.03027    TT    rs5972592    X    32544255
S-4NXQF BB      2.220446E-15    370.76147    1828.4547    CC    rs228347    X    32548066
S-4FZYD AA      6.340586E-5    3566.6406    1767.4355    TT    rs17318147    X    32552118
S-4AZFO BB      7.19913E-12    526.51654    1727.5386    GG    rs7888911    X    32561313

```

**Figure 11.19 Example of exported genotypes from a selected region of chromosome X**

The exported text file includes information about the analysis (for example, array type, NetAffx annotation database, hg version, and chromosome).

 **Note:** If the option “Separate File for each CYCHP File” was not selected, many of the headers will be repeated for each CYCHP file. The header titles will be appended with the CYCHP file name to indicate which file the column belongs to.

The column headers report the following information:

Column Header	Description
Probe Set ID	Probe set identifier
Call Codes	Genotype call for the SNP.
Confidence	Confidence value for the call.
Signal A	Raw signal value for Signal A on the probe set.
Signal B	Raw signal value for Signal B on the probe set.
Forward Strand Base Calls	Base calls for the forward strand.
dbSNP RS ID	dbSNP RS ID value
Chromosome	Chromosome associated with the probe sets.
Chromosomal Position	Chromosome position of the SNP.

QC and Sample Info Tab

The QC and Sample Info tab in the lower pane displays information about the loaded Data and Region files.

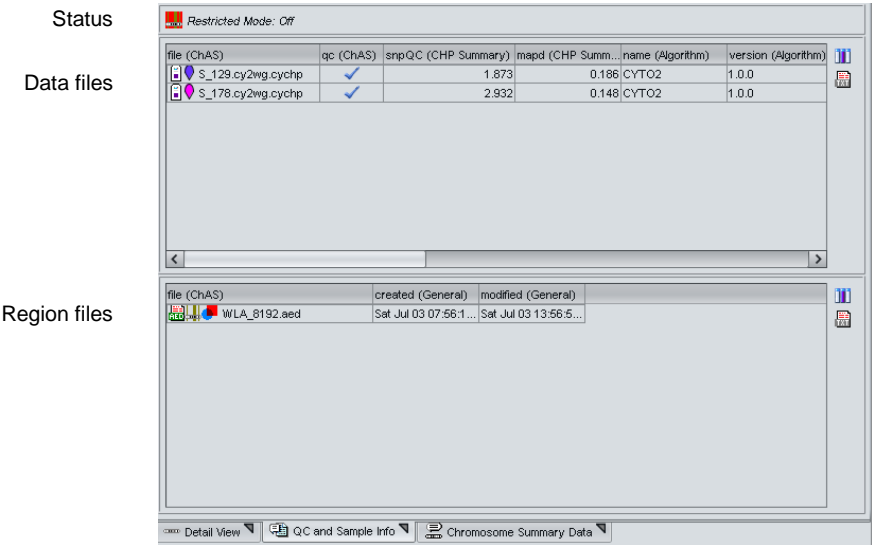


Figure 11.20 QC and Sample Info table

The top section displays Status for Restricted Mode (see [Using Restricted Mode](#), page 142)

The tables display information on:

- [Loaded Data Files](#) (below)
- [Loaded Region Files](#) (page 198)

Data Files Table

File	QC	SNP QC	MAPD	Gender	Genome LOH	Array Type	Parameter File
Beta10_F_01_Cyto_VH.cychp	✓	2.677	0.17	Female	0.262	Cytogenetic...	
Beta15_M_01_Cyto_VH.cychp	⚠	2.777	0.173	Male	0.228	Cytogenetic...	
Beta13_F_01_Cyto_VH.cychp	✓	2.544	0.171	Female	0.248	Cytogenetic...	

Figure 11.21 Data Files table

The table displays the following data by default.

File	File Name.
QC	In or out of QC bounds.
SNP QC	SNP QC score
MAPD	Median Absolute Pairwise Difference score.



<b>wavinessSd</b>	Waviness standard deviation (Waviness SD) is a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation.
<b>Name</b>	Algorithm name
<b>Version</b>	Algorithm version
<b>Sex Determination</b>	Gender call for the sample. See <a href="#">Gender Call Algorithms</a> below.
<b>AutosomeGenomeLOH</b>	The proportion of LOH on chromosome 1 to 22.
<b>GenomeLOH</b>	The proportion of LOH on chromosomes 1 to 22, chromosome X, and chromosome Y.
<b>Array type</b>	Type of array used in analysis.
<b>Parameter File</b>	Name of the chasparam file used to create the CYCHP file
<b>Annotation File</b>	Name of the Annotation file used to create the CYCHP file
<b>Reference File</b>	Reference file used in the single sample analysis
<b>Created</b>	Date file was created.
<b>Modified</b>	Date file was last modified.
<b>Sample ID</b>	Sample identifier in the ARR sample file.

Other data from the header of the Sample Data file or the ARR file can also be selected for display in the Select Columns dialog box.

You can only hide or display columns by using the Column Select dialog box, at the right of the table.

### Gender Call Algorithms

The table below explains which algorithm is used to make the gender calls for the different arrays.

The CytoScan™ HD Array and Cytogenetics Whole-Genome 27M Array use the call “Y-gender” which gives a male/female call.

Depending on the version they were created under, various GTC 2.x and 3.x SNP6 CNCHP files use other gender calls present in their CNCHP file header.

These calls used from the CNCHP file header are NOT the same gender calls used for those files in GTC, since the GTC-displayed gender calls were stored in GQC or CN\_SEGMENTS files which are not supported in ChAS.



**Note:** For more details how the array-specific algorithms call LOH segments for the X or Y chromosome, see [LOH Segments on the X and Y Chromosomes on page 40](#).

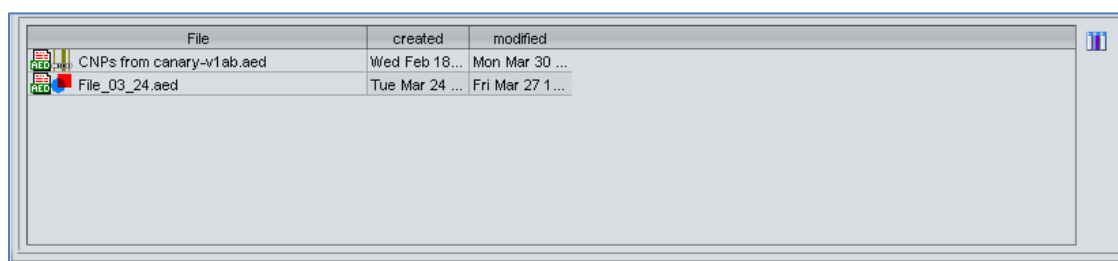
**Table 11.1 Gender call algorithms**



Software/Array Type	Gender Call Algorithm	Call	Gender Call Confidence
---------------------	-----------------------	------	------------------------

Software/Array Type	Gender Call Algorithm	Call	Gender Call Confidence
ChAS 1.0, ChAS 1.1, ChAS 1.2 / CytoScan™ HD Array, Cytogenetics Whole-Genome 2.7M Array	Y-gender	male/female	yes
GTC 3.0 to GTC 4.1/ Genome-Wide SNP Array 6.0	affymetrix-chipsummary-Gender	male/female/unknown	no
GTC 2.1/ Genome-Wide SNP Array 6.0	affymetrix-chipsummary-hasY	male/female	no
GTC 2.0/ Genome-Wide SNP Array 6.0	affymetrix-chipsummary-hasY	male/female	no

## CytoRegion Files Table

This table displays information about the loaded Region (AED or BED) files.



File	created	modified
 CNPs from canary-v1ab.aed	Wed Feb 18...	Mon Mar 30 ...
 File_03_24.aed	Tue Mar 24 ...	Fri Mar 27 1 ...

**Figure 11.22 CytoRegion Files information**

The Region information files section displays information on:

**File** File Name with Icons displayed if selected as Overlap File or CytoRegions File).

**Created** Date and time file was created.

**Modified** Date and time file was last modified.

You can only hide or display columns by using the Column Select dialog box, at the right of the table.

## Chromosome Summary Data

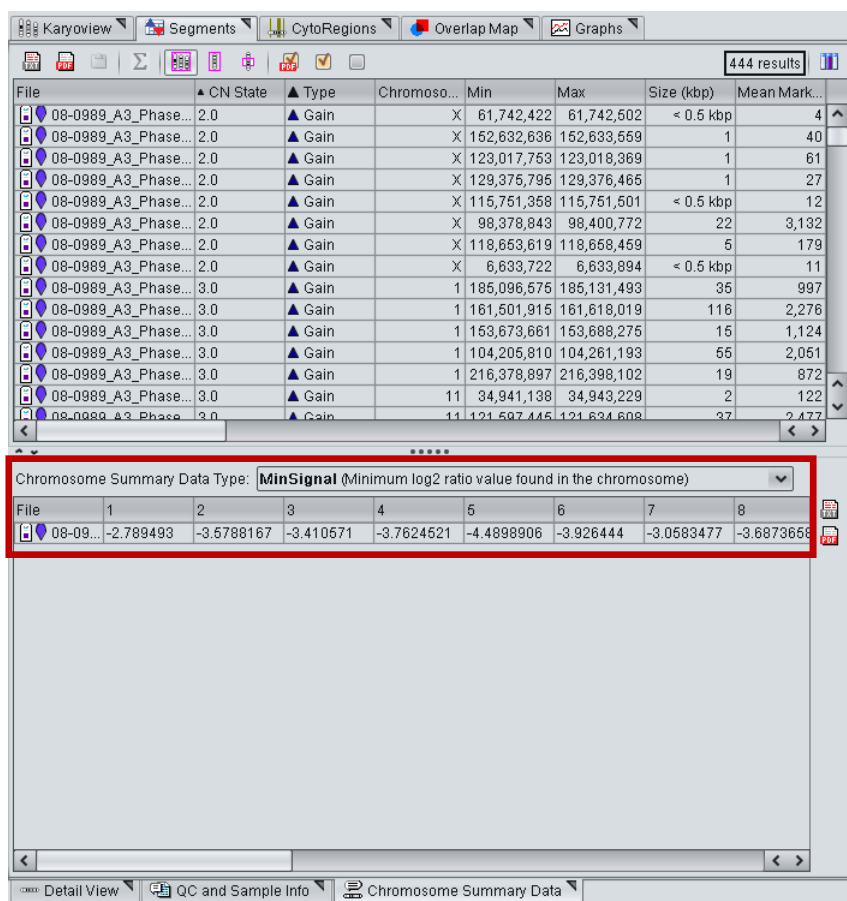
The Chromosome Summary Data table summarizes particular data across each chromosome in the loaded sample data files. The available data types are:

- Min Signal – minimum log2 ratio value found in the chromosome
- Median Signal – median log2 ratio value found in the chromosome
- Max Signal – maximum log 2 ratio value found in the chromosome
- Median CN State – median calibrated log2 ratio
- Mosaicism – median mosaicism mixture value

**Note:** The mosaicism CN state value is not an integer due to cell populations with different CN state values. In the Chromosome Summary Data table, the mosaicism value indicates how much the median CN state value is above or below two, the normal CN state value. For example, a median mosaicism mixture CN state value of 2.48 is displayed as 0.48 in the Chromosome Summary table.

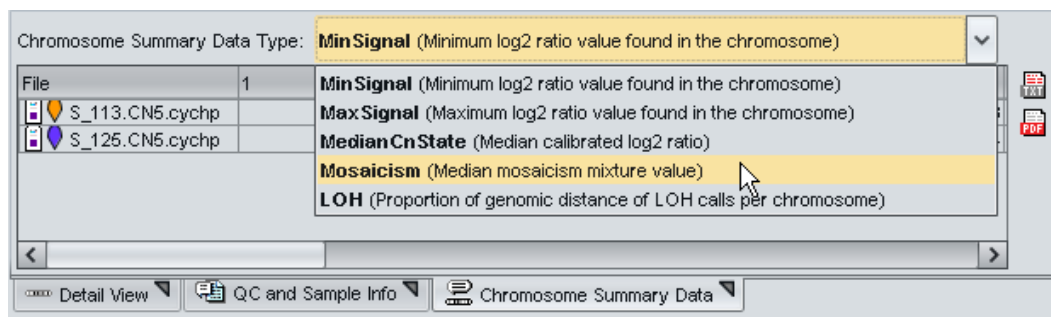
**Note:** Mosaicism (median mosaicism mixture value) for normal males is -1.0 for chromosome X and -1.0 for chromosome Y. Mosaicism for normal females is 0.0 for chromosome X and -2.0 for chromosome Y. A mosaic XO female is treated the same as a mosaic autosomal monosomy (i.e., the mosaicism level of chromosome X will be between -1.0 and 0). A mosaic XXY male is also given a mosaicism level between -1 and 0 for chromosome X.

- LOH – proportion of genomic distance of LOH calls per chromosome



**Figure 11.23 Chromosome Summary Data, Min Signal in each chromosome**

To choose the data type, make a selection from the drop-down list.



**Figure 11.24 Selecting the data type for the Chromosome Summary table**

## Searching Results

The Search function allows you to search:

- Detected Segments
- Reference Annotations
- Loaded Region Information Files

The search can find:

- Names of Reference Annotations
- BED and AED file elements, including those in files designated as CytoRegions or Overlap Maps
- Loaded and displayed segments

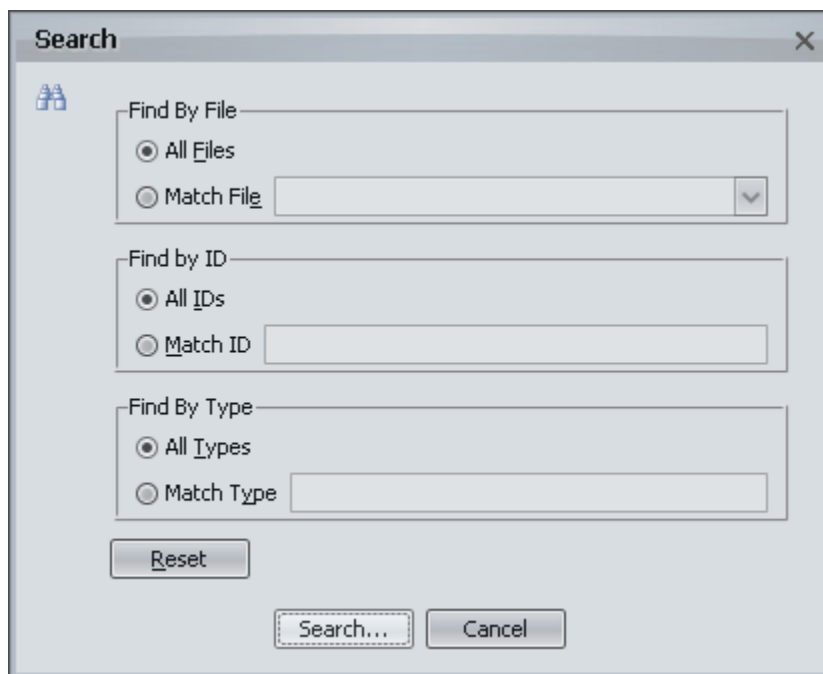
You can search by:

- File (select the files to be searched)
- ID label
- Type

### To perform a search:

1. From the View menu, select **Search**.

The Search dialog box opens.



**Figure 11.25 Search dialog box**

2. Search all files in the File List.

Alternatively, click **Match File** and select a file to be searched from the drop-down list.



**Note: Only files that are check marked in the Files List appear in the Match File drop-down list. Only files that are check marked in the Files list will be searched.**

3. Select Match ID and enter an ID (not case sensitive) for the search.

You can use the “\*” character as a wild card for the search.

4. Select Match Type and enter a type name (not case sensitive) for the search.



**Note: You can use the asterisk symbol \* as a wildcard. Using the wildcard is very useful when you want to search for the type “OMIM ®” because the symbol ® does not need to be entered. Another convenient way is to simply choose the OMIM file from the “Find By” File list.**

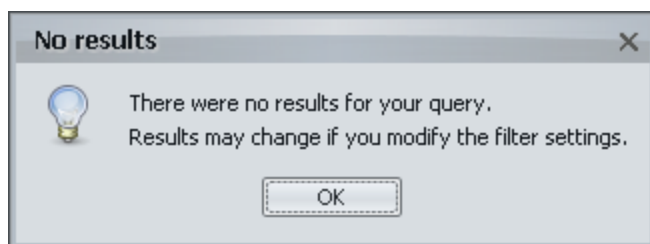
You can enter:

- Names for types of reference annotation features (Genes, DGV, etc.)
- Names for types of segments (Mosaicism, Loss, Gain, etc.)

Click **Reset** to clear the search entries.

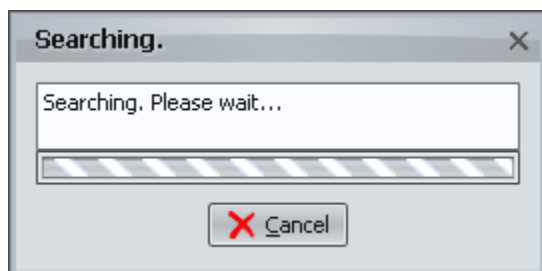
5. Click **Search...** to begin the search.

If no results are found, the following notice appears.



**Figure 11.26 No results notice.**

If the search takes more than a few seconds, an In Progress notice appears.



**Figure 11.27 Searching notice**

If results are found, the Search Results table opens.

Name/ID	Chromosome	Min	Max	Size (kbp)	Type
<a href="#">KIAA1804</a>	1	231,530,136	231,587,517	57	Genes
<a href="#">KCNJ9</a>	1	158,317,983	158,325,836	7	Genes
<a href="#">KLHDC9</a>	1	159,334,777	159,336,760	1	Genes
<a href="#">KPRP</a>	1	150,997,129	151,001,153	4	Genes
<a href="#">KIAA0859</a>	1	170,017,383	170,033,479	16	Genes
<a href="#">KISS1</a>	1	202,426,091	202,432,242	6	Genes
<a href="#">KCNK2</a>	1	213,245,507	213,477,059	231	Genes
<a href="#">KLF17</a>	1	44,357,108	44,373,396	16	Genes
<a href="#">KIAA1383</a>	1	231,007,260	231,012,715	5	Genes
<a href="#">KCNA3</a>	1	111,015,832	111,019,178	3	Genes
<a href="#">KCNK2</a>	1	213,323,182	213,477,059	153	Genes
<a href="#">KIAA1751</a>	1	1,874,611	1,925,136	50	Genes

**Figure 11.28 Search Results**

**To highlight features in the views or the table:**

- Double-click in a row of the table to zoom to the feature in the Selected Chromosome and Detail Views.
- Click on a feature in the Selected Chromosome or Detail View to highlight the feature in the Search Results table (the feature must be listed in the table to be highlighted).

You can perform the common table operations in the Search Results table (see [Common Table Operations](#), page 178).

The table displays the following information for each overlapping set of items:

<b>Name/ID</b>	Name or ID of the Item.
<b>Chromosome</b>	Chromosome the items are located in.
<b>Min</b>	Starting position of the item.
<b>Max</b>	Ending position of the item.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>Size (kbp)</b>	Size of the item.
<b>Type</b>	Type of Item.

## Finding Intersections

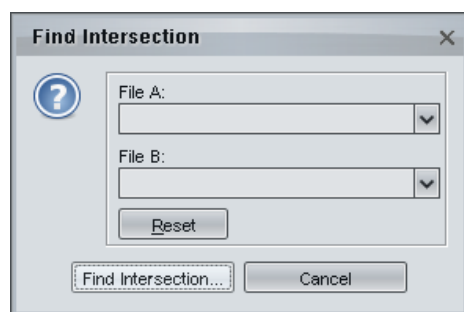
The Find Intersection function enables you to find segments and regions that overlap for different:

- Detected Segments
- Reference Annotations
- Loaded Region Information Files

**To find the intersection of items in selected files:**

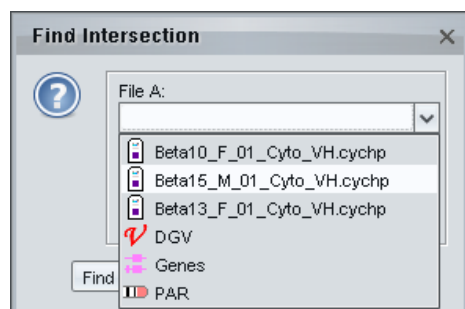
1. From the View menu, select **Find Intersections...**

The Find Intersection dialog box opens.



**Figure 11.29 Find Intersection dialog box.**

2. Select the first file for the comparison from the File A drop-down list.



**Figure 11.30 Drop-down list**

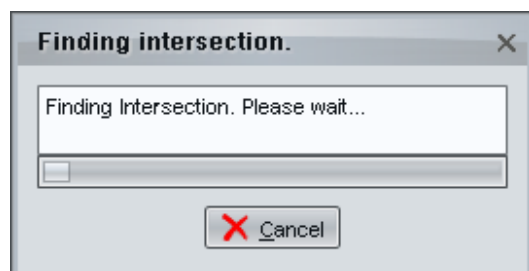
The list shows the available Sample files, Region Information Files, and Reference Annotations.



**Note:** Only files that are check marked in the Files List appear in the Match File drop-down list.

3. Select the second file from the File B drop-down list.
4. Click **Find Intersection...**

The Finding Intersection notice opens.



**Figure 11.31 Finding Intersection notice**

When the comparison is finished, the Intersection Results table opens.

A Type	A	Chromosome	A Min	A Max	A Size (kbp)	A CN State	B Type	B	B Min	B Max
▼ Loss	smoothed25...	1	25,467,089	25,543,873	76	1	Genes	RHD	25,471,567	25,521,567
▼ Loss	smoothed25...	1	25,467,089	25,543,873	76	1	Genes	RHD	25,471,567	25,521,567
▼ Loss	smoothed25...	1	25,467,089	25,543,873	76	1	Genes	TMEM50A	25,537,397	25,561,397
▲ Gain	seg16	1	12,960,917	13,028,391	67	3	Genes	PRAMEF22	12,968,129	12,992,129
▲ Gain	seg114	1	194,989,498	195,064,665	75	3	Genes	CFHR3	195,010,552	195,024,552
▲ Gain	seg114	1	194,989,498	195,064,665	75	3	Genes	CFHR1	195,055,483	195,069,483
▲ Gain	seg1014	10	46,383,964	46,453,447	69	3	Genes	GPRIN2	46,413,551	46,427,551
▲ Gain	seg1014	10	46,383,964	46,453,447	69	3	Genes	SYT15	46,378,532	46,392,532
▲ Gain	seg1014	10	46,383,964	46,453,447	69	3	Genes	SYT15	46,375,450	46,389,450
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20A	48,357,047	48,371,047
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20A	48,357,047	48,371,047
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20B	48,357,047	48,371,047
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20B	48,357,047	48,371,047
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20B	48,357,047	48,371,047
▲ Gain	seg1040	10	48,373,168	48,454,852	81	3	Genes	PTPN20B	48,357,047	48,371,047

**Figure 11.32 Intersection Results table**

The table displays the names of the A and B files above the toolbar.



### To highlight features in the views or the table:

- Double-click in a row of the table to zoom to the feature for File A in the Selected Chromosome and Detail Views.
- Click on a feature in the Selected Chromosome or Detail View to highlight the feature in the Intersection Results table (the feature must be listed in the table to be highlighted).

You can perform the common table operations in the Intersection Results table (see [Common Table Operations](#), page 178).

The table displays the following information for each overlapping set of items:

<b>A Type</b>	Type of item in A file with overlap.
<b>A</b>	Identifier used for item in A file.
<b>Chromosome</b>	Chromosome the items are located in.
<b>A Min</b>	Starting position of the A item.
<b>A Max</b>	Ending position of the A item.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>A Size (kbp)</b>	Size of the A item.
<b>A CN State</b>	Copy number of the segment in file A.
<b>B Type</b>	Type of Item in B file with overlap.
<b>B</b>	Identifier used for item in B file.
<b>B Min</b>	Starting position of the B item.
<b>B Max</b>	Ending position of the B item.  For all segments, the segment start coordinates are always lower by one bp from the coordinate for the starting probe of the segment as reported in the graphs table while the end coordinate matches the coordinate for the ending probe as reported in the graphs table (see <a href="#">Appendix E</a> , page 253).
<b>B Size (kpb)</b>	Size of the B item.
<b>B CN State</b>	Copy number of the segment in file B.
<b>% A Touching B</b>	How much of the A item is covered by the B item.
<b>% B Touching A</b>	How much of the B item is covered by the A item.
<b>Shared Size (kbp)</b>	Size of the overlap.

## Chapter 12: Reporting Results

Chromosome Analysis Suite has tools for reporting results. It provides options for:

- [Exporting Graphic Views](#) (see below): Export the Karyoview, Selected Chromosome View, and Detail View as a PDF or PNG file.
- [Exporting Table Data](#) (page 212): export the table data as a PDF or TXT file, or copy selected data to the clipboard.
- [Combining PDFs into a Single Report](#) (page 219)



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**



**Note: If you have trouble displaying non-English characters on screen or in exported PDF files, make sure that the font “Arial Unicode MS” is installed on the machine. Look in “C:\Windows\Fonts” for the file ARIALUNI.TTF.**

### Exporting Graphic Views

You can export the Karyoview, Selected Chromosome View, and Detail View as:

- PDF reports (see below)
- [PNG graphic file](#) (page 211)

#### Exporting as PDF Reports

ChAS provides a variety of options for exporting graphic views as PDFs. The PDF Report displays the graphic with basic information about data files and settings.



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

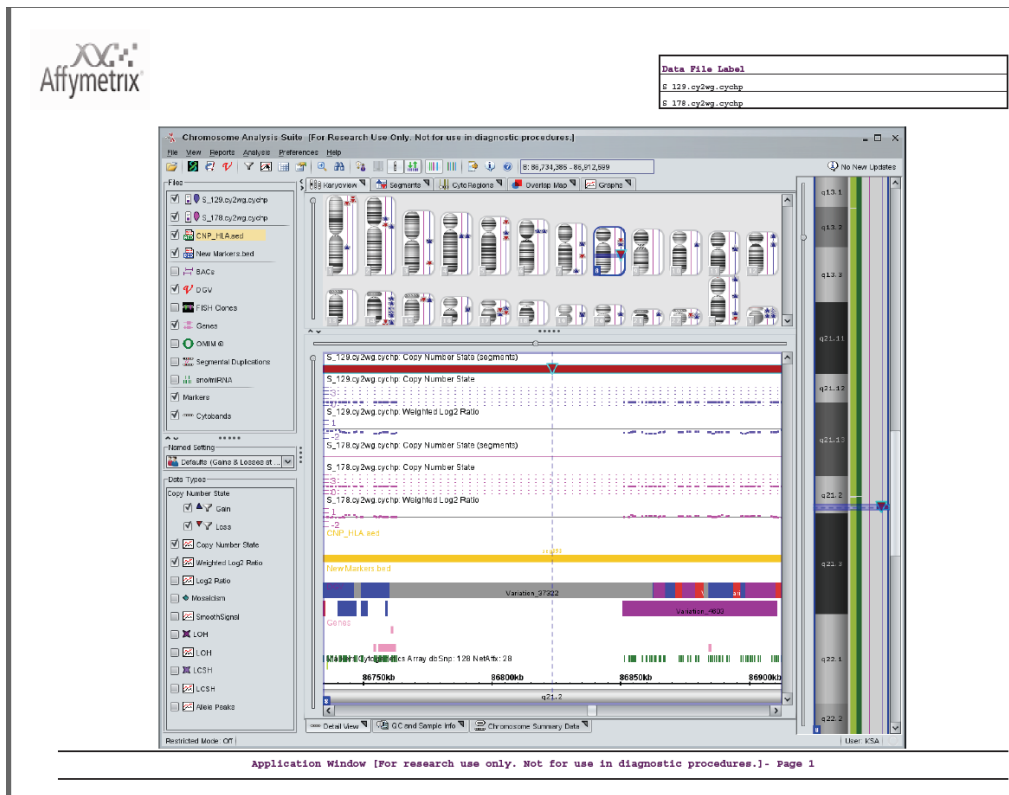


Figure 12.1 Whole screen exported to PDF, p 1

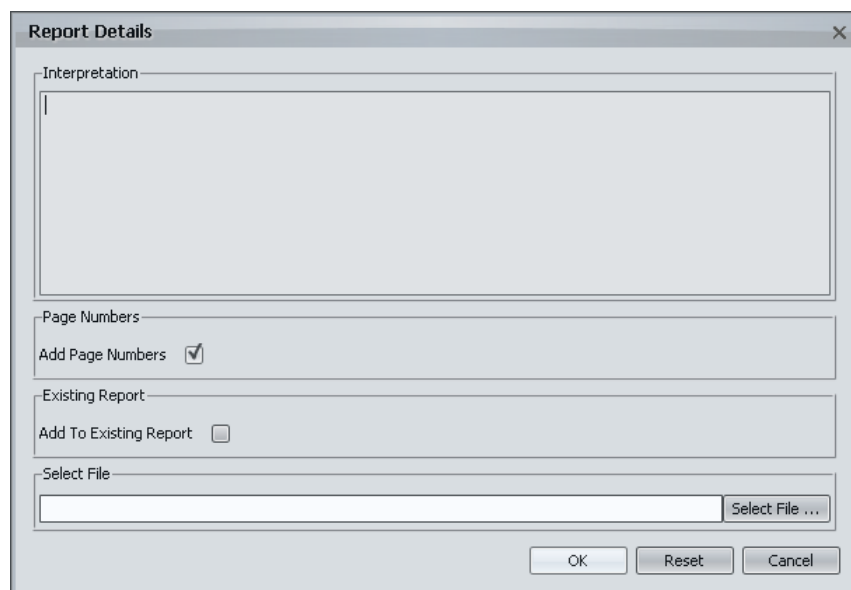


Figure 12.2 Whole screen exported to PDF, p 2

### To export graphics information:

1. From the **Reports** menu, select the PDF option you wish to use.
  - Export application window PDF - Creates PDF with entire software screen and information about data files
  - Export Karyoview PDF - Creates PDF with Karyoview and information about data files.
  - Export Selected Chromosome PDF - Creates PDF with Selected Chromosome View and information about data files.
  - Export Detail View PDF - Creates PDF with selected Detail View and information about data files.
  - Export Segments Table PDF - Creates PDF with Segment Table
  - Export QC and Sample Info PDF - Creates PDF with the QC and Sample Information
  - Export Chromosome Summary Data PDF - Creates PDF with the Chromosome Summary Data

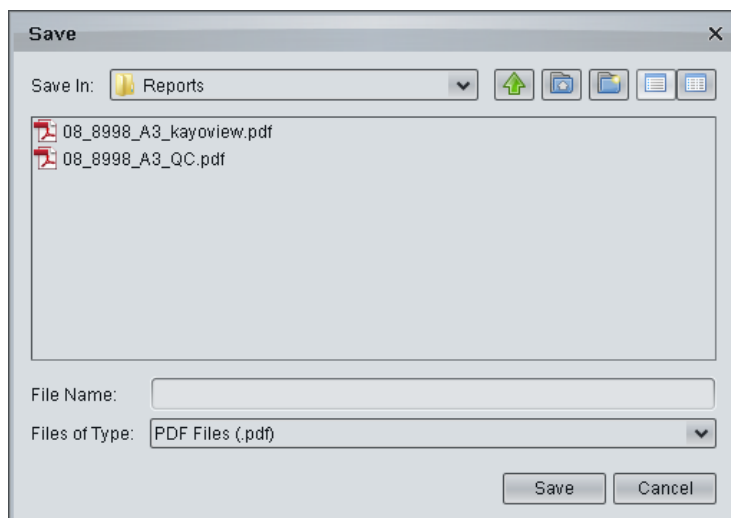
The Report Details dialog box appears.



**Figure 12.3 Report Details dialog box**

2. Enter text into Interpretation box, if desired.
3. Select the option for adding page numbers, if desired.
4. Select the option for adding to an existing report, if desired. See [Combining PDFs into a Single Report](#) (page 219).
5. Click **Select File**.

The Save dialog box opens.



**Figure 12.4 Save dialog box**

6. Select a folder location for the file using the navigation tools.  
This folder location is automatically selected when exporting other reports during a session.
7. Enter a name for the PDF file.  
If you are adding the graphic to an existing PDF, select the PDF file.
8. Click **Save**.
9. Click **OK** in the Report Details dialog box  
A PDF report is created with the selected data type saved.

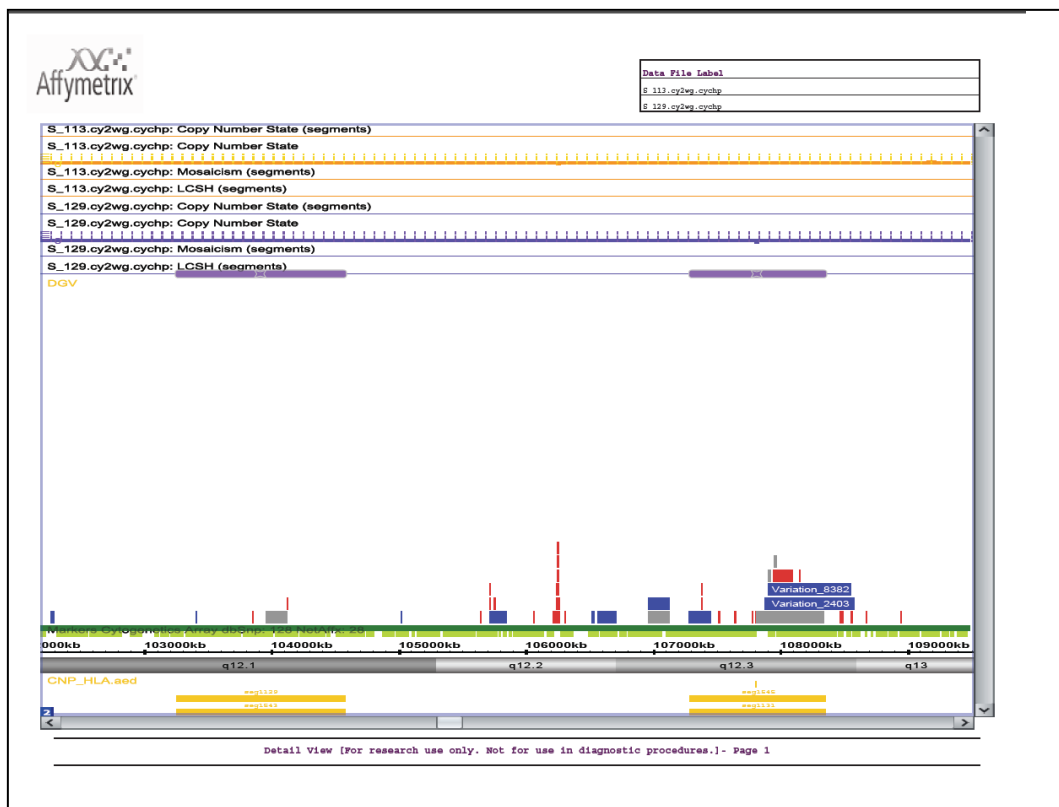


Figure 12.5 Detail View report, p 1

Data Files In View	
File Label	File Name
S_113.cy2wg.cychp	S_113.cy2wg.cychp
S_129.cy2wg.cychp	S_129.cy2wg.cychp

Genome Filters		
Segment Type	Filter Type	Minimum Filter Value
Gain	Segment Length	500000
Loss	Segment Length	200000

Settings for S_129.cy2wg.cychp	
Setting Type	Setting
Smoothing	On
Smoothing Max Jump Limit	Off
Joining	50 markers
Joining Max Jump Limit	Off
Restricted Mode	Off

Settings for S_113.cy2wg.cychp	
Setting Type	Setting
Smoothing	On
Smoothing Max Jump Limit	Off
Joining	50 markers

Detail View [For research use only. Not for use in diagnostic procedures.] - Page 2

Figure 12.6 Detail View report, p 2

Setting Type	Setting
Joining Max Jump Limit	Off
Restricted Mode	Off

Overlap Map - CNP\_HLA.aed

Detail View (For research use only. Not for use in diagnostic procedures.)- Page 3

Figure 12.7 Detail View report, p 2

## Export as PNG

You can also create a PNG screenshot of the entire software screen.



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

To export a PNG screenshot:

1. From the **Reports** menu, select **Export application window PNG**.

The Save As PNG dialog box opens.

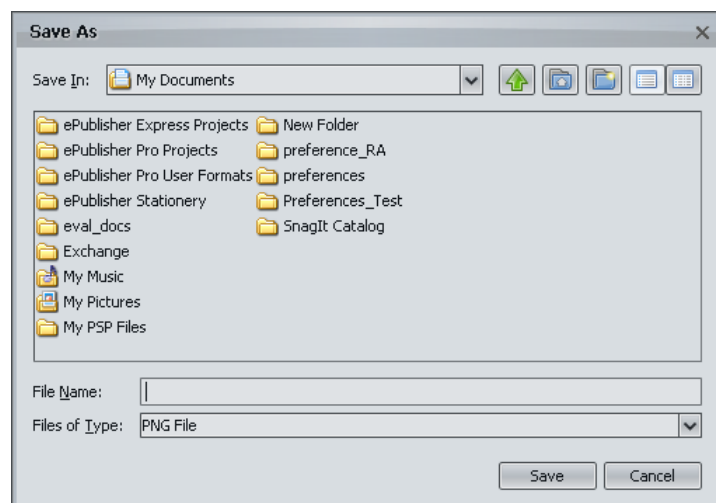


Figure 12.8 Save As PNG dialog box

2. Select a folder location for the file using the navigation tools.

This folder location is automatically selected when exporting other screenshots during a session.

3. Enter a name for the PNG file in the File Name box.
4. Click the **Save** button.

The PNG file screen shot is saved in the selected location.

The PNG file can be cropped in a graphics program like Paint and inserted into a word processing document if desired.

## Exporting Table Data

ChAS provides several options for exporting table data:

- [Export Table Data in a PDF](#) (below)
- [Export Tables as TXT File](#) (page 217)
- [Transfer to Clipboard](#) (page 218)

### Export Table Data in a PDF File



**Important:** The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.



**Note:** You can also export data from the Segment Table by selecting Export Segment Table PDF from the Reports Menu (see page 206).



**Note:** You can't export Graph table data as a PDF report.

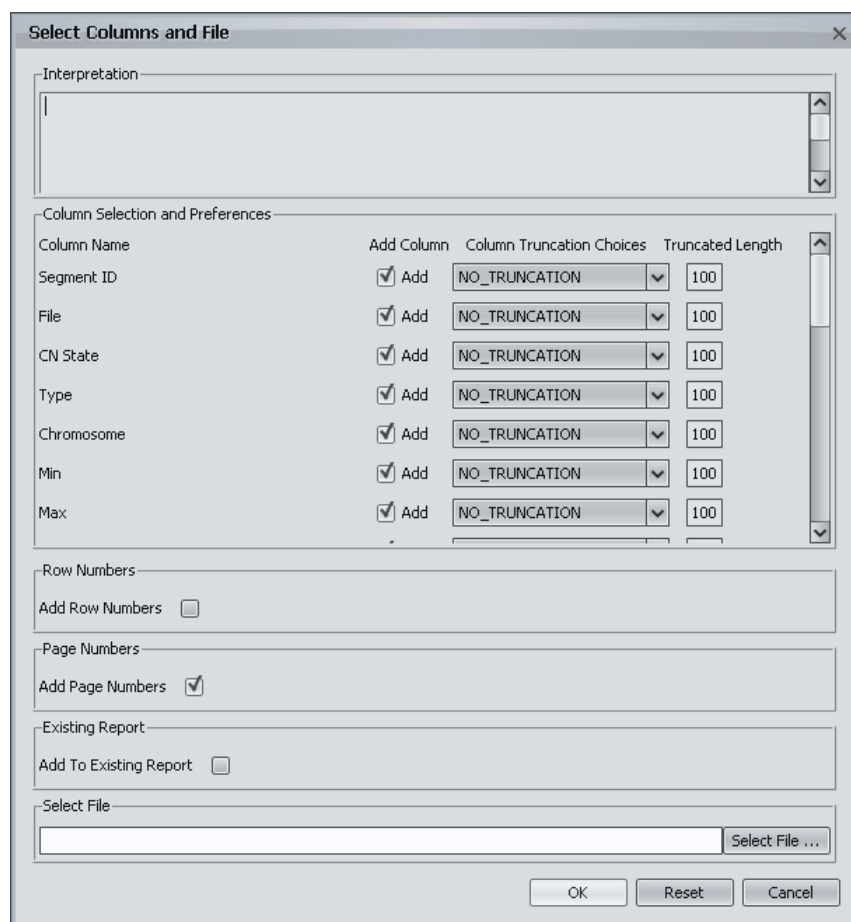
To export Table information in a PDF file:

Segment ID	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marke...	Max % Cov...	Max % Ove...	Nui
seg2024	Sample_0...	3	▲ Gain	3	131,253,287	131,321,447	68	1,286			
seg2100	Sample_0...	0	▼ Loss	3	164,007,533	164,019,098	11	1,051			
seg2014	Sample_0...	3	▲ Gain	3	127,155,074	127,159,387	4	862			
smoothed164019930	Sample_0...	5	▲ Gain	3	164,019,930	164,100,879	80	1,245			
seg1830	Sample_0...	3	▲ Gain	3	75,531,222	75,600,607	69	1,176			
smoothed163887385	Sample_0...	3	▲ Gain	3	163,887,385	164,006,897	119	1,475			
seg1832	Sample_0...	3	▲ Gain	3	75,624,861	75,637,504	12	1,053			
seg2096	Sample_0...	3	▲ Gain	3	75,838,844	75,962,472	123	1,188			
seg2515	Sample_0...	1	▼ Loss	3	190,842,432	190,853,556	11	855			
seg2204	Sample_0...	3	▲ Gain	3	116,133,749	116,146,833	13	817			
seg2557	Sample_0...	3	▲ Gain	3	196,822,294	196,914,807	92	1,492			
seg2252	Sample_0...	3	▲ Gain	3	126,917,179	126,991,856	74	1,037			
smoothed75549695	Sample_0...	3	▲ Gain	3	75,549,695	75,633,780	84	1,356			
seg2094	Sample_0...	3	▲ Gain	3	75,716,050	75,800,870	84	1,101			
seg2088	Sample_0...	3	▲ Gain	3	75,459,878	75,536,466	76	1,160			
seg2010	Sample_0...	3	▲ Gain	3	47,804,033	47,876,603	72	1,133			
seg2264	Sample_0...	4	▲ Gain	3	129,883,492	129,904,220	20	1,727			
smoothed164002633	Sample_0...	1	▼ Loss	3	164,002,633	164,091,128	88	1,229			

Figure 12.9 Segments table with data filtered

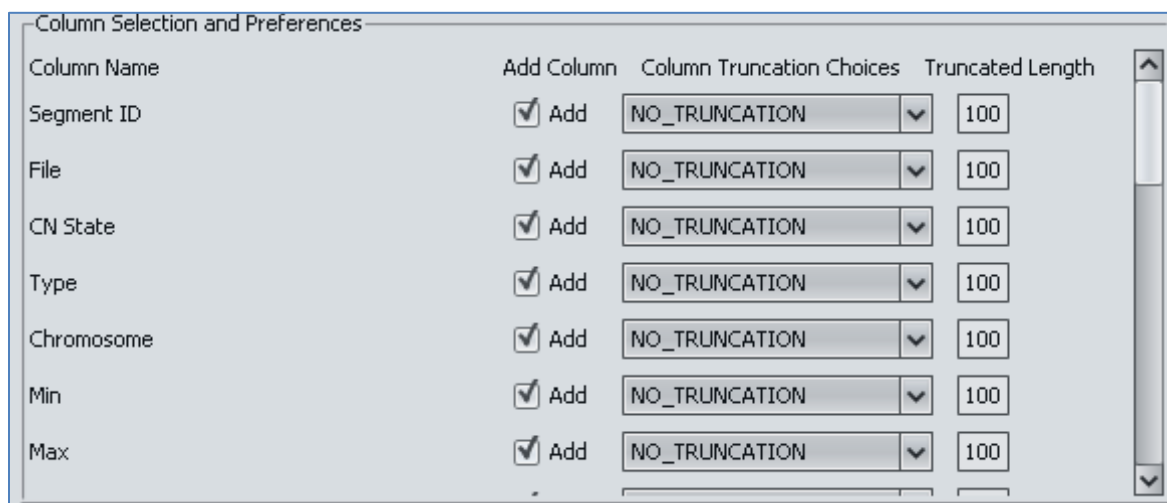
1. Click the PDF toolbar button
2. In the dialog box that appear, enter text in the Interpretation box.





**Figure 12.10 Select Columns and File dialog box for Segment Table PDF**

3. Select the columns to be displayed and the truncation rules for the content of the columns.



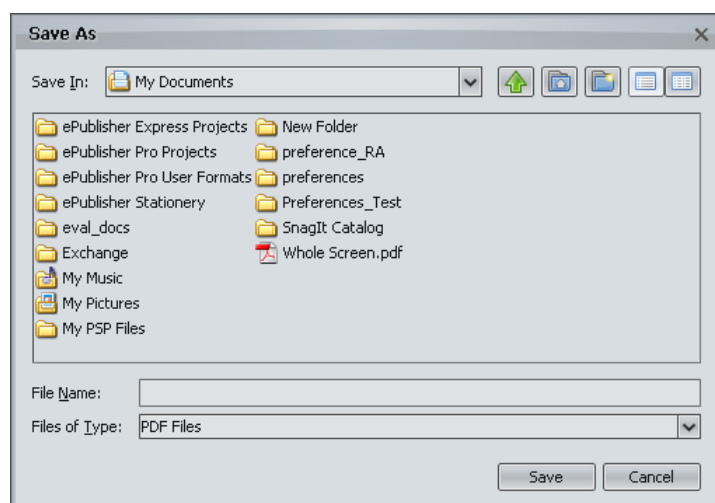
**Figure 12.11 Column Selection and Preferences**

**Column Name** Header of column in table.

<b>Add Column</b>	Select checkbox to display column in PDF.
<b>Column Truncation Choices</b>	<p>Select Truncation Options:</p> <ul style="list-style-type: none"> <li>• No_Truncation: field is exported as is, using wrap-around if necessary.</li> <li>• Truncate_Beginning: Truncates content at the beginning of the field, leaving as many characters as specified in Truncated Length box.</li> <li>• Truncate_Middle: Truncates content in the middle of the field, leaving characters at the beginning and end, with ellipses (...) to mark the truncated characters.</li> <li>• Truncate_End: Truncates content at the end of the field, leaving as many characters as specified in Truncated Length box.</li> </ul>
<b>Truncated length</b>	Number of characters displayed after truncating the data.

4. Select the option for adding page and row numbers, if desired.
5. Select the option for adding to an existing report, if desired. See [Combining PDFs into a Single Report](#) (page 219).
6. Click the **Select File** button.

The Save As dialog box opens.



**Figure 12.12 Save As for Segment Table PDF**

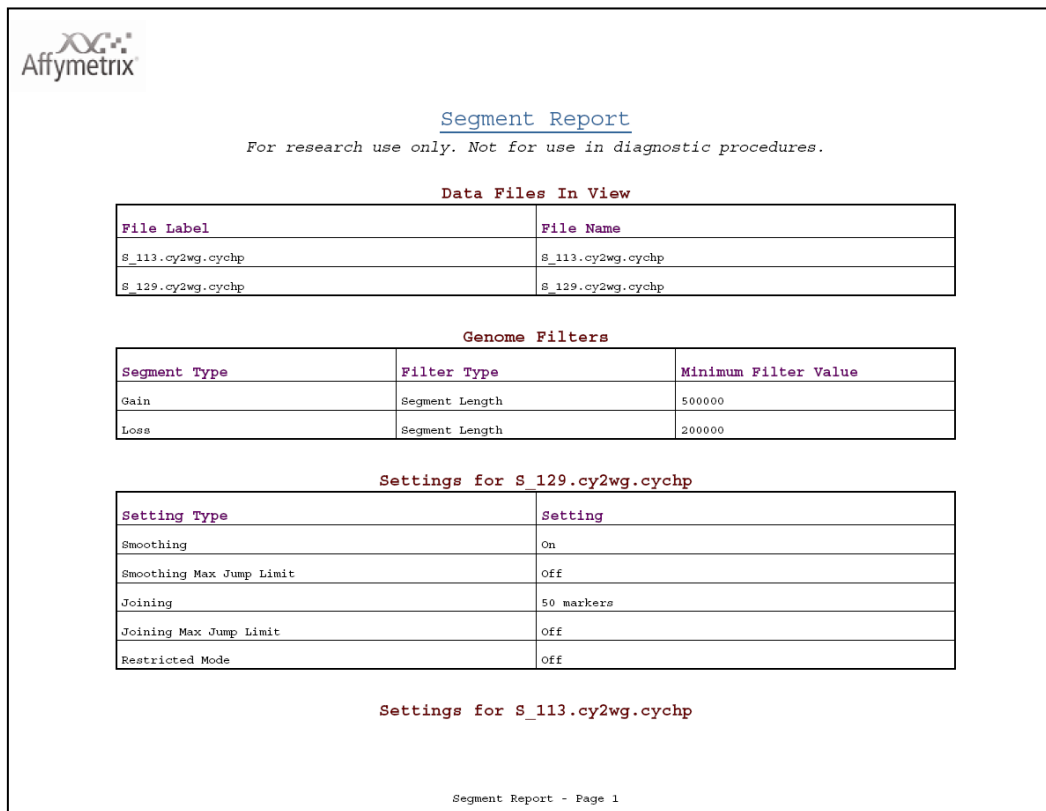
7. Select a folder location for the file using the navigation tools.
8. Enter a name for the PDF file, or select a file for the information to be appended to.
9. Click **Save** in the Save As dialog box.
10. Click **OK** in the Select Columns and File dialog box.

A PDF file is created with the selected data type saved.

The report displays:

- Table type
- Information on chromosome and genome region
- Interpretation
- Data files

- Genome or CytoRegion Segment Filters used
- Settings for Data Processing
- Details of the table data



**Figure 12.13 Segments Table report, page 1**

## Segments

**Figure 12.14 Segments Table, page 2**

**Figure 12.15 Segments Table Report, page 3**

## Tips for Exporting Tables

Follow the tips below to improve the export of table data in a PDF report:

- Use truncation
- Select only columns you need
- Use filtering options (Segment filters, displaying only results for a chromosome or area in the detail view, etc.) to limit the number of values being exported

## Export Tables as TXT File

The TXT file format allows you to transfer data to other software for analysis.



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

**To export table information as a text file:**

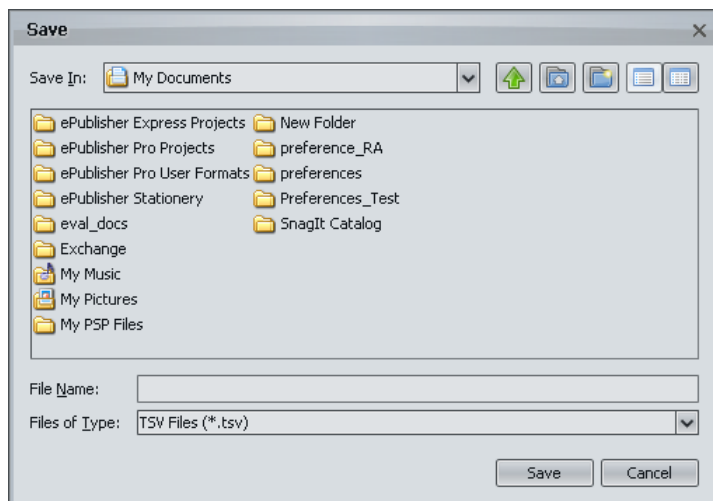
1. Perform pre-filtering on the data in the table.

Segment ID	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marke...	Max % Cov...	Max % Ove...	Nul
seg2024	Sample_0...	3 ▲ Gain		3	131,253,287	131,321,447	68	1,286			
seg2100	Sample_0...	0 ▼ Loss		3	164,007,533	164,019,098	11	1,051			
seg2014	Sample_0...	3 ▲ Gain		3	127,155,074	127,159,387	4	862			
smoothed164019930	Sample_0...	5 ▲ Gain		3	164,019,930	164,100,879	80	1,245			
seg1830	Sample_0...	3 ▲ Gain		3	75,531,222	75,600,607	69	1,176			
smoothed163887385	Sample_0...	3 ▲ Gain		3	163,887,385	164,006,897	119	1,475			
seg1832	Sample_0...	3 ▲ Gain		3	75,624,861	75,637,504	12	1,053			
seg2096	Sample_0...	3 ▲ Gain		3	75,838,844	75,962,472	123	1,188			
seg2515	Sample_0...	1 ▼ Loss		3	190,842,432	190,853,556	11	855			
seg2204	Sample_0...	3 ▲ Gain		3	116,133,749	116,146,833	13	817			
seg2557	Sample_0...	3 ▲ Gain		3	196,822,294	196,914,807	92	1,492			
seg2252	Sample_0...	3 ▲ Gain		3	126,917,179	126,991,856	74	1,037			
smoothed75549695	Sample_0...	3 ▲ Gain		3	75,549,695	75,633,780	84	1,356			
seg2094	Sample_0...	3 ▲ Gain		3	75,716,050	75,800,870	84	1,101			
seg2088	Sample_0...	3 ▲ Gain		3	75,459,878	75,536,466	76	1,160			
seg2010	Sample_0...	3 ▲ Gain		3	47,804,033	47,876,603	72	1,133			
seg2264	Sample_0...	4 ▲ Gain		3	129,883,492	129,904,220	20	1,727			
smoothed164002633	Sample_0...	1 ▼ Loss		3	164,002,633	164,091,128	88	1,229			

**Figure 12.16 Segments table with data filtered**

2. In the Table tool bar, click the Export **TXT** button

The Save as TXT dialog box opens.



**Figure 12.17 Save as TXT dialog box**

3. Select a folder location for the file using the navigation tools.
4. Enter a name for the TXT file.
5. Click **Save**.

The TXT file is saved in the selected location.

It can be opened using a text editing or spreadsheet program, or in other software designed to use Tab Separated Value TXT format.

	A	B	C	D	E	F	G	H	I	J
	Segment ID	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marker Distance	Max
1	seg2024	Sample_01.cyto2.cychp	3	Gain	3	131253287	131321447	68	1286	
2	seg2100	Sample_01.cyto2.cychp	0	Loss	3	164007533	164019098	11	1051	
3	seg2014	Sample_01.cyto2.cychp	3	Gain	3	127155074	127159387	4	862	
4	smoothed164019930	Sample_01.cyto2.cychp	5	Gain	3	164019930	164100879	80	1245	
5	seg1830	Sample_01.cyto2.cychp	3	Gain	3	75531222	75600607	69	1176	
6	smoothed163887385	Sample_01.cyto2.cychp	3	Gain	3	163887385	164006897	119	1475	
7	seg1832	Sample_01.cyto2.cychp	3	Gain	3	75624861	75637504	12	1053	
8	seg2096	Sample_03.cyto2.cychp	3	Gain	3	75838844	75962472	123	1188	
9	seg2515	Sample_03.cyto2.cychp	1	Loss	3	190842432	190853556	11	855	
10	seg2204	Sample_03.cyto2.cychp	3	Gain	3	116133749	116146833	13	817	
11	seg2557	Sample_03.cyto2.cychp	3	Gain	3	196822294	196914807	92	1492	
12	seg2252	Sample_03.cyto2.cychp	3	Gain	3	126917179	126991856	74	1037	
13	smoothed75549695	Sample_03.cyto2.cychp	3	Gain	3	75549695	75633780	84	1356	
14	seg2094	Sample_03.cyto2.cychp	3	Gain	3	75716050	75800870	84	1101	
15	seg2088	Sample_03.cyto2.cychp	3	Gain	3	75459878	75536466	76	1160	
16	seg2010	Sample_03.cyto2.cychp	3	Gain	3	47804033	47876603	72	1133	
17	seg2264	Sample_03.cyto2.cychp	4	Gain	3	129883492	129904220	20	1727	
18	smoothed164002633	Sample_03.cyto2.cychp	1	Loss	3	164002633	164091128	88	1229	
19										
20										
21										
22										

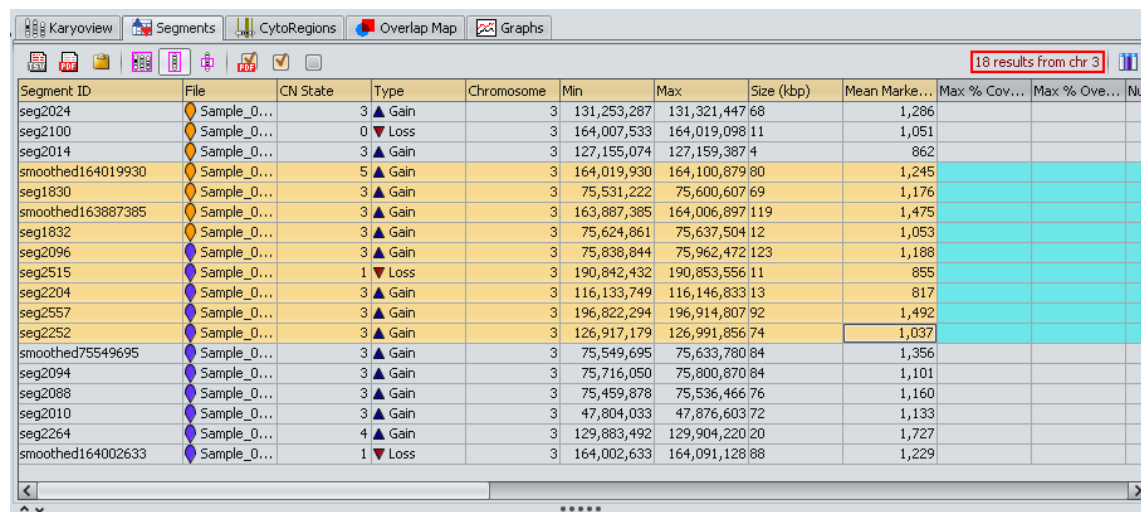
**Figure 12.18 TXT file opened in Excel**

### Transfer to Clipboard

You can copy data from selected cells to the clipboard for pasting into a text or spreadsheet file.

## To copy table data to the Clipboard:

1. Select the cells you wish to copy in the table.



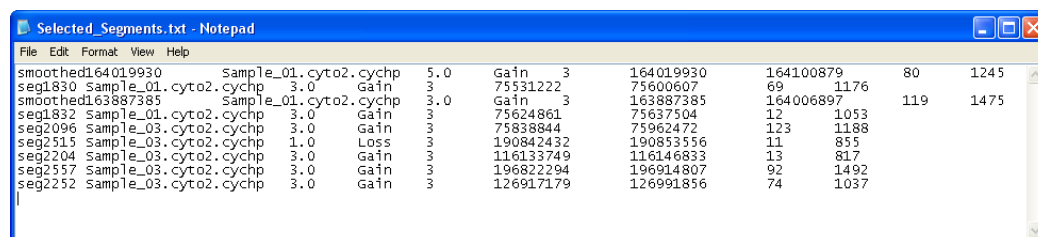
Segment ID	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marke...	Max % Cov...	Max % Ove...	Null
seg2024	Sample_0...	3 ▲ Gain		3	131,253,287	131,321,447	68	1,286			
seg2100	Sample_0...	0 ▼ Loss		3	164,007,533	164,019,098	11	1,051			
seg2014	Sample_0...	3 ▲ Gain		3	127,155,074	127,159,387	4	862			
smoothed164019930	Sample_0...	5 ▲ Gain		3	164,019,930	164,100,879	80	1,245			
seg1830	Sample_0...	3 ▲ Gain		3	75,531,222	75,600,607	69	1,176			
smoothed163887385	Sample_0...	3 ▲ Gain		3	163,887,385	164,006,897	119	1,475			
seg1832	Sample_0...	3 ▲ Gain		3	75,624,861	75,637,504	12	1,053			
seg2096	Sample_0...	3 ▲ Gain		3	75,838,844	75,962,472	123	1,188			
seg2515	Sample_0...	1 ▼ Loss		3	190,842,432	190,853,556	11	855			
seg2204	Sample_0...	3 ▲ Gain		3	116,133,749	116,146,833	13	817			
seg2557	Sample_0...	3 ▲ Gain		3	196,822,294	196,914,807	92	1,492			
seg2252	Sample_0...	3 ▲ Gain		3	126,917,179	126,991,856	74	1,037			
smoothed75549695	Sample_0...	3 ▲ Gain		3	75,549,695	75,633,780	84	1,356			
seg2094	Sample_0...	3 ▲ Gain		3	75,716,050	75,800,870	84	1,101			
seg2088	Sample_0...	3 ▲ Gain		3	75,459,878	75,536,466	76	1,160			
seg2010	Sample_0...	3 ▲ Gain		3	47,804,033	47,876,603	72	1,133			
seg2264	Sample_0...	4 ▲ Gain		3	129,883,492	129,904,220	20	1,727			
smoothed164002633	Sample_0...	1 ▼ Loss		3	164,002,633	164,091,128	88	1,229			

Figure 12.19 Segment table with cells selected

2. Click the **Copy to Clipboard** button  in the table tool bar.

The selected data is copied to the clipboard.

You can paste the data on the clipboard into a text or spreadsheet file.



Segment ID	File	CN State	Type	Chromosome	Min	Max	Size (kbp)	Mean Marke...	Max % Cov...	Max % Ove...	Null
smoothed164019930	Sample_01.cyto2.cychp	5.0	Gain	3	164019930	164100879	80	1245			
seg1830	Sample_01.cyto2.cychp	3.0	Gain	3	75531222	75600607	69	1176			
smoothed163887385	Sample_01.cyto2.cychp	3.0	Gain	3	163887385	164006897	119	1475			
seg1832	Sample_01.cyto2.cychp	3.0	Gain	3	75624861	75637504	12	1053			
seg2096	Sample_03.cyto2.cychp	3.0	Gain	3	75838844	75962472	123	1188			
seg2515	Sample_03.cyto2.cychp	1.0	Loss	3	190842432	190853556	11	855			
seg2204	Sample_03.cyto2.cychp	3.0	Gain	3	116133749	116146833	13	817			
seg2557	Sample_03.cyto2.cychp	3.0	Gain	3	196822294	196914807	92	1492			
seg2252	Sample_03.cyto2.cychp	3.0	Gain	3	126917179	126991856	74	1037			

Figure 12.20 Data pasted into text file

## Combining PDFs into a Single Report

You can combine different PDFs into a single report with multiple pages and content.

You can do this by:

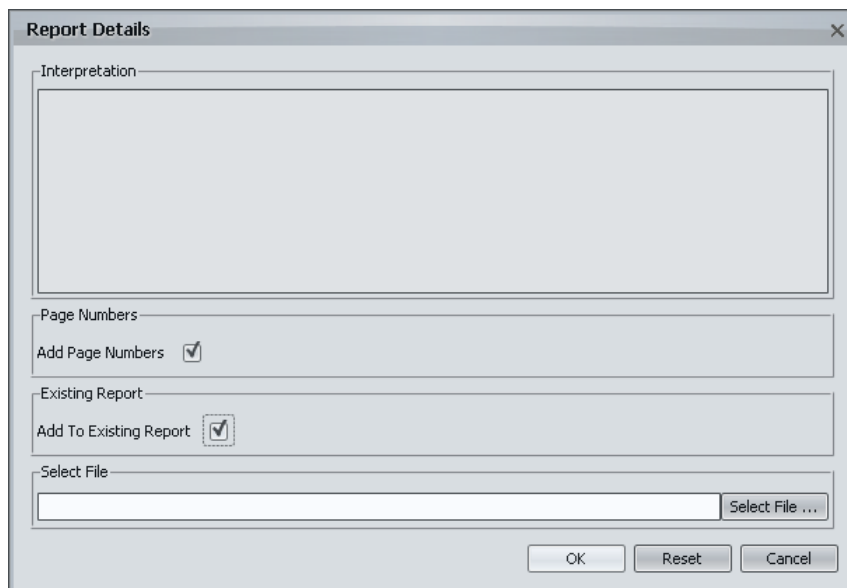
- Adding new data to an existing PDF Report
- Merging two or more existing PDF reports



**Important: The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.**

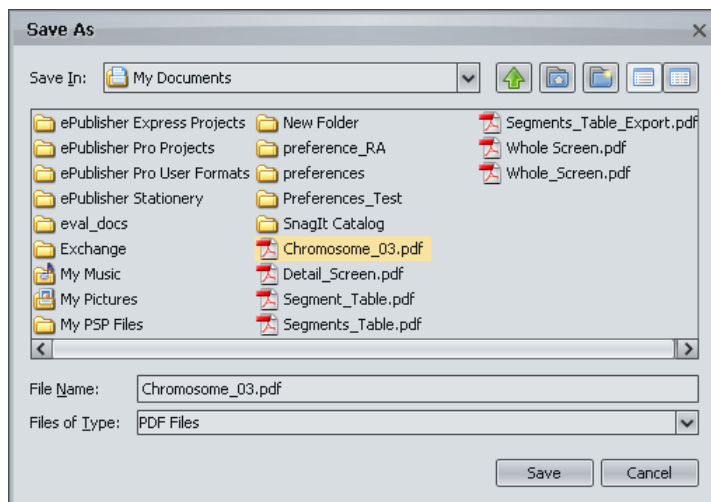
### To add new PDF report data to an existing PDF Report:

1. When creating a report, select the Add to Existing Report option in the Select Columns and File or Report Details dialog box.



**Figure 12.21 Report Details dialog box with Add to Existing Report selected**

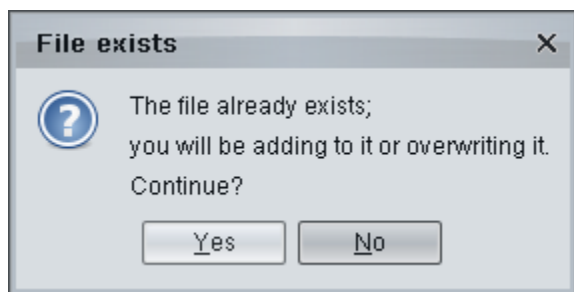
2. Click **Select File** and select an existing PDF Report for the data to be added to.  
The Save As dialog box opens.



**Figure 12.22 Save As dialog box**

3. Select a PDF file and click **Save** in the Save As dialog box.  
A confirmation message appears, asking if you want to overwrite or add to the data.





**Figure 12.23 Confirmation message**

4. Click Yes in the Confirm Rewrite notice to append the data in the selected file.
5. Click **OK** in the Select Columns and File or Report Details dialog box  
The new report data is combined with the existing report.

## Segments Table

For research use only. Not for use in diagnostic procedures.

Selected Results From: Chr 3

Data Files In View	
File Label	File Name
sample_01.cyt02.cychp	sample_01.cyt02.cychp
sample_01.cyt03.cychp	sample_01.cyt03.cychp

## Genome Filters

Segment Type	Filter Type	Minimum Filter Value
Gain	marker count	63
Loss	confidence	0.15
Loss	marker count	163

## Settings

Setting Type	Setting
Smooth	on
Join	off
Separated Mode	off

Page 1

Segments Table - Page 1

Overlap Map - File 03 24.aed

### Segments Details

Site Name	Segment ID	FLC	# of States	Type	Chromosome	Site	Size	Site (kb)
1	mpg104	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13351087	13351647	66
2	mpg110	chrX:11, cyt10 2-cytop	3, 3	SNP	8	13429793	13431938	11
3	mpg104	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13370974	13371937	4
4	emod3bda4410 188	chrX:11, cyt10 2-cytop	3, 3	SNP	5	14051490	14041974	87
5	mpg110	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13373122	13369007	63
6	emod3bda397 188	chrX:11, cyt10 2-cytop	3, 3	SNP	3	14288700	14260697	119
7	mpg110	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13372481	13367704	12
8	mpg100	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13370844	13362472	113
9	mpg110	chrX:11, cyt10 2-cytop	3, 3	SNP	8	13380432	13383576	11
10	mpg1104	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13433764	13434493	1
11	mpg107	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13265294	13265497	9
12	mpg110	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13265177	13265184	74
13	emod3bda36 188	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13374300	13373763	44
14	mpg1104	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13433480	13430070	64
15	mpg108	chrX:11, cyt10 2-cytop	3, 3	SNP	3	13365870	13354444	74

Page 2

Segments Table - Page 2

Row Num	Segment ID	File	CW State	Type	Chromosome	Min	Max	File (kbp)
16	seg3010	swg1a_03_cyto 2_cy2img	3.0	Gain	3	47894000	47896000	72
17	seg3204	swg1a_03_cyto 3_cy2img	4.0	Gain	3	129883802	129906200	20
18	880053e5d144702	swg1a_03_cyto 3_cy2img	1.0	LOH	3	1447032633	144791128	88

Page 3

Segments Table - Page 1

## Data Files In View

File Label	File Name
sample_01.cyto2.cychp	sample_01.cyto2.cychp
sample_01.cyto2.cychb	sample_01.cyto2.cychb

## Genome Filters

Segment Type	Filter Type	Minimum Filter Value
Gain	Marker Count	63
Loss	Confidence	0.15
Loss	Marker Count	163

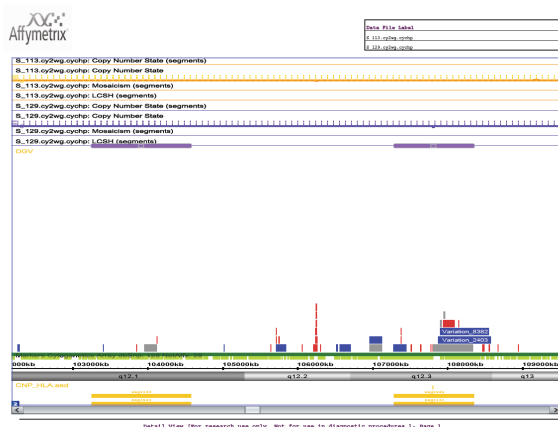
## Settings

Setting Type	Setting
smooth	on
Join	off
Restricted Mode	off

Overlap Map - File 03 24.aed

Page 5

Detail View [For research use only. Not for use in diagnostic procedures.] - Page 2



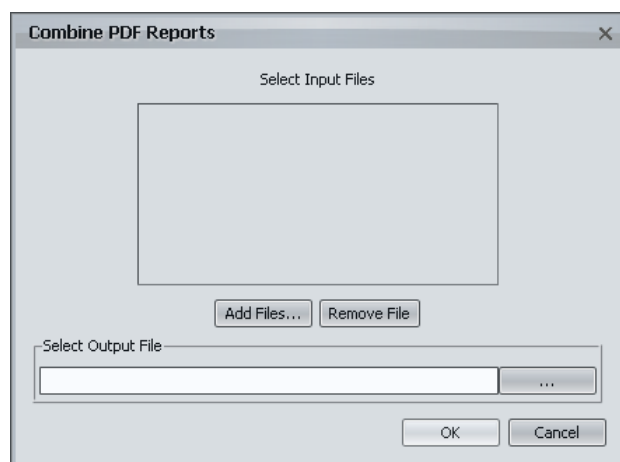
Detail View [For research use only. Not for use in diagnostic procedures.] - Page

### Figure 12.24 Combined PDF reports

**To combine two existing PDF reports:**

1. From the Report menu, select Combine PDFs into Report.

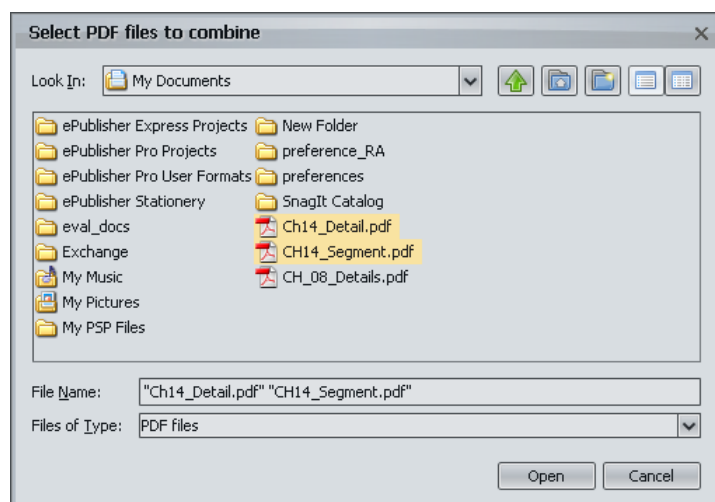
The Combine PDF Reports dialog box opens.



**Figure 12.25 Combine PDF Reports dialog box**

2. Click Add Files....

The Select PDF Files to Combine dialog box opens.



**Figure 12.26 Select PDF Files to Combine dialog box**

3. Select the PDF files to combine and click **Open** in the Select PDF Files dialog box.

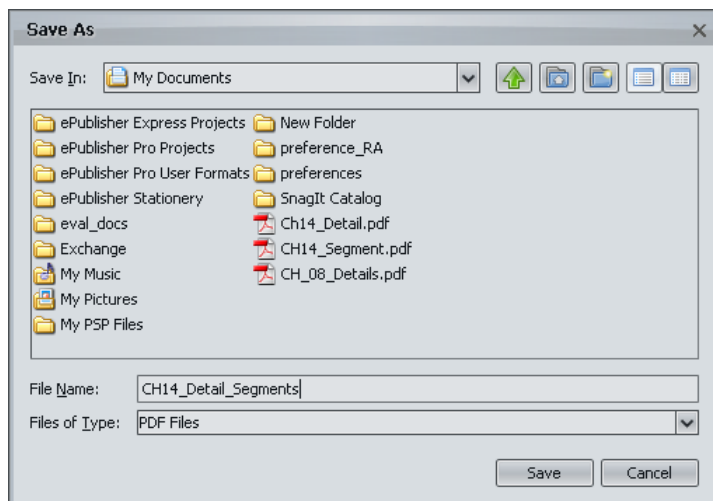
The selected files are displayed in the Select Input Files list.

You can use the Remove File button to remove a selected input file.

Click and drag on a file in the list to change the order of data in the Combined PDF.

4. Click the **Select Output File** button.

The Save As dialog box opens.



**Figure 12.27 Save As dialog box**

5. Enter a file name for the combined file and click **Save** in the Save As dialog box.
6. Click **OK** in the Combine PDF Reports dialog box.

The Selected PDFs are combined.

## Chapter 13: User Profiles and Named Settings

ChAS provides many options for customizing the display of data and annotations. The User Profiles and Named Settings functions allow you to save your analysis and display settings. These functions are described in the following sections:

- [Types of Settings](#) (below)
- [Creating and Using User Profiles](#) (page 227)
- [Creating and Using Named Settings](#) (page 228)
- [Exporting and Importing Preferences](#) (page 230)

### Types of Settings

ChAS provides two ways to store setup information:

- User Profiles
- Named Settings

They work differently and perform different functions.

#### User Profiles

A User Profile stores your selections for various display settings as they were when the software was last shut down while using that user profile.

A new user profile can be created or selected only when starting the software.

The user profile saves the following display settings:

- Screen size, displayed tabs, and sizing of display areas
- The views displayed in ChAS, and the size of the display panes
- Available Named Settings: Different users can have different lists of named settings to choose from
- Name of the currently selected named setting
- Copies of the user's custom (not shared) named settings
- Data Display Configurations
- Region information files selected for CytoRegions and Overlap Map
- Which types of graph and segment data are turned on or off
- Display options for graph data (height, grid, values, etc.)
- Chromosome and area displayed.
- Selected Reference Annotation database (ChAS Browser NetAffx Genomic Annotations file)
- Loaded AED and BED files
- The files and Reference Annotations (Genes, DGV, etc.) that are checked or unchecked
- Custom color rules

## Named Settings


A Named Setting stores the user's choices for:

- Which types of graph and segment data are turned on or off
- Segment Filter Settings
- Restricted Mode on/off

The Named Setting doesn't save a particular CytoRegion file, but does inform you if no file is selected when you select a setting with restricted mode on.

It is possible to apply a Named Setting with restricted mode using a different cytoregion file than was selected for the initial creation of the setting.

You can switch between different Named Settings in the same user profile to look at different types of data.

ChAS provides pre-configured (shared) Named Settings indicated by the  icon (Table 13.1). These named settings cannot be deleted.

**Table 13.1 Shared named settings**

Named Setting	Genome Segment Filters	CytoRegion Segment Filters	Data Types
Standard (400kbp and 50 markers, Gains and Losses)	<b>Gain:</b> Marker Count = 50 Size (kbp) = 400; <b>Loss:</b> Marker Count = 50 Size (kbp) = 400	<b>Gain:</b> Marker Count = 50 Size (kbp) = 400 <b>Loss:</b> Marker Count = 50 Size (kbp) = 400	Gain Loss Copy Number State Weighted Log2Ratio Allele Peaks Allele Difference
High Resolution	<b>Gain:</b> Marker Count = 50 Size (kbp) = 100; <b>Loss:</b> Marker Count = 50 Size (kbp) = 100	<b>Gain:</b> Marker Count = 25 Size (kbp) = 50 <b>Loss:</b> Marker Count = 25 Size (kbp) = 50	Gain Loss Copy Number State Weighted Log2Ratio
LOH only (3Mb and 50 SNPs)	<b>LOH:</b> Marker Count = 50 Size (kbp) = 3000	<b>LOH:</b> Marker Count = 50 Size (kbp) = 3000	LOH Genotype Calls Allele Peaks Allele Difference
Differential Gains and Losses	<b>Gain:</b> Marker Count = 50 Size (kbp) = 400; <b>Loss:</b> Marker Count = 50 Size (kbp) = 100	<b>Gain:</b> Marker Count = 25 Size (kbp) = 50 <b>Loss:</b> Marker Count = 25 Size (kbp) = 50	Gain Loss Copy Number State Weighted Log2Ratio

## Creating and Using User Profiles

You can only select or create user profiles when starting ChAS.

### To create a user profile

1. Double-click on the ChAS icon on the desktop; or

From the Windows Start Menu, select **Programs > Affymetrix >Chromosome Analysis Suite > Chromosome Analysis Suite**.

The ChAS Splash Screen and the Select User dialog box open.

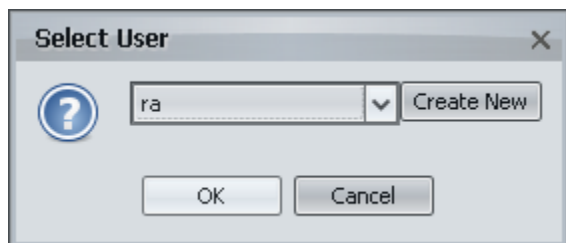


Figure 13.1 Select User dialog box

2. Click **Create New** in the Select User dialog box.

The Create New User dialog box opens.

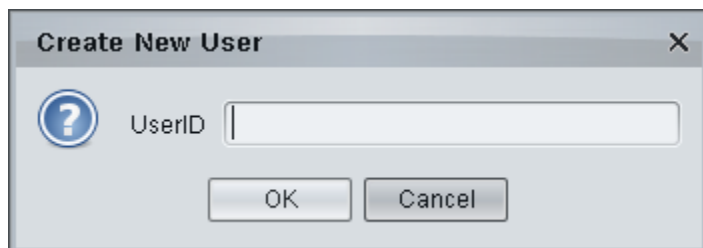


Figure 13.2 Create New User dialog box

3. Enter a name for the new profile in the User ID box.
4. Click **OK** in the Create New User dialog box.

The new user appears in the drop-down User list in the Select User dialog box.

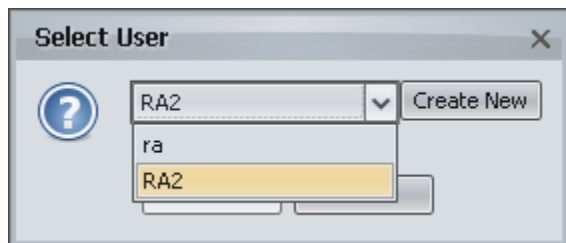


Figure 13.3 Select User dialog box with new user profile

5. Select the new user and click **OK**.

Any changes you make to the setup of the software that is listed in User Profiles (page 225) will be saved when you shut down the software and used the next time the software is opened with this user profile.

### To delete a user profile:

- Go to the Windows folder where the user profiles are stored and delete the folder with the profile name you wish to delete.

You can see the location of the folder in the About window, as described in [Analysis File Locations](#) (page 18).

## Creating and Using Named Settings

You can save a snapshot of your favorite settings as a Named Setting. To apply a particular Named Setting to the active data (check marked in the Files List), make a selection from the Named Setting drop-down list. Some pre-configured Shared Named Settings may be available for use by all users. Only an administrator can add or remove Shared Named Settings, but any user can apply them to their data.

### To save a named setting:

1. Set the display data settings as desired.

These can include:

- Which graphs and segment types are turned on or off
- Segment Filter Settings
- Restricted mode on/off

2. From the Preferences menu, select **Save Named Setting**.

The Setting Name dialog box opens.

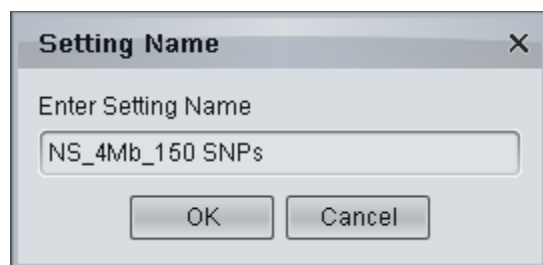


Figure 13.4 Named Setting List

3. Enter a name for the setting you wish to create and click **OK**.

The setting is saved and appears in the Named Setting dropdown list.

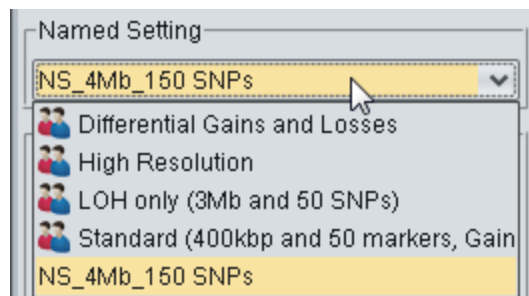


Figure 13.5 Named Setting dropdown list



**Note:** The Named Setting saves the settings at the time it was created. Subsequent changes to the settings will not be saved in the Named Setting.



### To select a Named Setting:

1. From the Named Setting drop-down list, select the setting;

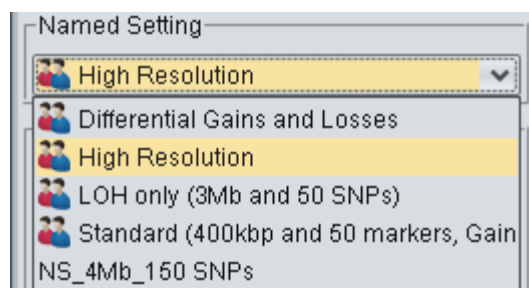


Figure 13.6 Named Setting dropdown list

Alternatively, from the Preference menu, select **Apply Named Setting...**

The Select Named Setting dialog box opens.

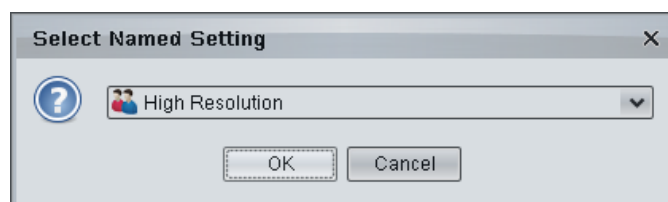


Figure 13.7 Select Named Setting dialog box

2. Select the Named Setting from the drop-down list and click **OK**.

The selected setting is applied.



**Note:** A Named Setting is not modified by any changes that you make to the settings in ChAS. If you want to keep a copy of your new settings, you will need to save them as a new Named Setting.

### To delete a Named Setting:

1. From the Preferences menu, select Delete Named Setting.

The Delete Setting dialog box opens.



**Note:** Shared Named Settings (denoted by the  icon in the Named Setting list) do not appear in the Delete Setting drop-down list. Users cannot delete or modify a shared Named Setting.

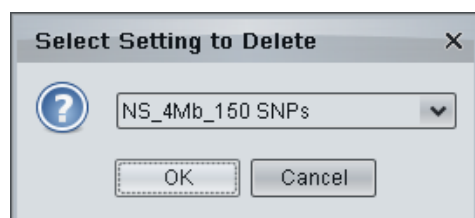


Figure 13.8 Delete Setting dialog box

2. Select the setting you wish to delete from the drop-down list and click **OK**.

The setting is deleted.

## Exporting and Importing Preferences

Preferences functions allow you to transfer most of the settings in a User Profile between one system and another.

- **Note:** If you import “exported” preferences that reference a Shared Named Setting which no longer exists, such as a Shared Named Setting from ChAS 1.0.1 or ChAS 1.1, the profile will be changed to point to the default Shared Name Setting.

To export preferences:

1. From the Preferences menu, select **Export Preferences...**

The Select Directory to export preferences to dialog box opens.

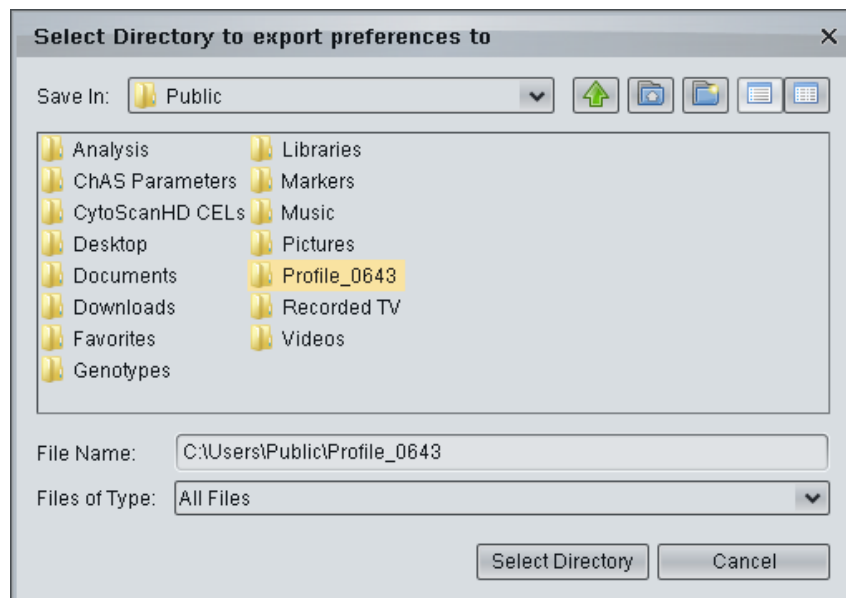



Figure 13.9 Select Directory to export preferences to

2. Use the navigation features of the dialog box to select or create a directory for the preferences.

- **Note:** The software will create a folder called “preferences” in the directory you select or create. If you select a directory that already contains a “preferences” folder, it will be overwritten. When you want to import the preferences, select the directory that contains the “preferences” folder that is indicated by the  icon.

3. Click **Select Directory**.

If the directory already contains a “preferences” folder, the Overwrite notice appears.

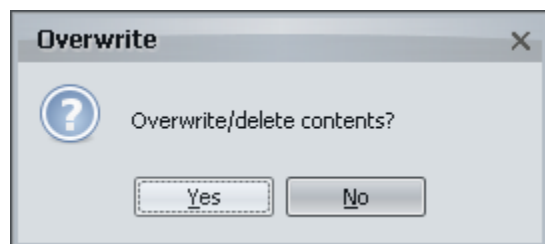


Figure 13.10 Overwrite notice

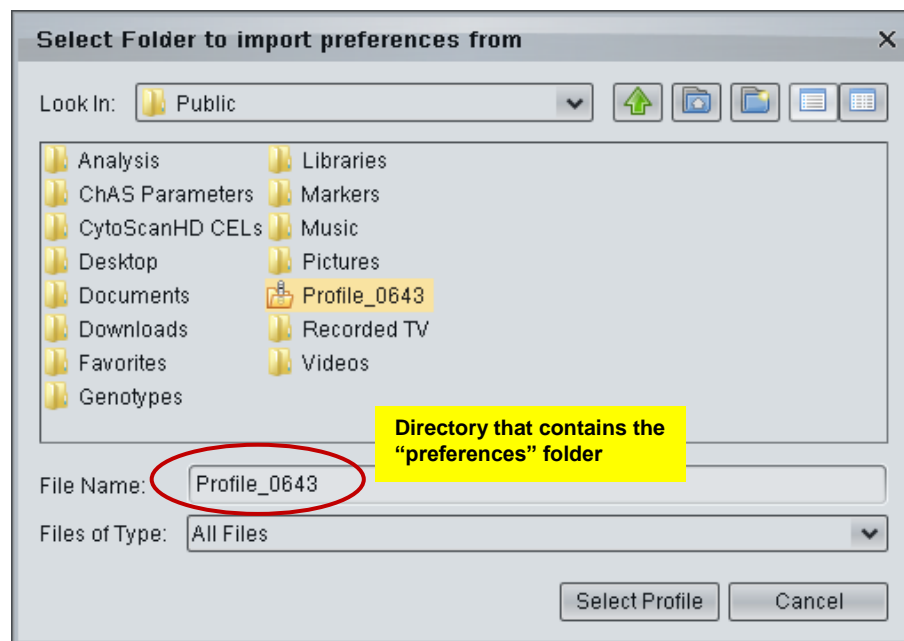
4. Click **Yes** to export the preference files to the directory that you selected.

You can then transfer the preferences to another user profile or system.


#### To import preferences:

1. From the Preferences menu, select **Import Preferences...**

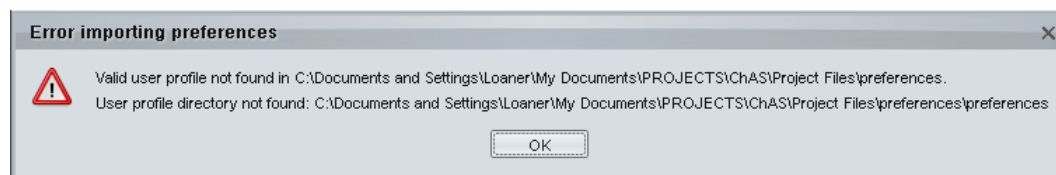
The Select Folder to import preferences from dialog box opens.



**Figure 13.11 Select Folder to import preferences from**

2. Use the navigation features of the dialog box to select the directory that the preferences were exported to (directories that contain a "preferences" folder are indicated by the ) icon.)
3. Click **Open** to import the preference files.

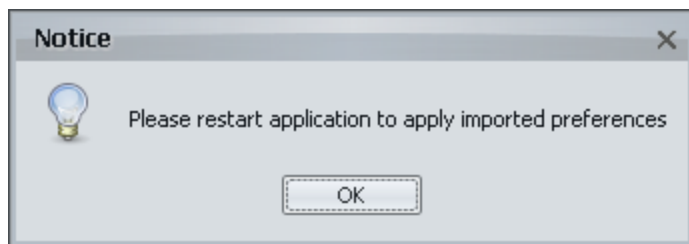
If you selected a directory that doesn't contain the "preferences" folder, the following notice appears.



**Figure 13.12 Error notice**

Click **OK** and repeat steps 1 through 3, selecting the correct folder

The Restart notice appears.



**Figure 13.13 Restart notice**

The imported preferences will not be applied until you restart ChAS.

## Appendix A: Analysis Parameters

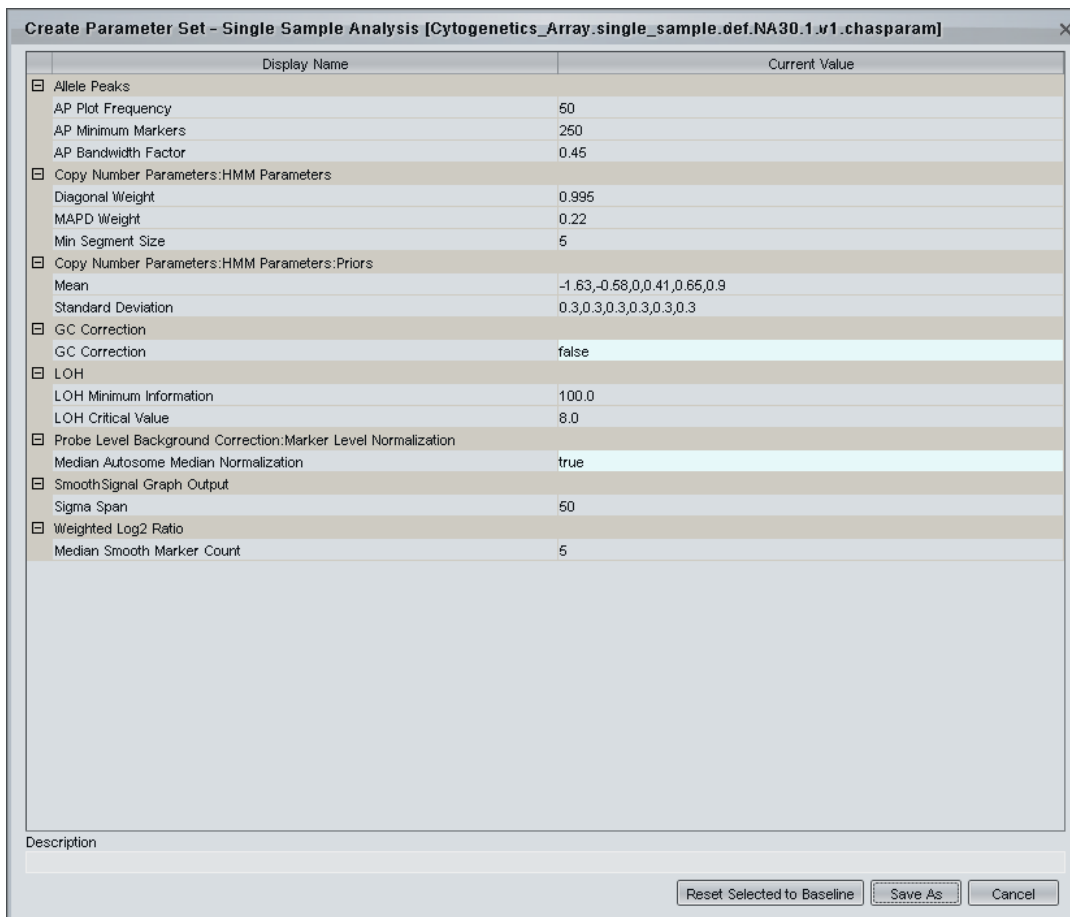
 **Note:** The CytoScan™ HD Array has no user-modifiable parameters for Single Sample Analysis.

Analysis parameters can be changed for the different CN/LOH analysis types:

- Analysis Parameters for Single Sample Analysis

The actual process of changing the parameters is described in [Creating a New Analysis Parameters File \(Customizing\)](#) (page 57).

### Analysis Parameters for Single Sample Analysis



Display Name	Current Value
<input checked="" type="checkbox"/> Allele Peaks	
AP Plot Frequency	50
AP Minimum Markers	250
AP Bandwidth Factor	0.45
<input checked="" type="checkbox"/> Copy Number Parameters:HMM Parameters	
Diagonal Weight	0.995
MAPD Weight	0.22
Min Segment Size	5
<input checked="" type="checkbox"/> Copy Number Parameters:HMM Parameters:Priors	
Mean	-1.63,-0.58,0,0.41,0.65,0.9
Standard Deviation	0.3,0.3,0.3,0.3,0.3,0.3
<input checked="" type="checkbox"/> GC Correction	
GC Correction	false
<input checked="" type="checkbox"/> LOH	
LOH Minimum Information	100.0
LOH Critical Value	8.0
<input checked="" type="checkbox"/> Probe Level Background Correction:Marker Level Normalization	
Median Autosome Median Normalization	true
<input checked="" type="checkbox"/> Smooth Signal Graph Output	
Sigma Span	50
<input checked="" type="checkbox"/> Weighted Log2 Ratio	
Median Smooth Marker Count	5

Description

**Figure A.1** User-modifiable parameters for the Cytogenetics Whole-Genome 2.7M Array

You can adjust the following types of algorithm parameters for the Cytogenetics Whole-Genome 2.7M Array:

- [Allele Peaks](#) (page 234)
- [Copy number parameters: HMM Parameters](#) (page 234)
- [Copy Number Parameters: HMM Parameters: Priors](#) (page 235)
- [GC Correction](#) (page 235)
- [LOH](#) (page 235)

- [Probe Level Background Correction: Marker Level Normalization](#) (page 235)
- [Smooth Signal Graph Output](#) (page 236)
- [Weighted Log2 Ratio](#) (page 236)



**Important:** The actual analysis software uses different terms to identify the parameters. You may see these terms in various error messages. The internal names are listed in parenthesis at the end of the description and in the material below.

## Allele Peaks

### AP Plot Frequency

Tells the algorithm how many SNPs to skip up for each allele peak. It should be a positive integer starting from 1. The maximum allowable input for this parameter is 200,000 for the Cytogenetics Array. At 100, allele peaks are calculated at every 100th SNP (using the window size defined by **AP Minimum Markers**). At 1, allele peaks are calculated at every SNP. The default value is 100 for the Cytogenetics Array.

### AP Minimum Markers

The number of markers in each window be used to find allele peaks. It cannot be too small. Recommended at least 30 or higher. The larger this parameter is, the less variation of allele peaks would be. However, it should not be set to be too large, otherwise, the markers that represent the local biological events, such as LOH, will be averaged out by their neighboring non-LOH markers. The maximum allowable input for this parameter is 5000 for the Cytogenetics Array. The default value is 250 for the Cytogenetics Array.

### AP Bandwidth Factor

A variation of bandwidth choice that is suggested by Scott (1992) is used for the density estimation for the allele plot. Since this bandwidth is likely to be too large for the density estimator to pick up small local biological events, we recommend that you adjust the bandwidth by this factor. In general, larger **AP Bandwidth Factor** values result in fewer peaks and vice versa. This input must be a decimal number between 0.01 and 1. The default value is 0.45 for the Cytogenetics Whole-Genome 2.7M Array.

Scott, D. W. (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley

### Clean Threshold

Controls with noisy markers are removed from the visualization because they are indeterminately far from adjacent peaks. A value of 0 will result in removal of no markers. Min=0.0, max=1.0, the default value is 0.3.

### Symmetry

This is a logical value ('false' = off, 'true' = on) that controls whether or not the data is mirrored about the X axis before fitting the density curves. A setting of true will tend to improve the estimation of peak locations for copy number gain regions. The default value is 'true'.

## Copy number parameters: HMM Parameters

### Diagonal Weight (Internal name=diagonal-weight)

Diagonal weight refers to the values on the diagonal of the HMM transition matrix. This weight can take on a value in the range (0,1) and it acts as a penalty for changing copy number state from one marker to the next in either direction. A diagonal weight of 0 would force a change in copy number to always happen. Likewise a diagonal weight of 1 would never allow a change to occur. These extremes are not accepted by APT which executes the HMM. The default value is close to 1 and discourages frequent changes in copy number but short runs of markers with values indicating a copy number change will overcome this weight as will large outliers. Given the noise level in copy number data, it makes no sense to use a diagonal weight below, say, 0.9. This will overload output with too many segments. Pushing the weight too close to 1 in order to not produce short segments can cause spurious segments of non-normal copy number. It is simplest to eliminate short segments by post processing the segmentation.

### **MAPD Weight (Internal name=map-weight)**

MAPD weight is used to add the Median Absolute Pairwise Difference statistic to the dispersion parameter Standard Deviation found in Copy Number Parameters:HMM Parameters:Priors. If the MAPD weight is increased from the default, it makes sense to decrease the Standard Deviation.

### **Min Segment Size (Internal name=min-segment-size)**

Min Segment Size is used to control reporting of small segments by APT. The markers in any segments shorter than the minimum size specified will have their copy number assignments imputed to conform with neighboring segments that are at least as large as the minimum size. Changing the default to a value of 1 turns this off.

## **Copy Number Parameters: HMM Parameters: Priors**

### **Mean (Internal name=hmmCN\_mu)**

Mean lists the expected values of the log base 2 ratios with respect to the reference sample corresponding to each copy number state. It is best to have the means as accurate as possible, however, it is difficult to estimate the copy number means of any sample from within the sample. The defaults have been determined out of sample.

### **Standard Deviation (Internal name=hmm\_CNsigma)**

Standard Deviation lists the corresponding expected standard deviations in the log2 ratio data corresponding to each copy number state. Note that MAPD is computed for each sample and added to these standard deviations after MAPD is multiplied by the MAPD weight.

## **GC Correction**

### **GC Correction (Internal name=gc\_correction)**

Some samples are subject to variation that is linked to regional variation in chromosomal GC content. The gc-correction algorithm removes such variation using a non-parametric method that matches probe set intensities given chromosomal GC content to the reference.

## **LOH**

### **LOH Minimum Information (Internal name=minInformation)**

LOH Minimum Information (in my algorithm documentation this is coded as T\_min) controls the size of the windows used by the algorithm. An attempt is made to make a call for LOH or not LOH for each window. The larger this value, the larger the size of the windows used (i.e. it can be used to reduce the number of small regions detected). Most of the time, the algorithm will actually dynamically select the value that is used for optimal performance. The value that the user inputs here is only used if the algorithm selects an even smaller value (i.e. if the user says 100 and the algorithm selects 75, then 100 will be used. The units of this quantity are based on giving each marker a weight based on its expected level of performance; as such, this quantity is the total information content required in a region to be able to make a decision.

### **LOH Critical Value (Internal name=lambdaCritical)**

LOH Critical Value (lambdaCritical) is the cut-off value for the test statistic used to decide between LOH and non-LOH. A value of 0 for the test statistic means that each are equally likely. A value higher than 0 means the LOH is more likely. Increasing this value reduces the number of false positives but will decrease resolution. The scale on this quantity is logarithmic. So a value of 1 means LOH is 2.71 times more likely than non-LOH, and a value of 10 means LOH is approximately 22,000 times more likely than non-LOH.

## **Probe Level Background Correction: Marker Level Normalization**

### **Median Autosome Median Normalization (Internal name=median-autosome-median-normalization)**

There is a secondary optional normalization ("MedianAutosome") done after log2 ratios are formed: the log2 ratios are adjusted by subtracting the median of the median log2 ratio of all the autosomes. This adjustment can be useful for samples with primarily diploid autosomes when probe-level normalization may be affected by an

aneuploidy such as high CN gain in Chromosome X. Note that this adjustment is not a probe-level background correction. This is only recommended for samples where most of the chromosomes are relatively normal.

### **Smooth Signal Graph Output**

#### **Sigma Span (Internal name=sigma\_span)**

A sigma span is the number of markers spanned by one standard deviation of the Gaussian smoothing kernel that is used for the SmoothSignal graph's smoothing.

#### **Weighted Log2 Ratio**

Weighted log2 ratio is calculated differently for different array types. For the Cytogenetics Array, weighted log2 ratio is a running median of a genomic window of log2 ratios that filters outliers. For the CytoScan™ HD Array, weighted log2 ratio is the Log2 ratios processed through a Bayes wavelet shrinkage estimator (not a running median smooth of the data). These processed values are input to the CNState algorithm HMM for the CytoScan™ HD Array.

#### **Median Smooth Marker Count (Internal name=median-smooth-marker-count)**

Weighted log2 ratio Median Smooth Marker Count is used in calculating the weighted log2 ratio track for the Whole-Genome Cytogenetics 2.7M Array. This track is constructed by taking a window size of "Weighted log2 ratio Median Smooth Marker Count" around each marker and calculating the median log2 ratio.

## **Reference Model File Creation**

Reference Model File Creation is done with fixed parameters in the Reference Creation workflow.

It is essential that the input for Reference Model Creation include at least 44 total CEL files and at least 20 males and 20 females.

There are no user-adjustable parameters for the Reference Model File Creation



## Appendix B: AED File Format

Affymetrix Extensible Data files contain data that annotate positions on a genome. AED allows custom, typed fields, can be edited in Excel, and supports internationalization.

AED files can be created by Chromosome Analysis Suite (ChAS) and loaded into ChAS as region information files to:

- Define CytoRegion and Overlap Map regions
- Record information of interest about features in the genome

They can be produced and edited in:

- ChAS (recommended)
- Text-editing software
- Spreadsheet software such as Microsoft Excel.

This appendix explains the formatting and use of Affymetrix Extensible Data (AED) files with ChAS.

- [AED File Description](#)
- [Property Name Elements](#) (page 239)
- [Compatibility](#) (page 243)
- [References](#) ( page 245)

### AED File Description

An AED file contains a list of annotations, descriptions of features on a biological *sequence* such as a chromosome. This description is comprised of several *properties*—either properties defined by this specification, such as the annotation start and stop positions; or properties defined by users or third parties.

An AED file may also provide *metadata* which describe the particular group of annotations in the file as a whole, such as the author of the file or the genome assembly for which the annotations were produced.

Properties and metadata have certain *types* which define the semantics and constrain the range of values they may have. Properties should begin with a lowercase letter, while types should begin with an uppercase letter.

The AED file format uses a tab-delimited text format with the \*.aed file extension. It uses Unicode character sets.



**Important: AED supports only Unicode, which can be stored in one of various encodings (*charsets* such as UTF-8, UTF-16LE, and UTF-16BE). The AED file indicates the charset with an initial Byte Order mark (BOM). An AED file with no initial BOM is not recommended. An AED file that does not begin with a BOM will be interpreted as containing only the ASCII subset of Unicode, resulting in an error if any characters lie outside the range of ASCII. (With no indication of a charset, it is not possible to determine which non-ASCII characters were intended. File formats such as BED that make assumptions about non-ASCII characters have the potential to corrupt data when transported between systems.)**

It has the following components:

- Header Row: names the properties that can be used in the annotations
- Metadata (optional): provides information about the AED file itself and the group of annotations it contains.
- Annotations: for each feature annotated, an annotation row provides values for the properties listed in the header rows.

	A	B	C	D	E
Header Row	1 bio:sequence(aed:String)	bio:start(aed:Integer)	bio:end(aed:Integer)	aed:name(aed:String)	aed:value(aed:String)
	2			aed:application(aed:String)	Chromosome Analysis Suite CytoB-N1.0.0.284 (r1764)
	3			aed:created(aed:DateTime)	2009-03-02T11:06:40.517-08:00
	4			aed:modified(aed:DateTime)	2009-03-02T11:06:40.517-08:00
	5 chr1	110034485	110041690	seg10	
	6 chr1	112497197	112507652	seg12	
	7 chr4	66574165	66587983	seg126	
Annotations	8 chr4	69102606	69592855	seg130	
	9 chr4	68970721	68970726	seg128	
	10 chr4	70177951	70276607	seg132	
	11 chr5	102163811	102287067	test_variation	
	12 chr5	102622340	102642260	New_Gene	
	13				

**Figure B.1 AED file in Excel with required header fields for properties and metadata**

## Header row

The header row of an AED file is a tab-delimited list of the properties that can be used to describe a region of the genome.

Each AED file header represents a property. Normal records in the file represent annotations, and the record fields represent annotation properties. Special metadata records represent metadata properties for the file as a whole, rather than for a particular annotation.

A property name has the following format:

`namespacePrefix:propertyIdentifier (namespacePrefix:TypeIdentifier)`

- `namespacePrefix` (**optional**) assigns the property or type to a vocabulary grouping called a namespace. The lack of a namespace prefix indicates that the property has been created by a user and is not part of the formal AED specification.

The lack of a namespacePrefix indicates that the property is in the default/custom namespace; this namespace enables users to add properties to an annotation just by adding new columns, such as `foo(aed:String)` or `bar(aed:Integer)`.

See Namespaces for more information.

- `propertyIdentifier` names the property that the values in the column are for.  
Each annotation can be assigned an unlimited number of properties. Each property has a certain meaning, and this meaning is usually defined by the documentation for the property namespace. The purpose of the AED file is to indicate values for certain properties for each annotation. For example, by use of the `aed:name(aed:String)` column, the AED file indicates a string value to be used as the name each annotation).
- `TypeIdentifier` (always capitalized) specifies the data type of the value to be used for the property in the AED file.

Examples include:

- `bio:sequence(aed:String)`
- `aed:value(aed:String)`
- `medianMarkerDistance(aed:Integer)`

## Required Fields

Fields may appear in any order, except that the following predefined fields must always appear in the following order at the beginning of the header:

- `bio:sequence(aed:String)`
- `bio:start(aed:Integer)`
- `bio:end(aed:Integer)`
- `aed:name(aed:String)`
- `aed:value(aed:String)`

(**optional:** you need to use this property if you are including metadata information in the file).



**Important: ChAS validates the property types when importing an AED file. If a file header specifies a known property, but includes an incorrect data type for the property, the file will not be loaded. For example, “fish:score” is a known property with “whole number” data type. An AED file header that specifies “fish:score(aed:String)” would be treated as an error.**

## Metadata Records

Some records, instead of providing annotation about a location on a genome assembly, provide metadata information about the AED file itself. These *metadata records* are identified by the presence of an empty string in the `bio:sequence` field. The `bio:start` and `bio:end` fields must also be left blank for metadata records. If there are metadata records present, the `aed:value` field is required.

In a metadata record, the value in the `aed:name` field is interpreted as the name of the metadata property, with type identification rules identical to those of the header fields. The value in the `aed:value` field is interpreted as the value of the metadata property, and the characters that make up its string value must follow the lexical and semantic rules specified by the type indicated in the `aed:name` field.

<code>bio:sequence(aed:String)</code>	<code>bio:start(aed:Integer)</code>	<code>bio:end(aed:Integer)</code>	<code>aed:name(aed:String)</code>	<code>aed:value(aed:String)</code>
			<code>aed:application(aed:String)</code>	Chromosome Analysis Suite CytoB-N1.0.0.284 (r1764)
			<code>aed:created(aed:DateTime)</code>	2009-03-02T11:06:40.517-08:00
			<code>aed:modified(aed:DateTime)</code>	2009-03-02T11:06:40.517-08:00

Blanks for `bio:sequence` and other properties

Metadata Property  
names

Metadata Property values

**Figure B.2 Metadata entries**

All other fields in a metadata record should be blank.

## Annotations

The rows below the Metadata properties are the annotations. Each row is a tab-delimited list of values. Each value must have the correct data type, as described in the property name for that value.

## Property Name Elements

The property name elements are described in more detail in the following sections:

- [Namespaces](#) (page 240)
- [Properties](#) (page 240)

- [Types](#) (page 242)

## Namespaces

The name of each type and property in AED is considered part of a vocabulary grouping called a *namespace*. Namespaces prevent clashes between names defined by disparate parties, as well as unambiguously identify commonly used types and properties so that identical semantics may be assured. A namespace is identified by a Uniform Resource Identifier (URI) as defined in **RFC 3986**.

A type or property identifies its namespace by a namespace prefix followed by a colon character. If no namespace prefix is present, the type or property is considered part of the AED default namespace. The part of the type or property after the namespace prefix is considered its *simple name*.

AED has several build-in namespaces, with predefined namespace URIs and prefixes:

Namespace Prefix	Description	Examples
	Custom and experimental properties not yet established as standard.	<ul style="list-style-type: none"> <li>• <code>foo</code></li> <li>• <code>bar</code></li> <li>• <code>myProperty</code></li> </ul>
<code>aed</code>	AED-specific properties and types.	<ul style="list-style-type: none"> <li>• <code>aed:name</code></li> <li>• <code>aed:Integer</code></li> </ul>
<code>bio</code>	Descriptions of biological entities.	<ul style="list-style-type: none"> <li>• <code>bio:sequence</code></li> <li>• <code>bio:Strand</code></li> </ul>
<code>style</code>	Information related to information representation, visually or otherwise.	<ul style="list-style-type: none"> <li>• <code>style:color</code></li> </ul>

If any other namespace is used in an AED file, it must be declared in the metadata section of the file using the special namespace prefix. The simple name of the metadata header indicates the prefix being declared, and the value (of type `aed:URI`) indicates the namespace to be associated with the prefix. For example, to associate the prefix “*example*” with the URI `http://example.com/namespace/`, and the “*affx*” with the URL <http://affymetrix.com/ontology/>, use the following metadata record:

```
namespace:example (aed:URI)    http://example.com/namespace/
namespace:affx (aed:URL)      http://affymetrix.com/ontology
```

## Properties

This section describes the properties defined by the AED specification. By convention property names begin with lowercase letters. A predefined property is only required if indicated. Some properties are only useful as metadata, and these are so indicated.

### Affymetrix Properties

This section describes the properties defined by the AED specification. By convention, property names begin with lowercase letters. A predefined property is only required if indicated. Some properties are only useful as metadata, and these are so indicated.

Name	Type	Description
<code>Affx:ucscGenomeVersion</code> ( <code>aed:String</code> )	<code>aed:String</code>	<b>(Metadata)</b> The genome assembly version using UCSC names, for example “hg19”.

### AED Properties

These properties are parts of the AED namespace.

Name	Type	Description
aed:application	aed:String	<b>(Metadata)</b> The name of the application that produced the AED file, if metadata, or the annotation.
aed:category	aed:String	Identifies the group and optionally subgroups into which the resource is classified. Subcategories, if any, should be delimited using the forward slash character ' / ' (U+002F) with no whitespace. (Example: copynumber/gain)
aed:created	aed:DateTime	<b>(Metadata)</b> The point in time the data was created; this is not necessarily the time the file was created.
aed:counter	aed:Integer	A general purpose field to be incremented when user-defined circumstances occur. A common use for this field is to indicate, the number of samples in which the condition has been observed.
aed:modified	aed:DateTime	<b>(Metadata)</b> The point in time the data was modified; this is not necessarily the time the file was modified.
aed:name	aed:String	<b>(Required.)</b> Indicates the name of the record.  In a metadata record, this value is interpreted as the name and type of the metadata property.
aed:note	aed:String	A user-defined explanation or comment regarding the annotation.
aed:value	aed:String	<b>(Required only if metadata records are present.)</b> In a metadata record, this value is interpreted as the value of the metadata property. An AED processor must ignore this field for all non-metadata records.
aed:uuid	aed:UUID	<b>(Metadata)</b> The universally unique identifier of the resource. Although allowed, it is not always advised to identify user-editable resources such as AED documents with UUIDs, as copying and manually editing such resources can result in multiple such resources with identical UUIDs, negating the purpose of UUIDs.

## Biology Properties

Name	Type	Description
bio:assembly	aed:URI	<b>(Metadata)</b> A URI indicating the genome assembly used. Currently the <a href="#">DAS GlobalSeqIDs</a> are recommended.
bio:state	aed:Rational	The algorithm-determined state (e.g. copy number) of an annotation.
bio:confidence	aed:Rational	A value between 0.0 and 1.0, inclusive, indicating the confidence that an annotation call is accurate.

Name	Type	Description
bio:end	aed:Integer	<b>(Required.)</b> The zero-based ending position, exclusive, of the record along the sequence.
bio:markerCount	aed:Integer	The number of markers such as probes that intersect an annotation.
bio:sequence	aed:String	<b>(Required.)</b> The name of the chromosome (e.g. chr3, chrY), contig (e.g. ctg5), or scaffold (e.g. scaffold90210).  The special value of an empty string ("") indicates that the record is a metadata record, giving special meaning to values in other fields in the record.
bio:start	aed:Integer	<b>(Required.)</b> The zero-based starting position, inclusive, of the record along the sequence.
bio:strand	bio:Strand	The sequence strand on which a feature lies.

## Style Properties

Style properties are used to control the display of the annotation.

Name	Type	Description
style:color	aed:Color	The color to be used when visually depicting the annotation.

## Types

AED defines the following types. By convention type names begin with an uppercase letter. The lexical form is applied to the resulting character sequence derived after following the quoting rules of **RFC 4180**.

Type	Description	Lexical Form	Examples
aed:String	A sequence of Unicode characters.	<i>character*</i>	<ul style="list-style-type: none"> <li>• Abc</li> <li>• a "fun" test!</li> </ul>
aed:Integer	The positive whole numbers, the negative whole numbers, and zero.	<i>[-] digit+</i>	<ul style="list-style-type: none"> <li>• 123</li> <li>• 0</li> <li>• -5000</li> </ul>
aed:Rational	A rational number. Currently only rational numbers with finite decimal expansion are allowed. AED processors typically implement this type using floating point values. This type does not allow floating point not-a-number (NaN) values.	<i>[-]digit+.digit+ [e[+]-]digit+</i>  The lexical form "NaN" is explicitly prohibited.	<ul style="list-style-type: none"> <li>• -123.0</li> <li>• 0.0</li> <li>• 123.45</li> <li>• 1.2e+5</li> </ul>

Type	Description	Lexical Form	Examples
aed:Boolean	A binary true/false value.	"false"   "FALSE"   "true"   "TRUE"	<ul style="list-style-type: none"> <li>• false</li> <li>• FALSE</li> <li>• TRUE</li> </ul>
aed:URI	A Uniform Resource Identifier (URI).	See <b>RFC 3986</b> .	<ul style="list-style-type: none"> <li>• http://example.com/</li> <li>• urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6</li> <li>• mailto:jdoe@example.com</li> </ul>
aed:DateTime	A timestamp with the absolute date and time and an identified time zone. The lexical form is a subset of <b>ISO 8601</b> with required milliseconds and time zone offset.	YYYY-MM-DDThh:mm:ss.s+ (+ -)hh:mm	<ul style="list-style-type: none"> <li>• 2008-09-12T18:45:43.779-07:00</li> </ul>
aed:Color	A color. Currently only supports colors in the RGB color space. The lexical form represents red, green, and blue components, respectively, each supporting a decimal integer value 0-255)	rgb(digit+, digit+, digit+)	<ul style="list-style-type: none"> <li>• rgb(0, 0, 0)</li> <li>• rgb(200, 50, 128)</li> <li>• rgb(255, 255, 255)</li> </ul>
bio:Strand	Represents a sequence strand relative to the landmark or assembly. The lack of an indicated strand encompasses the semantics of "unknown", and "non-stranded."	One of the following characters: + (U+002B) Forward strand. - (U+002D) Reverse strand.	<ul style="list-style-type: none"> <li>• +</li> <li>• -</li> </ul>
aed:UUID	A Universally Unique Identifier (UUID) as specified by <b>RFC 4122</b> (in canonical form, not as a URN).	xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx	<ul style="list-style-type: none"> <li>• f81d4fae-7dec-11d0-a765-00a0c91e6bf6</li> </ul>

## Compatibility

### UCSC Browser Extensible Data (BED)

The BED file format, developed at UCSC, is widely used for transfer of simple region coordinates. However, the format has been interpreted and implemented in multiple ways by various software within and outside of UCSC. Some implementations require a TAB delimited format, others require a space-delimited format, and still others accept both. Characters outside of the ASCII character set are not well supported. We created the AED format with very strict and explicit definitions so as to avoid some of these compatibility issues.

Although the AED format is preferred, ChAS supports both the import and export of data in BED format. When exporting data in BED format, ChAS exports only the basic 4-column tab-delimited BED format containing the position and name of each item. If the names of any of your items contain spaces or non-ASCII characters, there is no guarantee that all programs will be able to interpret those names correctly.

When importing data in BED format, ChAS supports the reading of BED files with anywhere from 4 to 12 columns.

- The file must be TAB delimited
- Only ASCII characters should be used
- The values for thickStart and thickEnd will be ignored for display purposes
- The value for itemRgb will be honored for display purposes
- The values for blockCount, blockSizes and blockStarts can be used to import and display data with intron/exon structure, such as genes.
- Formatting rules in the BED header are ignored
- BED files containing multiple tracks are not supported; use a separate BED file for each track.

The UCSC Browser, as well as ChAS, uses the strict definition of BED where chromStart is not allowed to be greater than chromEnd. ChAS will accept import of BED files even if this convention is violated, but will auto-correct and export BED files properly with  $\text{chromStart} \leq \text{chromEnd}$ .

AED has been structured to facilitate as much as possible migration of data rows to and from BED. Starting with existing AED and BED files, data records from AED may be transferred to BED by using:

- The “Export” function from inside ChAS (recommended)
- A text editor if the AED files are first prepared in the following manner:
  - Remove all fields except for `bio:sequence`, `bio:start`, `bio:end`, and `aed:name`.
  - Ensure that no non-ASCII characters are included. (The treatment of non-ASCII characters by a BED processor is undefined.)
  - Ensure that no name contains whitespace characters.
  - Data rows from the first four columns of a BED file can be transferred to AED with no constraints as long as the columns are delimited by TAB.

## Microsoft Excel and Other Spreadsheet Applications

An AED file may be edited using most spreadsheet programs that support tab-separated value (TSV) files and that recognize a byte order mark (BOM). An AED file can be edited in Microsoft Excel, for example, using the following rules:

When loading an AED file into Microsoft Excel as a TSV file, make sure that the Unicode code page for the correct encoding is selected (preferred), or accept the default “Windows (ANSI)” code page (which should still recognize Unicode characters if the correct BOM is present in the file).

When saving an AED file from Microsoft Excel, make sure the “Unicode Text” type is selected. This will result in a file encoded in UTF-16LE, which is still a valid AED file as it begins with the appropriate BOM.

## Microsoft Notepad and Other Text Editors

An AED file may be edited by any text editor that supports Unicode and that uses a byte order mark (BOM) to indicate the charset. The version of Microsoft Notepad in Windows XP, for example, will both correctly read text files marked with a BOM and save text files using the appropriate BOM if the following rules are followed:

When saving an AED file from Microsoft Notepad, make sure the encoding is set to “UTF-8” or “Unicode”.

For other text editors, make sure the correct preferences are set both to recognize and write BOMs for files.

### Text Editors

EmEditor <<http://www.emeditor.com/>> is a commercial text editor that has extremely good Unicode and BOM support, and is able to open up gigantic text files.

PSPad <<http://www.pspad.com/>> is a free text editor that has particularly extensive Unicode and BOM support and is available in many localizations.



UniPad <<http://www.unipad.org/>> is a shareware text editor that correctly handles Unicode and BOM, and provides a wide range of built-in glyphs for representing Unicode code points that cannot be viewed on most other text editors.

## References

**ISO 8601:** *ISO 8601:2004(E): Data elements and interchange formats — Information interchange — Representation of dates and times*. International Organization for Standardization, 2004-12-01.

**Microsoft Byte Order Mark:** [http://msdn.microsoft.com/en-us/library/ms776429\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/ms776429(VS.85).aspx)

**RFC 3986:** *RFC 3986: Uniform Resource Identifier (URI): Generic Syntax*. T. Berners-Lee, R. Fielding, and L. Masinter. Internet Engineering Task Force, 2005. <http://tools.ietf.org/html/rfc3986>

**RFC 4122:** *RFC 4122: A Universally Unique Identifier (UUID) URN Namespace*. P. Leach, M. Mealling, and R. Salz. Internet Engineering Task Force, 2005. <http://tools.ietf.org/html/rfc4122>

**RFC 4180:** *RFC 4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files*. Y. Shafranovich. Internet Engineering Task Force, 2005. <http://tools.ietf.org/html/rfc4180>

**Unicode Byte Order Mark FAQ:** [http://unicode.org/faq/utf\\_bom.html](http://unicode.org/faq/utf_bom.html)

# Appendix C: ChAS Properties and Types

Starting with version 1.1.0, ChAS has adopted the framework underlying AED files as its native framework for identifying and storing properties and value types. Every property of annotations, files, and other objects within the software is now identified by a URI behind the scenes. Furthermore, the types of values given to these properties are the same types available within AED files. The AED framework therefore provides a consistent and pervasive approach to describing entities throughout the application and seamlessly across AED files and other types of files such as CxCHP files.

## Identifying Properties within ChAS

### Standard AED Property Style

Because every user-accessible property within ChAS complies with the AED framework, any property may be entered in standard `prefix:simpleName` style, just as it would appear in an AED file. For example, the creation date of an entity may be entered using the `aed:created` property name.

The predefined AED prefixes defined for AED files may always be used. Unlike AED files, which allow declaration of arbitrary prefixes with additional namespaces, ChAS has an additional list of predefined namespace prefix associations valid only within the context of the ChAS user interface. These prefixes may be used to refer to the corresponding namespaces with no explicit namespace declaration. For example, the `fish` prefix may be used to refer to FISH namespace properties (e.g. `fish:labs`) with no need to explicitly associate the FISH namespace URI with the namespace. Table C.1 lists the namespace prefixes recognized within the ChAS user interface.

Table C.1 ChAS namespace prefixes

Namespace URI	Prefix	Label
---------------	--------	-------

<a href="http://affymetrix.com/ontology/aed/">http://affymetrix.com/ontology/aed/</a>	aed	General
<a href="http://affymetrix.com/ontology/aed/biology/">http://affymetrix.com/ontology/aed/biology/</a>	bio	Biology
<a href="http://affymetrix.com/ontology/aed/style/">http://affymetrix.com/ontology/aed/style/</a>	style	Style
<a href="http://affymetrix.com/ontology/aed/default/">http://affymetrix.com/ontology/aed/default/</a>	(none)	Custom
<a href="http://affymetrix.com/ontology/">http://affymetrix.com/ontology/</a>	affx	Affx
<a href="http://affymetrix.com/ontology/algorithm/">http://affymetrix.com/ontology/algorithm/</a>	alg	Algorithm
<a href="http://affymetrix.com/ontology/algorithm/option/">http://affymetrix.com/ontology/algorithm/option/</a>	algot	Algorithm Option
<a href="http://affymetrix.com/ontology/algorithm/state/">http://affymetrix.com/ontology/algorithm/state/</a>	algstate	Algorithm State
<a href="http://affymetrix.com/ontology/arr/">http://affymetrix.com/ontology/arr/</a>	arr	ARR
<a href="http://affymetrix.com/ontology/chp/">http://affymetrix.com/ontology/chp/</a>	chp	CHP
<a href="http://affymetrix.com/ontology/netaffx/">http://affymetrix.com/ontology/netaffx/</a>	netaffx	NetAffx
<a href="http://affymetrix.com/ontology/genome.ucsc.edu/bed/">http://affymetrix.com/ontology/genome.ucsc.edu/bed/</a>	bed	BED
<a href="http://affymetrix.com/ontology/projects.tcag.ca/variation/">http://affymetrix.com/ontology/projects.tcag.ca/variation/</a>	dgv	DGV
<a href="http://affymetrix.com/ontology/www.genome.ucsc.edu/fishClones/">http://affymetrix.com/ontology/www.genome.ucsc.edu/fishClones/</a>	fish	FISH
<a href="http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/omim/">http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/omim/</a>	omim	OMIM
<a href="http://affymetrix.com/ontology/www.pubmed.gov/">http://affymetrix.com/ontology/www.pubmed.gov/</a>	pubmed	PubMed
<a href="http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/RefSeq/">http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/RefSeq/</a>	refseq	RefSeq
<a href="http://affymetrix.com/ontology/www.genome.ucsc.edu/genomicSuperDups/">http://affymetrix.com/ontology/www.genome.ucsc.edu/genomicSuperDups/</a>	superdup	Segmental Duplications

## ChAS Property Style

Properties may also be entered by the user in a more attractive format specific to ChAS, consisting of the property simple name followed by a space and a special label in parentheses to identify the namespace. The namespace label is presented in the table above. For example, the `fish:labs` property may also be entered as `labs (FISH)`.

The ChAS style is the default format for display within ChAS. Regardless of what format you use to enter a property, it will be displayed in its ChAS style.

## Full URI Property Style

Any property, including those with predefined namespace prefixes, may be entered using its full canonical URI. This allows entry of unknown third-properties in other namespaces. If the property is in a known namespace, it will be displayed using the ChAS property style described above. For example, the `fish:labs` property may be entered using the full URI

`http://affymetrix.com/ontology/www.genome.ucsc.edu/fishClones/labs`; the property will be displayed as `labs` (FISH).

Properties may be entered in ChAS three different ways:

- **ChAS style (default):** `labs` (FISH)
- **AED style:** `fish:labs`
- **AED URI:** `http://affymetrix.com/ontology/www.genome.ucsc.edu/fishClones/labs`

## Identifying Value Types within ChAS

With the incorporation of the AED framework, ChAS now uses the same AED types used within AED files. Because there is a small, fixed set of types, within the application user interface these types are always selected using a prepopulated list such as a drop-down list control. The types are displayed using special user-friendly names, even though they refer to the same corresponding types within AED files. These types and their ChAS style labels appear below.

- `aed:Boolean` - Boolean
- `aed:Color` - Color
- `aed:DateTime` - DateTime
- `aed:Integer` - Whole Number
- `aed:Rational` - Decimal Number
- `aed:Strand` - Strand
- `aed:String` - Text
- `aed:URI` - URI
- `aed:UUID` - UUID

Some types may not be available in some contexts.

## Automatic Conversion of CxCHP Headers to Properties

The file format used by CxCHP files uses a separate approach for describing properties and value types. However, there exist equivalent AED types for the types used by CxCHP files. Moreover, the header names used in CxCHP files are logically grouped into categories by certain prefixes. ChAS uses these groupings to place CxCHP headers into AED namespaces when loading CxCHP files. ChAS also modifies the names of the headers to make them more readable and to provide consistency with other AED properties used throughout the application.

### Header Name Conversion

ChAS uses the following namespaces when converting CHP header names based on the header name prefix.

**Table C.2** CHP header name conversion to AED namespace

CHP Header Name Prefix	AED Namespace
affymetrix-algorithm-	<a href="http://affymetrix.com/ontology/algorithm/">http://affymetrix.com/ontology/algorithm/</a>
affymetrix-algorithm-param-	<a href="http://affymetrix.com/ontology/algorithm/">http://affymetrix.com/ontology/algorithm/</a>
affymetrix-algorithm-param-option-	<a href="http://affymetrix.com/ontology/algorithm/option/">http://affymetrix.com/ontology/algorithm/option/</a>
affymetrix-algorithm-param-state-	<a href="http://affymetrix.com/ontology/algorithm/state/">http://affymetrix.com/ontology/algorithm/state/</a>
affymetrix-chipsummary-	<a href="http://affymetrix.com/ontology/chp/summary/">http://affymetrix.com/ontology/chp/summary/</a>
(all others)	<a href="http://affymetrix.com/ontology/chp/">http://affymetrix.com/ontology/chp/</a>

After determining the appropriate namespace to use, ChAS removes the header name prefix and modifies the remaining characters according to the following rules (simplified):

1. All beginning uppercase letters are converted to lowercase.
2. All separator characters (such as '-' and '\_') are removed.
3. The characters immediately following separators are converted to uppercase.

For example, the CHP header `affymetrix-algorithm-param-option-gender-override-file` will be converted to `optionGenderOverrideFile` and placed in the <http://affymetrix.com/ontology/algorithm/> namespace.

### Converted Properties

ChAS performs further special conversions on the following header parameters for historical and consistency reasons.

CHP Header Name	Property
affymetrix-array-type	<a href="http://affymetrix.com/ontology/arrayType">http://affymetrix.com/ontology/arrayType</a>
affymetrix-chipsummary-snp-qc	<a href="http://affymetrix.com/ontology/chp/summary/snpQC">http://affymetrix.com/ontology/chp/summary/snpQC</a>
affymetrix-chipsummary-MAPD	<a href="http://affymetrix.com/ontology/chp/summary/mapd">http://affymetrix.com/ontology/chp/summary/mapd</a>

### Derived Properties

The following properties are each assigned to a file property derived from one or more header parameters, CxCHP file attributes, or other information, in the given order:

<http://affymetrix.com/ontology/aed/created> (aed:created)

1. The file creation time in the Calvin generic data header.
2. The `create_date` header parameter.

3. The create-date header parameter.

`http://affymetrix.com/ontology/aed/modified (aed:modified)`

1. File system last modified date.

`http://affymetrix.com/ontology/algorithm/annotationFile (alg:annotationFile)`

1. The affymetrix-algorithm-param-state-annotation-file header parameter.
2. The affymetrix-algorithm-param-cn-annotation-file header parameter.
3. The affymetrix-algorithm-param-mapfile header parameter.
4. The affymetrix-algorithm-param-option-annotation-file header parameter.

`http://affymetrix.com/ontology/algorithm/parameterFile (alg:parameterFile)`

1. The affymetrix-algorithm-param-state-config-file header parameter.
2. The affymetrix-algorithm-param-config-file header parameter.
3. The affymetrix-algorithm-param-paramfile header parameter.
4. The affymetrix-algorithm-param-option-config-file header parameter.

`http://affymetrix.com/ontology/algorithm/referenceFile (alg:referenceFile)`

1. The affymetrix-algorithm-param-state-reference-file header parameter.
2. The affymetrix-algorithm-param-reference-file header parameter.
3. The affymetrix-algorithm-param-reference-mdlfile header parameter.

## **Appendix D: Confidence of Segment Calculation**

### **Copy Number Gain and Loss Confidence Calculation**

For CN Segments, confidence is determined on a marker by marker basis by evaluating the concordance of the log2ratio at each marker with the copy number state assigned by the HMM. The average confidence score of markers in Gain and Loss segments determines the confidence score of that segment.

### **LOH Confidence Calculation**

Each LOH segment is associated with a confidence score that reflects the degree to which the segment can be believed to be in one state or another. The confidence score value ranges from zero (absolute certainty that a given segment is not LOH) to one (absolute certainty that a given segment is LOH).

To assign a confidence score, the LOH algorithm for the Cytogenetics Whole-Genome 2.7M Array computes the ratio of the probability of the SCAR measurements in the segment under the LOH model to the sum of the probabilities under the LOH and non-LOH models. This quantity will be near zero when the data in a given segment is much more likely to be non-LOH, and near one when the data is much more likely to be LOH. A value near 0.5 indicates both models are equally likely for the segment.

For the CytoScan™ HD Array, the LOH confidence score is computed based on the proportion of markers with heterozygous calls within a given segment. When this proportion is low, the confidence value will be near one (100%). As the proportion of heterozygous calls increases, the confidence will decrease. When the proportion of heterozygous calls in the segment reaches typical sample heterozygosity, the confidence score will be zero.

### **Mosaicism Confidence Calculation**

#### **Mosaicism Segment Algorithm**

The algorithm for detection of copy number aberrations in the presence of mosaicism considers single copy deletions and gains. The algorithm is tuned to be most accurate when the normal/expected Copy Number State is two. The algorithm targets detection of changes of approximately 5MB or more in size. Copy number change events less than this size may be detected; however estimates of the mosaicism level will be less accurate. The algorithm considers only a discrete number of mosaicism levels which are set at 30%, 50% and 70%.

To detect mosaicism, it is assumed that the range of log ratios has been broken into a series of bands according to the detection level (30% or greater, 50% or greater, 70%-100% bands) and that log ratios within each band denote a specific copy number change (mosaic) event. The band from 0% change to 30% change serves as the copy number neutral region. Percentage bands can be positive or negative, depending on whether the change is a gain or loss, respectively. Given this information, the Mosaicism algorithm:

1. Computes the running means of log2 ratios over N markers for each chromosome. This window of N markers moves one marker at a time for each new mean. The algorithm considers marker number  $N = 6000$  (corresponding roughly to 5MB).
2. Detects places where copy number change regions are entered as mosaic events (i.e. left the copy number neutral region and into at least the 30% band's boundary). Band ranges were determined empirically via titration experiments.
3. Identifies the band in which the most extreme value of the mosaic event falls, and uses this band as the level of mosaicism to report.
4. Determines the segment boundaries of the mosaic events. Estimates of the boundaries are computed based on the level of mosaicism, and the exit and entry points from the copy number neutral (normal) region into the mosaicism band regions.

## **Mosaicism Confidence Score**

The Mosaicism Segment confidence score addresses the question, “How confident am I that this segment is truly Mosaic, rather than no change from normal?”. To compute a confidence score for each reported mosaic event, the algorithm takes the running medians of the log2ratio values using a window moving one marker at a time, and of a default span of 251 markers, across each called Mosaicism segment. Of these median log2ratios spanning the segment, the algorithm then determines the proportion of markers with running median log2 ratio above or below the thresholds required to make any CN change call.

For example, if the segment is called a mosaic gain, the confidence score is the percentage of median smoothed values above the threshold required to make a CN gain call (i.e. in at least the 30% gain band class). This means that if 95% of the markers are in the 30% gain class or above, the segment is given a Confidence score of 95%.

If the segment is called a mosaic loss, the Confidence score is the proportion or percentage of median smoothed values below the threshold required to make a CN loss call (i.e. in the -30% loss band class or lower).



## Appendix E: Genomic Position Coordinates

There are multiple conventions and file formats to describe locations in chromosomal DNA sequences. Here we describe a few of the issues as they relate to ChAS.

### Genome Assemblies

First, it is important to know which set of DNA sequences is being used as the reference. For the human genome, the reference assembly is available for download from public sources such as UCSC and Ensembl. Those two sites currently use identical genome assemblies, but refer to them by different names. UCSC uses names such as “hg17”, “hg18” and “hg19”. The identical genome assemblies are known as “NCBI35”, “NCBI36” and “GRCh37” at Ensembl. Assemblies at NCBI can have a decimal point as well, for example, “36.3” or “37.1”. For positions on the chromosomes 1-22, X and Y, there is no difference between assemblies “36.1”, “36.2” and “36.3” and we expect the same will be true for future “point” releases.

### SNP and Marker Positions

When referring to individual positions on a chromosome, such as the positions of SNPs, it is sufficient to give a single coordinate. There are different conventions about whether to consider the first DNA base pair on the chromosome as position 0 or position 1.

For SNP marker positions, all of the following consistently use a 1-based index position coordinate: CYCHP files, CNCHP files, NetAffx detail pages for SNP markers, NCBI pages for SNP positions of dbSNP entries, and the Graphs Table in ChAS 1.0.

Consider the (randomly-chosen) SNP marker “S-2MOG4” from the Cytogenetics Array. This marker is designed to correspond to the SNP with ID “rs7641618” in the dbSNP database. The NCBI website reports the position as chr3:186255049 on NCBI37 (hg19) or chr3:187737743 on NCBI36 (hg18). On the NetAffx website, currently still based on NCBI36 (hg18), the identical coordinate chr3:187737743 is also given for S-2MOG4. The same coordinate value is given in CYCHP files and in the ChAS graphs table. Refer to [http://www.ncbi.nlm.nih.gov/SNP/snp\\_ref.cgi?type=rs&rs=7641618](http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?type=rs&rs=7641618) for this particular example.

The Cytogenetics Array also contains copy-number markers which are not based on SNP positions. For these markers, we continue to use a 1-based index position. Unlike the case for SNPs, there is no particular single base pair that the marker corresponds to. The convention in CYCHP files is to use the position of the first DNA pair corresponding to the position where the marker hybridizes with the DNA. When two or more markers on the Cytogenetics Array have the same start position on a chromosome, the coordinate of one of them will be shifted by one or occasionally a few more bases such that each marker is reported at a unique position.

### Segment Positions

Affymetrix uses the BED or AED file format for storing and sharing region files between software. The BED format was created by UCSC for use with their genome browser, and is also used in other software. The AED format was created by Affymetrix for use with ChAS and possible future software, but used the BED format as a starting point.

The BED file format is explicitly defined to use a 0-based coordinate system where the second column (chromStart) in the file is the position of the first base-pair and the third column (chromEnd) is the position of the last base-pair plus one. Another way of saying this is that the start index is inclusive and the end index is exclusive. As an example, to refer to the first 100 based on the chromosome, you would use chromStart=0 and chromEnd=100. The length of any region is always given simply by chromEnd minus chromStart.

The UCSC browser strictly requires that chromStart not be larger than chromEnd. In order to support file outputs from non-conforming programs, ChAS will accept BED files where chromStart > chromEnd. It will simply switch those two coordinates and act as if the coordinates were given in the correct order.

Since a SNP has, by definition, a length of one base-pair, the proper way to represent a SNP position is with  $\text{chromEnd} = \text{chromStart} + 1$ . The UCSC browser does allow  $\text{chromStart}$  to be equal to  $\text{chromEnd}$ . But this is used for representing insertion points, and is not used to represent SNP positions. Because the AED format was intended to be compatible with BED format, we use the same coordinate system.

In order to make the ChAS Segments Table (and other segments-based tables) resemble the BED and AED output files, the Segments Table uses these same BED-like coordinates. Each segment is defined to include the DNA position of its starting array marker and its ending array marker.

For example, suppose there are three markers with the following positions on a chromosome given in the CYCHP file: Marker A at 1000, Marker B at 2000, Marker C at 3000. Marker positions in the CYCHP file are 1-based index positions. To represent these in a BED file, we would need a file like this:

```
Chr3 999 1000 markerA
```

```
Chr3 1999 2000 markerB
```

```
Chr3 2999 3000 markerC
```

If there were a segment starting at markerA and ending at markerC, we would need to represent it in a BED or AED file as:

```
Chr3 999 3000 segment_1
```

It would be represented in exactly the same way in the ChAS Segments Table. The length of this segment is 2001, because it includes the DNA bases at markerA and the marker at markerC.

Thus there will appear to be a discrepancy between the Segments Table and the Graphs Table for the start of the segment because the Graph Table will say markerA is at position 1000 (1-based index) while in the Segment Table, the segment starts at position 999. The end position of 3000 for the segment agrees with the position of markerC. This is not a mistake; it reflects the fact the Segments Table uses BED/AED-style coordinates.

## **Difference Between ChAS 1.0 and ChAS 1.1**

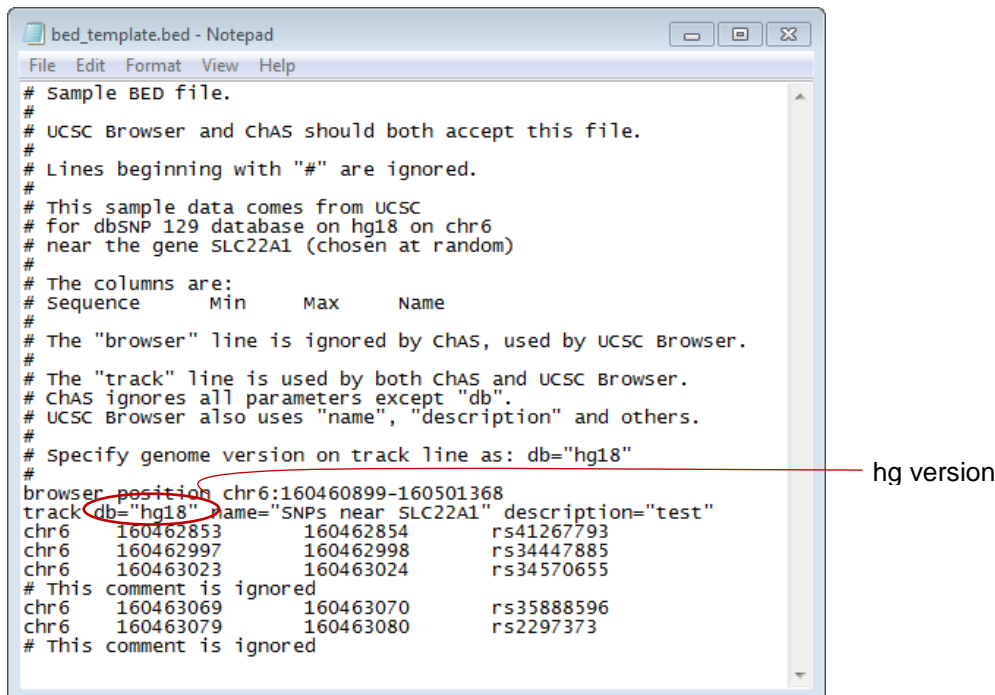
ChAS 1.0 unintentionally failed to correct the CxCHP coordinates to match BED conventions. Thus the Segments Table and AED files exported from ChAS 1.0 were shifted one position higher on the chromosome. Since ChAS 1.1 allows the direct export of BED files which can be loaded directly into the UCSC browser, this correction has been made so that the segment borders line-up precisely with SNP positions.

Comparing ChAS 1.0 with ChAS 1.1, you will see no difference in the Graphs Table, since the 1-based coordinate convention is retained for compatibility with NCBI and NetAffx websites, but you will see a difference of 1 base position for the Segments Table and AED outputs since we have corrected our coordinates to match the BED-style 0-based coordinates.

There are no additional changes in ChAS 1.2.

## Appendix F: Editing BED Files

A BED file is essentially a tab-delimited file.



```
bed_template.bed - Notepad
File Edit Format View Help
# Sample BED file.
#
# UCSC Browser and ChAS should both accept this file.
#
# Lines beginning with "#" are ignored.
#
# This sample data comes from UCSC
# for dbSNP 129 database on hg18 on chr6
# near the gene SLC22A1 (chosen at random)
#
# The columns are:
# Sequence      Min      Max      Name
#
# The "browser" line is ignored by ChAS, used by UCSC Browser.
#
# The "track" line is used by both ChAS and UCSC Browser.
# ChAS ignores all parameters except "db".
# UCSC Browser also uses "name", "description" and others.
#
# Specify genome version on track line as: db="hg18"
#
browser position chr6:160460899-160501368
track db="hg18" name="SNPs near SLC22A1" description="test"
chr6 160462853 160462854 rs41267793
chr6 160462997 160462998 rs34447885
chr6 160463023 160463024 rs34570655
# This comment is ignored
chr6 160463069 160463070 rs35888596
chr6 160463079 160463080 rs2297373
# This comment is ignored
```

Figure F.1 Example BED file

Using a text editor (not a spreadsheet application like Microsoft Excel) to edit BED files is highly recommended. Text editors such as WordPad and NotePad work well.

Editing a BED file using a spreadsheet application such as Excel is not recommended because these programs may not preserve the correct BED file format. For example, when exporting data from Excel into tab-delimited text, Excel may add quotation marks around some text. As a result, if you import a BED file in Excel, and then export the file, the exported BED file may not be a valid BED file that is usable in ChAS or other programs which accept BED files.

Some examples of specific problems include:

- A BED file that includes a track line which contains quotation marks or certain other special characters, for example:  
Track name="My Track" description="My special data" db=hg19  
When exported from Excel, extra quotation marks are added like this:  
"Track Name="My Track" description="My special data" db=hg19"
- A BED file that contains the optional **itemRgb** column (which always contains commas) will be corrupted during output because quotation marks will be added around all of the values.

There is no easy way to prevent Excel from adding extra quotation marks which corrupt the output. Advanced Excel users can use macro programming to create special output formats. Other options include:

- Do not use Excel to edit BED files. Use a text editor, but be certain to separate the columns with TAB characters and avoid non-ASCII characters. ChAS may display non-ASCII characters from the BED file correctly, but the BED format was not designed with such characters in mind; therefore, problems may occur when you try to share these files with others.

- After exporting a BED file from Excel, edit the file in another application to remove extra quotation marks
- Use AED format for the data, then use ChAS to export to BED format if needed
- Be careful to create a BED file that does not cause Excel to add quotation marks. Do not include the **itemRgb** column, quotation marks, or special characters in a track line. For example, the following is an acceptable track line:

Track db=hg19 name=My\_Track description=This\_is\_my\_data

## **Appendix G: CytoScan™ HD Algorithms and QC Metrics**

### **Algorithm Overview**

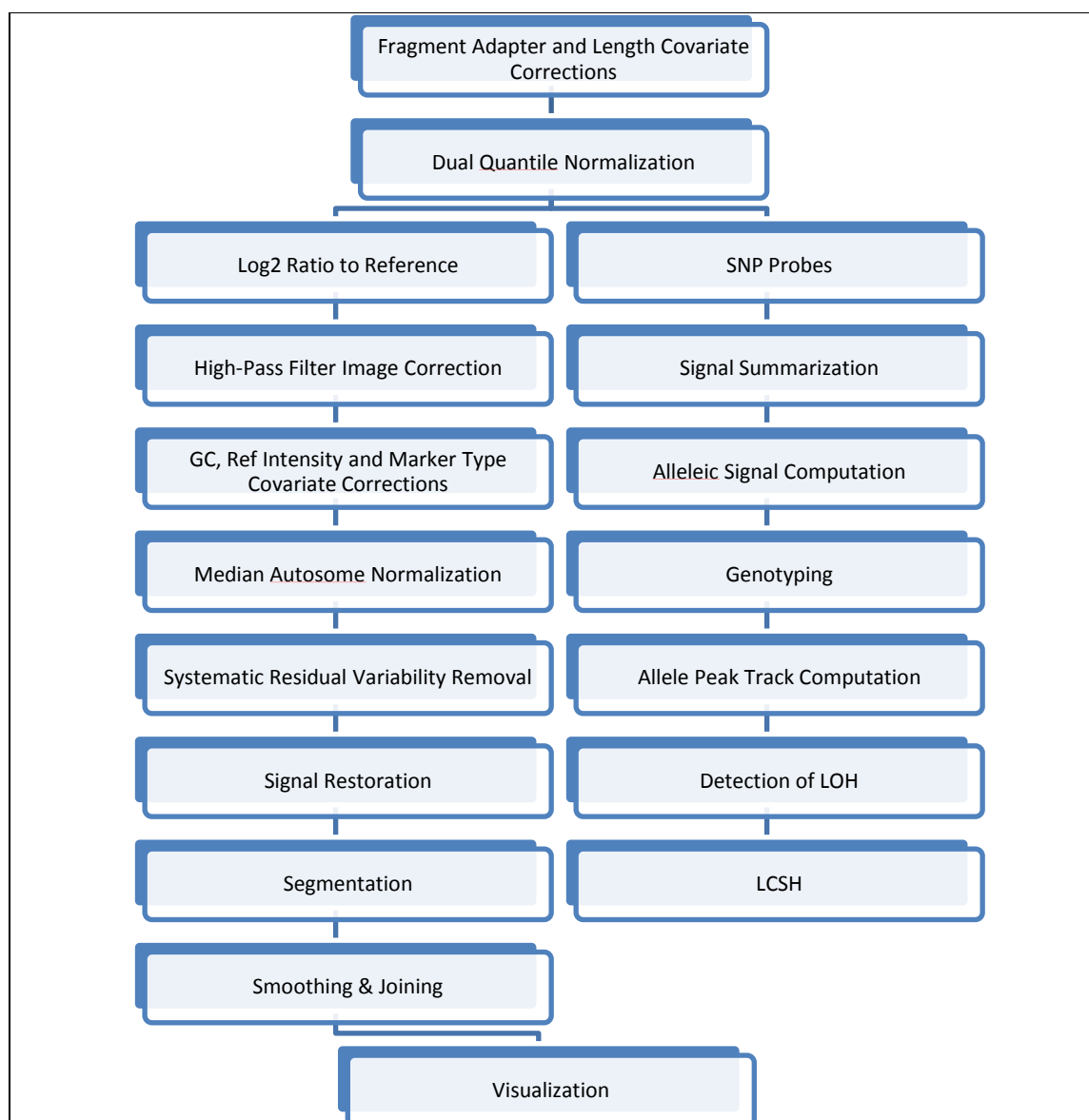
This section provides a high level overview of how copy number calls are generated within the software. The copy number workflow starts with the intensities on the array, include normalization and scaling, reference set ratios, log2 transformation, CN state segmentation, and how CN segment calls are made.

#### **Feature Identification and Signal Extraction**

Affymetrix® GeneChip® Cartridge Microarrays are scanned on the GeneChip® Scanner and processed by the AGCC scanner software package. AGCC aligns a grid on the DAT file (the original scanned image) to identify each microarray feature and calculates the signal from each feature. This process uses the .DAT file, containing the raw signal, and creates a .CEL file, which contains a single signal intensity for each feature. The .CEL file is used for all downstream analyses.

#### **Single Sample CytoScan™ Workflow**

Beginning with the raw signal data in the CEL file, the Single Sample CytoScan™ Workflow implements a series of steps that perform probe set summarization, normalization, removal of variation caused by known properties and residual variation, and completing with calling genotypes, copy number segments and LOH segments. The complete detail of the steps performed by the CytoScan™ Workflow is shown in the following figure and each step is briefly described in the subsequent text.



**Figure G.1 CytoScan™ Workflow Overview.** Steps related to probes used for copy number determination run down the left and those steps used for SNP probes run down the right side of the diagram

## Signal-Level Covariate Adjustors

The first level of covariate adjustors operate on the raw signal.

### Fragment Adapter Covariate Adjustor

After the Nsp I restriction digest, an Nsp I-specific adaptor is ligated onto the cohesive end termini. Since Nsp I is a 6-nucleotide cutter with degenerate sites, meaning that they contain one or more base pairs that are not specifically defined, these ends are of various sequences and the ligated adaptors are a variety of sequences. The exact sequences of the cut site and ligation adaptor have an effect on the overall efficiency of ligation and subsequent PCR amplification. The Adaptor Covariate Adjustor corrects for these differences by normalizing the signals for each adaptor/cut site sequence class to an overall median.

## Fragment Length Covariate Adjustor

The length of each Nsp I fragment impacts the efficiency of PCR amplification and therefore the signal. Fragments of 300-500 bp are amplified with the highest efficiency and the degree of amplification tapers off as the fragments get longer. The Length Covariate Adjustor corrects for these differences by normalizing the signals for a series of fragment size bins to an overall median.

## Dual Quantile Normalization

Dual quantile normalization is simply a two-phase process where probes used for copy number detection and probes used for SNP genotype detection are normalized separately. In both cases, a normalization sketch is built using the autosomal probes in the reference set. The normalization sketch is the prototype distribution of probe intensities that defines what this distribution looks like for all arrays. The single sample autosomal probes are fit to the sketch and the X and Y probes are interpolated into the distribution.

Quantile normalization makes the assumption that the distribution of probes on the array is fairly consistent from array to array. Since the X-chromosome is one of the largest chromosomes (155Mbp, ~5% of the genome), differences between males and females would stretch this assumption. That is why the quantile normalization focuses on creating an autosomal sketch and normalizing the autosome to it. The X and Y chromosome probes are then handled in a special way. Each of them is matched to the closest pre-normalization signal value. Based on that match, their normalized signal should be close to the signal for the very same autosomal probe. So the normalized values for X and Y probes are simply “looked-up” in the pre-normalization autosomal sketch, and transformed to the post-normalization value.

## Copy Number Workflow

### Log2 Ratio Calculation

Log2 ratios for each marker are calculated relative to the reference signal profile. The log2 ratio is simply  $\log_2(\text{sample}_m) - \log_2(\text{reference}_m)$ , for each marker, “m”.

### High Pass Filter Image Correction

Since most probes map to genomic markers associated with a normal copy number, most log2 ratios should be centered at a value of zero. Also, since markers from any genomic region are scattered across the surface of the microarray, regions of altered copy number will not appear as regional changes on the microarray image.

Some samples do reveal spatial trends away from zero that are gradual and this spatial bias when scattered back across the genome exhibits itself as added noise in the log2 ratios. The High Pass Filter Image Correction identifies these gradual spatial trends and adjusts log2 ratios to remove the spatial bias and lower the level of noise. Log2 Ratio-Level Covariate Adjustors

### Log2 Ratio-Level Covariate Adjustors

#### Super GC Covariate Adjustor

The GC content of genomic DNA sequence impacts probe signal dose-response and therefore probe log2 ratios. The sequence GC content of the microarray probe impacts hybridization kinetics. In addition, the genomic GC content of the Nsp I fragment and the 500 kbp surrounding the probe (local GC) all impact the efficiency of target preparation in the genomic region of each probe. The super GC covariate adjustor combines the probe GC content, the fragment GC content and the local GC content into one covariate that corrects for log2 ratio differences based on the combination of GC contents associated with each probe.

#### Reference Intensity Covariate Adjustor

Probes in different intensity categories have different dose responses in log2 ratio space. Using the Reference Set probes to define bins based on probe intensity, the single sample probes are binned and the median of the

distribution of log2 ratios within each bin is adjusted to the median log2 ratio of the corresponding bin from the reference set.

### **Marker Type Covariate Adjustor**

Polymorphic probes designed for SNP detection and non-polymorphic probes designed for copy number detection have different properties and different dose responses. The Marker Type Covariate Adjustor normalizes the median log2 ratios of SNP and CN markers to account for differences in log2 ratios between the two groups.

### **Median Autosome Normalization**

This final level of normalization simply shifts the median log2 ratio of the autosomes to a copy-number state equal to 2, i.e. a log2 ratio of 0.

### **Systematic Residual Variability Removal**

Even after all of the Covariate Adjustors, there is some residual variation with unknown origins. During product development we have introduced variation into the protocol in an attempt to capture other forms of unanticipated variation. The Systematic Residual Variability Removal step matches sample variability to the residual variability of the reference set, and when matched, corrects the data to remove the residual.

### **Signal Restoration**

Signal restoration is an application of Bayes wavelet shrinkage to the log2 ratios associated with markers. The result of this transformation is an overall reduction in variation around the local mean of log2 ratios. In this context, “local” means a region consisting of a small set of markers upstream and downstream from a given marker. The resulting data can be viewed as the Weighted Log2 Ratio in ChAS and serves as the input to the segmentation algorithm. The wavelet shrinkage method is augmented by a reduction of influence of outliers when local means are computed.

### **Segmentation**

#### **Copy Number Calls for each Marker based on Log2 Ratios**

For CytoScan™ HD, markers are individually assigned a copy number call by a Hidden Markov model (HMM). The sample specific inputs to the HMM are the Weighted Log2 Ratios generated by the Signal Restoration module.

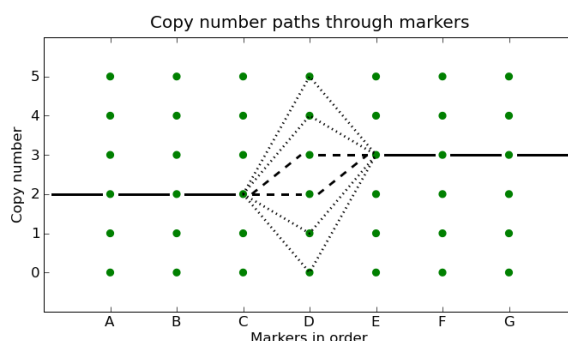
The weighted log2 ratios are centered on copy number (CN) = 2. In theory, when log2 ratio = 0 then CN = 2, when log2 ratio = -1 the CN = 1, etc. In truth, microarrays, or any hybridization-based technology, exhibit log2 ratio compression due to many factors, so the log2 ratios never exhibit the amplitude expected by the math. The following table shows theoretical and actual log2 ratios for different copy number states.

Copy Number Truth	Theoretical Log2 Ratio	Actual Log2 Ratio
1	-1	-0.45
2	0	0
3	0.58	0.3

The actual log2 ratios observed are best derived from a very large data set with well-characterized copy number changes. To this end, we have analyzed over 1400 samples that have copy number changes across 75% of the genome and have established stable empirical values for these expected log2 ratios. These values, as well as the dispersion characteristics of the log2 ratio data, are used as inputs to the HMM along with the weighted log2 ratios of the sample data.



The HMM uses these inputs to convert observed log2 ratios into a CN state for each marker. It uses a table of transition probabilities that express the probability of changing from any CN state to another. As can be seen in the following example, there are many potential paths through the possible CN states of a set of markers.



**Figure G.2 Potential CN paths through a set of markers**

The HMM uses the Viterbi algorithm to calculate the most probable path through the set of markers using the transition probabilities between each pair of CN states. Essentially, the graph of potential CN states is the “hidden” layer of the HMM, and the measure log2 ratios are the observed layer. The HMM algorithm finds the most probable CN states given the observed log2 ratios.

### Segment Formation

Once markers are assigned copy number states by the HMM, contiguous stretches of adjacent markers ordered by chromosome position having the same state are aggregated into segments. These segments are described in a segment table within the resulting CYCHP file that provides for each segment, the common copy number state, the number of markers in the segment, the genomic marker position that initiates the segment and the genomic marker position that terminates the segment.

### Enforce Minimal Segment Length

Default parameters enforce a minimum segment length of 5 markers. This is a subjective choice of parameter that implicitly states that the user is not interested in segments with fewer markers than the minimum. The algorithm that enforces minimum segment size distributes markers from any segment with fewer than the minimum to its larger neighboring segments by changing the copy number call on the modified markers to conform with those of the neighbors.

### Smoothing & Joining

To stabilize the calling of copy number gains or losses, the ChAS software implements a smoothing step. Smoothing will combine adjacent segments that are both gains, even if they are not the same copy number state. For example, smoothing will combine a set of adjacent segments of copy number state 3 and 4 into one segment and assign it the most prevalent copy number state of the markers in the original segments, (rounding up for gains and rounding down for losses in case of a tie). Smoothing will also combine copy number states 0 and 1. But smoothing will not combine gains with losses or either with normal segments.

Joining combines segments of gains or losses if they are separated by small spans of normal copy number segments. The default value defining “small spans” in ChAS is  $\leq 50$  markers and  $\leq 200$  kbp. Small segments of less than 50 markers/200kbp of normal copy number are removed, and the adjacent gain segments are joined. Likewise for flanking loss segments. This is a dynamic process in ChAS, in that smoothing and joining can be turned on and off, and parameters altered, resulting in modifications of the displayed segments, but not altering the underlying CNState graph.

### Segment Table Output

The final result of the copy number pipeline is a table of segments identified in the sample. The table in the CYCHP file includes segments of normal and non-normal copy number. Segments called on the X- and Y-

chromosomes are characterized as normal or non-normal using gender information and adjusting for the Pseudo Autosomal Regions (PAR) that are present on the X and Y. In ChAS, the segment table display only shows segments of non-normal copy number.

## **SNP Marker Workflow**

### **Signal Summarization**

CytoScan™ HD contains 6 probes for each SNP probe set, 3 targeting each allele. The first step of the SNP-specific workflow is to summarize the previously-normalized probe intensities for the A and B alleles, yielding allelic signal values.

### **Allelic Signal Computation**

For each marker, the Allelic Difference is calculated as the difference between the summarized signal of the A allele minus B allele, standardized such that an A-allele genotype is scaled to a positive value, and the B allele is scaled to a negative value. The standardization is determined based on median values for this difference under different genotype configurations determined by the reference set. In this way a homozygous AA maps to approximately +1, and a homozygous BB allele maps to approximately -1, with the heterozygote mapping to approximately 0. Additionally, single A and B allele signals will map to 0.5 and -0.5, respectively. This scaling provides a useful way of discerning two copies of an A allele from a single copy, enabling detection of regions of copy-neutral LOH (e.g. IBD) from hemizygous LOH.

### **Genotyping**

Genotyping for CytoScan™ HD is accomplished using the BRLMM-P algorithm described in the Affymetrix White Paper: BRLMM-P: A Genotype Calling Method for the SNP Array 5.0 (2007).

### **Allele Peak Track Computation**

The allele peak tracks for CytoScan™ HD are generated by signal processing the allele differences for each marker to improve the resolution of the tracks. The first step is the identification of local structure for each region by using a sliding window of markers. Within each window, a set of peak estimates are determined. Allelic difference signals for each marker are then shrunk toward the nearest peak, reducing some individual marker variability.

### **Detection of LOH**

The LOH algorithm frames the problem in terms of a statistical hypothesis test. Given a specific region containing  $N$  SNP markers with heterozygous and homozygous genotype calls, decide between the following two hypotheses:

Null Hypothesis: Region is LOH

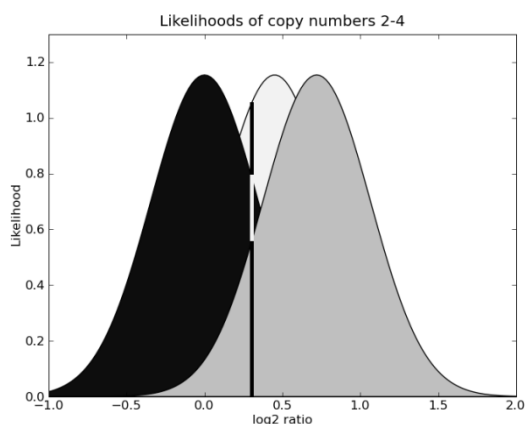
Alternative Hypothesis: Region is non-LOH

To decide between the two hypotheses the number of heterozygous calls is compared with a critical value that is computed for each sample. When the number of heterozygous calls is above the critical value, then the alternative hypothesis is favored, i.e. region is not LOH. If there are not a sufficient number of heterozygous calls then the decision is made in favor of LOH. The algorithm moves the region of  $N$  markers along the genome to determine LOH events. Further details are provided in the Affymetrix White Paper: The Loss of Heterozygosity (LOH) Algorithm in Genotyping Console 2.0.

## Copy Number Confidence Score

The copy number confidence score is a number between 0 and 1 that represents the combined confidence of a copy number call for a marker, incorporating a probability based on where the marker falls in the distributions of log<sub>2</sub> ratio values for each copy number state, and the confidence of the HMM call.

The first factor is the probability that a single marker's log<sub>2</sub> ratio falls into a distribution of log<sub>2</sub> ratios for markers of all possible copy number states.



Comparing a measured log<sub>2</sub> ratio value of 0.3 (vertical black line) with the normal distributions for copy number states 2 (black), 3 (light gray) and 4 (medium gray), there are respective likelihoods of that log<sub>2</sub> ratio coming from any of those three overlapping distributions. The probability that a log<sub>2</sub> ratio comes from any one distribution is its likelihood over the sum of all likelihoods.

The second factor is the certainty of the HMM call for that marker. The HMM call has the added value that it takes adjacent markers into account when calculating the most probable copy number state. For example, if a marker with a given positive log<sub>2</sub> ratio is surrounded by copy number 3 markers, it is more likely for the HMM to call it as copy number 3. By focusing on the transition probabilities between adjacent markers, the HMM provides a robust estimate of the underlying copy number state of each marker. The HMM also provides discrete state information for each marker. Each marker is called unambiguously and therefore the probability of a copy number call based on the HMM will be 1.0 for the called copy number state, and 0.0 for all other states.

Therefore, for each marker, we have a probability based on the distributions of log<sub>2</sub> ratio in each copy number state as well as a probability based on the HMM call. These are combined with a weighted average of 40% distributional probability and 60% HMM probability. The weighted average becomes the confidence score for a marker.

The calculation of confidence score for a segment is a simple arithmetic mean of the confidence scores of each of its comprising markers.

## Array Data QC Metrics

This section provides a high level overview of the key QC metrics used with the CytoScan™ HD Array..

### Median of the Absolute values of all Pairwise Differences (MAPD)

MAPD is a global measure of the variation of all microarray probes across the genome. It represents the median of the distribution of changes in log<sub>2</sub> ratio between adjacent probes. Since it measures differences between adjacent probes, it is a measure of short-range noise in the microarray data. Based on an empirical testing dataset, we have determined that array data with MAPD > 0.25 has too much noise to provide reliable copy number calls.

## Waviness SD

Waviness-SD is a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. Based on an empirical testing dataset, we have determined that array data with Waviness-SD > 0.12 has either sample or processing batch effects that will reduce the quality of the copy number calls. Elevated Waviness-SD is not always an indication of too much noise. Elevated Waviness with good MAPD and SNPQC metrics can occur in samples with many copy number changes or very large regions of change. It is therefore advised to check the data when observing elevated Waviness with good MAPD and SNPQC.

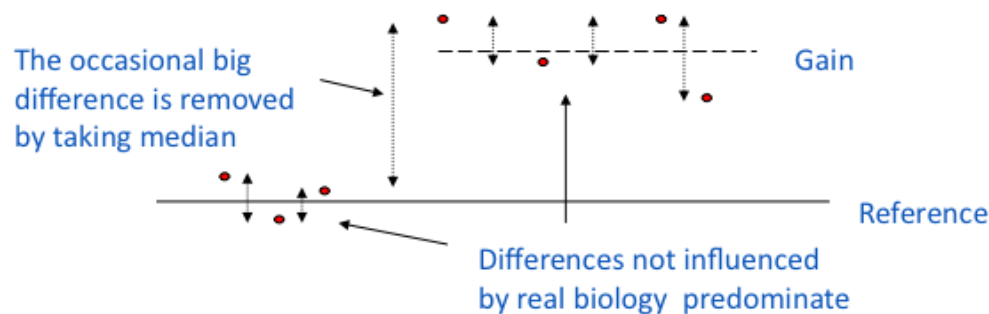
## SNPQC

SNPQC is a measure of how well genotype alleles are resolved in the microarray data. Based on an empirical testing dataset, we have determined that array data with SNPQC < 15 is of poorer quality than is required to meet genotyping QC standards.

## MAPD – Detailed Description

For quality assessment purposes, we define metrics that assess whether the microarray data is useful for copy number (CN) analysis. One of these metrics is Median of the Absolute values of all Pairwise Differences (MAPD).

MAPD is defined as the Median of the Absolute values of all Pairwise Differences between log2 ratios for a given chip. Each pair is defined as adjacent in terms of genomic distance, with SNP markers and CN markers being treated equally. Hence, any two markers that are adjacent on the genome are a pair. Except at the beginning and the end of a chromosome, every marker belongs to two pairs



Formally, if  $x_i$  is the log2 ratio for marker  $i$ :

$$\text{MAPD} = \text{median}(|x_{i-1} - x_i|, \text{ with } i \text{ ordered by genomic position})$$

MAPD is a per-microarray estimate of variability, like standard deviation (SD) or interquartile range (IQR). If the log2 ratios are distributed normally with a constant SD, then  $\text{MAPD}/0.96$  is equal to SD and  $\text{MAPD} \cdot 1.41$  is equal to IQR. However, unlike SD or IQR, using MAPD is robust against high biological variability in log2 ratios induced by conditions such as cancer.

Variability in log2 ratios in a microarray arises from two distinct sources:

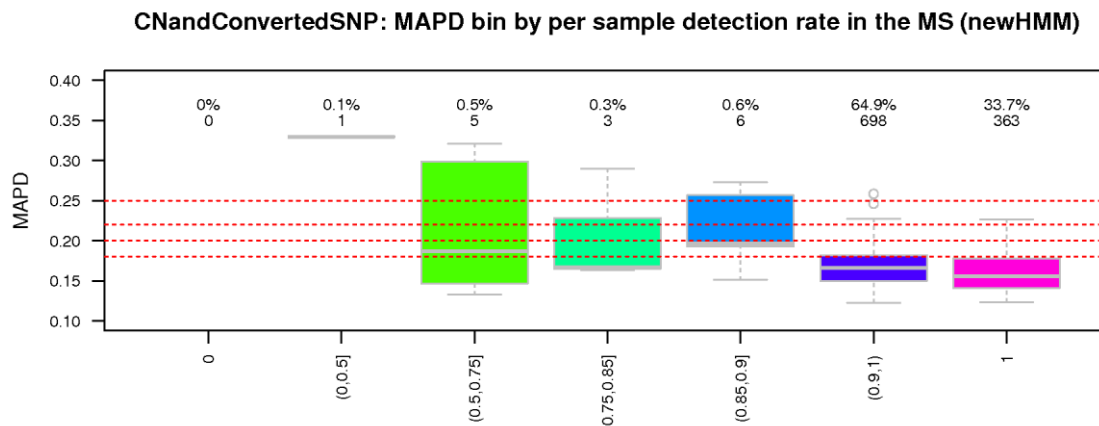
- Intrinsic variability in the starting material, hybridization cocktail preparation, microarray or scanner
- Apparent variability induced by the fact that the reference may have systematic differences from this microarray

Regardless of the source of the variability, increased variability decreases the quality of CN calls.

## Effect of MAPD on Functional Performance

As a measure of performance, we measured copy number gain and loss using samples with large chromosome aberrations that spanned approximately 70% of the genome. With this dataset of nearly 1500 microarrays we measured the sensitivity for detecting regions of copy number change across all of these regions. The sensitivity

of detecting an aberration on each array was binned into groups of varying sensitivities, and plotted versus MAPD for each array in the following graph.



The bins of detection sensitivity are displayed as coordinates along the x-axis, with 0% detection at the left and 100% at the right. The number of arrays is listed above each box plot. The majority of the arrays had sensitivities above 90%. Based on this analysis, we established a QC cutoff for MAPD of 0.25. Arrays with MAPD above 0.25 cannot be reliably used to determine copy number.

## Waviness-SD – Detailed Description

For quality assessment purposes, we define metrics that assess whether the microarray data is useful for copy number (CN) analysis. In addition to MAPD (above) we define an alternate form of measurement of variance in the array data that is called Waviness-SD, where SD stands for Standard Deviation.

Waviness refers to an effect seen in all genomic microarrays (see Maroni et al. (2007) Genome Biology 8:R228) where long-range variation is observed, often associated with regional genomic differences like local GC-content changes.

Waviness-SD is a QC metric that focuses on measuring these long-range effects. As described separately, MAPD is a metric that measures short-range variation, the variation of adjacent probes. The long-range variation measurement is accomplished by calculating the variation in log<sub>2</sub> ratios across the whole genome and subtracting out the short-range variation, specifically, for autosomal probes:

Define:

$X_i$  as the log<sub>2</sub> ratios of autosomal probes

And  $Z_i$  as the variance between adjacent probes:

$$Z_i = X_{2i+1} - X_{2i}$$

Waviness-SD is the total variance ( $X_i$ ) minus the local variance ( $Z_i$ ):

$$\text{Waviness-SD} = \sqrt{(\text{Var}(X_i) - \text{Var}(Z_i))/2}$$

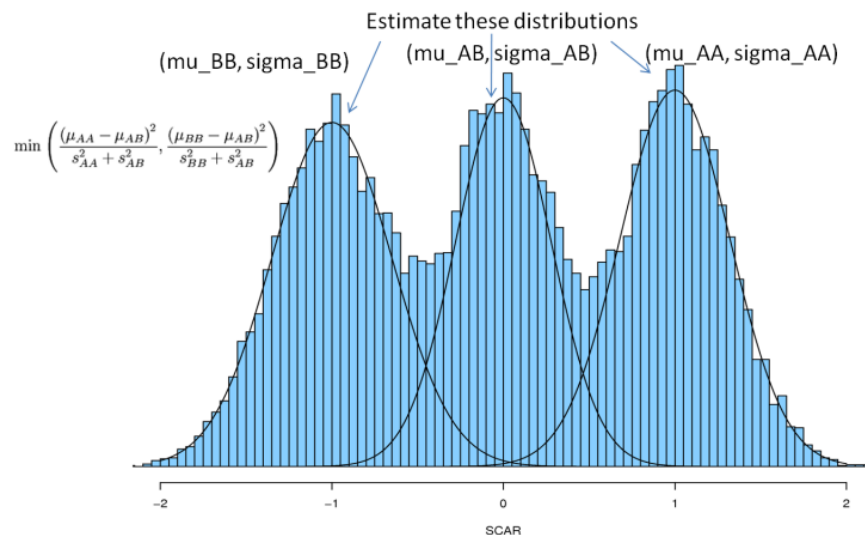
While this metric is useful in most cases, it does make the assumption that most of the genome is of normal copy number. This assumption may not be reasonable for some types of cancer samples with large amounts of genomic copy number variations, or for multiple-chromosome constitutive trisomies, where a considerable fraction of the genome is duplicated.

For most samples, a Waviness-SD value below 0.12 for CytoScan™ HD arrays indicates that the long-range variation is within levels that can be accommodated by the CytoScan™ HD algorithms. But a high Waviness-SD measure on a sample with good MAPD and SNPQC metric values should be checked for the presence of large regions of copy number change to assess whether it is a sample effect or a QC failure.

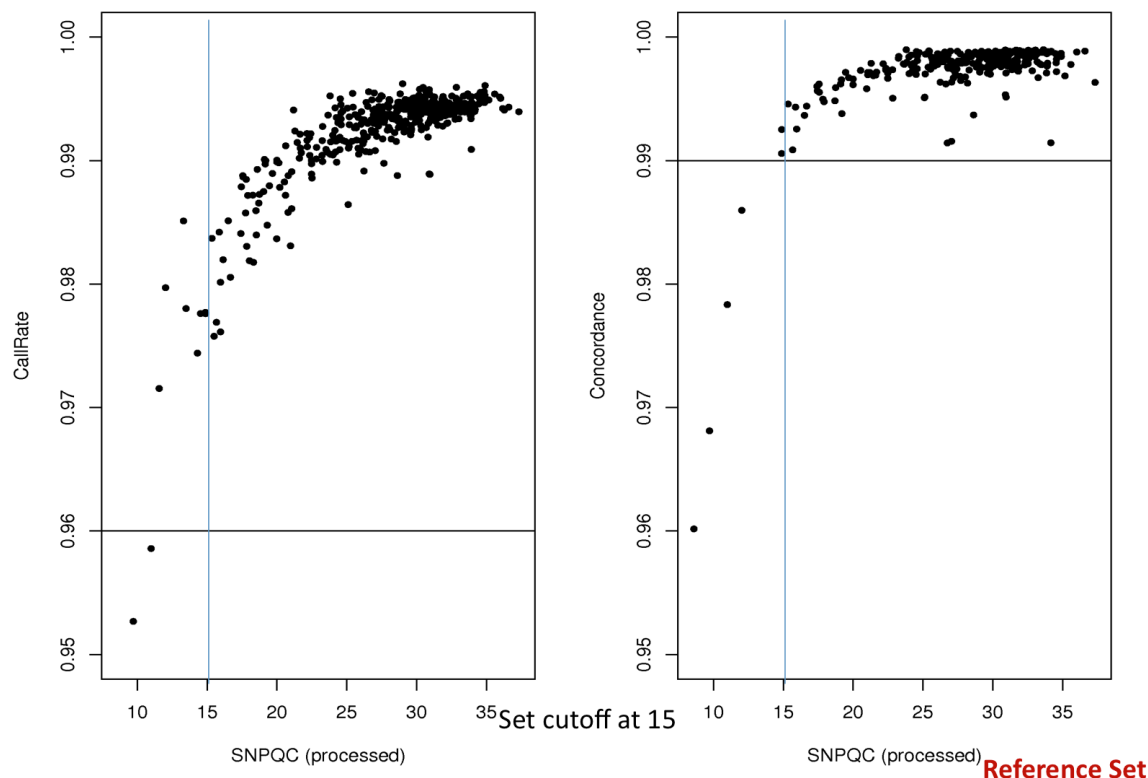
Waviness-SD can be a good indicator of process drift since it measures long-range variation relative to the CytoScan™ HD reference profile. A general rise of Waviness-SD for all samples coming from your laboratory may be an indication of a change of protocol, technique or reagents.

## SNPQC – Detailed Description

SNPQC is a metric that estimates the distributions of homozygous AA, heterozygous AB and homozygous BB alleles and calculates the distance between them. The better the separation of these distributions, the better the ability to identify a genotype based on its cluster position.



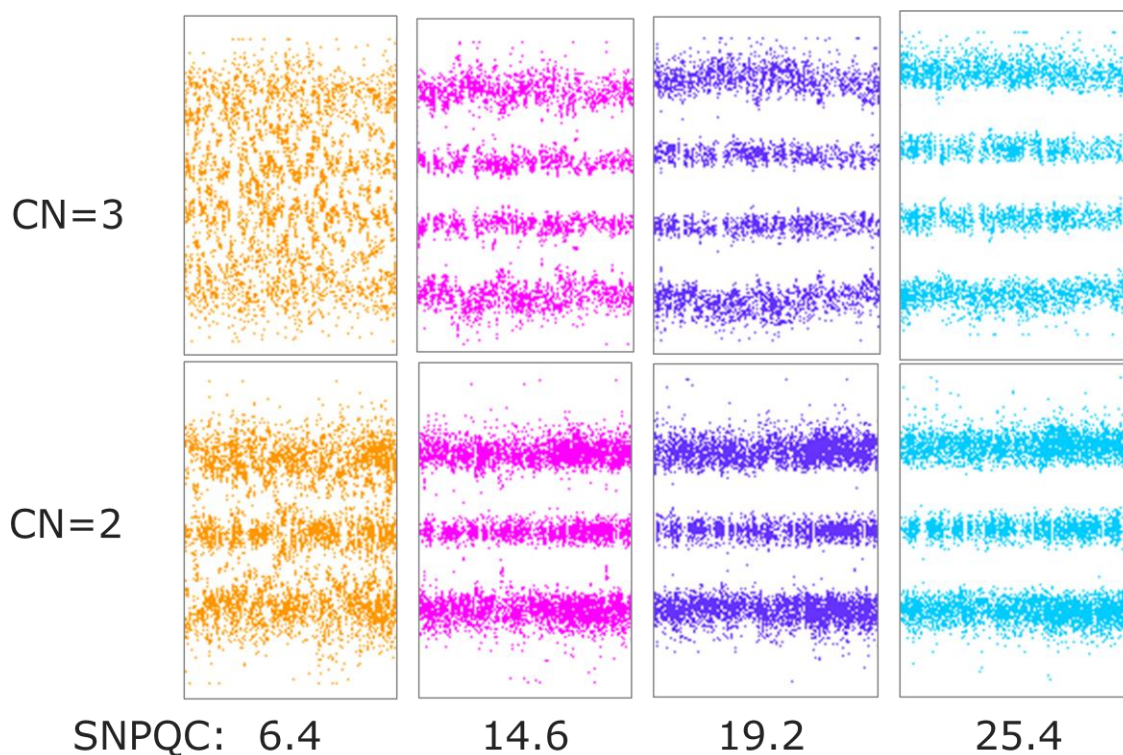
SNPQC correlates well with genotype performance, as measured by Call Rate and Concordance to published HapMap genotypes. To establish this relationship, we scored 380 microarrays from the Reference Set by calculating SNPQC, Call Rate and Concordance. The following graphs show the relationships between SNPQC and the other two metrics.



The left panel shows that when  $\text{SNPQC} > 15$ , Call Rate is above 98%. The right panel shows that when  $\text{SNPQC} > 15$ , Concordance is above 99%. This functional mapping of SNPQC has allowed us to set a functional threshold for this QC metric at 15. Microarrays with  $\text{SNPQC} \geq 15$  are considered of high quality and interpretation of the data is possible.

## Effect of SNPQC on Functional Performance

SNPQC provides insight into the overall level of data quality from a SNP perspective. The key consideration when evaluating the SNPQC value is to ensure the threshold is exceeded. The quality of the SNP allele data is compromised, and is noisier and more difficult to interpret when the SNPQC values are below the recommended acceptance threshold as illustrated in the figure below. When the SNPQC value is below 15, the noise within the array is higher than normal which compromises the overall data quality and clarity of results. However, when the SNPQC value is above 15, the data is of excellent quality and can be relied upon as robust with regard to performance.



**Figure G.3 Examples of Allele Track data quality at various levels of SNPQC. The lower row of figures show data for a CN=2 and the upper for CN=3 regions. The panels from left-to-right represent increasing SNPQC quality. The functional threshold for SNPQC is 15, so all values above 15 show excellent data quality.**

The key consideration is whether the SNPQC value is above or below the threshold value and not the absolute magnitude. As long as the SNPQC value exceeds the threshold there is a retention in the data quality as illustrated by the graphs to the right which demonstrate clear allelic data across a broad range of SNPQC values which exceed the recommended threshold. SNPQC is one of the metrics used to assess array quality and should be helpful towards determining which experimental data sets are of satisfactory quality to continue with subsequent interpretation.